

Forecasting carbon price in Hubei Province using a mixed neural model based on mutual information and Multi-head Self-Attention

Youyang Ren^a, Yiyuan Huang^a, Yuhong Wang^{a,*}, Lin Xia^a, Dongdong Wu^b

^a School of Business, Jiangnan University, Wuxi, Jiangsu Province, 214122, PR China

^b College of Tourism and Service Management, Nankai University, Tianjin, 300350, PR China

ARTICLE INFO

Handling Editor: Xin Tong

Keywords:

Mutual information regression

Multi-head self-attention

LSTM

Carbon price forecasting

ABSTRACT

Accurate carbon price forecasting holds significant practical importance for the Chinese government in formulating region-specific carbon reduction policies, balancing supply, and achieving regional green and low-carbon development. Considering carbon price data's uncertainty and nonlinear characteristics, this paper constructs a mixed artificial intelligence model to forecast Hubei's carbon price development. This paper proposes a novel decomposition and reconstruction method by integrating mutual information regression with mode decomposition. It integrates multiple deep learning modules to form the main forecasting framework, incorporating Multi-head Self-Attention with the mixed neural network. Bayesian optimization is applied to determine the modeling hyperparameters. The result reflects that the test group's Mean Absolute Percentage Error, Mean Absolute Error, Mean Squared Error, and R-squared are 1.3176%, 0.5889, 0.7455, and 0.9256, surpassing all the comparison models. Compared with the baseline model, the Diebold-Mariano value is 3.7503, and the improvement ratio is 34.7118%, 44.7655%, 33.1303%, and 6.9810%, which reveals that the model's forecasting performance and generalization ability improve significantly. This paper forecasts that Hubei's carbon price may initially decrease, rise considerably in the mid-term, and gradually stabilize over the next 180 days. The highest carbon price may reach 50.09 Yuan/tCO₂e, and the lowest may reach 39.98 Yuan/tCO₂e. The findings indicate that the Hubei government may set the price corridor for the Hubei carbon market based on the forecasting value, flexibly formulate dynamic carbon reduction policies, and optimize market mechanisms to gradually ensure that the supply and demand of the carbon market reach a new balance.

1. Introduction

Since the 21st century, the greenhouse gas emissions generated by human industry and production activities have increased, and the global greenhouse effect has intensified. As a result, the human living environment faces the threat and challenge of global warming and extreme climate. For the global climate problem, all countries have taken the initiative to control the excessive emission of greenhouse gases, led by CO₂. Regarding technical governance, governments increase clean energy consumption and reduce the dependence on fossil energy for industrial development and economic construction within the country. They also continue developing innovative technologies to reduce carbon emissions through energy efficiency, Carbon capture (Soeppan et al., 2024), and Carbon storage (Yang et al., 2023a). From the economic perspective, with the launch of the Kyoto Protocol, the European Union

Emission Trading Scheme, and the California Cap-and-Trade Program, the carbon emission market system led by carbon price has gradually formed a mature carbon emission trading mechanism. The Paris Agreement recognizes carbon price as one of the essential governance modalities to achieve control over the rise in global temperature (Zhang et al., 2023b).

As the cost price of CO₂ or other equivalent greenhouse gas emissions, the carbon price is related to the stability of the market supply and demand mechanism under carbon emission allowance trading. It is a core reference and economic signal for relevant enterprises considering purchasing carbon emission allowances or developing carbon emission reduction technologies. Nowadays, the European Union, the United States, and China have gradually implemented the carbon emission trading system and strive to promote global coordinated emission reduction and fulfill emission reduction commitments. Considering that

* Corresponding author.

E-mail addresses: 7220912008@stu.jiangnan.edu.cn (Y. Ren), 6220910033@stu.jiangnan.edu.cn (Y. Huang), wyh2003@gmail.com (Y. Wang), 7220912010@stu.jiangnan.edu.cn (L. Xia), dww@mail.nankai.edu.cn (D. Wu).

<https://doi.org/10.1016/j.jclepro.2025.144960>

Received 16 September 2024; Received in revised form 11 January 2025; Accepted 5 February 2025

Available online 8 February 2025

0959-6526/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

the Chinese economy is a major carbon emitter, the Chinese government has established seven carbon emission trading markets in different provinces to implement carbon emissions trading schemes, promoting comprehensive low-carbon transformation of regional markets. This conduct aims to encourage local enterprises to take the lead in participating in green development investment and realize carbon emission reduction through market control. In the Emissions Trading System (ETS), market participants must purchase carbon allowances or permits based on carbon prices to satisfy their emission demand. Under this mechanism, the carbon price is an incentive or constraint on companies' decisions regarding energy conservation, environmental protection, and carbon emissions (Fleschutz et al., 2021). Cleaner production is a vital link between sustainable growth and energy-saving and emission-reduction goals. It is a green development mode that reduces resource waste, lowers pollutant emissions, and enhances energy efficiency. Usually, carbon price development guides enterprises to make reasonable investments and decisions according to low-carbon development requirements (Bompard et al., 2022). Reasonable adjustments to carbon prices could encourage enterprises to adopt cleaner and more environmentally friendly production methods. By investing in energy-saving and cleaner production, enterprises could actively utilize clean energy and develop emission-reduction technologies. This could help them reduce their expenditures on carbon allowances, thereby supporting global efforts to reduce carbon emissions.

Hubei Province is an important economic and industrial town in central China and a vital hub connecting the eastern coast and the western interior. Its carbon emission reduction plays an essential role in China's green transformation to achieve the dual carbon commitment. In addition, as one of the seven pilot carbon markets in China, the change in carbon price of the China Hubei Carbon Emission Exchange can reflect the rationality and effectiveness of carbon emission reduction policies and market mechanisms in the regions. Forecasting and researching its carbon price development is of great practical significance for formulating more accurate environmental policies and optimizing China's regional carbon trading mechanisms. Considering the carbon price is an essential reference and evaluation index for government market supervision and enterprise market investment, both government and enterprise decision-makers urgently need to consider the development of carbon prices from a forward-looking perspective. Therefore, a scientific carbon price forecasting system is significant for the government and enterprises to avoid financial risks and formulate carbon market control policies reasonably. However, it is difficult for the traditional forecasting model to capture the overall change trend of the carbon price. This paper proposes a mixed model for carbon price forecasting based on deep learning.

Regarding innovation and contribution, the proposed model incorporates a novel modal decomposition and reconstruction module that combines the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) with mutual information (MI) regression. The novel module reconstructs the Intrinsic Mode Functions (IMF) by correlating MI interval distributions, providing input features for the subsequent forecasting module. In this paper, the LSTM is the baseline model. It further integrates CNN and Transformer to enhance the functionality and robustness of the deep learning module, thereby establishing a mixed forecasting framework that simultaneously considers local patterns, long-term dependencies, and global features. This paper enhances the training efficiency and accuracy of carbon price forecasting in Hubei Province by integrating advanced decomposition and denoising techniques with artificial intelligence forecasting models and optimization algorithms. As a result, it introduces a novel approach to handling dynamic changes in complex nonlinear data and provides an innovative analytical tool for carbon price forecasting.

The following constitutes the literature review as the second section, the methodology as the third section, and the empirical analysis as the fourth section. The fifth section is the conclusion.

2. Literature review

Different from the traditional time series, the carbon price development is a unique and complex time series affected by different qualitative and quantitative factors, which have prominent nonlinear, fluctuating, and uncertain characteristics (Wang et al., 2022a). In terms of the modeling method, scholars have conducted relevant research on time series forecasting and gradually formed three main forecasting modeling methods. They are the statistical models and their derivative models, the artificial intelligence forecasting models (Gao and Shao, 2022), and the mixed forecasting models based on deep learning (Liu et al., 2022).

2.1. Research on the statistical model

The statistical models conduct the forecasting process through mathematical relationships estimating based on historical data. They combine with regression models or stochastic equations to express the characteristics of the carbon price. Under the premise of linear assumption, the statistical models have ideal forecasting performance and reasonable interpretability. Wang et al. (2020) constructed three combined statistical models based on ARIMA to forecast the carbon emission trend in China, the United States, and India. Qin et al. (2022) combined Hodrick-Prescott and ARIMA to deal with the structural characteristics and trend reconstruction of carbon price forecasting. Liu and Huang (2021) applied GARCH to optimize the fluctuation defects of fractional Brownian motion and improved the referability of forecasting results. Alkathery and Chaudhuri (2021) proposed three multi-form GARCH for continued carbon market volatility and verified the diagonal BEKK GARCH's validity for carbon price forecasting. However, the statistical models are established based on regression equations and statistical law. These models' forecasting performances are inaccurate when facing high-complexity samples. Therefore, scholars mainly apply classical statistical models to assist machine learning or deep learning in combinatorial modeling (Zhang and Wu, 2022). Some studies have verified that these combination models could reduce modeling complexity and enhance the interpretability of black-box models (Chen et al., 2022).

2.2. Research on the artificial intelligence model

With the continuous development of data technology, scholars gradually favor artificial intelligence forecasting models dominated by data-driven technology (Hong et al., 2024). They optimize the multi-variate configuration of neural networks by encompassing training techniques, hidden neurons, delays, and data segmentation strategies to capture better the nonlinear characteristics of time series than statistical models (Jin and Xu, 2024a). The artificial intelligence models have acceptable training and learning ability for nonlinear time series, resulting in better robustness and forecasting accuracy (Ye et al., 2024). They also could process vast amounts of data and break the sample size limit of the statistical models. The research of artificial intelligence forecasting models focuses on practical applications and the continuous optimization of performance. Li et al. (2020) selected the BPNN model to simulate the carbon price trend under the carbon trading market in six scenarios. Salinas et al. (2020) proposed the DeepAR, a novel probabilistic forecasting model leveraging autoregressive RNN, and verified its superiority based on five datasets. Zhang et al. (2023a) combined cosine function and whale optimization algorithm to design a novel optimized extreme learning machine. Their results proved that the optimization algorithm could improve the stability by optimizing the structure of traditional machine learning methods. Van Belle et al. (2023) introduced a composite loss function that considers the forecasting accuracy and stability to extend the N-BEATS architecture. Qin et al. (2024) forecasted the carbon price in the Guangdong and Hubei carbon markets with the bidirectional extended LSTM. Jin and Xu (2024b) applied

Bayesian Optimization (BO) and cross-validation over various kernels and basic functions to conduct the complex commodity price forecasting problem. Although artificial intelligence forecasting technology is maturing, it still has some limitations. The complex mapping between the layers leads to difficulty in optimizing node weight and structure parameters. Faced with complex nonlinear samples, the single artificial intelligence models may carry out excessively complex training and calculation operations with high time and space complexity. They tend to fall into local optimality, resulting in overfitting phenomenon. Most of these models lack explainability, and it is not easy to reasonably characterize the dynamic information.

2.3. Research on the mixed model

The single artificial intelligence model still has limitations in processing the time series data with high complexity and nonlinear characteristics. Scholars gradually recognize the mixed forecasting model to fully explore the dynamic characteristics and hidden rules (Mao and Yu, 2024). The main idea of these mixed models is to decompose and reconstruct the complex data based on the principle of divide and rule (Wang et al., 2022b). Empirical mode decomposition (EMD) is the primary decomposition strategy at present. The decomposition strategy could decompose any nonlinear data into IMF and stabilize the nonlinear data effectively. The improved EMD derivative algorithms, including Variational Mode Decomposition (VMD), Ensemble EMD (EEMD), Complete EEMD (CEEMD), and CEEMDAN, are optimized to deal with the problem of mode mixing. Scholars usually combine multiple sets of measured values of different scales to form final forecasting results (Zhao et al., 2024). By contrast, the mechanism of the mixed forecasting model considers the sensitivity of data preprocessing and reduces the computational complexity of the forecasting process.

Existing research has also gradually verified that these mixed models have better forecasting performance than single models and provide convincing results (Li et al., 2023). Sun et al. (2021) proposed a forecasting framework based on the selection and matching strategy containing data preprocessing and VMD. Wang et al. (2023) proposed a novel multi-objective forecasting model to enhance the convergence speed and forecasting accuracy, which significantly increases the forecasting model's generalization. Li and Liu (2023) proposed a mixed forecasting model of multiple data set decomposition and multiple feature screening for carbon prices. They modified and optimized the model's forecasting accuracy by combining historical data with relevant factors. Yang et al. (2023b) applied multiscale fuzzy entropy to optimize the decomposition method in studying carbon price forecasting in China. They adopted the secondary data decomposition strategy to form complexity components for forecasting and reconstruction. Liu et al. (2024) applied the secondary decomposition with the sparrow search algorithm to enhance carbon price forecasting performance. Hao et al. (2024) established a mixed forecasting framework based on the peak and trough values combined with the optimal solutions of five machine learning methods with multi-objective optimization.

Meanwhile, the mixed model architecture of CEEMDAN and Transformer has gradually become a hotspot for mixed modeling (Chen et al., 2024). Wang et al. (2024) proposed a mixed model to forecast wind power with high and low frequency. Their mixed framework includes the CEEMDAN, sample entropy, Transformer, and BiGRU-Attention modules. The CEEMDAN and sample entropy divide the wind power sequence into high and low-frequency sequences, while Transformer and BiGRU-Attention forecast the two sequences. Zhou and Zhang (2024) applied CEEMDAN to decompose the battery capacity degradation data and further introduced VMD to reduce the remaining noise. They combined the CNN with the Transformer to forecast the component sequences and form the lithium-ion batteries' remaining useful life forecasting result. Yin and Zhou (2024) proposed a novel modal decomposition integrated model that combined CEEMDAN with other deep learning models for the reheater tube temperature. They designed

the CEEMDAN and Transformer architecture to conduct the moderate complexity component. Yao et al. (2024) proposed a mixed model for total phosphorus forecasting. Their mixed model framework comprises CEEMDAN, fuzzy entropy, LSTM, and Transformer. They applied CEEMDAN to decompose water quality data, reconstructed them as trend-term components by fuzzy entropy, and integrated LSTM and Transformer to extract sequence features and improve the model's forecasting performance. These researches reasonably combine CEEMDAN's multi-frequency decomposition and Transformer's Multi-head Self-Attention mechanism to extract multi-scale features and capture critical patterns and trends in the sequences from a global scope, improving the final forecasting effect.

However, mixed models have complex structures, which increases the calculation and modeling complexity while improving the forecasting accuracy. As the complexity of the hybrid model increases, so does the demand for computational time and resources, which complicates the handling of higher-dimensional data series. These models must rely on more extensive data samples to maintain the overall performance and generalization ability. The hyperparameters of the sub-models and the hybrid models should be optimally tuned within the context of the overall forecasting framework. It inevitably contains multiple sub-models and layers, leading to the risk of overfitting. Table 1 provides a detailed summary of the literature reviewed above.

2.4. Innovations and contributions

Recent research increasingly highlights the application of mixed models that combine decomposition techniques and hybrid forecasting frameworks in complex nonlinear time series forecasting. These mixed models represent a modeling strategy that divides and conducts the nonlinear time series. After decomposing the target time series, the mixed models typically employ entropy to reconstruct IMFs or further decomposition and denoising for high-frequency data. In the forecasting module, the mixed models could apply different individual statistical and artificial intelligence models or mixed forecasting frameworks to forecast different components of varying complexities and fuse them to form the final forecasting results. However, these mixed models also have certain limitations and gaps.

Reconstruct IMFs by sample entropy, fuzzy entropy, or other entropy methods for different frequency sequences after decomposition may incur information loss due to the simplification of information and discrepancies in the time scale selection. Although multiple decompositions could reduce noise in high-frequency sequences, the decomposition process may also result in new information loss. Both modeling ideals carry the risk of introducing unnecessary noise and generating redundant complex computations. Additionally, employing individual forecasting models to handle different decomposed sequences may weaken the model's training efficiency and generalization ability in the forecasting module. While a mixed framework with multiple modules enhances performance, the model's overall complexity also increases, diminishing its interpretability. Without optimization algorithms, the workload and difficulty of parameter adjustments for the individual forecasting models or mixed forecasting framework may become more complex. Therefore, the specific problems are primarily these aspects. One aspect is developing a sequence decomposition and reconstruction module that decreases information loss while reducing noise and computational complexity. Another aspect is effectively integrating multiple deep learning techniques into a mixed forecasting module, ensuring improved forecasting accuracy without introducing excessive noise or sacrificing efficiency. Finally, this paper constructs a novel carbon price forecasting model using the optimization algorithm to combine these modules. The novelty and contributions of the proposed model are summarized as follows.

Considering the high nonlinearity and dynamic variability of the carbon price time series, this paper applies CEEMDAN to decompose the original series. Then, it introduces MI and determines their interval

Table 1

Summary table of the related studies.

Category	Models	Refs	Advantage	Disadvantage
Statistical model	ARIMA	Wang et al. (2020)	Fit forecasting in the short or medium term.	Conduct the nonlinear features hard.
	HP-ARIMA	Qin et al. (2022)	The complexity is lower than that of artificial intelligence models or mixed models.	The modeling relies on strict model assumptions.
	GARCH	Liu and Huang (2021)	The modeling and training process is convenient and interpretable.	Have no ability to identify the extreme fluctuations.
	Multi-form GARCH	Alkathery and Chaudhuri (2021)		
Artificial intelligence model	BPNN	Li et al. (2020)	Fit forecasting in the global term.	Have a high demand for large data sample sizes.
	DeepAR	Salinas et al. (2020)	Easy to capture complex nonlinear features.	Easy to trap into local optimal or overfitting.
	ELM	Zhang et al. (2023b)	Effectively handle forecasting studies with big data.	Many parameters and parameter adjustments are complex.
	N-BEATS	Van Belle et al. (2023)	Less constrained by data non-stationarity.	
Mixed model	LSTM	Qin et al. (2024)		
	GPR	Jin and Xu (2024b)		
	HPG-VMD-RE-BPNN	Sun et al. (2021)	Fit forecasting time series with high complexity and prominent nonlinear characteristics.	Have high time complexity and space complexity.
	CEEMDAN-SE-(BPNN/ELM/ENN/ET2QFNN/LSTM)	Wang et al. (2023)	Combine the advantages of multiple models to capture data characteristics comprehensively.	The forecasting accuracy depends on the performance of the internal component models.
	ICEEMDAN-DWT-(SVR/MLP)	Li and Liu (2023)	Strong multi-scale information processing capability.	The parameter adjustments depend on the optimization algorithm.
	ICEEMDAN-MFE-CEEMD-(SSARF/CSBP/WOAEIM)-ELM	Yang et al. (2023b)	The mixed model modeling adopts the idea of decomposition, forecasting, and integration, which improves the explainability.	
	SSA-BEMD-MPE-MVND-IMLP	Liu et al. (2024)	Robustness and stability are higher than that of single artificial intelligence models.	
	MMVMD-(BPNN/ELM/ENN/BiLSTM/CNN)-MZOA-LSTM	Hao et al. (2024)		
	CEEMDAN-SE-(Transformer/BiGRU-Attention)	Wang et al. (2024)		
	CEEMDAN-VMD-(CNN-Transformer/BO-Transformer)	Zhou and Zhang (2024)		
	CEEMDAN-SE-(ResNet-CBAM/Transformer/GRU/TCN)-MLP	Yin and Zhou (2024)		
	CEEMDAN-FE-LSTM-Transformer	Yao et al. (2024)		

distribution to reconstruct new IMFs. In the forecasting module, this paper constructs the mixed deep learning framework based on the modeling ideal that extracts local patterns, captures long-term dependencies, and learns global features, progressing from local to global levels. The framework integrates CNN, LSTM, and Transformer to extract and forecast the trend of carbon price fluctuations at different levels. During the forecasting process, the Multi-head Self-Attention mechanism from the Transformer dynamically weighs critical information. It enhances the frequency of its contribution to the model parameters based on the importance of the information. This paper employs BO to optimize the global hyperparameters of the entire mixed framework, enhancing the model's adaptability and forecasting performance for carbon price. In empirical research, this paper applies the proposed mixed model to forecast and evaluate the future carbon price development trend in Hubei province. It provides target reference and data support for the formulation of the Hubei government's dynamic carbon market regulation policy.

Unlike prominent time-series models such as DeepAR, N-BEATS, and Temporal Fusion Transformer (TFT) (Du et al., 2024), the model proposed in this paper constructs a novel module combining CEEMDAN and MI before the forecasting module. This module provides better-processed input sequences to the forecasting framework by denoising and reconstructing IMFs. After reconstruction, each IMF represents a characteristic component of the original data with a specific frequency characteristic. According to their basic components, DeepAR relies on probability and RNN. N-BEATS constructs its forecasting modules based on the Feedforward Neural Network. TFT employs a predictive architecture formed by LSTM and Transformer. In contrast, the mixed framework proposed in this paper further synthesizes the advantages of CNN, LSTM, and Transformer modules, resulting in a more comprehensive architecture that more effectively handles the local

and global features of complex nonlinear data. It presents a more enriched forecasting architecture offering greater functionality, adaptability, and flexibility in the forecasting process compared to the models above.

Compared to the architectures integrating CEEMDAN and Transformer, the proposed model replaces the entropy-based reconstruction of IMFs or multiple sequence decompositions for denoising applied in previous studies with MI regression. This improvement retains the target sequence's dynamic information and extracts more crucial components for future forecasting. It could reduce the adverse effects of noise and irrelevant factors on forecasting performance and simplify modal sequences, reducing the complexity of subsequent computations and optimizing operational efficiency. The proposed model incorporates CNN and LSTM to process the sequence's local and long-term data features. Compared to a single Transformer or an overly complex combination of modules in deep learning frameworks, the proposed model maintains forecasting performance while optimizing the complexity and computational cost associated with excessive module integration.

3. Methodology

The carbon price forecasting model proposed in this paper is a multi-module combination. The progress comprises decomposition and reconstruction considering MI, CNN-LSTM-Transformer construction with Multi-Head Self-Attention, and final carbon price forecasting. BO determines the hyperparameters according to the MSE as the optimization target. Fig. 1 shows the overall framework of the proposed mixed forecasting model.

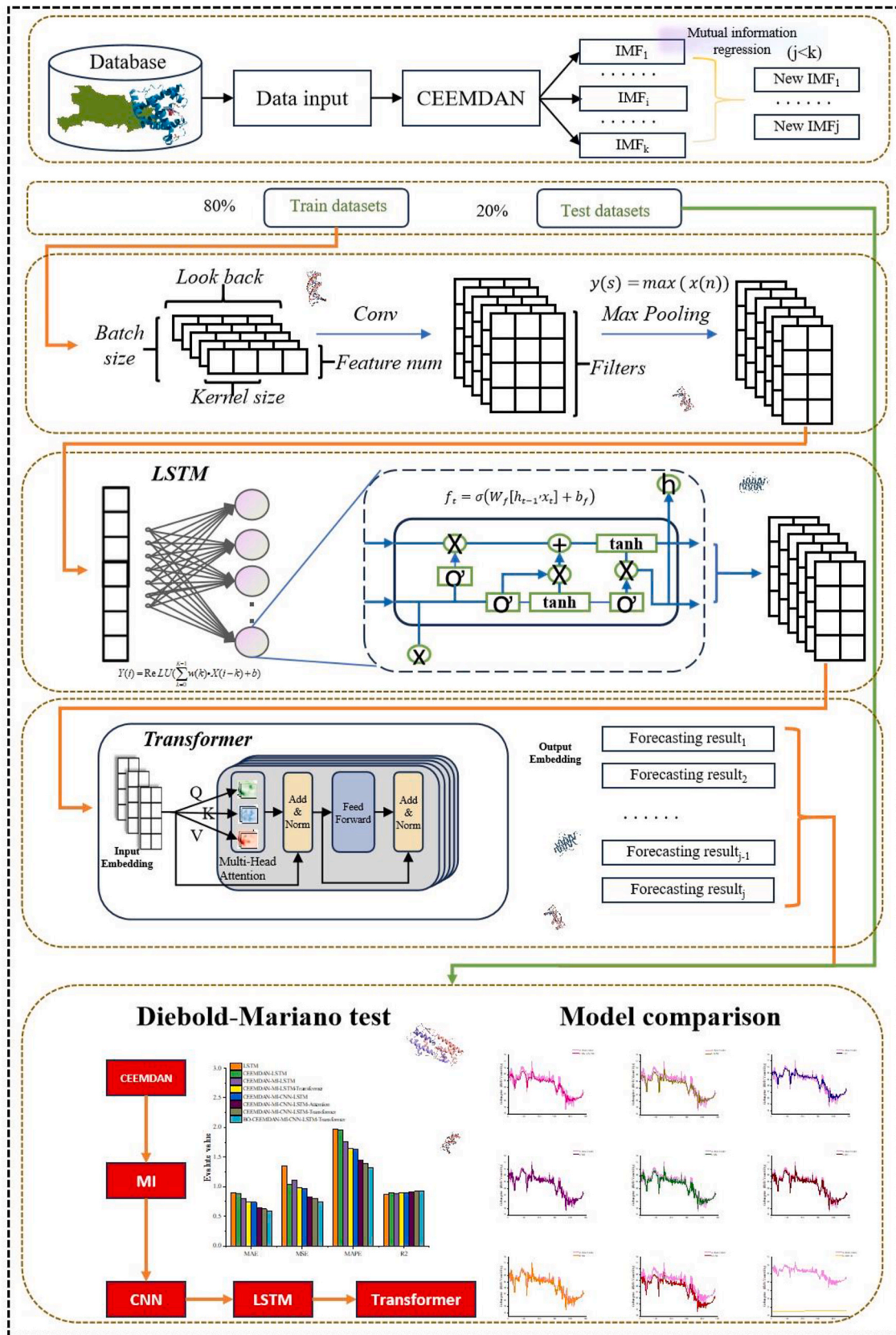


Fig. 1. The overall framework of the proposed mixed model.

3.1. The CEEMDAN based on mutual information regression

The CEEMDAN is an improved version of EEMD. It inserts adaptive white noise into each stage of nonlinear sequence decomposition, effectively controlling the noise level during each decomposition and avoiding redundant noise problems during reconstruction. This method makes the noise level proportional to the standard deviation of the residual signal, thus improving the accuracy and robustness of the decomposition. Considering the importance of the internal dynamic structure and sequence information of the actual nonlinear sequences, this section introduces MI regression to reconstruct the essential components of carbon price decomposition. This paper selects IMFs in the same MI value interval for fusion to construct new IMFs and provide input features for the subsequent forecasting sections, which are as follows.

Noise $n_1(t)$ obeying normal distribution is added to the carbon price sample dataset $A(t)$ to form a new sequence, $A_1(t)$, as shown in Eq. (1).

$$A_1(t) = A(t) + \sigma_0 \alpha n_1(t) \quad (1)$$

σ_0 is the standard deviation of $A(t)$, and α is the noise intensity control coefficient.

Perform EMD on the new sequence $A_1(t)$ to obtain the residual $r_1(t)$, as shown in Eq. (2).

$$r_1(t) = A(t) - \frac{1}{k} \sum_{i=1}^k IMF_{1,k}(t) \quad (2)$$

Take $r_1(t)$ as the target sequence $A_2(t)$, add normal distributed white noise, and proceed with EMD according to the above steps to get residual $r_2(t)$. The cycle is repeated until $A_i(t)$ becomes monotonous and cannot be decomposed to form new IMFs. Finally, the decomposed IMFs and residuals are obtained, as shown in Eq. (3).

$$r_i(t) = A_{i-1}(t) - \frac{1}{k} \sum_{k=1}^k IMF_{i-1,k}(t) \quad (3)$$

Based on all the decomposed IMFs, this section further calculates their MI values for $A(t)$ as Eq. (4).

$$\begin{cases} MI(X; Y) = H(X) + H(Y) - H(X, Y) \\ H(X) = -\sum_{x \in X} p(x) \log p(x) \\ H(Y) = -\sum_{y \in Y} p(y) \log p(y) \\ H(X, Y) = -\sum_{y \in Y} p(x, y) \log p(x, y) \end{cases} \quad (4)$$

Y represents $A(t)$, and X represents IMFs. $H(X)$ and $H(Y)$ are the entropy. $H(X, Y)$ is the joint entropy.

According to the MI value results of each IMF, this section sets the classification interval threshold. Suppose k IMFs with different frequency characteristics are distributed at the same interval. This paper constructs the new IMF as shown in Eq. (5).

$$NewIMF = \sum_{k=1}^k IMF_k \quad (5)$$

Through the improved CEEMDAN, this paper not only retains the internal dynamic structure of carbon price data but also refines the important components describing the change in carbon price. The reconstruction preserves the vital information that contributes the most to the uncertainty of carbon price and reduces the complexity of subsequent calculations.

3.2. The framework of CNN-LSTM-transformer

Based on the preliminary input features in Section 3.1, this part establishes the proposed mixed framework consisting of the CNN, LSTM, and Transformer to forecast the reconstructed IMFs. Fig. 2 shows the overall process in this section. The whole mixed framework construction must carefully consider the Number of Filters and Kernel Size in LSTM-unit and CNN and control the speed of updating model weights in each step of the optimization algorithm, batch size, and look-back steps in the training and testing process. In the Transformer module, the dimensions of the attention mechanism, Feedforward Network (FFN), and the number of Multi-heads must be determined.

(1) CNN

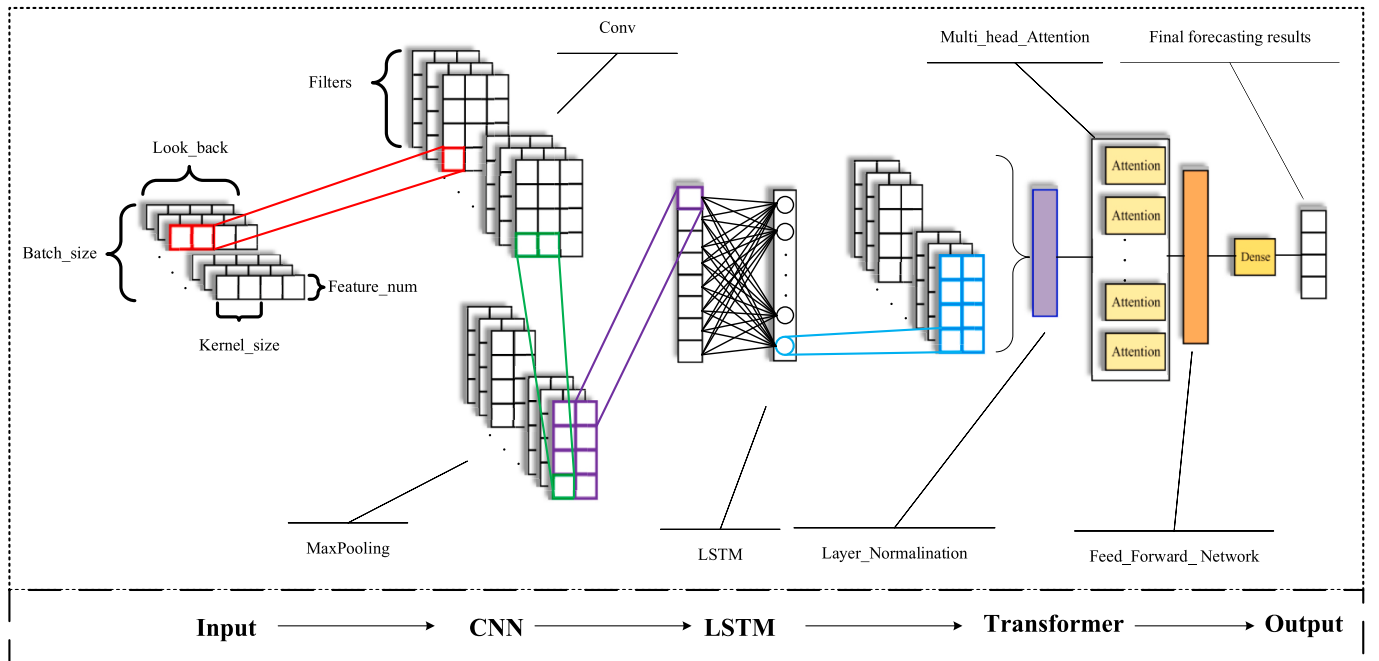


Fig. 2. The process of the proposed CNN-LSTM-Transformer.

In the proposed mixed forecasting framework, the CNN module mainly extracts carbon price data's local patterns and trend characteristics. It mainly includes the Input Layer, Hidden Layer, and Output Layer. The Hidden Layer comprises the Convolutional Layer, Pooling Layer, and Activation Layer. This paper applies the Convolutional Layer to perform convolution operations as Eq. (6).

$$Y(i) = \text{ReLU}\left(\sum_{k=0}^{K-1} w(k) \cdot X(i-k) + b\right) \quad (6)$$

X is the input, i is the time step, w is the convolutional kernel weight, and b is the bias term. In the Activation Layer, the activation function ReLU is applied for nonlinear mapping to improve the overall expression ability. It sets the input value from less than 0 to 0 to reduce the gradient disappearance during training. ReLU could be expressed as follows.

$$\text{ReLU}(x) = \max(0, x) \quad (7)$$

The Pooling Layer reduces the computational complexity by compressing the feature map and reducing the parameters. It retains the most significant feature values and extracts the key factors by the maximum Pooling Layer.

$$y(s) = \max(x(n)), n \in [sT, sT + F] \quad (8)$$

$y(s)$ is the element of the output sequence, T is the step of the pooling layer, and F is the size of the pooling window.

Finally, the Output Layer outputs the CNN feature sequence to provide feature input for subsequent LSTM and Transformer modules.

(2) LSTM

This section applies the LSTM module to the long-term information dependency in the time series data based on the CNN module output. It evaluates the information value of input sequences for long-term memory and selective forgetting of corresponding information by LSTM. The LSTM comprises the Forget Gate, Input Gate, Output Gate, and Cell State. The function of the Forget Gate is to determine the historical information that LSTM chooses to retain and other information that needs to be discarded, as shown in Eq. (9).

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (9)$$

f_t is the output vector between 0 and 1, indicating the degree to which the cell needs to be forgotten. σ represents the Sigmoid activation function, compressing the output value to 0 and 1. W_f is the weight matrix of the Forget Gate. $[h_{t-1}, x_t]$ is the connection vector between the state h_{t-1} of the previous time step and the input x_t of the current time step. b_f is the bias vector.

The Input Gate decides to update the Cell State with new information. Its expression and meaning are similar to those of the Forget Gate. As shown in Eq. (10), i_t is the output of the Input Gate, indicating the degree to which it needs to be updated. \tilde{C}_t represents the new candidate Cell State. \tanh is the activation function that scales the output value between -1 and 1. W_i and b_i represent the Input Gate's weight matrix and bias vector. W_c and b_c represent the candidate Cell State's weight matrix and bias vector.

$$\begin{cases} i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \end{cases} \quad (10)$$

Update the Cell State according to the output of the Forget Gate and Input Gate. It updates the new information and removes unnecessary information. The Cell State C_t of the current time step is shown in Eq. (11).

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (11)$$

The Output Gate determines the hidden state value at the current time step, generating the output based on the current input and Cell

State.

$$\begin{aligned} o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\ h_t &= o_t \tanh(C_t) \end{aligned} \quad (12)$$

As shown in Eq. (12), o_t indicates the hidden state output, and h_t is the hidden state of the current time step. W_o and b_o represent the weight matrix and bias vector of the Output Gate, respectively.

(3) Transformer and Multi-head Self-Attention

In this section, the three-dimensional tensor output of the previous CNN-LSTM module is further introduced into the Transformer module for key information extraction under the Multi-head Self-Attention mechanism. It optimizes the shortcomings of LSTM in processing distant position dependencies in sequences, effectively improves the performance of the mixed framework in capturing global dependencies, and enhances the mixed model's ability to capture complex patterns. The Transformer module comprises the Multi-head Self-Attention mechanism and the FFN.

The core of the Transformer module is Self-Attention. This mechanism generates corresponding weights for each position by measuring the dependencies of different positions in the sequence. According to the input sequence X , the vector Query Vector(Q), Key Vector(K), and Value Vector(V) could be obtained by the linear transformation.

$$\begin{cases} Q = XW_Q \\ K = XW_K \\ V = XW_V \end{cases} \quad (13)$$

In Eq. (13), W_Q , W_K , and W_V are the trainable weight matrices, respectively. Subsequently, the attention weights output could be obtained by Eq. (14) through the Softmax function.

$$\text{Attention}(X) = \text{Attention}(Q, K, V) = \text{soft max}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (14)$$

However, to better capture the global dependence of the carbon price sequence and optimize the diversified processing of complex internal data, this paper further applies the Multi-Head Attention to improve the global feature capture capability of the mixed framework and the robustness and expression capability. The Multi-head Self-Attention mechanism measures attention in different subspaces using multi-head attention simultaneously. Eq. (15) illustrates the process, where W_o is the weight matrix of the output.

$$\begin{cases} \text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \\ \text{MultiHead}(X) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_i)W_o \end{cases} \quad (15)$$

The Transformer applies the FFN for linear transformation and nonlinear activation based on the multi-head attention output. In order to stabilize the training model, Eq. (16) conducts the output results of multi-head attention and FFN by residual connection and layer normalization.

$$\begin{cases} \text{FFN}(X) = \text{ReLU}(XW_1 + b_1)W_2 + b_2 \\ \text{AttentionOutputNorm} = \text{LayerNorm}(X + \text{MultiHead}(X)) \\ \text{FFNOutput} = \text{FFN}(\text{AttentionOutputNorm}) \end{cases} \quad (16)$$

The output after entering the fully FFN is as follows.

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (17)$$

In this paper, the selection of hyperparameters plays a vital role in the accuracy and stability of the proposed mixed forecasting framework. It applies the BO algorithm to improve the accuracy and training efficiency while considering computing power and time cost. The BO adopts the Gaussian Process (GP) as a probability proxy model. It applies Expected Improvement (EI) as a collection function to efficiently search for the optimal combination of hyperparameters on a global scale for the whole

model. The code settings of the BO algorithm are according to Eq. (18) to Eq. (19).

$$\begin{cases} f(x) \sim GP(m(x), k(x, x')) \\ k(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2l^2}\right) \end{cases} \quad (18)$$

$$EI(x) = \int_{-\infty}^{\infty} \max(0, f(x) - f(x^*)) p(f|x) df \quad (19)$$

In GP, $m(x)$ is the mean function, typically set to zero. x and x' are the input variables representing the hyperparameters of the proposed model. The $k(x, x')$ is the squared exponential kernel, where σ^2 and l^2 are parameters that control the vertical variation and smoothness of the function. In EI, $f(x)$ represents the predicted value of the target function at x , and $f(x^*)$ is the currently known optimal value of the target function. The collection function helps determine the next set of hyperparameters most likely to improve model performance, integrating the trade-off between global exploration and local exploitation.

3.3. The modeling steps

To sum up, Fig. 3 shows the modeling steps of the mixed model. This paper optimizes the coupling complexity and computing cost of different modules' hyperparameter sets and fully realizes the performance potential of these intelligent models.

- Step 1.** Collect the carbon price history data and form the mixed model's input tensor according to the time step.
- Step 2.** Apply the CEEMDAN based on MI regression to decompose the historical carbon price data and reconstruct new IMFs according to MI value.
- Step 3.** Effectively integrate the CNN, LSTM, and Transformer to construct the mixed forecasting framework. BO is applied to determine the proposed model hyperparameters. Apply the mixed forecasting framework to train and test the reconstructed IMFs.
- Step 4.** Reconstruct the carbon price training set and test set results based on the IMFs' training set and test set. Test and compare the forecasting results with other models according to the evaluation

indexes. It aims to verify the effectiveness and superiority of the proposed model.

Step 5. Conduct the subsequent rolling carbon price forecasting based on the proposed mixed forecasting model. Formulate policy suggestions and references for carbon price development according to the forecasting results for the future.

This paper selects the Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2) to measure and evaluate the accuracy of the forecasting model. This paper further employs the Diebold-Mariano (DM) test (Zhang et al., 2024) and the Improvement Ratio (IR) test (Wu and Du, 2024) to analyze the degree of improvement of the proposed model in the ablation experiments and model comparisons.

4. Empirical analysis and results

4.1. Data collection and description

This section selects the publicly available carbon price dataset from the Hubei Carbon Emissions Exchange in China as the research object. It aims to verify the effectiveness and superiority of the proposed mixed model. The sample consists of 2441 carbon price data points from April 2, 2014, to June 19, 2024, for forecasting purposes. The entire modeling process divides the dataset into training and test sets in the 8:2 ratio. Fig. 4 illustrates the daily carbon price development trend over the past decade.

Regarding the development trend, the carbon price in Hubei remained relatively low between 2014 and 2017, mostly fluctuating between 10 and 20 Yuan/tCO₂e. The price began to rise, reaching a peak of nearly 70 Yuan/tCO₂e by the end of 2018. Subsequently, the carbon price in Hubei declined but showed an upward trend. Regarding volatility, it experienced considerable fluctuations between 2019 and 2021, generally staying within 30–50 Yuan/tCO₂e. A new spike occurred in mid-2021, with prices briefly exceeding 60 Yuan/tCO₂e before rapidly declining. From 2022 to 2024, the carbon price has stabilized, maintaining a range of 20–30 Yuan/tCO₂e. Overall, the current carbon price in Hubei exhibits apparent nonlinearity and uncertainty, making the

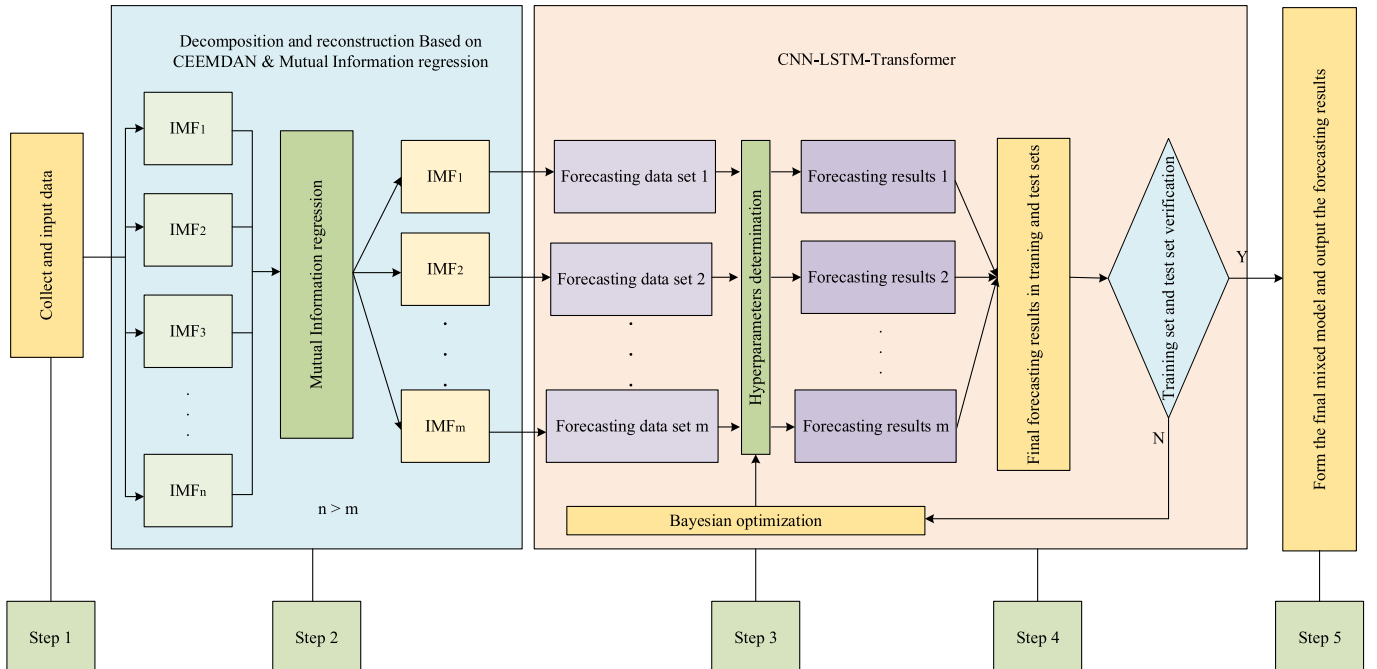


Fig. 3. The modeling process of the BO-CEEMDAN-MI-CNN-LSTM-Transformer model.



Fig. 4. The development trend of the carbon price in Hubei.

single model forecasting complex. Table 2 reveals the descriptive statistical characteristics of Hubei carbon price.

Based on the perspective of descriptive statistical analysis, this paper further studies Autocorrelation, seasonality, trend, and cycle components used to describe the characteristics of carbon price data in Hubei Province. Fig. 5 contains the box, scatter, HP filter decomposition, and ACF plots of carbon price data in this paper.

Fig. 5(a) displays a scatter plot illustrating the daily fluctuations in Hubei carbon prices. The data show a declining trend from Day 1 to approximately Day 800, followed by a general upward trend. In Fig. 5(b), this paper applies the first-order differencing of the original carbon price data to remove trends, showing most data fluctuating around zero, with only a few points exhibiting significant volatility. The box plot in Fig. 5(c) indicates that 25%–75% of carbon prices range between 20 and 40 Yuan/tCO₂e, with a median value of approximately 27 Yuan/tCO₂e. Fig. 5(d) reveals that the autocorrelation coefficients of the carbon price samples exhibit a first-order truncation, with the maximum autocorrelation coefficient at lag = 1 being −0.190, indicating a significant negative correlation between the current day's carbon price and the previous day's carbon price. Fig. 5(e) and (f) present the trend, cyclical, and seasonal components extracted from daily and monthly average carbon prices using the HP filter. Similar to the scatter plot, the trend component shows an initial decrease followed by an increase. However, the cyclical and seasonal components fail to display clear, repetitive patterns, exhibiting irregular fluctuations instead. It indicates almost no evident presence of distinct cyclicity or seasonality in Hubei's overall carbon price dynamics.

4.2. Carbon price decomposition and reconstruction

This paper first applies CEEMDAN to decompose carbon price data and form 9 IMFs. Fig. 6 reveals the trends and fluctuations from IMF₁ to IMF₉.

By methodically extracting different frequency components from the carbon price data, the nine IMFs display distinct oscillation characteristics and frequency ranges. The first three IMFs are characterized by

Table 2
The descriptive statistical characteristics of Hubei carbon price (2014–2024).

Empirical Sequence	Units	Minimum	Maximum	Mean	Standard Deviation
Daily carbon price	RMB (Yuan/tCO ₂ e)	10.48	61.89	29.63	11.16

significantly higher frequencies and amplitude fluctuations, identifying them as high-frequency components. In contrast, IMFs four through six show a noticeable reduction in frequency and amplitude, indicating more stable mid-frequency oscillations. IMF₇ and IMF₈ exhibit apparent, low-frequency, periodic changes. Lastly, IMF₉ predominantly presents a single-frequency waveform, effectively reflecting the long-term trend in carbon prices within the Hubei Carbon Emissions Exchange.

To further improve the carbon price decomposition quality and reduce the mode mixing phenomenon, this paper restructures the decomposition results of CEEMDAN with MI regression. This section calculates the MI value of the nine IMFs for the original carbon price data. It applies the MI value to evaluate the direct interdependence and correlation between nine IMFs and original carbon. Fig. 7 clearly shows the MI result distribution interval of the nine IMFs, among which the MI values of the first six IMFs are distributed between 0 and 1, while the MI values of IMF₇ and IMF₈ are between 1 and 2. IMF₉ has the slightest oscillation performance and has an MI value between 2 and 3. The purpose is to reasonably divide the MI value into three intervals for integration to ensure the noise reduction of carbon price data while retaining their data characteristics as much as possible. Therefore, in this paper, the first six IMFs, the mutual information values lower than 1, similar and low, are reconstructed as the high-frequency sequences of the original carbon price sequences. It also fuses IMF₇ and IMF₈ with values of 1–2 to form the mid-frequency sequences. The IMF₉ constitutes the low-frequency sequence alone.

As shown in Fig. 8, this paper further reconstructs the nine IMFs into three new IMFs. It further demonstrates the multi-scale characteristics of carbon price data significantly, simplifying the input features of the subsequent mixed forecasting framework and reducing computational complexity.

4.3. Forecasting and ablation study discussion

Based on the three reconstructed IMFs, this section adopts the proposed BO-CEEMDAN-MI-CNN-LSTM-Transformer to train and forecast the high-frequency, mid-frequency, and low-frequency sets. Fig. 9 reflects the five high degrees of influence time steps on forecasting during training. Among them, for reconstructing the new IMF₁, the positive contribution values of the 15th time step are more significant, and the response sensitivity to features is the highest. Similarly, for reconstructing the new IMF₂ and IMF₃, the 15th and the 20th time steps have the most substantial positive influence on the output of the corresponding IMFs. Therefore, considering that the eigenvalues of these time steps have a more significant positive influence and contribution to the output of the proposed mixed model, it could further determine the number of the look-back steps in the forecasting process.

Table 3 describes the final specific modeling parameters of forecasting by BO. Fig. 10 shows the proposed mixed model's forecasting effect of the three reconstructed IMFs in the training and test groups. It reflects that the mixed model simulates the development trend and fluctuation amplitude of the three new IMFs, and the forecasting values are consistent with their development and changes.

To further verify the validity of the proposed mixed model, this paper fuses the forecasting results of the three new IMFs to form the final carbon price forecasting results in the training and test groups. Based on the forecasting effect and the concentration degree of the actual and forecasting carbon price values in Fig. 11, the actual and forecasting values are concentrations in the training group. Meanwhile, Fig. 11 shows that the test group maintains a high concentration of actual and forecasting values. In the test group, the actual and forecasting values of the carbon price are mainly concentrated near the approximate baseline, indicating that the forecasting value and the actual value have a strong positive correlation. Among them, the test group's carbon price fluctuation is consistent with the development trend, proving the model has high forecasting accuracy and reliability.

Finally, Fig. 12 shows the overall forecasting effect of the mixed

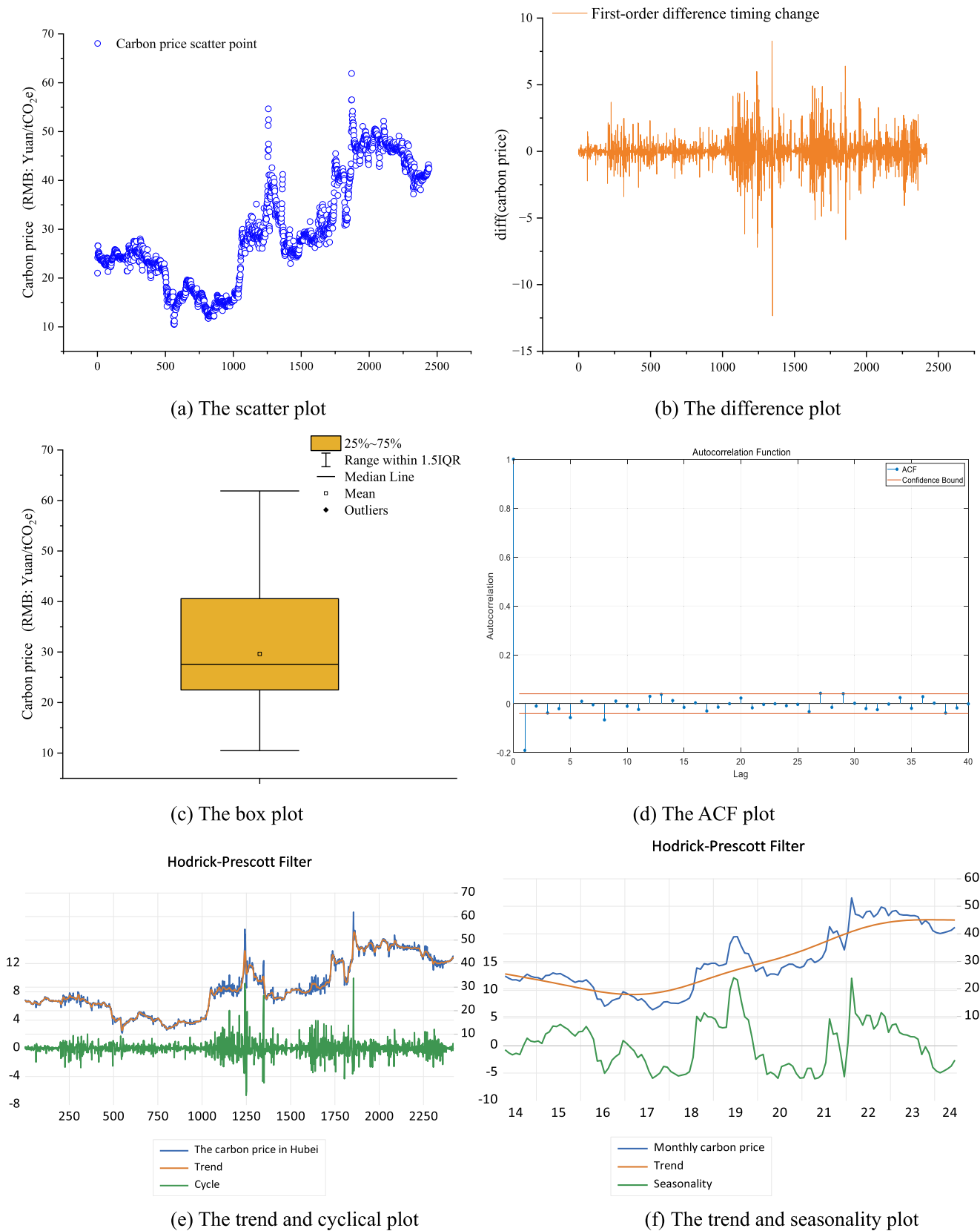


Fig. 5. The descriptive statistical plots for the carbon price in Hubei.

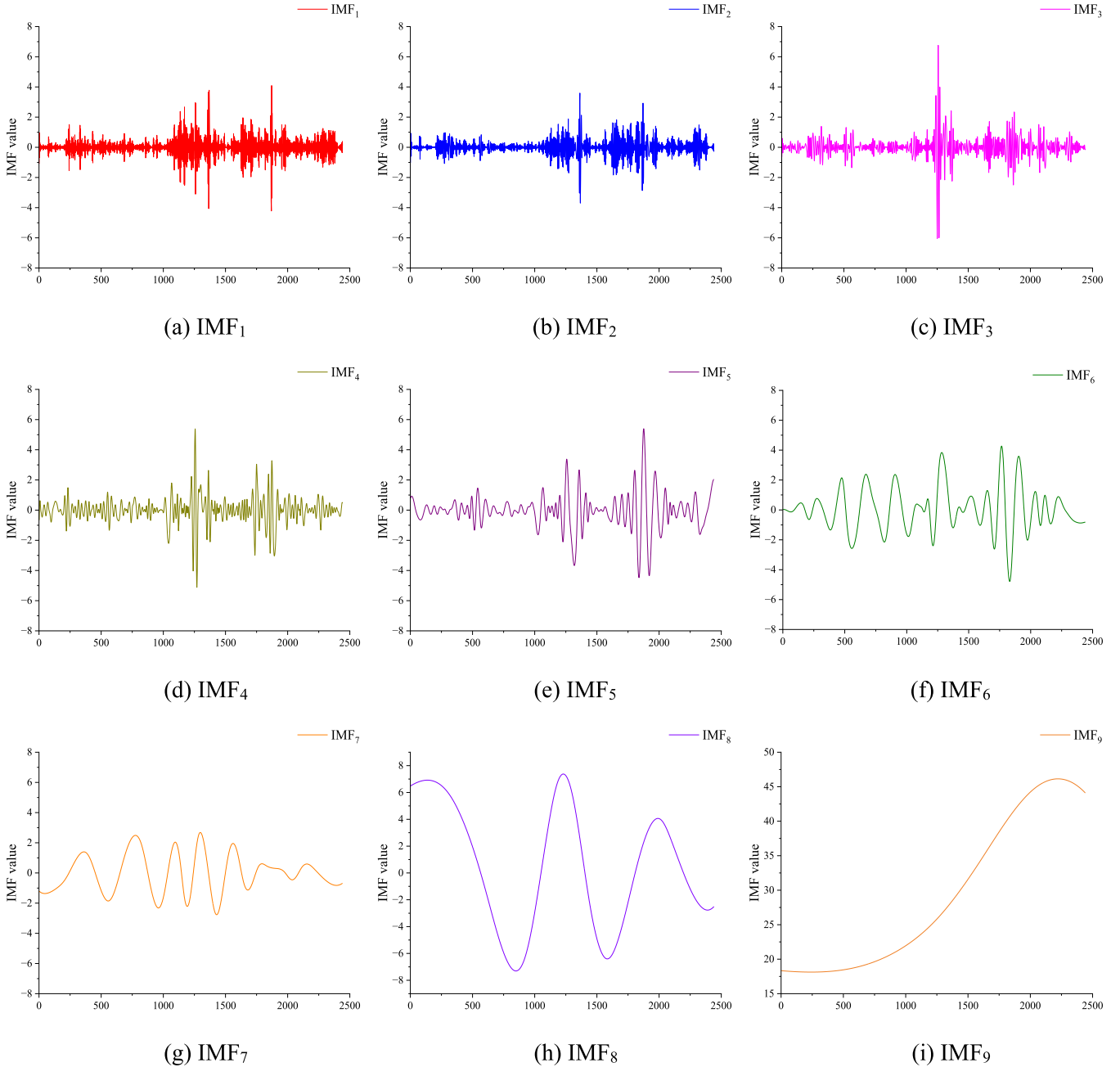


Fig. 6. The IMF results based on CEEMDAN.

model. Table 4 illustrates the results under different evaluation indicators, which convinces the training and test groups to have acceptable forecasting accuracy. The results of the test group are almost better than those of the training group, reflecting that there is no overfitting problem in the forecasting calculation. In contrast, MSE and MAE in the test group are better than those in the training group, which shows that the mixed model has better generalization ability in the subsequent forecasting period. The test group also shows a significant increase in MAPE, reaching 1.3176%, and R^2 is higher than 0.9, indicating that the proposed model is adaptable to the carbon price development in Hubei. It could be rational and reliably applied to the subsequent continuous update of carbon price data forecasting.

(2) Ablation experiment

Based on the data of the test group, this paper conducts the ablation experiment to verify further improvement in the effect of different modules in the mixed model on forecasting ability. This paper applies LSTM as the baseline model to forecast the Hubei carbon price.

This section constructs CEEMDAN-LSTM, CEEMDAN-MI-LSTM, CEEMDAN-MI-LSTM-Transformer, CEEMDAN-MI-CNN-LSTM, CEEMDAN-MI-LSTM-Attention, and CEEMDAN-MI-CNN-LSTM-Transformer as comparison models with different module combinations applied in this paper. Table 5 shows the experimental results.

As shown in Fig. 13, with the continuous addition of MI regression, CNN, and Attention mechanism, different mixed modules could improve the forecasting effect gradually. In Table 5 and Fig. 13, each added module optimization has different degrees of impact on the forecasting performance. According to the forecasting effect results of CEEMDAN-MI-LSTM, MI regression optimizes the CEEMDAN-LSTM's forecasting accuracy. The MAPE increases from 1.9588% to 1.7627%, and MAE

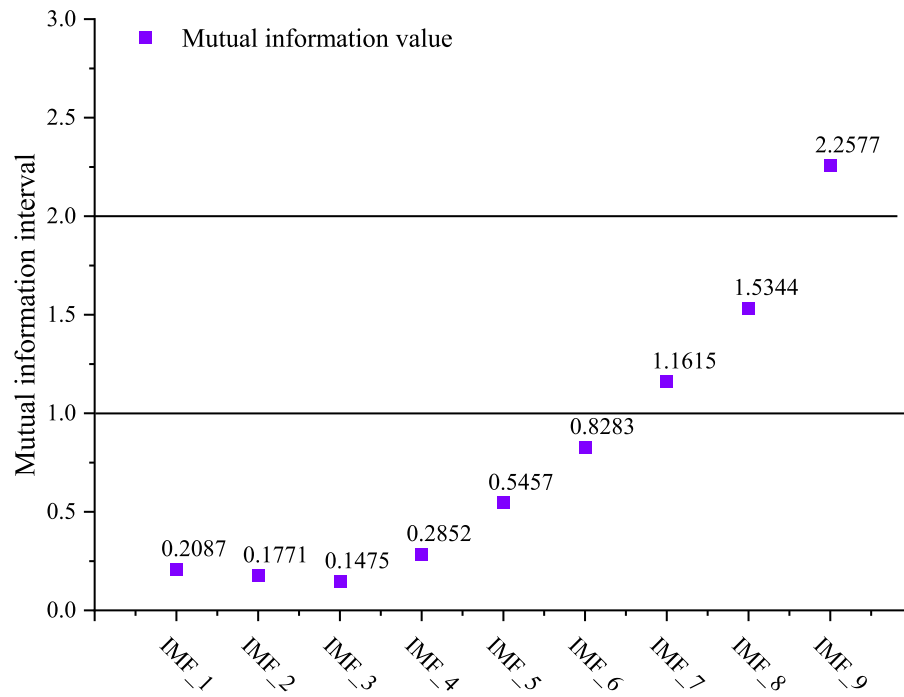


Fig. 7. The results and distribution of the MI value.

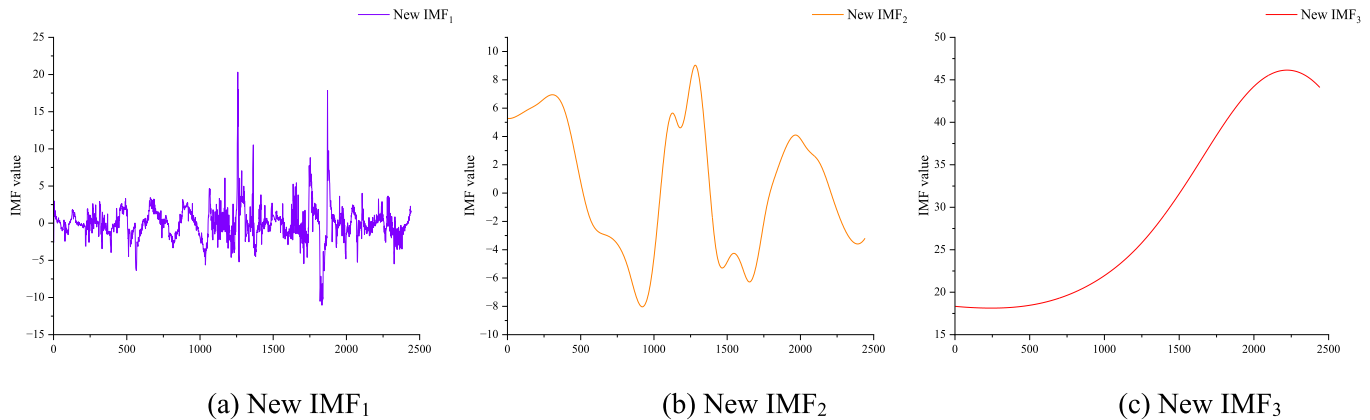


Fig. 8. The new IMF reconstruction based on MI regression

Notes: New IMF₁ represents the high-frequency sequence, New IMF₂ represents the mid-frequency sequence, and New IMF₃ New IMF₃ represents the low-frequency sequence.

increases from 0.8890 to 0.7975. However, it negatively affects the MSE, as shown in Fig. 13, where its MSE is higher than the other test results. In reconstructing IMFs based on the MI value, some data features may be affected while reducing the noise. From the perspective of feature retention, this paper preliminarily considers the trade-off between noise reduction and feature loss. It constructs the trade-off perspective by quantifying the changes in MI and sample entropy. Before reconstruction, the total MI value is approximately 7.1461, while the total MI value of the new IMFs after reconstruction is 4.8274. The feature retention may reach 67.55%. However, this paper quantifies the effect of noise reduction through sample entropy reduction. It finds that the sum of the sample entropy of all IMFs before reconstruction is approximately 2.4394. After reconstruction, the new IMFs' sample entropy sum is about 0.6471, about 26.53% of the value before reconstruction. The sum of sample entropy decreases significantly, which indicates that the IMFs' uncertainty and randomness are weakened, and the noise influence is reduced. Therefore, the MI reconstruction may reach a referenceable

trade-off that loses about 32.45% of feature information and reduces approximately 73.47% of noise in this paper. Compared with the IMFs produced by CEEMDAN decomposition, the range of the new high-frequency IMF after MI reconstruction becomes extensive, and the short-term fluctuation is more substantial, increasing MSE. Therefore, regarding the new high-frequency sequences containing more short-term fluctuations and noise, this paper considers enhancing the forecasting module's capacity for local feature extraction to improve the overall forecasting performance of the mixed model. Based on the LSTM, this paper further adds the CNN module. In this improvement, the CNN applies convolutional filters that slide across the input feature data for localized computations, while max-pooling reduces feature dimensions and preserves critical feature information. This enhancement strengthens the new module's ability to extract local features in the time series data. Meanwhile, this paper also tries to weaken the influence of MI on MSE by improving the ability of global feature extraction and introduces the Transformer module after the LSTM module for testing.

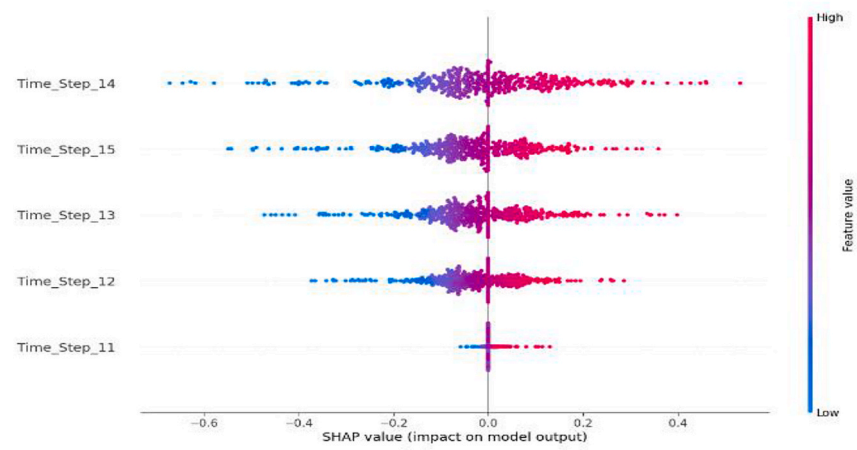
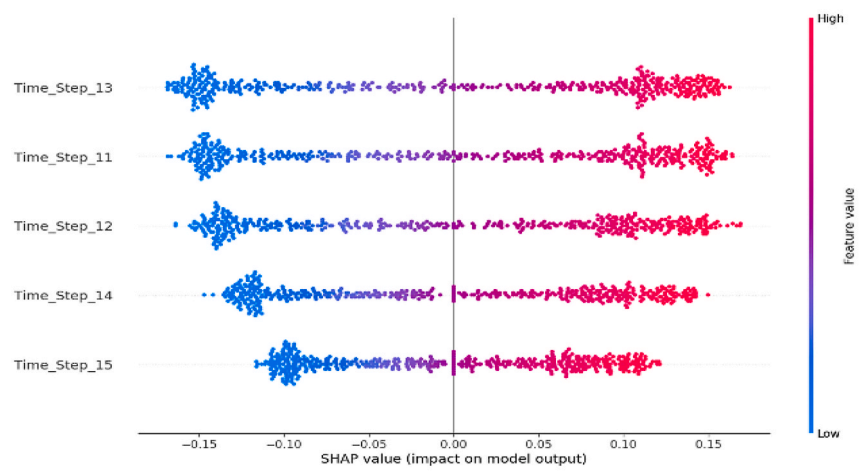
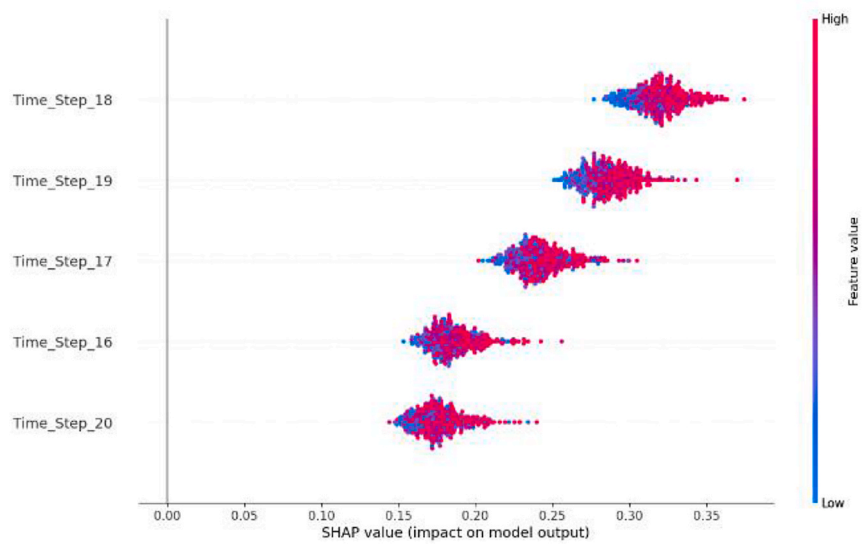
(a) New IMF₁(b) New IMF₂(c) New IMF₃**Fig. 9.** The SHAP plot for the impact on model outputs.

Table 3

The determined modeling parameters.

Best Parameters/IMF	New IMF ₁	New IMF ₂	New IMF ₃
LSTM Unit	70	20	70
CNN Filter	128	128	128
CNN Kernel Size	5	5	3
Learning Rate	1e-05	1e-02	1e-04
Batch Size	16	32	32
Look-Back	15	15	20
Transformer Head Size	32	32	128
Transformer Muti-heads	8	8	2
FFN Dim	32	64	128
Best MSE	0.7285	0.0009	0.0108

With the addition of CNN and Transformer, CEEMDAN-MI-CNN-LSTM or CEEMDAN-MI-LSTM-Transformer could further optimize CEEMDAN-MI-LSTM. The changes in Fig. 13 reflect that the selected error indicators are all reduced, and their R^2 is also slightly improved compared with the CEEMDAN-MI-LSTM. Improving the mixed model's local or global feature extraction ability through CNN or Transformer could effectively alleviate the adverse effects of the MI regression.

In order to further prove the advantages and effectiveness of the BO-CEEMDAN-MI-CNN-LSTM-Transformer, this paper conducts the statistical hypothesis test DM test and calculates IR further to discuss the improvement effect of the different forecasting modules.

This paper applies the DM test to determine if there are significant differences between the baseline LSTM in the ablation study. The evaluation criteria refer to whether the DM of the two comparison models is more significant than $Z_{\alpha/2}$ at a given significance level α . Table 6 shows the DM values for comparing their improvement with the LSTM.

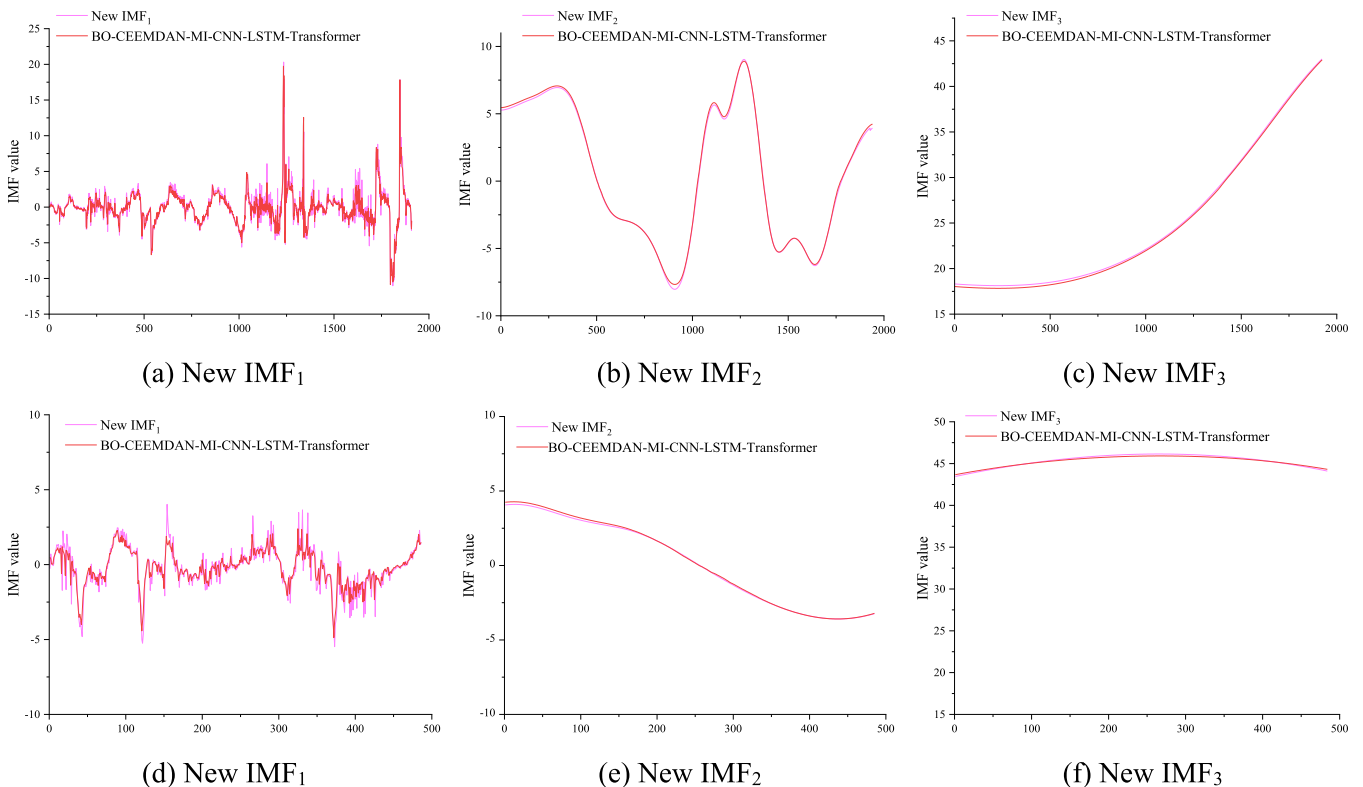
According to the DM test results, each model improvement could improve the forecasting performance of the baseline model. DM values of all improved models are significantly higher than 1.96, indicating that the added modules (CEEMDAN, MI, CNN, Attention, and Transformer)

contribute significantly to the model's predictive performance at the 10% significance level. In contrast, based on CEEMDAN-MI-LSTM, CNN, Transformer, and BO optimization are further added, and the DM values of these models are higher than 2.58. The DM values mean that at the significance level of 1%, the predictive performance of these models is significantly different from the LSTM. Further introducing artificial intelligence forecasting models to optimize the mixed forecasting module could improve forecasting performance significantly. Among them, either CNN or Transformer integration has greatly enhanced forecasting performance, with CEEMDAN-MI-CNN-LSTM showing a more significant improvement over LSTM than CEEMDAN-MI-LSTM-Transformer.

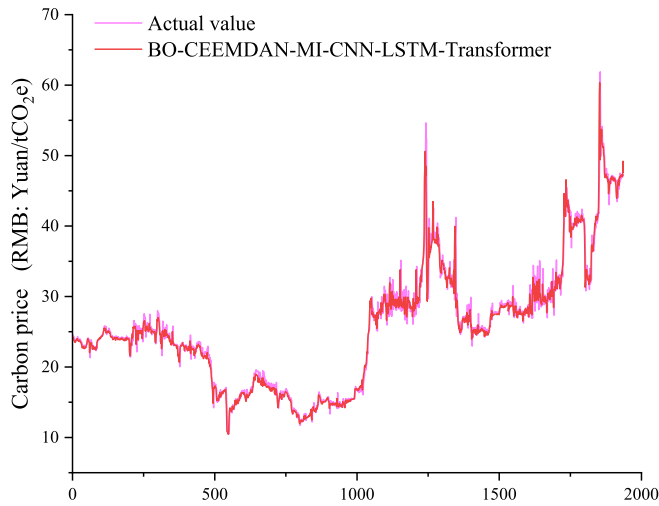
Regarding the impact of individual modules on forecasting performance, CNN outperforms Transformer. However, when both CNN and Transformer are incorporated simultaneously into the LSTM, resulting in the CEEMDAN-MI-CNN-LSTM-Transformer model, the DM value further escalates to 3.7486. This DM value substantiates the synergy between CNN's local feature extraction and the Transformer's global information integration, which collectively elevates the predictive module's performance, which is better than that of each CNN or Transformer. Finally, the BO-CEEMDAN-MI-CNN-LSTM-Transformer has the highest DM value of 3.7503, which has significant advantages in forecasting performance compared with other models with different modules. It indicates that optimizing hyperparameters through BO could improve the model's operational efficiency and its overall forecasting performance.

Based on the DM test, this paper further calculates the IRs of different models' MAE, MSE, MAPE, and R^2 in the ablation experiment. Table 7 reveals the IRs in Dataset 1 and Dataset 2.

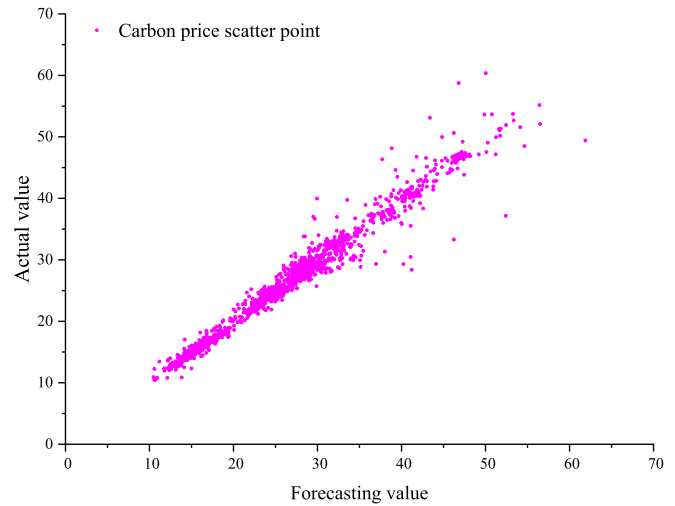
The results from Table 7 indicate that mere decomposition and reconstruction based on CEEMDAN and MI do not significantly boost the LSTM's forecasting capabilities. As LSTM progressively integrates with CNN and Transformer within the forecasting module, the IR exhibits an increasing trend, with the first three metrics exceeding 15%, demonstrating that adding new modules could directly enhance the forecasting

**Fig. 10.** The forecasting effect of the new IMFs in the training and test groups

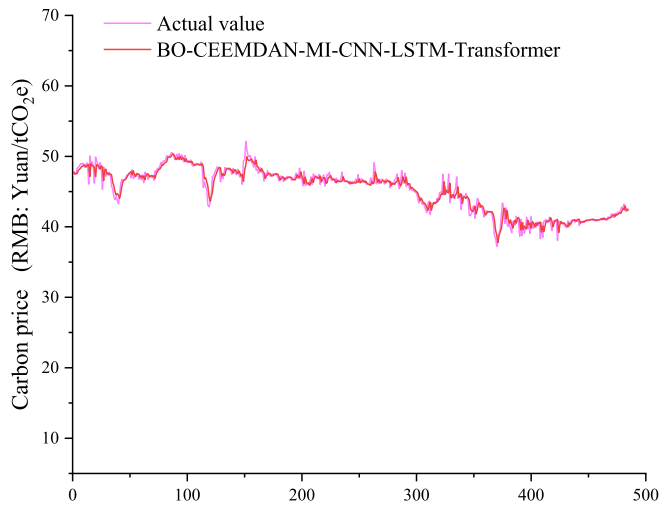
Notes: Fig.(a), Fig.(b), and Fig.(c) show the forecasting effect in the training group. Fig.(d), Fig.(e), and Fig.(f) show the forecasting effect in the test group.



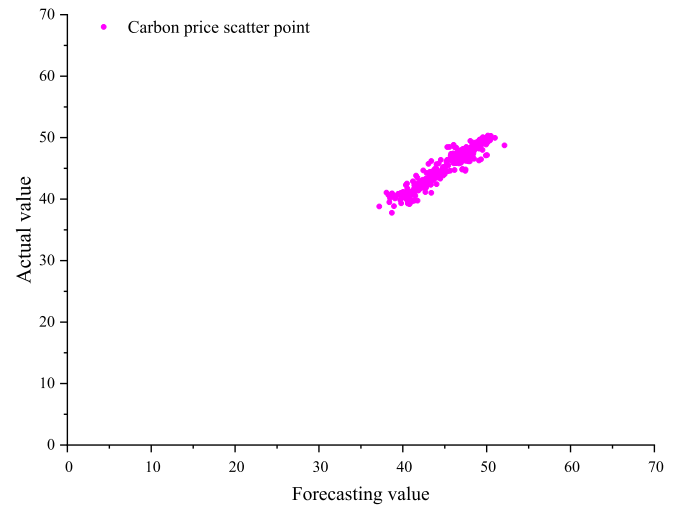
(a) Forecasting effect



(b) Concentration



(c) Forecasting effect



(d) Concentration

Fig. 11. The forecasting effect of the carbon price in the training and test groups

Notes: Fig.(a) and Fig.(b) show the forecasting effect and concentration of actual values and carbon prices in the training group. Fig.(c) and Fig.(d) show the test group's forecasting effect and concentration of actual values and carbon prices.

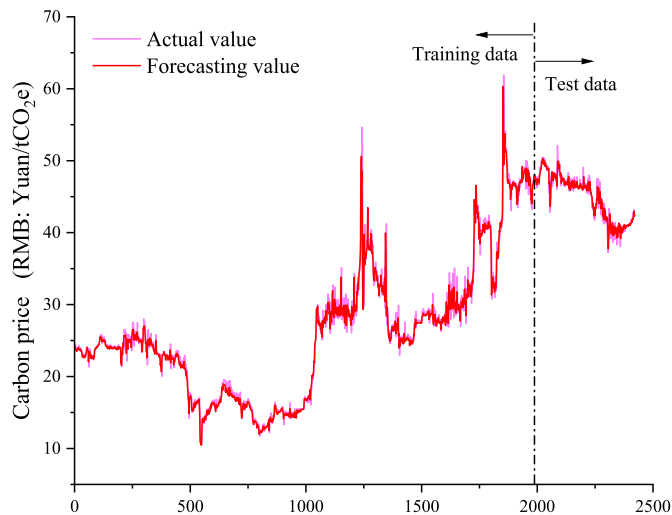


Fig. 12. The overall forecasting effect of the carbon price.

Table 4

The evaluation results of the proposed model.

Evaluation Metrics/Groups	Training Group	Test Group
MAE	0.7452	0.5889
MSE	2.0034	0.7455
MAPE	2.7487	1.3176
R ²	0.9741	0.9256

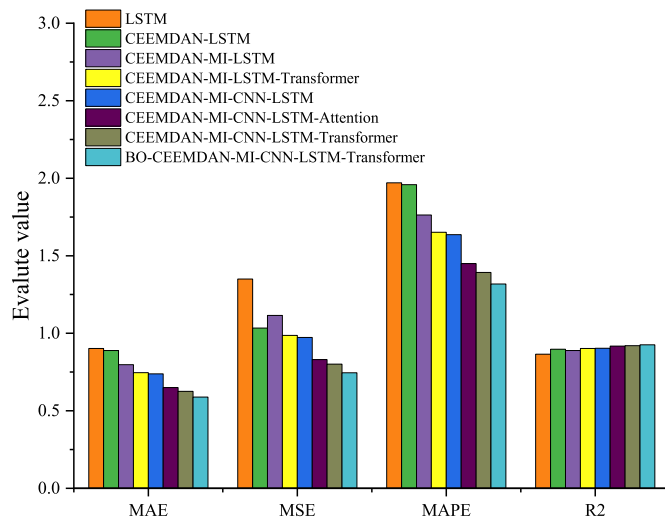
performance. On the other hand, IRs from Dataset 1 and Dataset 2 further reveal slight differences in the degree of model optimization between CNN and Transformer.

Therefore, this paper further investigates the roles of the relative effectiveness of CNN and Transformer on datasets with different statistical properties. This paper applies the new IMFs that retain some original carbon price data features. Fig. 8 shows that the three reconstructed new IMFs exhibit different noise levels: New IMF₁ has the highest noise, New IMF₂ has moderate noise, and New IMF₃ has the least noise. This paper applies CNN and Transformer to test the forecasting performance of these new IMFs, and Table 8 shows the test results.

Table 5

The evaluation results of the models with different module combinations.

Models/Evaluation Metrics	MAE	MSE	MAPE	R ²
LSTM	0.9020	1.3497	1.9704	0.8652
CEEMDAN-LSTM	0.8890	1.0337	1.9588	0.8970
CEEMDAN-MI-LSTM	0.7975	1.1151	1.7627	0.8889
CEEMDAN-MI-LSTM-Transformer	0.7457	0.9860	1.6517	0.9017
CEEMDAN-MI-CNN-LSTM	0.7385	0.9733	1.6361	0.9031
CEEMDAN-MI-CNN-LSTM-Attention	0.6499	0.8303	1.4495	0.9173
CEEMDAN-MI-CNN-LSTM-Transformer	0.6256	0.8007	1.3924	0.9201
BO-CEEMDAN-MI-CNN-LSTM-Transformer	0.5889	0.7455	1.3176	0.9256

**Fig. 13.** The forecasting effect under the influence of different modules.**Table 6**

The evaluation results of the models with different module combinations.

Models	DM value
LSTM	–
CEEMDAN-LSTM	2.0561
CEEMDAN-MI-LSTM	2.3434
CEEMDAN-MI-LSTM-Transformer	2.6698
CEEMDAN-MI-CNN-LSTM	3.5200
CEEMDAN-MI-CNN-LSTM-Attention	3.6725
CEEMDAN-MI-CNN-LSTM-Transformer	3.7486
BO-CEEMDAN-MI-CNN-LSTM-Transformer	3.7503

Notes: $Z_{0.01/2} = 2.58$, $Z_{0.05/2} = 1.96$, $Z_{0.10/2} = 1.64$. DM value represents the DM test results of different models and the baseline model LSTM.

Table 7

The improvement ratios of the BO-CEEMDAN-MI-CNN-LSTM-Transformer with other models.

Models	Dataset 1				Dataset 2			
	IR _{MAE} (%)	IR _{MSE} (%)	IR _{MAPE} (%)	IR _{R²} (%)	IR _{MAE} (%)	IR _{MSE} (%)	IR _{MAPE} (%)	IR _{R²} (%)
LSTM	–	–	–	–	–	–	–	–
CEEMDAN-LSTM	1.44%	23.41%	0.59%	3.68%	–	–	–	–
CEEMDAN-MI-LSTM	11.59%	17.38%	10.54%	2.74%	10.29%	7.87%	10.01%	0.90%
CEEMDAN-MI-LSTM-Transformer	17.33%	26.95%	16.17%	4.22%	6.50%	11.58%	6.30%	1.44%
CEEMDAN-MI-CNN-LSTM	18.13%	27.89%	16.97%	4.38%	7.40%	12.72%	7.18%	1.60%
CEEMDAN-MI-CNN-LSTM-Attention	27.95%	38.48%	26.44%	6.02%	12.00%	14.69%	11.41%	1.57%
CEEMDAN-MI-CNN-LSTM-Transformer	30.64%	40.68%	29.33%	6.35%	15.29%	17.73%	14.90%	1.88%
BO-CEEMDAN-MI-CNN-LSTM-Transformer	34.71%	44.77%	33.13%	6.98%	5.87%	6.89%	5.37%	0.60%

Notes: Dataset 1 contains the IR of different modules added in the ablation experiment with the baseline model LSTM. Dataset 2 contains the IR between the adjacent improved models with different modules gradually added. The IRs of CEEMDAN-MI-CNN-LSTM and CEEMDAN-MI-LSTM-Transformer are according to the CEEMDAN-MI-LSTM. The IRs of CEEMDAN-MI-CNN-LSTM-Attention and CEEMDAN-MI-CNN-LSTM-Transformer are according to the CEEMDAN-MI-CNN-LSTM.

As shown in Tables 8 and in the comparison of the New IMF₁ dataset, Transformer outperforms CNN. This paper suggests that when forecasting datasets with higher noise levels, such as New IMF₁, CNN may be negatively affected by excessive and useless noise that interferes with local features, which weakens the model's performance. In contrast, Transformer, with its superior global modeling capability, may effectively capture long-term dependencies in the data and reduce noise interference. Therefore, Transformer may be preferred over CNN when forecasting high-noise datasets. However, for New IMF₂ and New IMF₃, CNN and Transformer shift their roles in terms of forecasting performance, with CNN outperforming Transformer on moderate-noise and low-noise datasets. Specifically, New IMF₂ has less noise than New IMF₁, and New IMF₃ shows more pronounced trend components with even less noise and more prominent local features. Therefore, CNN's ability to capture local features becomes increasingly evident in both cases, and its performance generally outperforms Transformer. Thus, this paper proposes that the roles of relative effectiveness between CNN and Transformer in different datasets may shift according to their noise level variation. In this paper, the performance of Transformer may be superior to CNN when forecasting high-noise datasets. When forecasting moderate-noise and low-noise data sets, CNN may outperform Transformer.

Additionally, the IRs for CEEMDAN-MI-CNN-LSTM-Attention and CEEMDAN-MI-CNN-LSTM-Transformer in Datasets 1 and 2 confirm that Multi-head Self-Attention has a more substantial optimizing effect compared to regular Attention mechanisms. In Dataset 2, the IRs for BO-CEEMDAN-MI-CNN-LSTM-Transformer relative to CEEMDAN-MI-CNN-LSTM-Transformer are below 10%, indicating that its performance optimization is less significant than the introduction of different artificial intelligence models. However, BO addresses the determination of hyperparameters for the proposed model, and similar to MI, its impact and application are more on enhancing computational efficiency and reducing computational costs rather than improving model performance.

4.4. Model comparison

This section further verifies the superiority of the proposed mixed model and the referability of subsequent forecasting results by

Table 8

The performance of CNN and Transformer on the new IMFs.

Model	IMFs	MAE	MSE	R ²
CNN	New IMF ₁	0.6572	0.8958	0.5560
	New IMF ₂	0.0937	0.0134	0.9983
	New IMF ₃	0.2718	0.0965	0.8191
Transformer	New IMF ₁	0.5959	0.7487	0.6289
	New IMF ₂	0.1777	0.0434	0.9944
	New IMF ₃	0.3457	0.1252	0.7653

comparing the forecasting ability of the proposed model with other models. This paper selects these benchmark models: GARCH, SVM, ANN, CNN, RNN, LSTM, GRU, CNN-LSTM, and TFT as the comparison models. Table 9 and Fig. 14 illustrate the comparison results from the perspective of visualizing and measuring results.

Fig. 14 and Table 9 show that SVM and GARCH have the lowest overall performance among all the compared models. In Fig. 15(a) and (b), their accuracy distribution deviates significantly from the other models. As traditional statistical and machine learning models, these two methods have limited capability in capturing nonlinear and complex dynamic features, making them inadequate for effectively handling the strong nonlinearity and temporal dependencies in Hubei's carbon data. As shown in Fig. 14, all the neural network forecasting models have good adaptability to the nonlinearity and uncertainty of carbon price. However, in terms of data differences, in Fig. 14, (a), (b), and (e) reflect the unsatisfactory forecasting effect of LSTM, CNN, and CNN-LSTM. The numerical results in Table 9 also demonstrate the accuracy shortcomings of the three models, and the forecasting accuracies of these three models are weaker than those of other models. The distribution of evaluation indicators in Fig. 15 (c) shows that the accuracy of these three models is similar, and the average amplitude and percentage deviation of their errors are higher than those of other neural network models.

The CNN-LSTM optimizes LSTM by introducing the CNN module to improve the forecasting performance, proving that combined forecasting could effectively conduct the data complexity in the forecasting process. However, its forecasting effect is still inferior to others. On the contrary, the forecasting effect of CNN is higher, indicating that the direct combination of CNN and LSTM cannot fundamentally optimize the forecasting effect on carbon price data and may even produce adverse effects. The CNN is adept at quickly capturing local features within the data, but this may not align well with the need for long-term dependencies characteristic of time series data. When the local features of Hubei's carbon price data are directly inputted to LSTM for processing, they may be insufficient or unsuitable for representing the long-term dynamic changes in the time series. It could lead to the mixed model's inability to effectively utilize this information or even result in overfitting due to feature mismatch and increased complexity, ultimately impacting overall performance. In contrast, the BO-CEEMDAN-MI-CNN-LSTM-Transformer first preprocesses and decomposes the time series using CEEMDAN-MI, extracting multiple IMFs and integrating only those IMF components that contribute most to forecasting performance through MI regression. It enhances the quality and relevance of the input features, which helps CNN more accurately extract useful local features, thus enhancing the quality of local feature inputs. Following the CEEMDAN-MI-CNN allows LSTM to process and maintain these features' temporal dependencies more effectively. Finally, the Transformer at the end of the model comprehensively models' relationships and integrates features from the information processed by the preceding network layers. As shown in Table 9, the forecasting accuracy and robustness of the BO-CEEMDAN-MI-CNN-LSTM-Transformer

are superior to CNN and CNN-LSTM.

This paper also applies the TFT to forecast carbon price data for further exploration. Unlike CNN-LSTM, TFT's primary structure consists of the LSTM and Transformer. The Python and the TensorFlow-based deep learning environment in VSCODE are applied to implement the TFT. This paper uses grid search and stepwise manual tuning to determine the TFT's hyperparameter settings and optimize its training process. In this paper, the applied TFT's Look-Back is 30, Batch-Size is 32, Transformer Multi-heads is 2, FFN Dim is 128, and the Vector Dim is 64.

As shown in Table 9, TFT's performance surpasses most models among the comparisons. It indicates that the basic framework of TFT demonstrates good forecasting performance. However, compared to the BO-CEEMDAN-MI-CNN-LSTM-Transformer, the TFT applied in this paper does not include the decomposition and reconstruction module. It may lead to some noise components remaining in the forecasting framework, which affects the TFT's performance when handling the original carbon price sequences. On the other hand, the proposed model adopts an integrated framework combining CNN, LSTM, and Transformer. Compared to the TFT's core framework, which consists of LSTM and Transformer, the BO-CEEMDAN-MI-CNN-LSTM-Transformer's framework further benefits from CNN's ability to capture local features. Therefore, these may lead to the BO-CEEMDAN-MI-CNN-LSTM-Transformer's function being more powerful and outperforming TFT in Hubei carbon price forecasting.

On the other hand, RNN, ANN, GRU, and TFT all have reasonable forecasting effects on the carbon price data provided in this paper. Their forecasting accuracy is better than that of the above five models. Fig. 15 (c) clearly shows that in terms of the distribution of MAPE, MSE, and MAE, the error evaluation indicators of RNN are significantly higher than those of ANN and GRU. The accuracy gap between ANN and GRU in MAPE is noticeable, while MSE and MAE are close. The data from Table 9 more intuitively show that ANN is superior to GRU.

In contrast, according to Table 9 and Fig. 15(c), the proposed BO-CEEMDAN-MI-CNN-LSTM-Transformer's MAPE is significantly superior to other comparison models. This paper further validates the proposed model performance through the DM test. Table 10 reveals the DM test values between BO-CEEMDAN-MI-CNN-LSTM-Transformer and other models. The DM test values are greater than 1.64 and exceed 1.96 for all models. It indicates a significant difference in forecasting performance at the 10% significance level when forecasting Hubei carbon prices between the BO-CEEMDAN-MI-CNN-LSTM-Transformer and the comparison models. Considering the evaluation metrics and DM test results, the BO-CEEMDAN-MI-CNN-LSTM-Transformer model shows at least a 10% significance level difference in forecasting performance compared to other models. It confirms that the proposed model achieves higher prediction accuracy in forecasting the Hubei carbon price than the other models from multi-angle validation. This paper verifies that the BO-CEEMDAN-MI-CNN-LSTM-Transformer can effectively explain the variability of carbon price data in Hubei Province than other models.

Therefore, the mixed model BO-CEEMDAN-MI-CNN-LSTM-Transformer is superior to the comparison models in the MAE, MSE, MAPE, R^2 , and DM test results. It demonstrates the model's superiority and verifies the validity of the proposed modeling framework. The result shows that the interaction of data decomposition technology, optimization algorithm, and attention mechanism could comprehensively improve the overall forecasting performance of the model under the same basic model. However, there is still uncertainty about the forecasting effect of the direct combination with different modules. When constructing the mixed mode, it is necessary to focus on the adaptability and interaction between the optimized modules during the parameter determination.

4.5. Additional analysis and discussion

(1) Subsequent forecasting

Table 9

The comparison between the proposed model and other forecasting models.

Models/Evaluation Metrics	MAE	MSE	MAPE	R^2
BO-CEEMDAN-MI-CNN-LSTM-Transformer	0.5889	0.7455	1.3176	0.9256
CNN-LSTM	0.8560	1.3313	1.8775	0.8670
LSTM	0.9020	1.3497	1.9704	0.8652
TFT	0.6750	0.8994	1.5134	0.9102
ANN	0.6207	0.7719	1.3850	0.9229
CNN	0.7862	1.0056	1.7311	0.8995
GRU	0.6618	0.8656	1.4673	0.9135
RNN	0.7116	0.9861	1.5817	0.9015
SVM	1.9737	4.6194	4.4019	0.5399
GARCH	22.2111	503.6488	48.8360	-49.2608

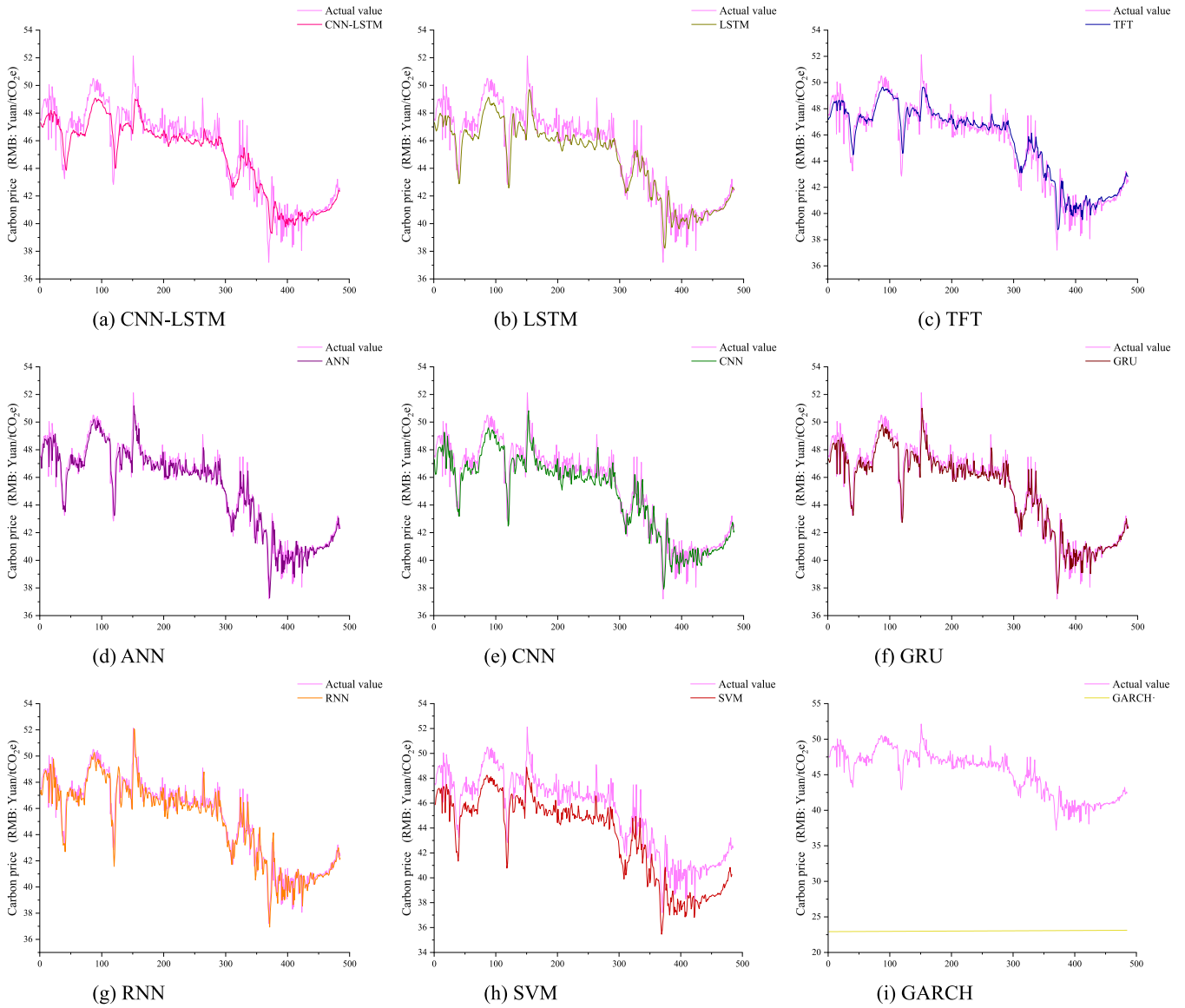


Fig. 14. The visualization of forecasting effects of comparison models.

This paper further applies the proposed BO-CEEMDAN-MI-CNN-LSTM-Transformer to forecast the carbon price development trend of the Hubei Carbon Emission Exchange for the next 180 days. Fig. 16

shows the forecasting trend. Table 11 describes the statistical characteristics of the forecasting.

Fig. 16 indicates that, over the forthcoming period, the carbon price

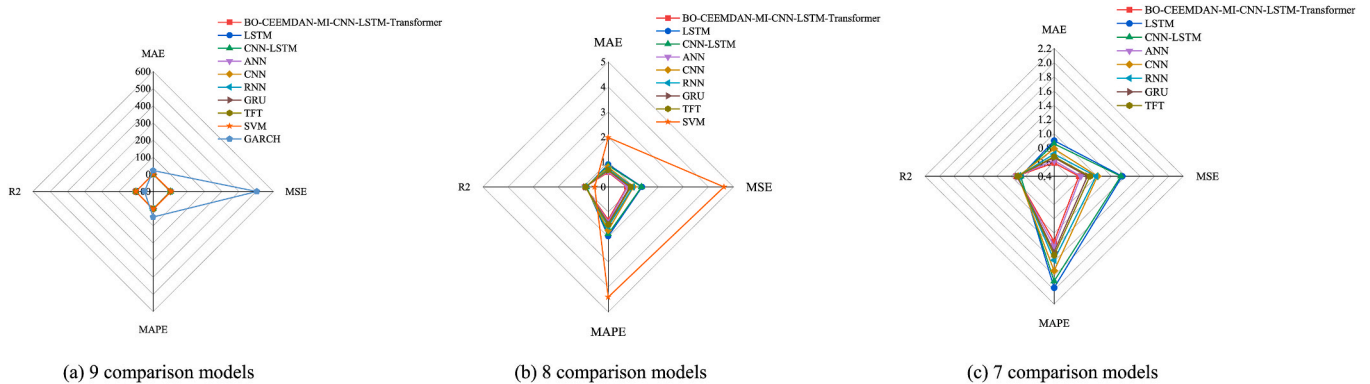


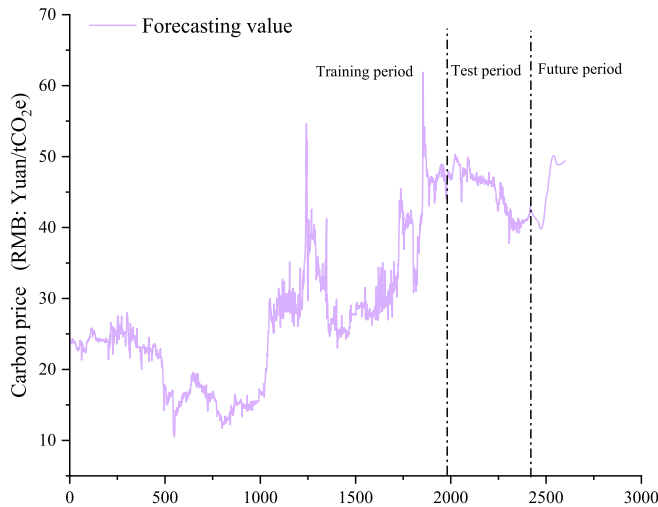
Fig. 15. The visualization of forecasting effects of comparison models

Notes: (a) contains all models, (b) contains all models except GARCH, (c) contains all models except GARCH and SVM.

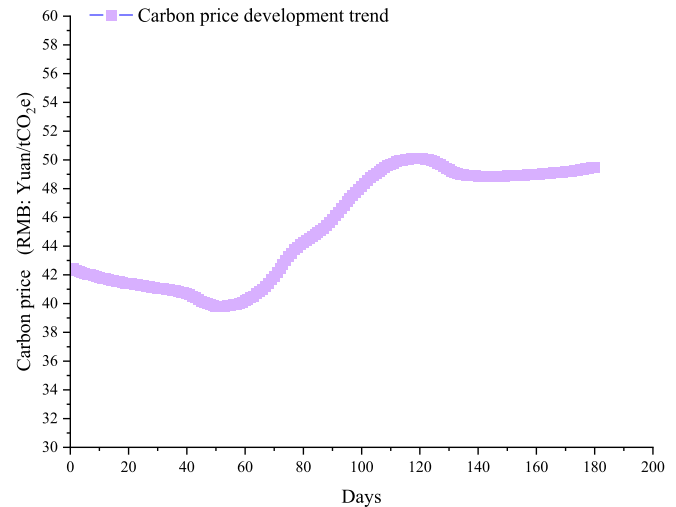
Table 10

DM test values of the proposed model and comparison forecasting models.

Models	CNN-LSTM	LSTM	TFT	ANN	CNN	GRU	RNN	SVM
BO-CEEMDAN-MI-CNN-LSTM-Transformer	2.7698	3.7503	2.3121	1.8842	3.7744	3.3208	4.0937	4.9872



(a) The total trend



(b) The forecasting trend

Fig. 16. The forecasting trends of Hubei carbon price after 180 days.

in Hubei exhibits relatively gradual multi-stage variations. Initially, there is a continuous downward trend, followed by a significant increase after 66 days, peaking at approximately 50.09 Yuan/tCO₂e on the 119th day. Subsequently, the carbon price will slightly decline and stabilize at 49 Yuan/tCO₂e by December. Compared to the carbon price trends over the past decade, the statistical characteristics in Table 11 indicate that in the near term, the carbon price level in Hubei may remain at a relatively high average of 45.37 Yuan/tCO₂e.

As shown in Fig. 17, this paper further conducts the descriptive statistical discussion to explore the development and changes in Hubei carbon prices over the next 180 days. Fig. 17(a) demonstrates that the nonlinear characteristics of the predicted carbon prices resemble the actual historical fluctuations, indirectly confirming that the forecasting results are similar to the historical development traits of Hubei's carbon prices. Fig. 17(b) reveals that the interquartile range (25%–75%) of the forecasting carbon prices may be between 42 and 46 Yuan/tCO₂e. Fig. 17(c) illustrates that over the next 180 days, Hubei may continue to exhibit an upward trend in carbon prices, indicating a consistency within the historical character, yet with noticeable periodic fluctuations. These cycles display consistent amplitudes and frequencies, objectively reflecting periodic characteristics in Hubei's short-term carbon prices. Moreover, Fig. 17(d) indicates that the predicted carbon price series exhibits significant positive autocorrelation in the short term, which diminishes with increasing lags and eventually shifts to negative correlation, further substantiating the short-term periodic influence on

Hubei's carbon prices. Therefore, this paper speculated that while the carbon price in Hubei Province would maintain an increasing historical trend in the short term, seasonal factors or other cyclical policies might be present based on the statistical analysis.

(2) Subsequent forecasting comparison

This paper further applies the comparison model adopted in Section 4.4 to verify and demonstrate the reference and reliability of the subsequent forecasting. It compares their subsequent forecasting results with the proposed model. Fig. 18 reflects the forecasting trends of these models over the next 180 days.

Among them, the SVM presents a sharp downward trend, starkly contrasting with Hubei's historical carbon price movements. Given that this paper does not incorporate additional factors to adjust the model, it implies that SVM could not adequately capture the potential volatility in future carbon prices. In contrast, the GARCH exhibits a singular linear trend with stable progression. The TFT's forecasting reflects a smooth upward trend. Although its forecast results align with the developmental patterns of carbon prices compared to the previous two models, it still fails to capture the significant volatility characteristics of carbon prices. Regarding the forecasting results, the parameters of these models need to be further adjusted to form forecasting results with reference.

The forecasting results of the LSTM show a sharp decline trend, and its carbon price prediction result even reaches 0.31 yuan/CO₂e. These results deviate significantly from the historical carbon price range. In contrast, the subsequent forecasting results of CNN-LSTM and GRU develop slowly, reflecting the insufficient response of the two models to the fluctuations of the carbon market and failing to capture the possible cyclical fluctuations in the later period. Although the forecasting results of ANN and CNN show different upward and downward trends at the initial stage, they may oversimplify the complexity of the carbon market under certain conditions. It may lead to a disconnect between forecasting results and the reality of the market in practical application.

The subsequent forecasting of RNN reflects slightly more volatile

Table 11

The descriptive statistical characteristics of Hubei carbon price forecasting after 180 days.

Empirical Sequence	Units	Minimum	Maximum	Mean	Standard Deviation
Daily carbon price	RMB (Yuan/tCO ₂ e)	39.82	50.09	45.37	3.88

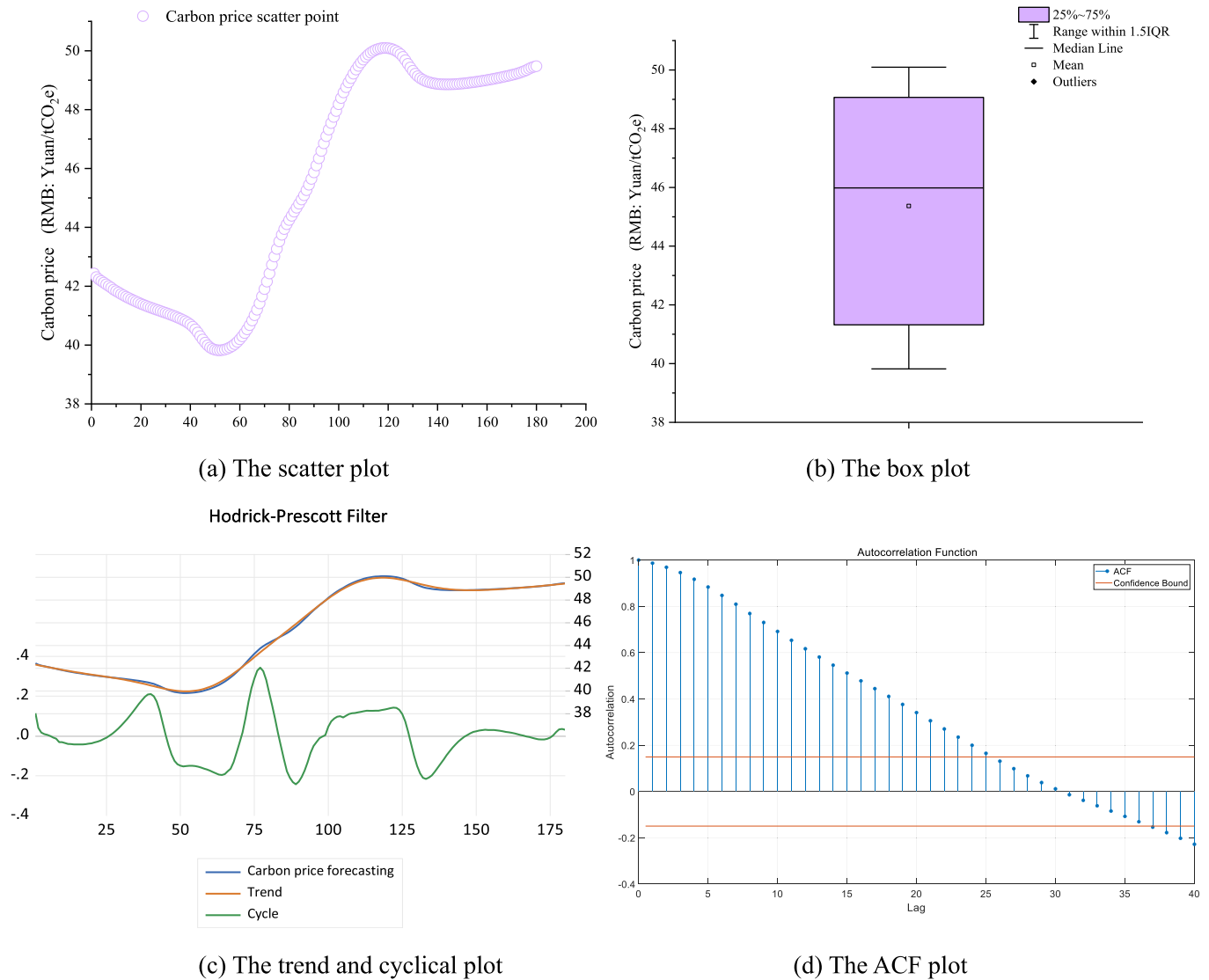


Fig. 17. The descriptive statistical plots for Hubei carbon price after 180 days.

carbon price changes, indicating that it can capture some carbon market dynamics. However, the overall development of RNN is still changing smoothly, which reflects that it is difficult for RNN to effectively reflect subsequent market changes after reaching a particular carbon market state. Although RNN is relatively effective in processing local carbon price forecasting sequences, as the number of forecasting steps increases, it may not be sufficient to deal with the nonlinear and complex characteristics of subsequent medium and long-term development in the carbon market in Hubei.

However, the forecasting development of the BO-CEEMDAN-MI-CNN-LSTM-Transformer contains both trend and periodic fluctuation characteristics. It significantly captures the future rising and stable trend of the carbon price in Hubei Province and identifies seasonal fluctuations. For the next 180 days, Hubei Province may be in summer, autumn, and early winter. The subsequent forecasting trends indicate that although electricity demand typically rises during the summer months, there may have been an oversupply of carbon allowances in the Hubei carbon market, leading to a short-term decline in carbon prices from June to August. It also reflects the carbon market's early response to the summer demand peak. From August to October, the peak agricultural production in autumn and storage energy production for winter may contribute to the significant increase in the carbon price. It effectively depicts the carbon market's expectation of energy demand in autumn

and winter and the corresponding increase in carbon quota purchasing behavior. Finally, in the early winter, the advanced storage of autumn and the energy consumption required for winter heating may increase the development trend gradually. Although the carbon price does not reach the historical high of 61.89 Yuan/tCO₂e, it remains at a mid-to-high price level.

(3) Policy recommendations

This section proposes the following policy recommendations based on the anticipated forecasting results. First, a reasonable short-term price corridor for Hubei's carbon market should be established to provide a calibrated buffer against sudden market fluctuations. The Hubei government could consider the forecasting limits (between 39.82 and 50.09 Yuan/tCO₂e) to scientifically determine the upper and lower bounds of the price corridor. In implementing the government regulation, the Hubei provincial government could refer to the EU ETS Market Stability Reserve (MSR) framework (Perino, 2024) by adjusting carbon allowance supplies to stabilize prices and mitigate potential short-term extreme volatility. The carbon price corridor could be set as the regulatory benchmark of the MSR. When market prices approach either boundary, adjustments in carbon allowance supply could counter excessive market swings, maintaining price fluctuations within this

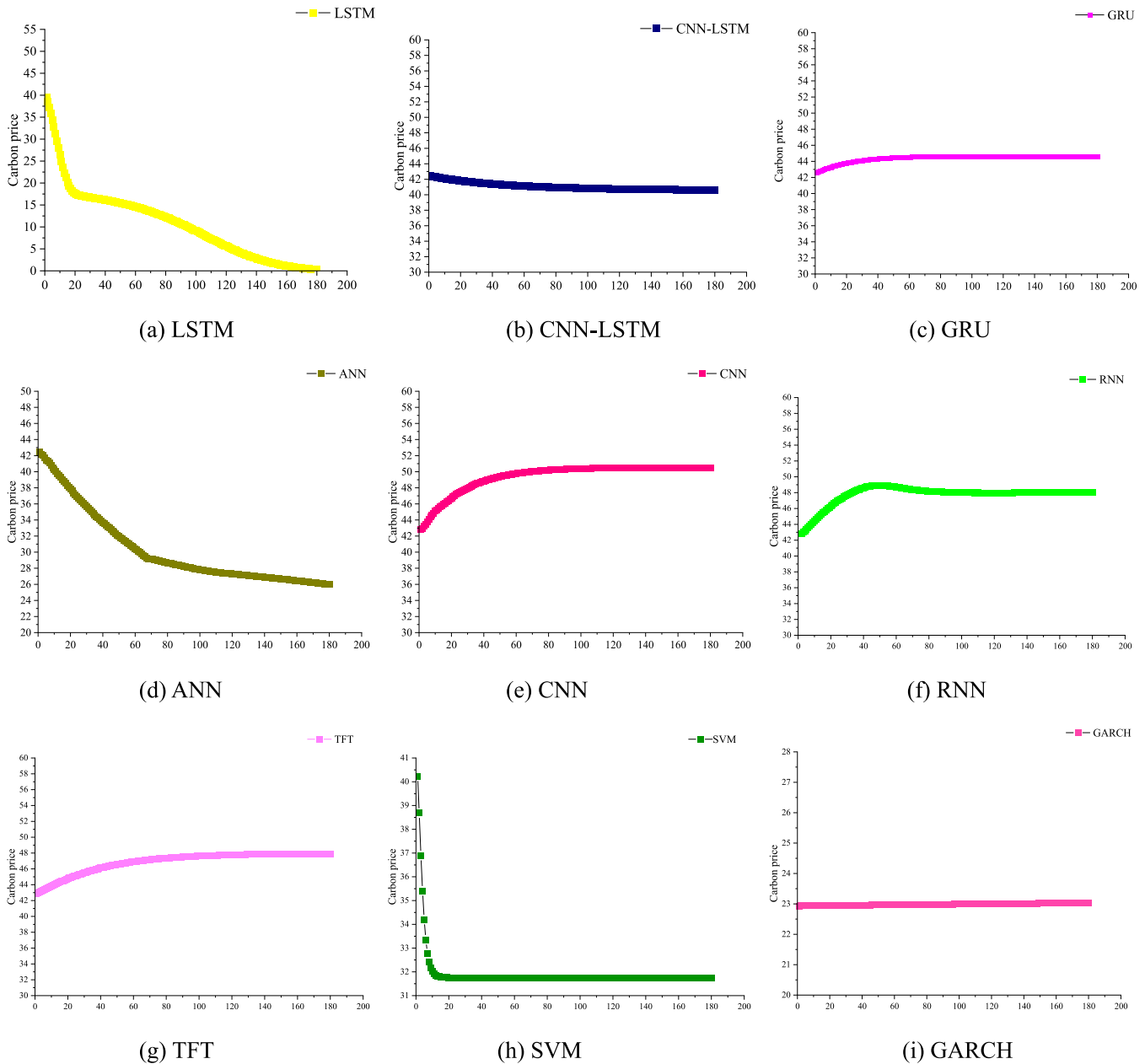


Fig. 18. The subsequent forecasting results of the comparison models.

range. The Hubei government's short-term upper and lower carbon price limits could deliver clear carbon cost signals, encouraging enterprises to invest in green technologies and process improvements proactively to guide low-carbon and cleaner production practices, fostering green and low-carbon regional development.

Second, regional carbon reduction policies should be adjusted dynamically based on the dual intervention of the “visible hand” and the “invisible hand.” The Hubei government should dynamically assess enterprises’ carbon “price tolerance walls” under Low-Carbon Following, Business-As-Usual, and Economic Following scenarios to further adjust the carbon quota allocation and carbon tax rates in Hubei’s carbon market (Sun et al., 2024). It aims to provide market participants and policymakers with a standard short-term carbon price reference. Combining the “price tolerance wall” with forecasted trends, the government could proactively regulate supply and demand in the carbon market with government intervention and enterprise alignment. The government could increase the supply of carbon allowances to stabilize

the market and balance supply and demand based on the capacity of local enterprises. In addition, Feng et al. (2024) suggest that rising carbon prices stimulate both the quantity and quality of green patents and that reasonable carbon pricing can incentivize enterprises to invest in green and low-carbon initiatives. They also point out that transparency of information is crucial for market stability. Therefore, the Hubei provincial government may consider incentivizing enterprises to increase energy-saving efforts when carbon prices are high through tax incentives and subsidies, thus enhancing the carbon market’s supply capacity. The government could disclose province carbon trading information, regularly publish market data, and further promulgate relatively complete carbon price risk prevention and control policies to establish a risk early warning and prevention and control system for reducing supply-demand imbalances caused by panic trading.

Third, promote carbon emission financing innovation and adjust the China Certified Emission Reduction (CCER) offset ratio. The average carbon price in China’s carbon market in June 2024 is 90.66 Yuan/

tCO₂e, with fluctuations typically ranging between 90 and 100 Yuan/tCO₂e. However, in Hubei Province, the carbon price forecasting may change to 39.82–50.09 Yuan/tCO₂e in the next 180 days, indicating that Hubei's carbon price is significantly lower than the national average. This presents an opportunity for an upward adjustment in Hubei's carbon price. The CCER offset ratio in Hubei Province is 5% now. Considering the potential future trend of the Hubei carbon market, the Hubei Provincial Government should further evaluate the carbon-intensive differentiation of different industries in the province and dynamically adjust the upper limits of the CCER offset ratio. Maintaining the existing 5% offsets for carbon-intensive power, steel, and chemical industries is feasible, maintaining the push for deeper emissions cuts from these high-emissions industries. However, for high-cost industries such as service and forest industries with small emission reduction potential in the province, the government can dynamically increase the offset ratio by more than 5% to alleviate and balance their cost pressure. In addition, Li et al. (2021) validated that the financing efficiency of the carbon market is crucial for green development. Therefore, it is necessary to introduce professional financial institutions into carbon trading and establish a carbon market with diversified investor participation. The Hubei government should encourage provincial financial institutions to promote carbon allowance mortgage loans and develop green bonds, sustainable development investment funds, and other financial products related to environmental protection and climate governance to form new local CCER projects. It aims to help enterprises leverage carbon assets for green financing. The government could further incentivize local enterprises to engage in low-carbon transformation and clean energy investment activities through such financial innovations.

Finally, the Chinese government could leverage the Hubei carbon market as a pilot demonstration to promote the above recommendations within the national ETS framework. The government could base short-term forecasts on projected trends in national carbon prices to determine upper and lower limits for future carbon prices, establishing a national carbon price corridor integrating the "price tolerance wall" concept to maintain the stability of the national carbon market. Additionally, the Chinese government could refer to Hubei's carbon price forecasting and the development trends of national carbon prices to establish a cross-regional Market Stability Reserve within the national ETS framework. This cross-regional adjustment mechanism could help balance supply and demand across different regions. Furthermore, the government could draw on Hubei's pilot experiences to construct a national risk management system for intensifying green and low-carbon financial activities and dynamically adjusting national carbon tax policies, thereby enhancing the rationality and effectiveness of China's ETS.

5. Conclusion and future studies

5.1. Conclusions

This paper considers carbon prices' complex nonlinear and non-stationary characteristics and proposes a novel mixed model called BO-CEEMDAN-MI-CNN-LSTM-Transformer for carbon price forecasting. Methodologically, this study improves data decomposition and reconstruction techniques through MI regression and CEEMDAN. It applies the BO algorithm to combine the CNN, LSTM, and Transformer to form the forecasting module, extracting information features from local to long-term to global levels. By applying the novel mixed model to the Hubei Province carbon price forecasting for validity verification, this paper finds that, based on ablation experiments and comparisons with popular forecasting models, introducing the MI to reconstruct the IMFs could effectively enhance the computational efficiency of the mixed model. Furthermore, the forecasting accuracy of the proposed mixed model significantly surpasses that of using a single model or direct forecasting with raw data. The DM test and IR test indicate that the reasonable synergy within the mixed forecasting module based on deep learning allows for helpful information capture from local to global

scales and enables effective multi-scale data handling, improving forecasting performance. Through the 180 days following forecasting, this paper further reveals that, unlike long-term time series where cyclical and seasonal features are insignificant, Hubei's carbon price shows evident cyclical fluctuations and seasonal characteristics in the short term, fluctuating between 39 and 50 Yuan/tCO₂e. The government should develop dynamic and flexible regulation and incentive policies that conform to the short-term cycle of carbon price in Hubei Province and cooperate with the market mechanism to stabilize the stable operation of the carbon market in Hubei.

5.2. Limitations and future studies

This paper still has some limitations that should be studied in future research. From the perspective of modeling limitations, the proposed model has preliminarily achieved effective forecasting of carbon price data. Nonetheless, there remains improved space for further development in the overall model framework. Currently, the proposed model forecasts solely from the perspective of carbon price trends, and the input data could be further added to enhance the model's adaptability to the carbon market. Future research could incorporate diversified data sources, such as market complexity, policy changes, news indices, and other relevant factors, to reduce the impact of market uncertainties and enable multi-scale forecasting. It would provide a foundation for diverse scenario forecasting and yield more comprehensive results. For the reconstruction of IMFs, this paper preliminarily determines a trade-off between feature retention and noise reduction by MI and sample entropy. However, the optimal trade-off between them still needs to be determined. Therefore, this paper may further consider the interactive relationship between feature loss and noise reduction in future work. Based on the demand scenarios focusing on noise reduction or feature retention, this paper will further explore the proposed model with the optimal trade-off between noise reduction and feature loss to improve forecasting performance. Additionally, the proposed model's forecasting architecture, which consists of CNN, LSTM, and Transformer, could be further replaced, optimized, and supplemented to enhance overall forecasting performance. This paper preliminarily discusses the roles of the relative effectiveness of CNN and Transformer in different noise levels through simulation experiments. However, the applicability of these modules to different features of various data sets remains to be demonstrated. The following research may conduct more empirical experiments around CNN and Transformer to further explore their relative effectiveness in role shifts at different noise levels through testing. Future work could continue from this perspective to optimize the integration patterns of each module and enhance the integration framework's forecasting potential. Meanwhile, this paper suggests that future research could further integrate frontier deep learning techniques, such as TFT, to replace the single LSTM and Transformer modules in the proposed forecasting architecture. Future research could focus on developing novel carbon price forecasting models by incorporating TFT with decomposition and reconstruction modules, integrating additional deep learning components, and utilizing advanced optimization algorithms.

From the perspective of application limitations, this paper needs to enhance further the applicability and generalizability of Hubei's carbon price forecasting results to other regions in China and globally. Carbon markets in different regions exhibit unique and regional variations in economic environments, policy changes, and industrial structures. Although the proposed model is adapted to the carbon price development in Hubei Province, there remain limitations in directly applying it to other regions. Therefore, in future research, this paper would further incorporate regional differences in policy, market, and economic conditions to optimize and adjust the proposed model's structure and parameters, making it adaptable to the characteristics of China's other regions to provide reasonable support for their formulation of economic policies for regional carbon markets.

Considering China's overall carbon reduction goals of peaking carbon emissions by 2030 and achieving carbon neutrality by 2060, although this paper presents a reference mixed model for carbon price forecasting, the model's capacity for handling high-dimensional data for long-time-span forecasting still requires further enhancement. The proposed model demonstrates effective short-term forecasting of 180 days, but the period is considerably shorter than the target period of 2030 and 2060. Therefore, this paper could incorporate the conditions of China's regional low-carbon development and further optimize the LSTM and Transformer modules in the model, specifically for capturing long-term and global features in the future. The future work aims to improve the model's ability to extract long-time-span trends, expand the forecasting period and time, and ultimately provide scientific guidance for carbon reduction from the carbon price and carbon market management perspective to support China's dual carbon goals in 2030 and 2060.

CRedit authorship contribution statement

Yuyang Ren: Writing – original draft, Visualization, Conceptualization. **Yiyuan Huang:** Methodology, Formal analysis. **Yuhong Wang:** Writing – review & editing, Supervision, Investigation. **Lin Xia:** Formal analysis, Data curation. **Dongdong Wu:** Software, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The research reported was partially supported by the National Natural Science Foundation of China (71871106); National Social Science Fund later funded projects (23FGLB051); the Fundamental Research Funds for the Central Universities (JUSRP1809ZD; 2019JDZD06; JUSRP321016); the Major Projects of Philosophy and Social Science Research of Guizhou Province (21GZZB32); Project of Chinese Academic Degrees and Graduate Education (2020ZDB2); Major research project of the 14th Five-Year Plan for Higher Education Scientific Research of Jiangsu Higher Education Association (ZDGG02); 2021 Wuxi Science and Technology Association key topics (KX-21-C025); Special Research Project on Education Digitalization by the Ministry of Education (CSDP24LF1G402); Key Project in Philosophy and Social Science of Wuxi (WXSK24-A-06); Special project on scientific and technological ethics (scientific research integrity) of Jiangsu Provincial Social Science Applied Research Excellent Project (24SLA-01); 2024 Wuxi Philosophy and Social Science Bidding Project (Special Project on Social Education Development) (WXSK24-JY-A01); Key-funded Project of Wuxi Association for Science and Technology (KX-24-A12); International Joint Research Laboratory for Artificial Forecasting and Decision Making Optimization at Jiangnan University.

Data availability

Data will be made available on request.

References

- Alkathery, M.A., Chaudhuri, K., 2021. Co-movement between oil price, CO₂ emission, renewable energy and energy equities: evidence from GCC countries. *J. Environ. Manag.* 297, 113350.
- Bompard, E.F., Corgnati, S.P., Grosso, D., Huang, T., Mietti, G., Profumo, F., 2022. Multidimensional assessment of the energy sustainability and carbon pricing impacts along the Belt and Road Initiative. *Renew. Sustain. Energy Rev.* 154, 111741.
- Chen, N., Liu, Z., Zhang, T., Lai, Q., Zhang, J., Wei, X., Liu, Y., 2024. Research on prediction of yellow flesh peach firmness using a novel acoustic real-time detection device and Vis/NIR technology. *LWT (Lebensm.-Wiss. & Technol.)* 209, 116772.
- Chen, P., Vivian, A., Ye, C., 2022. Forecasting carbon futures price: a hybrid method incorporating fuzzy entropy and extreme learning machine. *Ann. Oper. Res.* 313 (1), 559–601.
- Du, P., Yang, D., Li, Y., Wang, J., 2024. An innovative interpretable combined learning model for wind speed forecasting. *Appl. Energy* 358, 122553.
- Fleschutz, M., Bohlender, M., Braun, M., Henze, G., Murphy, M.D., 2021. The effect of price-based demand response on carbon emissions in European electricity markets: the importance of adequate carbon prices. *Appl. Energy* 295, 117040.
- Feng, T., Wang, X., Shi, Y., Tu, Q., 2024. The role of carbon price signal in green innovation: evidence from China. *J. Environ. Manag.* 370, 122787.
- Gao, F., Shao, X., 2022. A novel interval decomposition ensemble model for interval carbon price forecasting. *Energy* 243, 123006.
- Hong, J., Bai, Y., Huang, Y., Chen, Z., 2024. Hybrid carbon price forecasting using a deep augmented FEDformer model and multimodel optimization piecewise error correction. *Expert Syst. Appl.* 247, 123325.
- Hao, Y., Wang, X., Wang, J., Yang, W., 2024. A novel interval-valued carbon price analysis and forecasting system based on multi-objective ensemble strategy for carbon trading market. *Expert Syst. Appl.* 244, 122912.
- Jin, B., Xu, X., 2024a. Price forecasting through neural networks for crude oil, heating oil, and natural gas. *Measurement: Energy* 1 (1), 100001.
- Jin, B., Xu, X., 2024b. Forecasting wholesale prices of yellow corn through the Gaussian process regression. *Neural Comput. Appl.* 36 (15), 8693–8710.
- Liu, Z., Huang, S., 2021. Carbon option price forecasting based on modified fractional Brownian motion optimized by GARCH model in carbon emission trading. *N. Am. J. Econ. Finance* 55, 101307.
- Li, J., Liu, D., 2023. Carbon price forecasting based on secondary decomposition and feature screening. *Energy* 278, 127783.
- Li, Y., Liu, T., Song, Y., Li, Z., Guo, X., 2021. Could carbon emission control firms achieve an effective financing in the carbon market? A case study of China's emission trading scheme. *J. Clean. Prod.* 314, 128004.
- Li, D., Li, Y., Wang, C., Chen, M., Wu, Q., 2023. Forecasting carbon prices based on real-time decomposition and causal temporal convolutional networks. *Appl. Energy* 331, 120452.
- Liu, J., Wang, P., Chen, H., Zhu, J., 2022. A combination forecasting model based on hybrid interval multi-scale decomposition: application to interval-valued carbon price forecasting. *Expert Syst. Appl.* 191, 116267.
- Liu, S., Xie, G., Wang, Z., Wang, S., 2024. A secondary decomposition-ensemble framework for interval carbon price forecasting. *Appl. Energy* 359, 122613.
- Li, Z., Yang, L., Zhou, Y., Zhao, K., Yuan, X., 2020. Scenario simulation of the EU carbon price and its enlightenment to China. *Sci. Total Environ.* 723, 137982.
- Mao, Y., Yu, X., 2024. A hybrid forecasting approach for China's national carbon emission allowance prices with balanced accuracy and interpretability. *J. Environ. Manag.* 351, 119873.
- Perino, G., 2024. Carbon market design and market sentiment. *J. Environ. Econ. Manag.* 128, 103057.
- Qin, Q., Huang, Z., Zhou, Z., Chen, Y., Zhao, W., 2022. Hodrick–Prescott filter-based hybrid ARIMA–SLFNs model with residual decomposition scheme for carbon price forecasting. *Appl. Soft Comput.* 119, 108560.
- Qin, C., Qin, D., Jiang, Q., Zhu, B., 2024. Forecasting carbon price with attention mechanism and bidirectional long short-term memory network. *Energy* 299, 131410.
- Sun, Q., Chen, H., Long, R., Chen, J., 2024. Integrated prediction of carbon price in China based on heterogeneous structural information and wall-value constraints. *Energy* 306, 132483.
- Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T., 2020. DeepAR: probabilistic forecasting with autoregressive recurrent networks. *Int. J. Forecast.* 36 (3), 1181–1191.
- Soeppan, F.B., Habib, M., Zhang, Z., Nemetz, L.R., Haque, M.E., Esquino, A.M., Rivero, J. R., Bhattacharyya, D., Lipscomb, G.G., Matuszewski, M.S., Hornbostel, K.M., 2024. Optimization of a natural gas power plant with membrane and solid sorbent carbon capture systems. *Carbon Capture Sci. Technol.* 10, 100165.
- Sun, S., Jin, F., Li, H., Li, Y., 2021. A new hybrid optimization ensemble learning approach for carbon price forecasting. *Appl. Math. Model.* 97, 182–205.
- Van Belle, J., Crevits, R., Verbeke, W., 2023. Improving forecast stability using deep learning. *Int. J. Forecast.* 39 (3), 1333–1350.
- Wu, H., Du, P., 2024. Dual-stream transformer-attention fusion network for short-term carbon price prediction. *Energy* 311, 133374.
- Wang, Q., Li, S., Pisarenko, Z., 2020. Modeling carbon emission trajectory of China, US and India. *J. Clean. Prod.* 258, 120723, 2020.
- Wang, P., Liu, J., Tao, Z., Chen, H., 2022a. A novel carbon price combination forecasting approach based on multi-source information fusion and hybrid multi-scale decomposition. *Eng. Appl. Artif. Intell.* 114, 105172.
- Wang, S., Shi, J., Yang, W., Yin, Q., 2024. High and low frequency wind power prediction based on Transformer and BiGRU-Attention. *Energy* 288, 129753.
- Wang, J., Wang, Y., Li, H., Yang, H., Li, Z., 2023. Ensemble forecasting system based on decomposition-selection-optimization for point and interval carbon price prediction. *Appl. Math. Model.* 113, 262–286.
- Wang, M., Zhu, M., Tian, L., 2022b. A novel framework for carbon price forecasting with uncertainties. *Energy Econ.* 112, 106162.
- Yao, J., Chen, S., Ruan, X., 2024. Interpretable CEEMDAN-FE-LSTM-transformer hybrid model for predicting total phosphorus concentrations in surface water. *J. Hydrol.* 629, 130609.
- Ye, L., Du, P., Wang, B., 2024. Industrial carbon emission forecasting considering external factors based on linear and machine learning models. *J. Clean. Prod.* 401, 136701.

- Yang, G., Su, C., Zhang, H., Zhang, X., Liu, Y., 2023a. Tree-level landscape transitions and changes in carbon storage throughout the mine life cycle. *Sci. Total Environ.* 905, 166896.
- Yang, H., Yang, X., Li, G., 2023b. Forecasting carbon price in China using a novel hybrid model based on secondary decomposition, multi-complexity and error correction. *J. Clean. Prod.* 434, 140010.
- Yin, L., Zhou, H., 2024. Modal decomposition integrated model for ultra-supercritical coal-fired power plant reheater tube temperature multi-step prediction. *Energy* 292, 130521.
- Zhao, Y., Huang, Y., Wang, Z., Liu, X., 2024. Carbon futures price forecasting based on feature selection. *Eng. Appl. Artif. Intell.* 135, 108646.
- Zhang, W., Wu, Z., 2022. Optimal hybrid framework for carbon price forecasting using time series analysis and least squares support vector machine. *J. Forecast.* 41 (3), 615–632.
- Zhang, W., Wu, Z., Zeng, X., Zhu, C., 2023b. An ensemble dynamic self-learning model for multiscale carbon price forecasting. *Energy* 263, 125820.
- Zhang, K., Yang, X., Wang, T., The, J., Tan, Z., Yu, H., 2023a. Multi-step carbon price forecasting using a hybrid model based on multivariate decomposition strategy and deep learning algorithms. *J. Clean. Prod.* 310, 136959.
- Zhou, K., Zhang, Z., 2024. Remaining useful life prediction of lithium-ion batteries based on data denoising and improved transformer. *J. Energy Storage* 100, 113749.
- Zhang, X., Zong, Y., Du, P., Wang, S., Wang, J., 2024. Framework for multivariate carbon price forecasting: a novel hybrid model. *J. Environ. Manag.* 369, 122275.