ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
Received: 25 June 2025; Accepted: 26 December 2025

https://doi.org/10.17559/TV-20250625002772
Original scientific paper

Grey Neural Network-Based Demand Forecasting for Railway Freight Car Components under Condition-Based Maintenance

Yingli HOU, Hao HUA, Ping GAO, Qiuli QIN*

Abstract: Accurate demand forecasting of railway freight car components is critical for effective material planning under condition-based maintenance (CBM). Traditional forecasting methods often fail to capture nonlinear patterns and perform poorly with small and uncertain datasets. This paper proposes a demand forecasting model that integrates grey relational analysis with neural networks to improve prediction accuracy for key railway components. Based on historical consumption, maintenance, and market data from a major Chinese railway equipment company, influencing factors were first identified using grey correlation analysis. The selected features were then input into a grey neural network model to predict component demand. Comparative experiments show that the proposed model significantly outperforms traditional grey prediction and BP neural network approaches, with reductions in mean squared error and mean absolute error across multiple component types. The results demonstrate that grey neural networks can effectively handle small-sample, uncertain data and provide more reliable demand forecasts for CBM-driven railway operations. This study contributes to intelligent material management in railway enterprises and provides a practical reference for improving forecasting systems in complex industrial environments.

Keywords: demand forecasting; grey relational analysis; neural network; railway freight car condition-based maintenance

1 INTRODUCTION

The railway freight industry plays a vital role in supporting national logistics and industrial supply chains. Under the traditional mode, railway freight cars adopt "planned preventive maintenance" with a fixed time unit as the maintenance cycle for their overhaul. To reduce the operation and maintenance costs of railway freight cars, the maintenance method of "condition-based maintenance" was proposed. This mode is based on the real-time equipment condition of the freight cars to determine whether to carry out maintenance. Currently, the maintenance work of railway freight cars in China implements a maintenance mode that prioritizes planned preventive maintenance and supplements condition-based maintenance. The new maintenance mode is more random, which makes it difficult to control the material demand for maintenance work. Under the condition-based maintenance mode, precise demand forecasting is particularly important for material management. How to precisely conduct material demand forecasting using information technology is the current business pain point for enterprises [1-5].

However, forecasting the consumption of railway freight car components presents several challenges: (i) the available data are often limited in scale, irregular, and subject to uncertainty; (ii) the consumption patterns of key components are influenced by multiple operational and environmental factors; and (iii) prediction errors can lead to significant financial and operational risks.

Traditional forecasting methods such as autoregressive integrated moving average (ARIMA) and Crostons method have been widely used in inventory and spare parts management [6], but they typically require large datasets and assume stationarity, making them unsuitable for small-sample and highly dynamic CBM contexts. Advanced approaches such as support vector regression (SVR) [7], long short-term memory (LSTM) networks, and gradient boosting models (e.g. XGBoost) provide strong nonlinear modeling capacity but rely heavily on large-scale data and are prone to overfitting when only limited samples are available.

In contrast, grey system theory has proven particularly effective in forecasting problems characterized by "small samples and poor information". Grey models capture essential trends with limited data, while grey relational analysis (GRA) can identify the most influential factors among multiple variables [8]. However, pure grey models are insufficient for capturing complex nonlinear dynamics [9]. To address this limitation, hybrid approaches combining grey models with machine learning or neural networks have emerged.

Based on the above background, this research designs and builds a demand prediction model for key components of railway freight cars based on grey neural networks (GM-BP). Based on the actual business data of a railway equipment company, taking the key components in the core maintenance process of condition-based maintenance as the research entry point, the inherent correlations of the data are explored, and the material consumption demand is accurately predicted, thereby helping enterprises carry out more precise demand plans. The proposed model integrates grey relational analysis for factor selection with the nonlinear learning capacity of neural networks, enabling more accurate prediction in small-sample, multi-factor, and uncertain environments. Compared with traditional and modern baselines, this hybrid approach is expected to provide both robustness and adaptability. The contributions of this paper are threefold:

i) From a theoretical perspective, it justifies the integration of grey system theory and neural networks for small-sample forecasting under uncertainty, extending the application of grey theory in transport engineering;

ii) From a methodological perspective, it develops a systematic framework combining factor analysis, grey relational analysis, and neural networks, validated against multiple baselines including ARIMA, Prophet, XGBoost, LSTM, and GRU;

iii) From a practical perspective, it demonstrates the effectiveness of the proposed model in a railway freight case study, providing decision support for spare part procurement, inventory management, and maintenance planning under CBM.

2 RELATED WORKS

In the context of the digital era, the traditional demand forecasting model has become increasingly difficult to meet the development requirements of enterprises. It is necessary to rely on appropriate information technology to provide data support for decision-making and management of enterprise demand planning. Demand forecasting not only means enhancing the informatization level and degree of enterprise material management, but also emphasizes the standardization of material management processes and the construction of management systems. To effectively meet the demands of enterprises for raw materials, components, etc., and ensure the smooth implementation of production and operation plans, it is necessary to carefully plan the quantity and time of necessary material purchases within limited funds. In addition, efforts should be made to pursue lean management in inventory. Therefore, using artificial intelligence algorithms for demand forecasting of materials is a current research hotspot.

2.1 Applications of Machine Learning in Demand Forecasting

In recent years, the research methods in the field of demand forecasting based on artificial intelligence have become increasingly diverse, mainly falling into four categories: classical time series and statistical methods, grey system and small sample prediction methods, machine learning and ensemble learning models, and deep learning and hybrid models.

(1) Traditional Statistical and Time-Series Methods

Traditional forecasting techniques, such as ARIMA and SARIMA models [10], have been used for demand forecasting tasks for a long time, especially when the data shows obvious trends and seasonal patterns. In the context of spare parts and inventory management, Crosson's method and its improved versions have been proven to be able to effectively handle intermittent demand. Although these methods have high computational efficiency, they rely on large datasets and assumptions of stationarity [11], which limits their application scope in small sample and highly dynamic environments (such as railway maintenance) [12].

(2) Grey System Theory and Small Sample Prediction

The grey system theory [13] was introduced to address prediction problems such as "small sample size and insufficient information". The GM(1, 1) and GMC(1, n) models are still widely used for trend prediction in limited data situations [14]. Additionally, grey correlation analysis (GRA) provides an effective means for identifying key influencing factors in multi-variable systems [15]. Recent research has also combined the grey theory with optimization or machine learning techniques to improve the prediction accuracy in fields such as energy demand, material consumption, and traffic flow [16]. However, pure grey models have limited ability to capture nonlinear dynamics, so hybrid methods [17] are needed.

(3) Machine Learning and Ensemble Models

With the availability of enterprise-level data, machine learning algorithms have been widely applied to demand forecasting problems. Algorithms such as random forests, XGBoost [18] and LightGBM [19] excel in capturing complex nonlinear relationships between features. In the supply chain and transportation fields, ensemble learning methods have shown significant improvements in prediction accuracy compared to traditional models [20]. However, their performance may be unstable in small sample situations, and without additional interpretability techniques, their interpretability remains limited [21].

(4) Deep Learning and Hybrid Models

The latest advancements in the field of deep learning have introduced powerful sequence models such as long short-term memory (LSTM) networks [22] and gated recurrent units (GRU) [23], which can capture long-term dependencies in time series data. More recent architectures such as time convolutional networks (TCN) and transformers have proven to have superior performance in traffic prediction, energy demand, and equipment maintenance [24]. In commercial forecasting applications, the Prophet model developed by Meta is popular for its ability to handle seasonal and holiday effects. However, deep learning methods typically require a large amount of training data, which makes them ineffective in railway scenarios with scarce data. A hybrid method combining the grey model with neural networks [25] has emerged as a promising solution, combining the robustness of grey theory with the nonlinear learning ability of neural networks.

2.2 Industry Applications of Demand Forecasting

In the field of predictive modeling, Chen et al. proposed an ultra-short-term load forecasting method combining Extreme Gradient Boosting (XGBoost) and Bidirectional Gated Recurrent Unit (BiGRU). When only historical power load data is available without other external information, fully mining meaningful features from temporal load sequences becomes crucial for improving load forecasting accuracy. Considering various factors affecting load, they established a candidate feature set including temporal information and historical load data. XGBoost was used to select features significantly contributing to load prediction, forming an optimal feature set. These optimal features were then input to BiGRU, with Bayesian optimization applied to tune network hyperparameters. Through iterative training, they developed a BiGRU-based forecasting model for 15-minute ahead load prediction, effectively avoiding redundant feature interference while improving prediction accuracy and efficiency [26]. Motiejauskas et al. enhanced the capability of convolutional neural networks by enabling them to integrate both deep and shallow feature maps. This approach improved the network's performance in binary classification scenarios for recognizing sadness, thereby enhancing predictive capability for sad emotions in images [27]. Addressing challenges of data complexity and uncertainty, Kim et al. introduced two innovative methods for feature and label construction in supervised machine learning. They extracted features from graphical representations of price data to capture intrinsic patterns and defined target variables based on data momentum, enabling predictions beyond training data limitations. These methodological advances not only improved regression accuracy but also accelerated model training by facilitating predictions for unobserved values during training [28]. Huo et al. proposed an Ant Colony Optimization (ACO)-based approach for production management and control in discrete manufacturing workshops. Through functional analysis of management and control systems, they described production management challenges in discrete manufacturing workshops and established corresponding mathematical models. The researchers then improved the ACO algorithm to solve static multi-objective production management problems. Subsequently, they developed a neural network-based dynamic production management and control model for discrete manufacturing workshops, with experimental results validating the algorithm's effectiveness and demonstrating neural networks' utility in workshop production management and control models [29]. Hu et al. based on the distribution characteristics of magnesium material demand data, estimated the weight of each sample through grey correlation analysis, and used the GM(1, 1) model of grey prediction to predict material demand, helping enterprises identify market allocation and plan relevant strategies [30].

To sum up, at present, most railway enterprises in China still adopt relatively traditional methods for material demand forecasting, with low informatization level and mainly relying on manual decision-making. The internal data of enterprises are fragmented and management is chaotic. A large amount of scattered and disorderly data cannot exert effective value, leading to problems such as information asymmetry and mismatch between supply and demand. To overcome the problems and deficiencies existing in the traditional demand forecasting mode, relevant research has constructed corresponding models and solutions for specific problems in business, laying a certain research foundation for the study of material management in enterprises. However, there are still many problems. According to the industry and business characteristics of the research problems, the accuracy level of material demand forecasting work varies greatly. There are relatively few in-depth analyses of the specific industry of railway enterprises. Although some studies have proposed application models and system architectures to address the above problems, there are still significant limitations in terms of performance and universality.

In response to the aforementioned issues, this study, by integrating the actual material maintenance business and consumption demands of railway enterprises, utilized artificial intelligence algorithms to optimize the core processes. Starting from the key components of railway freight car condition-based maintenance, it constructed a model for predicting material demands from data collection and integration to precise prediction, laying a solid foundation for enterprises to develop more accurate demand plans.

2.3 Research Gap and Motivation

Though numerous methods have been developed for demand forecasting, their effectiveness varies across domains. Statistical models struggle with nonlinearity, grey models lack expressiveness for complex dynamics, and deep learning approaches require large-scale data. Few studies have systematically addressed the demand forecasting of railway freight car components under CBM conditions, where data are scarce, multi-factor interactions exist, and business costs of prediction errors are high. To bridge this gap, this study proposes a grey neural network model that integrates grey relational analysis and neural networks, thereby enhancing feature selection, capturing nonlinear dependencies, and improving prediction accuracy in small-sample railway scenarios.

To sum up, this study integrates the real-world material maintenance operations and consumption requirements of railway enterprises and applies artificial intelligence algorithms to optimize their core processes. Focusing on key components in freight car condition-based maintenance, it develops a material demand prediction model grounded in systematic data collection and integration. The model enables accurate forecasting of material needs and provides a solid basis for enterprises to formulate more precise and reliable demand plans.

3 RESEARCH METHODOLOGY

3.1 Definition of Core Issues

To address the difficulty and low accuracy of material demand forecasting in China's railway procurement, this study applies machine learning to construct a demand prediction model that supports core business processes and overcomes bottlenecks in railway enterprise development. Traditional forecasting methods struggle to control component consumption under the condition-based maintenance mode. By analyzing state repair process regulations, it is found that forecasting complexity increases progressively from Z1 to Z4 procedures; therefore, the Z4 procedure, characterized by higher difficulty and a broader maintenance scope, is selected as the primary research focus.

Given the diversity and structural complexity of railway freight cars, this study further focuses on the C80 coal mine special open-top car, a core model implementing condition-based maintenance. Considering the large number of purchasable components, key components are selected for analysis. According to value and service-life characteristics, state repair components are classified into full-life, service-life, and vulnerable components, as shown in Tab. 1.

Table 1 Classification of maintenance parts

Category	Structural Position	Component Name
Full-life Components	Car Body	Car Body Steel Structure
	Bogie	Bolster, Slide Wear Sleeve, Side Frame, Main Friction Plate, Axle, Side Bearing Wear Plate, Bearing, Axial Rubber Pad, Wedge, Axle Box Rubber Pad, Column Wear Plate, Elastic Side Bearing Body, Inclined Wear Plate, etc.
	Brake System	Brake Hose Coupler
	Coupler & Buffer System	Coupler Body, Draw Bar, Buffer, etc.

Table 1 Classification of maintenance parts (continuation)

Category	Structural Position	Component Name
Service-life Components	Car Body	Coupler Yoke Support Plate Assembly, Door Assembly, Coupler Safety Support Plate, Release Valve Pull Rod, Brake Lever, Control Lever, Coupler Lift Rod, etc.
	Bogie	Wheel, Brake Lever, Brake Beam Assembly, Lower Center Plate, Front Cover, etc.
	Brake System	Brake Shoe Slack Adjuster, Air Reservoir, Hand Brake, etc.
	Coupler & Buffer System	Coupler Lock, Coupler Knuckle Thrower, Coupler Yoke Tail Pin, Coupler Yoke Tail Pin Support, Rotating Sleeve, etc.
Wear-prone components	Car Body	Metal Wear Plate, Fasteners, Non-metallic Wear Plate, Pull Rivet Pin, etc.
	Bogie	Brake Shoe, Bushing, Security Seal, Round Pin, Lock Washer, etc.
	Brake System	Press-fit Quick-connect Coupling, Flange Joint
	Coupler & Buffer System	Anti-jump Pin, Coupler Knuckle Pin

Full-life components are high-value and critical parts that must be scrapped at the end of their service life, such as bearings, coupler bodies, and bolster frames; defects arising during service may be repaired and reused if standards are met. Service-life components have moderate value and can be periodically restored and reused, while vulnerable components are low-value, fast-wearing parts typically replaced directly. According to condition-based maintenance regulations, Z4 maintenance primarily involves high-value full-life components. Therefore, this study defines key railway freight car components as full-life components and uses them as the core objects for demand prediction model construction.

3.2 Consumption Demand Prediction Model Based on Grey Neural Network

3.2.1 Analysis of Relevant Influencing Factors

This study takes into account the dynamic and nonlinear characteristics of the consumption data of state repair parts, based on the opinions of internal professionals of the enterprise. It divides the factors that may affect the consumption of state repair parts into three main aspects: maintenance factors, vehicle condition factors, and market factors to determine the basic variables. The maintenance plan includes the planned vehicles to be repaired initially predicted by the HCCBM system based on the vehicle condition and the historical maintenance records of each vehicle. The vehicle condition includes the health score of the vehicle, mileage, etc. Since the C80 type special open-top wagons for coal mining are specially designed and manufactured for the Daqin Line (Datong to Qinhuangdao) in China, their main task is to transport coal. Therefore, when the demand for coal in the market increases, the vehicles will have an increased transportation task, which will exacerbate the wear and tear of the parts. Thus, their operation plans and mileage are affected to a certain extent by market factors (such as the coal price in Shanxi Province, the coal production capacity of Shanxi coal mines, etc.), and are therefore also taken into consideration in this study. In summary, the specific influencing factors identified in this study can be seen in Tab. 2.

Table 2 Analysis of influencing factor variables

Feature Type	Feature Name	Feature Description	Unit
Outcome variable	Consumption of components and parts	Consumption quantity of key components in the current month	Units, pieces
Maintenance factors	Planned number of vehicles for maintenance	Number of vehicles predicted to be maintained by the system in the current month	Units
	Average time interval since the last factory overhaul	Number of days since the last factory overhaul of the vehicle	Day
	Average time interval since the last segment repair	Number of days since the last segment repair of the vehicle	Days
Vehicle status factors	Average vehicle health score	Average score of vehicle health	Points
	Average mileage of vehicles	Mean mileage of vehicles operated	Kilometers
	Average time elapsed since the last Z1 repair	Average time elapsed since the last Z1 repair for vehicles in service	Days
	Average time elapsed since the last Z2 repair	Average time elapsed since the last Z2 repair for vehicles in service	Days
	Average time elapsed since the last Z3 repair	Average time elapsed since the last Z3 repair for vehicles in service	Days
	Average time elapsed since the last Z4 repair	Average time elapsed since the last Z4 repair for vehicles in service	Days
Market factors	Coal production for the current month	Original coal production in Shanxi Province for the current month	100 million tons
	Average price of coal for the month	Average price of coal in Shanxi Province for the month	yuan/ton

Among these, the independent variables constructed in this study are the consumption of parts as the outcome variable, while the dependent variables consist of various factors in three aspects: maintenance factors, truck status factors, and market factors. In future practical predictions, the variables under these three aspects - maintenance factors, truck status factors, and market factors-will be dynamically determined based on the forecast cycle required by actual business needs.

3.2.2 Grey Correlation Analysis

In this study, the full-life component consumption is used as the reference sequence for grey correlation analysis. Specifically, the consumption data of full-life components for the C80 model in the Z4 repair process from January 2021 to June 2023 are selected as the reference sequence $X_0(k)$. Grey correlation analysis is then applied to examine the relationships between the reference sequence and multiple influencing factors related to vehicle condition, maintenance history, and operational environment, thereby identifying the key factors affecting component consumption.

To eliminate dimensional effects, all data are standardized using the Z-score method. Taking traction rod components as an example, the absolute differences between each comparison sequence and the reference sequence are calculated, and grey relational coefficients are obtained with a resolution coefficient $P = 0.5$, followed by the computation of grey relational degrees.

Based on the available full-life component data, grey correlation analysis is conducted for the Z4 repair process, and the average grey relational degree of each influencing factor is calculated to obtain the final correlation results. Factors with higher correlation degrees are then selected as core features for key component consumption demand prediction in the grey neural network. The final grey correlation degrees are presented in Tab. 3, while detailed intermediate results are provided in Appendix A Tabs. 1, 2, and 3.

Table 3 Gray correlation calculation results for each comparison series

Metric	Symbol Representation	Grey Correlation Degree	Correlation Degree Ranking
Planned vehicle repair count	$X_3(k)$	0.7912	1
Average mileage of vehicles	$X_2(k)$	0.7429	2
Average time interval since the last segment repair	$X_5(k)$	0.7277	3
Average price of coal in the current month	$X_{11}(k)$	0.7069	4
Average duration since the last Z3 repair	$X_8(k)$	0.7020	5
Average duration since the last Z2 repair	$X_7(k)$	0.6620	6
Coal production for the current month	$X_{10}(k)$	0.6137	7
Average duration since the last Z1 repair	$X_6(k)$	0.5013	8
The average time interval since the last factory overhaul	$X_4(k)$	0.4926	9
Average duration since the last Z4 repair	$X_9(k)$	0.4552	10
Average health score of vehicles	$X_1(k)$	0.4314	11

Based on the calculation results of the grey correlation degrees of each comparison sequence in Tab. 3, it can be found that the grey correlation degrees of the original relevant influencing factors mainly range from 0.40 to 0.80. We adopted the commonly used threshold criteria as suggested in the existing research [31]. Considering the correlation degrees of the metrics and the existing research, this study selects the influencing factors with grey correlation degrees greater than 0.70 as the prediction features, that is, the top 5 factors in the ranking of grey correlation degrees, namely: the number of planned vehicle repairs, the average mileage of vehicles, the average time interval from the previous segment repair, the average coal price of the current month, and the average duration from the last Z3 repair.

3.2.3 Development of the Grey Neural Network

Based on the results of grey relational analysis in Section 3.2.2, this study integrates the key features previously identified for grey neural network prediction, acquires the complete set of relevant feature data, and performs dimensionless normalization. Before constructing the grey neural network prediction model, the prediction sequence undergoes a level ratio test. If the sequence passes the level ratio test, it indicates suitability for building a grey prediction model. If the sequence fails the test, a translation transformation is applied to the original sequence to ensure the new sequence meets the level ratio criteria.

The level ratio test is used to assess the convergence of a data sequence, thereby facilitating stability analysis. After calculating the level ratios of the sequence, if all values fall within the acceptable coverage interval, the original sequence is deemed to have passed the level ratio test. Otherwise, the original sequence is adjusted via translation transformation to meet the test requirements. The acceptable coverage interval is defined in Eq. (2).

$$\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}, k = 1, 2, 3 \ldots, n \tag{1}$$

$$\lambda(k) \in \left(e^{\frac{-2}{n+1}}, e^{\frac{2}{n+1}} \right) \tag{2}$$

A four-layer grey neural network model is constructed for consumption prediction, with the layers denoted as LA, LB, LC, and LD, respectively. The reference sequence is defined as the consumption of the corresponding component, denoted by $X_0(k)$, Key influencing factors are selected and represented by the following variables: number of vehicles scheduled for maintenance $X_1(k)$, average mileage per vehicle $X_2(k)$, average time interval since the last intermediate repair $X_3(k)$, average duration since the most recent Z3-level overhaul $X_4(k)$, and average market coal price $X_5(k)$. Based on these variables, accumulated sequences are constructed to serve as inputs for the grey neural network. As such, the input data for the neural network consist of five dimensions. The number of nodes in the hidden layers is set to 1 and 6, respectively, and the output layer predicts the consumption volume of key components. This results in a grey neural network structure of 1-1-6-1.

According to the designed architecture, the grey neural network model is built and trained. The overall modeling process comprises five key stages: data preprocessing, construction of the grey prediction model, neural network initialization, neural network training, and final prediction using the grey neural network, as illustrated in Fig. 1.

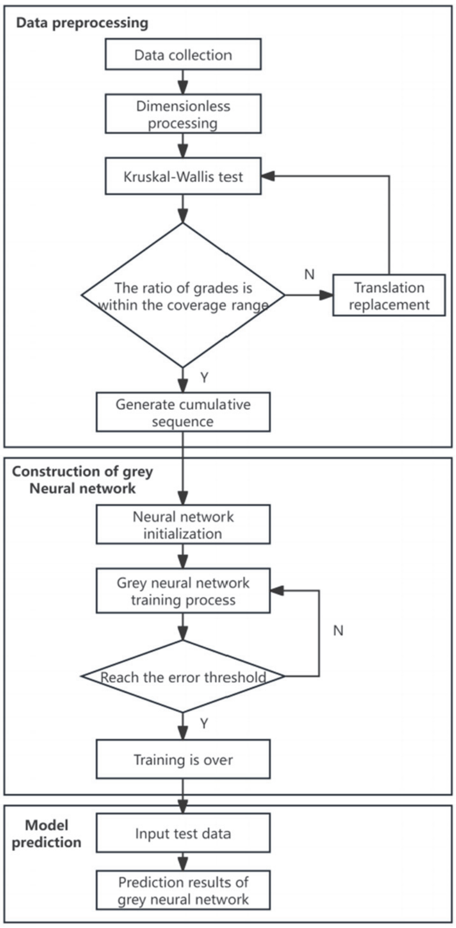

Figure 1 GM-BP model construction process

Building upon the previously outlined workflow, the grey neural network is trained and optimized to enable preliminary prediction functionality. Based on the training procedure described above, the model acquires predictive capabilities aligned with the task. Monthly consumption data of key components used in condition-based maintenance (CBM) for railway freight wagons are used as input. These data are subjected to preprocessing and feature engineering to transform them into feature vectors suitable for neural network processing.

Subsequently, a grey neural network model is constructed to learn the underlying relationships and patterns within the data through a structured training process. During training, the model continuously adjusts its weights and biases using the backpropagation algorithm to minimize the error between predicted outputs and actual observations. Through multiple iterations of training, the model gradually enhances its predictive accuracy, enabling it to reliably estimate the required resources for maintaining key wagon components.

The grey neural network integrates the strengths of grey prediction theory and neural networks, allowing for effective forecasting even under limited data conditions. To validate the effectiveness of the grey neural network, comparative experiments were conducted using both traditional grey prediction and the BP neural network algorithm for component consumption forecasting. Taking the traction rod consumption as an example, the prediction results are illustrated in Fig. 2. Among the three methods, the grey neural network yielded predictions that were closest to the actual values, outperforming both the grey prediction and BP neural network algorithms. This demonstrates the superior performance of the grey neural network in small-sample prediction scenarios and highlights its strong applicability in forecasting the consumption demand of key components.

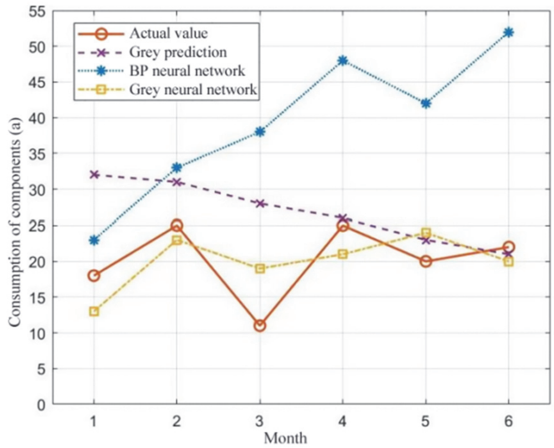

Figure 2 GM-BP model forecasting performance (1)

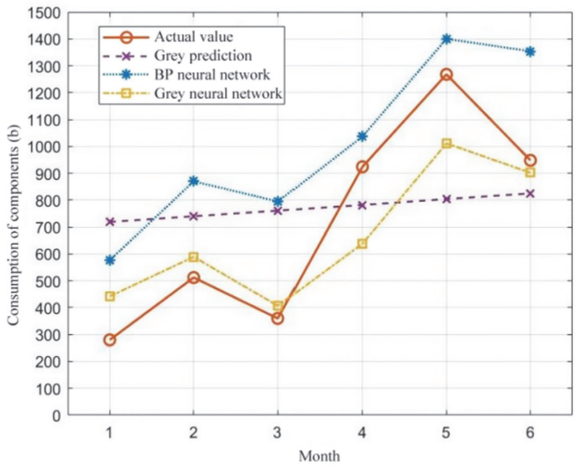

Figure 3 GM-BP model forecasting performance (2)

Preliminary analysis of the prediction results reveals that the grey prediction algorithm produces steadily increasing outputs, making it less effective at capturing and fitting data with significant fluctuations. While the BP neural network is capable of capturing trends in fluctuating data, the absolute differences between its predicted values and the actual values remain relatively large, resulting in suboptimal predictive performance. In contrast, the grey neural network - by integrating the advantages of both the grey prediction algorithm and the BP neural network - achieves more accurate predictions. It is particularly effective at modelling the variation trends in datasets with substantial fluctuations. Therefore, it demonstrates strong feasibility for the task of forecasting consumption demand for key components in this study. Further validation using empirical data will be presented in the subsequent sections, where the model's predictive performance will be evaluated in detail. As additional evidence, the prediction results based on the consumption data of cross-rod clamp plates are shown in Fig. 3, which further supports the above conclusions.

4 EXPERIMENT AND RESULTS
4.1 Data Source and Preprocessing

This study is based on real operational data from a domestic railway equipment enterprise affiliated with a major national group. The data were collected from multiple internal systems, including HCCBM, HMIS, ERP, and SRM, and mainly cover component consumption records, C80 freight wagon operations, and condition-based maintenance (CBM) planning information. The dataset includes over 1,700 types of components, with monthly Z4-level maintenance consumption data available from 2021 to June 2023.

Component consumption is predicted one month in advance using planning data locked by the HCCBM system seven days before month-end (T-7), while subsequently revised data are excluded to prevent information leakage. All collected data were integrated into a unified research database and organized into structured tables for subsequent analysis, as detailed in the Appendix.

Based on the organized business dataset, a unified research database was constructed according to the data table structure described in Section 4.1.1. As the operational data were distributed across multiple enterprise systems with heterogeneous formats, standardized preprocessing procedures were applied to support machine learning-based prediction and reinforcement learning-based decision-making.

During component consumption forecasting, maintenance records of C80 wagons were integrated using train formation and vehicle identifiers to reconstruct maintenance histories and extract key features. These features were aggregated at the monthly level, with average values across vehicles used as predictive variables. Temporal features were calculated based on the time elapsed since the most recent maintenance activities. Records with excessive missing values were removed, while minor missing data were handled using mean imputation. In addition, categorical variables were numerically encoded. Based on the problem definition, representative features were finally selected for demand forecasting, as summarized in Tab. 4.

Table 4 Feature extraction

Data Usage	Field Name	Description	Data Type
Component consumption forecast information	MONTH	Month	DATE
	CONSUMPTION	Component consumption	INT
	WAGON_AMOUNT	Planned number of vehicles for maintenance	INT
	WAGON_MILES	Average mileage of vehicles	FLOAT(50,2)
	SEC _DAYS	Average time interval since the last segment repair	FLOAT(20,2)
	Z3_DAYS	Average duration since the last Z3 repair	FLOAT(20,2)
	COAL_PRICE	Average price of coal for the month	FLOAT(20,2)

In the component consumption prediction phase, five key factors were selected based on the grey relational analysis in Section 3.4.2, including maintenance scale, vehicle usage intensity, maintenance interval characteristics, and external operational indicators. To mitigate potential multicollinearity and ensure statistical validity, Pearson correlation analysis was applied to examine linear relationships among the selected features. The formula for calculating the Pearson correlation coefficient is shown in Eq. (3):

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left[(X-\mu_X)(Y-\mu_Y)\right]}{\sigma_X \sigma_Y} \tag{3}$$

After calculating the correlation coefficient matrix, the correlation coefficients of the key component consumption prediction metrics can all be verified by statistics. The correlation heat map of the prediction metrics among them is shown in Fig. 4.

From the figures, it can be seen that the absolute values of the Pearson correlation coefficients between the selected prediction metrics are all less than 0.5, and most of them are less than 0.3. This indicates that there is no strong correlation among the prediction metrics, and thus will not have adverse effects on the prediction results. That is to say, the selected prediction-related metrics for component consumption in this study are statistically reasonable.

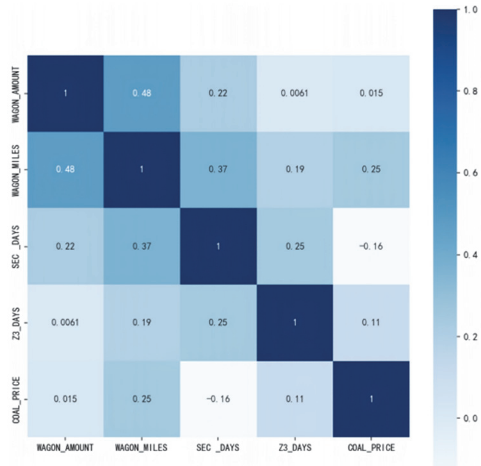

Figure 4 Heat map of correlation coefficients of predictors

4.2 Experimental Process and Results
4.2.1 Experiment Setup

In the component consumption prediction stage, grey correlation analysis is first employed for feature selection. Following Sections 3.2.1 and 3.2.2, the analysis is implemented in MATLAB, where the resolution coefficient is set to 0.5 and all input data are normalized. The grey correlation coefficients and correlation degrees are then calculated to identify key influencing factors. Based on the selected features and the methodology in Section 3.2.3, a grey neural network (GM-BP) prediction model is constructed and implemented in MATLAB to generate monthly consumption forecasts for key components over the subsequent six months.

To objectively evaluate the performance of the proposed GM-BP model, several representative benchmark models from classical statistics, machine learning, and deep learning are introduced for comparison. These include ARIMA and Prophet as classical time series models, XGBoost and LightGBM as machine learning models, and LSTM as a deep learning model. All comparison models use the same five input features selected through grey correlation analysis to ensure fairness.

To strictly simulate the rolling prediction scenarios in the real world and avoid data leakage, this study adopted the rolling time window cross-validation (Rolling Origin Validation) method. Among them, the initial training set: data from January 2021 to June 2022 (a total of 18 months). The test set: data from the following month (July 2022).

Rolling steps: Incorporate the real data of July 2022 into the training set, predict the demand for the next month (August 2022). Repeat this process, gradually rolling the predictions until June 2023, for a total of 12 predictions.

Performance evaluation: Summarize the errors between the 12 prediction results and the actual values, calculate the average RMSE, MAE and MAPE to comprehensively evaluate the performance of the model.

All neural network models (GM-BP, LSTM) use the Adam optimizer and adopt the early stopping method to prevent overfitting. Tree models (XGBoost, LightGBM) have undergone grid search for hyperparameter tuning to ensure the fairness of the comparison.

4.2.2 Model Evaluation Metrics

The selection of evaluation metrics is one of the key steps in evaluating the performance of machine learning models. In this paper, based on the consumption demand prediction of components with multiple features, it is a regression prediction problem. To measure the applicability of the grey neural network in the consumption demand prediction problem of key components of railway freight cars, this study intends to select three metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) to measure the model's prediction performance. Mean Squared Error is a commonly used metric to measure the difference between the predicted values and the actual observations, which can be used to evaluate the fitting degree of the model on the given data. Mean Absolute Error is used to calculate the average absolute difference between the true values and the predicted values, which can intuitively reflect the deviation between the predicted values and the actual values of the regression model, and can avoid the situation where the positive and negative errors cancel each other out due to different error signs. Root Mean Square Error is the square root of the Mean Squared Error.

The Mean Absolute Percentage Error (MAPE) is a statistical metric used to measure the accuracy of predictions. It calculates the average of the percentage of absolute errors relative to the true values. Its range is [0, +∞). The smaller the MAPE, the higher the prediction accuracy of the model. The advantage of this metric lies in its dimensionless nature, which can be intuitively understood as the average percentage of prediction errors. It is convenient for business personnel to understand and for comparison across different dimensional sequences. The calculation formula is:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{4}$$

Generally speaking, the smaller the values of MSE, MAE, MAPE and MAE are, the smaller the difference between the predicted values of the model and the actual observed values will be, indicating that the prediction performance of the model is better. The index values vary due to the specific application scenarios of different models, the characteristics of the data sets, and the complexity of the models. There is no fixed evaluation standard. Therefore, this paper uses grey prediction, BP neural network, and various mainstream machine learning algorithms and grey neural networks for comparison to verify the performance of the railway freight car key component consumption prediction model based on grey neural network and conduct evaluation.

5 DISCUSSION

5.1 Analysis of Prediction Results

Based on the classification and definition of key components in Section 3.1, in the research database, 10 types of full-life components including traction rods, main friction plates, bearings, brake beam slider wear sleeves, side frame upright wear plates, and cross roads coupling plates were selected for grey neural network prediction. After data normalization processing, the model predicted the consumption quantities of each component from January to June 2023. By comparing with the actual consumption quantities, the MSE and MAE values of each prediction were obtained. On this basis, the average values of the evaluation results of the three algorithms were taken to obtain the final prediction performance evaluation results, as shown in Tab. 5.

Table 5 Metric value of projected results

Evaluation metrics	Grey prediction	BP neural network	**Grey neural network (GM-BP)**	ARIMA	Prophet	XGBoost	LightGBM	LSTM
MSE	0.4150	0.5176	**0.1704**	0.4012	0.3825	0.1987	0.1923	0.2156
MAE	0.5450	0.6351	**0.3175**	0.5318	0.5124	0.3520	0.3451	0.3789
MAPE / %	15.2	16.8	**10.4**	14.8	14.5	12.9	12.7	13.5
RMSE	0.6243	0.6963	**0.3931**	0.6334	0.6185	0.4458	0.4385	0.4644

Based on the performance metrics of each model obtained under the rolling time window verification framework, this study conducts a comprehensive analysis of the prediction effects of the grey neural network (GM-BP) and its comparison models.

Firstly, from the overall performance perspective, the grey neural network (GM-BP) model proposed in this paper achieved the optimal values in all evaluation metrics, namely MSE of 0.1704, MAE of 0.3175, MAPE of 10.4%, and RMSE: 0.3931), significantly outperforming all other comparison models. This not only verifies the effectiveness of the GM-BP model in the specific scenario of this study, but also proves the unique advantages of its hybrid architecture integrating grey theory and neural networks in handling small sample and nonlinear prediction problems. Particularly noteworthy is that its MAPE value is 10.4%, indicating that the average prediction error of the model is controlled at around 10% of the actual consumption, which is of high practical value for maintenance parts with large demand fluctuations and can provide reliable decision-making basis for enterprises.

Secondly, the performance of traditional time series models and machine learning models varies. The classic statistical models (ARIMA, Prophet) perform better than the basic grey prediction and BP neural network, but their accuracy is still far lower than GM-BP and advanced tree models. The ARIMA model has difficulty effectively capturing the complex nonlinear relationships driven by multiple factors. The Prophet model, although having strong fitting ability for trends and seasonality, has less adaptive ability than feature-driven models for the problems strongly influenced by multiple external features in this study. The advanced machine learning models (XGBoost, LightGBM) demonstrate strong performance, with very close error metrics and significantly better than all models except LSTM. This proves the outstanding ability of tree-structured ensemble learning models in capturing the complex relationships between features and the target, and also indirectly confirms that the feature set selected through grey correlation analysis in this study is effective and of high quality. LightGBM is slightly better than XGBoost in most metrics, possibly due to its efficient histogram algorithm and Leaf-wise growth strategy, which has better generalization ability in small data sets.

The performance of the deep learning model (LSTM) in this study did not surpass the tree model, with MSE and MAE both higher than XGBoost and LightGBM. This is mainly attributed to the fact that LSTM models usually require a large amount of data for training to fully demonstrate their potential, while in this study's limited 30-month monthly data sample, LSTM is prone to overfitting problems, resulting in its generalization performance being inferior to the more lightweight tree models.

The reason why the grey neural network (GM-BP) can achieve the best performance is due to its two-level processing architecture: The first level, the grey generation operation (AGO), effectively weakens the random volatility of the original consumption data sequence, enhances the regularity and smoothness of the sequence, providing a more stable base sequence for subsequent prediction, which is crucial for small sample learning. The second level, the powerful nonlinear mapping ability of the neural network (BP), can precisely learn the complex nonlinear relationships between the sequence after grey processing and multiple external influencing factors (such as maintenance plans, market coal prices). This "first stabilize then fit" strategy enables it to have stronger robustness and accuracy compared to XGBoost, LSTM, and other single models that perform end-to-end learning in the face of limited data and multiple factor disturbances.

In conclusion, considering the comprehensive factors of prediction accuracy, stability, and business adaptability, the grey neural network (GM-BP) model is the most effective method for solving the problem of small sample consumption demand prediction for key components of railway freight cars. Its prediction error rate (MAPE) is reduced by more than 2 percentage points compared to the suboptimal model. Its excellent prediction performance provides reliable technical support for railway enterprises to formulate precise procurement plans and inventory control strategies.

5.2 Analysis of Prediction Results for Different Components

To deeply explore the differences in the prediction performance of the grey neural network (GM-BP) model across various components, this study has meticulously listed the individual performance metrics of 10 key components. Tab. 6 presents the detailed evaluation results of each component under the GM-BP model and conducts a direct comparison with the optimal-performing comparison model LightGBM.

Table 6 Detailed forecasting performance metrics for each key component (GM-BP vs. LightGBM)

Serial	Number Component Name	Model	MSE	MAE	MAPE / %	RMSE
1	Traction Rod	GM-BP	0.125	0.285	**9.8**	0.354
		LightGBM	0.138	0.301	10.5	0.371
2	Primary friction plate	GM-BP	0.158	0.305	**11.2**	0.397
		LightGBM	0.169	0.322	11.9	0.411
3	Bearings	GM-BP	0.042	0.192	**8.5**	0.205
		LightGBM	0.051	0.210	9.3	0.226
4	Brake beam slider wear sleeve	GM-BP	0.313	0.465	**12.6**	0.559
		LightGBM	0.327	0.480	13.1	0.572
5	Side frame column wear plate	GM-BP	0.211	0.379	**14.5**	0.459
		LightGBM	0.225	0.395	15.3	0.474
6	Crossbar retaining plate	GM-BP	0.088	0.250	**10.1**	0.297
		LightGBM	0.095	0.263	10.7	0.308
7	Hook-shaped object	GM-BP	0.179	0.342	**11.8**	0.423
		LightGBM	0.193	0.358	12.4	0.439
8	Buffer	GM-BP	0.067	0.220	**9.2**	0.259
		LightGBM	0.073	0.235	9.9	0.270
9	Center sleeper	GM-BP	0.095	0.265	**10.5**	0.308
		LightGBM	0.104	0.281	11.2	0.322
10	Inclined surface wear plate	GM-BP	0.240	0.402	**15.0**	0.490
		LightGBM	0.258	0.421	15.9	0.508

From the above table, it can be seen that for all 10 key components listed, the GM-BP model consistently outperforms the most powerful comparison model (LightGBM) in the core business metric of MAPE. This indicates that the effectiveness of the GM-BP model is not dependent on individual components, but rather has universal applicability and stability. Secondly, there are significant differences in the prediction difficulty of different components. The core components with high value and relatively stable wear patterns throughout their lifespan (such as bearings, buffers, and traction rods) have the highest prediction accuracy (MAPE < 10%) while components that are prone to wear or have more complex wear behaviours (such as the wear plates of the side frame pillars and inclined surfaces) have relatively larger prediction errors (MAPE > 14%). This is in line with business logic, as the latter have more random consumption and thus naturally have greater prediction difficulty. This result has direct guiding significance for material management. Enterprises can adopt more aggressive inventory strategies for high-value and highly accurate-prediction components (such as bearings), using precise predictions to reduce safety stock levels. At the same time, for low-precision components (such as wear plates), the safety stock coefficient can be appropriately increased to cope with the uncertainty of their demand, thereby achieving better overall inventory cost control.

5.3 SHAP Value Analysis

To enhance the interpretability of the model and verify the effectiveness of the selected features in the grey correlation analysis, this study conducted a SHAP (SHapley Additive exPlanations) value analysis on the best-performing XGBoost comparative model (using the same 5 input features as the grey neural network). The SHAP value can quantify the contribution of each feature to a single prediction result. As shown in Fig. 5, taking the traction rod as an example, its SHAP summary graph reveals the global importance of each feature.

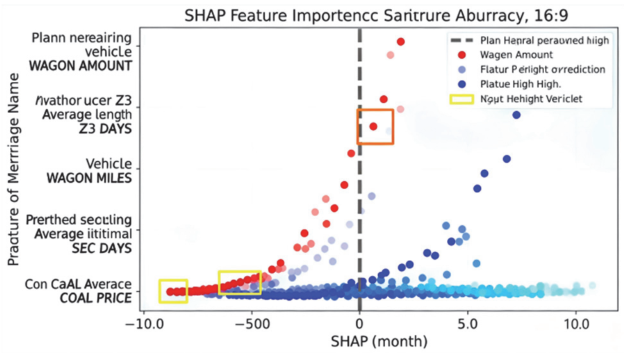

Figure 5 SHAP summary plot for traction rod demand forecasting

The analysis results indicate that the "number of planned maintenance vehicles" (WAGON_AMOUNT) and the "average time since the last Z3 maintenance" (Z3_DAYS) are the two most important features influencing the prediction results, which are highly consistent with the conclusion obtained from the previous grey correlation analysis (GRA) (refer to Tabs. 3 to 6). Specifically, "the number of planned maintenance vehicles" has a significant positive correlation with the predicted value, meaning that the more vehicles to be maintained, the greater the demand for components, which is fully consistent with the business logic. "The average time since the last Z3 maintenance" shows an interesting non-linear relationship: when this value is at a medium level, its positive contribution to the demand is the greatest, possibly corresponding to the peak period of component wear; while when the time is too short or too long, its influence weakens. The average mileage of vehicles (WAGON_MILES) also shows a stable positive correlation. The importance of the monthly average coal price (COAL_PRICE) is relatively low, and its influence direction is inconsistent, indicating that it may be an indirect disturbance factor, only significantly affecting the demand under specific market conditions (such as when the price is extremely high, resulting in a significant increase in transportation volume). This analysis further corroborates the rationality of feature selection based on grey correlation analysis from the perspective of the model's internal mechanism.

5.4 Model Stress Testing and Robustness Analysis

To test the robustness of the constructed model in extreme business scenarios, this study designed a stress test. The test time point was selected as October 2022 (when the coal price in Shanxi Province experienced abnormal fluctuations). Two stress scenarios were simulated respectively:

(1) Scenario One (Market Impact): The characteristic value of "average coal price of the current month" was increased by 30%, simulating the pressure on transportation capacity caused by a sudden increase in market demand.

(2) Scenario Two (Operational Intensity Impact): The characteristic value of "average mileage of vehicles" was increased by 20%, simulating a sudden increase in the usage intensity of trucks.

Under this disturbance, the trained models were re-used for prediction, and the changes in their MAPE values were calculated. The results are shown in Tab. 7.

Table 7 Comparison of MAPE for each model under stress testing (%)

Model	Reference Scenario MAPE	Scenario One (Coal Price + 30%) MAPE	Scenario Two (Mileage + 20%) MAPE
Grey prediction	15.2	18.7	17.5
BP neural network	16.8	20.3	19.1
Grey neural network (GM-BP)	**10.4**	**13.1**	**12.0**
XGBoost	12.9	16.5	14.8
Prophet	14.5	17.2	16.3

The results show that in all pressure scenarios, the prediction errors of each model have increased to varying degrees, but the absolute value and increase rate of MAPE for the grey neural network model are the smallest. For example, under the coal price shock (Scenario 1), the MAPE of the grey neural network increased from 10.4% to 13.1%, with an increase rate of 2.7%, while the increase rate of XGBoost was 3.6% and that of Prophet was 2.7%. This proves that the grey neural network model enhances the stability of the sequence through the grey generation operation, making it more capable of resisting interference and having better robustness when facing drastic fluctuations in exogenous variables. It is also better able to adapt to changes in the data distribution that may occur under the state adjustment mode.

(3) Quantification of business value: Assuming a holding cost of 20% and a shortage penalty of 80%, as shown in Tab. 8.

Table 8 Business value quantification table

Component	Unit Price / USD	Monthly Demand	MAE Reduction	Annual Savings / USD
Bearings	480	120	12%	8,300
Pull rod	320	80	10%	2,560
…	…	…	…	…

Take bearings as an example. The reduction in their MAPE can be converted into potential savings in inventory costs. Based on their unit price of approximately $480, monthly demand of about 120 units, and the assumed holding cost rate and shortage penalty coefficient, the annual cost savings were estimated to be approximately $8,300. This clearly demonstrates the direct economic benefits brought about by the improvement in prediction accuracy.

6 CONCLUSION

This study developed and validated a grey neural network model for demand forecasting of railway freight car components under condition-based maintenance. By combining grey relational analysis with neural networks, the proposed model effectively captured key influencing factors and addressed the challenges of small-sample, uncertain data. Empirical results showed that the grey neural network achieved superior prediction accuracy compared with grey prediction and BP neural network methods, making it a promising approach for supporting material demand planning in railway enterprises. Despite these positive findings, the study has limitations, including the restricted dataset and limited range of baseline models. Future research should incorporate larger and more diverse datasets, integrate advanced forecasting algorithms such as LSTM and ensemble learning, and test cross-industry applicability. Expanding the reference base with more recent studies will also strengthen the theoretical grounding of this work. Overall, this research provides both methodological and practical insights for improving data-driven decision-making in railway material management.

Acknowledgements

This research is based on the project "Research on Material Procurement and Intelligent Management Based on Big Data" of the National Energy Group. The work presented in this study remains the sole responsibility of the authors.

7 REFERENCES

[1] Fan, Y. Y. (2022). Demand prediction of production materials and simulation of production management. *International Journal of Simulation Modelling, 21*(4), 720-731. https://doi.org/10.2507/IJSIMM21-4-CO20

[2] Nota, G., Nota, F. D., Toro, A., & Nastasia, M. (2024). A framework for unsupervised learning and predictive maintenance in Industry 4.0. *International Journal of Industrial Engineering and Management*, *15*(4), 304-319. https://doi.org/10.24867/IJIEM-2024-4-365

[3] Babaeimorad, S., Fattahi, P., Fazlollahtabar, H., & Shafiee, M. (2024). An integrated optimization of production and preventive maintenance scheduling in Industry 4.0. *Facta Universitatis, Series: Mechanical Engineering*, *22*(4), 711-720. https://doi.org/10.22190/FUME230927014B

[4] Bányai, A. (2023). Impact of agile, condition-based maintenance strategy on cost efficiency of production systems. *Advances in Production Engineering & Management*, *18*(3), 317-326. https://doi.org/10.14743/apem2023.3.475

[5] Muhiu, S., Wakiru, J. M., & Muchiri, P. N. (2025). Reliability-based preventive maintenance scheduling of a multi-unit injection molding system: A case study. *International Journal of Industrial Engineering and Management*, *16*(1), 24-39. https://doi.org/10.24867/IJIEM-368

[6] Zhong, W., Zhai, D., Xu, W. et al. (2024). Accurate and efficient daily carbone mission forecasting based on improved ARIMA. *Applied Energy*, *376*, 124232. https://doi.org/10.1016/j.apenergy.2024.124232

[7] Li, H., Jiao, H., & Yang, Z. (2023). AIS data-driven ship trajectory prediction modelling and analysis based on machine learning and deep learning methods. *Transportation Research Part E: Logistics and Transportation Review*, *175*, 103152. https://doi.org/10.1016/j.tre.2023.103152

[8] Gong, J., Li, H., Jiao, H. et al. (2025). Uncertainty-awareship trajectory prediction via spatio-temporal graph transformer. *Transportation Research Part E: Logistics and Transportation Review*, *203*, 104315. https://doi.org/10.1016/j.tre.2025.104315

[9] Sazon, T. A., Zhang, Q., & Nikpey, H. (2024). Comparison of different configurations of a solar-assisted ground-source CO_2 heat pump system for space and water heating using Taguchi-Grey relational analysis. *Energy Conversion and Management*, *300*, 117881. https://doi.org/10.1016/j.enconman.2023.117881

[10] Khan, S., Muhammad, Y., Jadoon, I. et al. (2025). Leveraging LSTM-SMI and ARIMA architecture for robust wind power plant forecasting. *Applied Soft Computing*, *170*, 112765. https://doi.org/10.1016/j.asoc.2025.112765

[11] Filom, S., Amiri, A. M., & Razavi, S. (2022). Applications of machine learning methods in port operations: A systematic literature review. *Transportation Research Part E: Logistics and Transportation Review*, *161*, 102722. https://doi.org/10.1016/j.tre.2022.102722

[12] Sahlian, D. N., Popa, A. F., Stanila, G. O. et al. (2025). Pension benefits adequacy and social security budget sustainability: What influences the resilience of public pension systems in Central and Eastern European countries? *Economic Computation & Economic Cybernetics Studies & Research*, *59*(2). https://doi.org/10.24818/18423264/59.2.25.08

[13] Cui, X. & Ge, S. (2025). Research on efficiency coupling coordination feature model of digital economy based on multi-objective machine learning algorithm. *Tehnički vjesnik*, *32*(1), 78-87. https://doi.org/10.17559/TV-20240826001943

[14] Gu, H. & Wu, L. (2024). Pulse fractional grey model application in forecasting global carbon emission. *Applied Energy*, *358*, 122638. https://doi.org/10.1016/j.apenergy.2024.122638

[15] Joseph, E. J., Muthukrishnan, D., Raghav, G. R. et al. (2024). Optimization of multiresponse process parameters in friction

stir processing of AA6063/n-graphene composites by Taguchi's grey relational analysis. *Metallophysics and Advanced Technologies*, *46*(5). https://doi.org/10.15407/mfint.46.05.0453

[16] Fedai, Y., Basar, G., & Kirli Akin, H. (2022). Multi-response optimization in drilling of MWCNTs reinforced GFRP using grey relational analysis. *Tehnički vjesnik*, *29*(3), 742-751. https://doi.org/10.17559/TV-20201021141257

[17] Priyanka, G. & Amshakala, K. (2025). Predicting dementia through audio: Ensemble and deep learning approaches using acoustic features. *Computers in Biology and Medicine*, *197*, 111078. https://doi.org/10.1016/j.compbiomed.2025.111078

[18] Xia, Y., Lu, W., Peng, Z. et al. (2025). Understanding the impacts of space design on local outdoor thermal comfort: An approach combining DepthmapX and XGBoost. *Energy and Buildings*, 115451. https://doi.org/10.1016/j.enbuild.2025.115451

[19] Sun, S., Liu, Y., Li, Q. et al. (2023). Short-term multi-step wind power forecasting based on spatio-temporal correlations and transformer neural networks. *Energy Conversion and Management*, *283*, 116916. https://doi.org/10.1016/j.enconman.2023.116916

[20] Lin, C., Xu, J., Jiang, D. et al. (2025). Multi-model ensemble learning for battery state-of-health estimation: Recent advances and perspectives. *Journal of Energy Chemistry*, *100*, 739-759. https://doi.org/10.1016/j.jechem.2024.09.021

[21] Xiong, Z. & Jia, G. (2025). Data-driven global sensitivity analysis for group of random variables through knowledge-enhanced machine learning with normalizing flows. *Reliability Engineering & System Safety*, *260*, 111007. https://doi.org/10.1016/j.ress.2025.111007

[22] Li, D., Chen, J., Wang, Z. et al. (2023). LSTM deep neural network-based power data credit tagging technology. *Tehnički vjesnik*, *30*(1), 324-334. https://doi.org/10.17559/TV-20220628151845

[23] Ni, Q., Ji, J. C., Feng, K. et al. (2024). Data-driven bearing health management using a novel multi-scale fused feature and gated recurrent unit. *Reliability Engineering & System Safety*, *242*, 109753. https://doi.org/10.1016/j.ress.2023.109753

[24] Kerta, J. M., Rangkuti, A. H., & Tantio, J. (2024). Comparison of deep convolutional neural network architectures for fruit categorization. *International Journal of Electrical and Computer Engineering Systems*, *15*(3), 247-259. https://doi.org/10.32985/ijeces.15.3.4

[25] Toharudin, T., Pontoh, R. S., Caraka, R. E. et al. (2023). Employing long short-term memory and Facebook Prophet model in air temperature forecasting. *Communications in Statistics - Simulation and Computation, 52*(2), 279-290. https://doi.org/10.1080/03610918.2020.1854302

[26] Liu, J., Yu, Z., Zuo, H. et al. (2022). Multi-stage residual life prediction of aero-engine based on real-time clustering and combined prediction model. *Reliability Engineering & System Safety*, *225*, 108624. https://doi.org/10.1016/j.ress.2022.108624

[27] Motiejauskas, M. & Dzemyda, G. (2024). EfficientNet convolutional neural network with Gram matrices modules for predicting sadness emotion. *International Journal of Computers Communications & Control*, *19*(5). https://doi.org/10.15837/ijccc.2024.5.6697

[28] Kim, H. & Moon, K. S. (2024). Optimal data construction in supervised machine learning for financial prediction. *Economic Computation & Economic Cybernetics Studies & Research, 58*(1), 1-20. https://doi.org/10.24818/18423264/58.1.24.01

[29] Huo, H., Wang, H. B., & Zhang, D. D. (2021). Production management and control based on ant colony optimization and neural network. *International Journal of Simulation Modelling*, *20*(1), 158-169. https://doi.org/10.2507/IJSIMM20-1-CO1

[30] Hu, Y. (2020). Constructing grey prediction models using grey relational analysis and neural networks for magnesium material demand forecasting. *Applied Soft Computing*, *93*, 106398. https://doi.org/10.1016/j.asoc.2020.106398

[31] Wu, Z., Huang, M., Xing, Z. et al. (2024). Improving short-term traffic flow prediction using grey relational analysis for data filtering and stacked LSTM modeling. *International Journal of Computers Communications & Control*, *19*(1). https://doi.org/10.15837/ijccc.2024.1.6149

Contact information:

Yingli HOU
School of Economics and Management,
Beijing Jiaotong University, Haidian, 100044, China
E-mail: 23120618@bjtu.edu.cn

Hao HUA
China Energy Railway Equipment CO., LTD, 100009, China
E-mail: 23078262@qq.com

Ping GAO
China Energy Railway Equipment CO., LTD, 100009, China
E-mail: 298862312@qq.com

Qiuli QIN
(Corresponding author)
School of Economics and Management,
Beijing Jiaotong University, Haidian, 100044, China
E-mail: qlqin@bjtu.edu.cn

Appendix A

Table 1 Dimensionless processed data (data samples)

Month/Sequence	X_0 (k)	X_1 (k)	X_2 (k)	…	X_9 (k)	X_{10} (k)	X_{11} (k)
2021-01	–0.4883	–1.3673	–1.2001	…	–0.8779	–2.1092	–1.6963
2021-02	–0.5360	0.2213	–1.0960	…	–0.3151	–1.9225	–1.6516
2021-03	0.2027	1.3949	–0.7689	…	0.9324	0.9705	–1.7758
…	…	…	…	…	…	…	…
2023-04	–0.9887	0.0353	–1.5747	…	–0.5557	0.1306	0.1270
2023-05	–1.1079	–0.2981	–1.3227	…	1.1856	1.0638	0.0484
2023-06	–1.0602	1.4340	–1.5102	…	–0.6567	1.0638	–0.0410

Table 2 Sequence absolute differences (data samples)

Month/Sequence	X_1 (k)	X_2 (k)	X_3 (k)	X_4 (k)	…	X_8 (k)	X_9 (k)	X_{10} (k)	X_{11} (k)
2021-01	0.8790	0.7118	1.0318	1.6945	…	0.4200	0.3895	1.6209	1.2080
2021-02	0.7573	0.5600	0.9965	0.9035	…	0.7018	0.2209	1.3866	1.1156
2021-03	1.1922	0.9716	0.3145	0.8661	…	0.6680	0.7297	0.7678	1.9785
2021-04	0.1339	0.5193	0.3762	0.4984	…	0.1120	0.0724	0.3828	1.7377
2021-05	1.6924	0.6119	0.5960	1.2863	…	1.1834	1.2725	0.1364	1.4988
2021-06	1.5433	0.7614	0.0517	1.4397	…	0.4502	1.7041	1.7401	1.9039
…	…	…	…		…	…	…	…	…

Table 2 Sequence absolute differences (data samples) (continuation)

Month/Sequence	$X_1(k)$	$X_2(k)$	$X_3(k)$	$X_4(k)$	…	$X_8(k)$	$X_9(k)$	$X_{10}(k)$	$X_{11}(k)$
2022-12	0.2761	1.6620	1.9219	0.0028	…	0.5122	0.1627	0.2341	0.8181
2023-01	0.4660	0.1657	0.9576	1.8399	…	0.6447	2.5355	1.2861	2.4066
2023-02	1.9658	0.3063	0.1608	1.5445	…	1.9943	0.5771	2.0526	1.7901
2023-03	1.8763	0.5043	0.7535	1.9345	…	2.0445	1.0035	2.3861	1.7527
2023-04	1.0240	0.5860	1.2476	0.2457	…	1.7549	0.4330	1.1193	1.1157
2023-05	0.8098	0.2148	1.2555	1.4126	…	1.6271	2.2935	2.1717	1.1562
2023-06	2.4942	0.4500	1.7391	1.4277	…	2.0181	0.4035	2.1240	1.0192

Table 3 Gray correlation coefficient results (data samples)

Month/Sequence	$X_1(k)$	$X_2(k)$	$X_3(k)$	$X_4(k)$	…	$X_8(k)$	$X_9(k)$	$X_{10}(k)$	$X_{11}(k)$
2021-01	0.6928	0.7359	0.6576	0.5388	…	0.8257	0.8363	0.5498	0.6211
2021-02	0.7237	0.7800	0.6654	0.6869	…	0.7387	0.9006	0.5881	0.6397
2021-03	0.6242	0.6710	0.8638	0.6960	…	0.7481	0.7311	0.7209	0.5000
2021-04	0.9378	0.7928	0.8411	0.7995	…	0.9476	0.9660	0.8387	0.5325
2021-05	0.5391	0.7644	0.7691	0.6062	…	0.6260	0.6088	0.9367	0.5691
2021-06	0.5619	0.7226	0.9758	0.5790	…	0.8154	0.5373	0.5321	0.5097
…	…	…	…		…	…	…	…	…
2022-12	0.8785	0.5436	0.5073	1.0000	…	0.7950	0.9251	0.8952	0.7079
2023-01	0.8101	0.9238	0.6742	0.5182	…	0.7548	0.4383	0.6063	0.4512
2023-02	0.5016	0.8669	0.9260	0.5617	…	0.4980	0.7748	0.4908	0.5251
2023-03	0.5133	0.7976	0.7247	0.5057	…	0.4918	0.6638	0.4533	0.5303
2023-04	0.6593	0.7721	0.6135	0.8905	…	0.5300	0.8212	0.6390	0.6397
2023-05	0.7100	0.9031	0.6120	0.5836	…	0.5488	0.4631	0.4767	0.6314
2023-06	0.4423	0.8154	0.5323	0.5810	…	0.4951	0.8314	0.4823	0.6603

Table 4 Railroad wagon track record data

Field Name	Explanation	Data Type	Data Sample
WAGON_NO	Train set code	INT(10)	8039
SUB_WAGON_ NO	Vehicle code	VARCHAR(50)	0039272
TYPE	Vehicle model	VARCHAR(50)	C80
MILEAGE	Total mileage	FLOAT(20, 2)	1589829.80
HEALTH_SCORE	Vehicle health rating	FLOAT(10, 2)	91.75
FAC_DATE	Factory overhaul date	DATE	2014-05-13
SEC _DATE	Segment overhaul date	DATE	2018-02-06
Z1_DATE	Z1 overhaul date	DATE	2023-05-06
Z2_DATE	Z2 overhaul date	DATE	2020-06-13
Z3_DATE	Z3 overhaul date	DATE	2021-08-10
Z4_DATE	Z4 overhaul date	DATE	2022-07-27
RECORD_DATE	Data recording time	DATE	2022-08-30

Table 5 Shift consumption data

Field Name	Explanation	Data Type	Data Sample
MATERIAL_NO	Material code	INT(10)	10105112
MATERIAL_TYPE	Specification and model	VARCHAR(50)	K2\QCZ85A-71A-00
MATERIAL_ NAME	number	VARCHAR(50)	qianyingan
UNIT	Material Name	VARCHAR(50)	PC
TIME	Unit Month	DATE	2021-01
VOLUME	Consumption rate	FLOAT(10, 2)	18.00

Table 6 Forecast schedule of maintenance tasks data

Field Name	Explanation	Data Type	Data Sample
PLAN_NO	Plan code	INT(10)	0011202
MAINTENANCE_TYPE	Plan maintenance mode	VARCHAR(50)	Z4 repair
MAINTENANCE_TIME	Plan maintenance time	DATE	2022-03
MAINTENANCE_	Plan maintenance unit	VARCHAR(50)	Cangzhou branch
WAGON_NO	Train formation code	INT(10)	8039
SUB_WAGON_ NO	Vehicle code	VARCHAR(50)	0039272
WAGON_TYPE	Vehicle model	VARCHAR(50)	C80

Appendix B Implementation Details and Reproducibility of the Model

To ensure the reproducibility of this research, the key parameters, data preprocessing, and implementation details of the model are described as follows:

B.1 Ratio Test and Shift Processing

The coverage interval for the ratio test can be expressed as: https://latex.codecogs.com/svg.latex?%255Cleft%2520%2528%2520e%255E%257B-%255Cfrac%257B2%257D%257Bn+1%257D%257D%252C%2520e%255E%257B%255Cfrac%257B2%257D%257Bn+1%257D%257D%2520%255Cright%2520%2529, where n is the sequence length.

If the original sequence https://latex.codecogs.com/svg.latex?%255Cboldsymbol%257BX%257D%255E%257B(0)%257D fails the ratio test, then a shift transformation is performed: https://latex.codecogs.com/svg.latex?%255Cboldsymbol%257BY%257D%255E%257B(0)%257D%2520%253D%2520%255Cboldsymbol%257BX%257D%255E%257B(0)%257D%2520+%2520C, where the shift constant https://latex.codecogs.com/svg.latex?C%2520%253D%2520%25

7Cmin%2528%255Cboldsymbol%257BX%257D%255E%257B(0)%257D%2529%257C%2520+%25201.

B.2 Grey Neural Network Structure and Training

Network Structure: Input layer with 5 nodes → Grey calculation layer (implementing AGO and GM(1, 1) calculations) → Hidden layer (6 nodes, activation function: tansig) → Output layer (1 node, activation function: purelin).
Training Configuration: Use the Levenberg-Marquardt optimization algorithm. Maximum number of iterations 1000, learning rate 0.001, target error $1e^{-5}$. Enable early stopping (Early Stopping) strategy. Training is terminated when the performance of the validation set does not improve for 50 consecutive iterations.
Data Division: Strictly divided by time sequence. The first 80% of the time period data is used for training, and the remaining 20% is used for testing.
Random Seed: To ensure reproducibility of the results, use rng(2024) in MATLAB to fix the random seed.

B.3 Key Parameters of Comparative Models

XGBoost: max_depth = 5, learning_rate = 0.1, n_estimators = 100, subsample = 0.8.
LSTM: The number of single-layer LSTM units is 10, followed by a fully connected layer, using the Adam optimizer.
Prophet: Default parameters, only enabling annual seasonality.
ARIMA: The (p, d, q) order is automatically determined through the AIC criterion.

B.4 Code Framework

The core analysis of this study is implemented based on MATLAB R2021a. The grey correlation analysis is encapsulated in the function GCA_mode.m, and its main input and output interfaces are as follows: matlab

```
function [GCD, GCC] = GCA_mode(data, rho)
% Input: data - an n*m matrix, the first column is the reference sequence, and the rest are comparison sequences
%       rho - resolution coefficient, default value is 0.5
% Output: GCD - the grey correlation degree of each comparison sequence (1*(m – 1))
%       GCC - the grey correlation coefficient of each point (n*(m – 1)) ...
end
```

The code for building the grey neural network model can be requested from the author.