



## OPEN ACCESS

## EDITED BY

Yingchao Hu,  
Huazhong Agricultural University, China

## REVIEWED BY

Aytaç Altan,  
Bülent Ecevit University, Türkiye  
Yuandong Yang,  
Shandong University, China

## \*CORRESPONDENCE

Lue Li,  
✉ lueli126@126.com

RECEIVED 21 November 2023

ACCEPTED 21 December 2023

PUBLISHED 11 January 2024

## CITATION

Li Z, Long J and Li L (2024), A novel machine learning ensemble forecasting model based on mixed frequency technology and multi-objective optimization for carbon trading price.

*Front. Energy Res.* 11:1341881.

doi: 10.3389/fenrg.2023.1341881

## COPYRIGHT

© 2024 Li, Long and Li. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# A novel machine learning ensemble forecasting model based on mixed frequency technology and multi-objective optimization for carbon trading price

Zejun Li<sup>1</sup>, Jun Long<sup>1</sup> and Lue Li<sup>2,3\*</sup>

<sup>1</sup>School of Mathematics and Information, Guangxi University, Nanning, China, <sup>2</sup>School of Big Data and Artificial Intelligence, Guangxi University of Finance and Economics, Nanning, China, <sup>3</sup>Guangxi Key Laboratory of Big Data in Finance and Economics, Guangxi University of Finance and Economics, Nanning, China

Carbon trading prices are crucial for carbon emissions and transparent carbon market pricing. Previous studies mainly focused on data mining in the prediction direction to quantify carbon trading prices. Although the prospect of high-frequency data forecasting mechanisms is considerable, more mixed-frequency ensemble forecasting is needed for carbon trading prices. Therefore, this article designs a new type of ensemble prediction model to increase the scope of model research. The module is divided into three parts: data denoising, mixed frequency and machine learning, multi-objective optimization, and ensemble forecasting. Precisely, the data preprocessing technology enhanced by adopting a self-attention mechanism can better remove noise and extract effective features. Furthermore, mixed frequency technology is introduced into the machine learning model to achieve more comprehensive and efficient prediction, and a new evaluation criterion is proposed to measure the optimal submodel. Finally, the ensemble model based on deep learning strategy can effectively integrate the advantages of high-frequency and low-frequency data in complex datasets. At the same time, a new multi-objective optimization algorithm is proposed to optimize the parameters of the ensemble model, significantly improving the predictive ability of the integrated module. The results of four experiments and the Mean Absolute Percent Error index of the proposed model improved by 28.3526% compared to machine learning models, indicating that the ensemble model established can effectively address the time distribution characteristics and uncertainty issues predicted by carbon trading price models, which helps to mitigate climate change and develop a low-carbon economy.

## KEYWORDS

carbon trading price forecasting, machine learning model, mixed frequency forecasting, multi-objective optimization, ensemble forecasting

## 1 Introduction

This part presents the relevant research background, a literature review, and the main contributions and innovations.

### 1.1 Research background

At present, climate change and reducing carbon dioxide emissions have become significant issues worldwide, and how to effectively reduce greenhouse gas emissions and curb global warming has become a common challenge for all countries (Wang Y. et al., 2022). The carbon market derives from the theory of environmental property rights and the theory of ecological modernization, which advocates market-based means to solve environmental problems (James and Menzies, 2022; Zhao et al., 2022; Emam and Tashkandy, 2023). It is an essential institutional innovation to promote the green and low-carbon transformation of economic development (Alhakami et al., 2022). The essence is to encourage companies to make decisions about emissions reductions by forming effective prices for trading carbon emissions through the mechanism of supply and demand (Taborianski and Pacca, 2022). The international carbon market has grown rapidly since the Kyoto Protocol came into force. In the academic community, carbon market research has attracted the attention of an increasing number of scholars (Han et al., 2019). Therefore, proposing an effective carbon trading price forecasting mechanism is crucial. The carbon trading price forecasting mechanism is the core issue in carbon Emissions (Wang F. et al., 2022; Yang et al., 2023).

### 1.2 Literature review

With vast amounts of methods to predict the stock market, researchers have devoted themselves to improving the accuracy of their models, including four types of methods: statistical technique, artificial intelligence technique, combined technique, and ensemble technique. In the literature review, we will examine the strengths and weaknesses of each of these approaches.

#### 1.2.1 Statistical model

Statistical models have been widely used to predict stock markets. These models are based on historical data and use statistical techniques to identify patterns and correlations in the data (Li et al., 2022). For instance, Chai et al. employed an enhanced GARCH model for forecasting (Chai et al., 2023), and Guan et al. utilized an ARMA model to manipulate time series proficiently (Guan and Zhao, 2017). Although statistical models can still achieve good results, traditional statistical models need help capturing price changes' complexity and nonlinear trends. Therefore, statistical models do not provide satisfactory results.

#### 1.2.2 Artificial intelligence model

Artificial intelligence (AI) models are prevalent prediction models that have become increasingly popular in recent years. Gao et al. used multiple machine learning sub-models to predict crude oil price (Gao et al., 2022). Ruobin Gao et al. demonstrated the superiority of the ESN model in time series prediction (Gao et al.,

2021). While Sun et al. used the LSTM model to predict carbon trading price, and the experiment showed that LSTM, as a deep learning model, is superior to general machine learning models (Sun and Huang, 2020). In addition, Borup et al. introduced a strategy of mixed frequency in machine learning to process high-frequency data more efficiently (Borup et al., 2023). Nevertheless, they should have capitalized on the strengths of each sub-model comprehensively. Besides, Machine learning models can become too complex during the training process, leading to overfitting of the training data and thus affecting the model's generalization ability.

#### 1.2.3 Combined model

To improve predictive ability, combined models combine multiple models to provide a more accurate prediction of stock prices (Xia et al., 2023). For example, a combined model may use a physical model to capture long-term trends, a traditional statistical model to capture short-term trends, and an AI model to identify patterns in the data. Wang et al. used the technique of calculating combined weights to combine the benefits of various machine learning models, enhancing the effectiveness of power load forecasting (Wang et al., 2021). In addition, Dong et al. introduced an LSTM model, which seamlessly integrated three sub-models, including a BPNN model (Dong et al., 2023). This combination leveraged the advantages of each sub-model and effectively captured the intricate characteristics of wave energy. However, there are challenges in determining the combined weight of each model.

#### 1.2.4 Ensemble model

Ensemble prediction can provide a more comprehensive and reliable forecast of carbon trading prices by taking advantage of the strengths of different models and mitigating their weaknesses (Yang W. et al., 2022; Guo et al., 2022). Ensemble prediction is a powerful technique that can improve the accuracy of stock market prediction (Chullamonthon and Tangamchit, 2023). Wang et al. designed an ELM-based ensemble model optimized using the Sparrow Search Algorithm (SSA) for predicting carbon trading prices (Wang et al., 2022b). The model includes a VMD data preprocessing module, multi-objective optimization, and an ELM integration module. The prediction model Seçkin Karasu et al. proposed can handle the chaos and non-linear dynamics of WTI and Brent COTS. In recent years, multi-objective optimization techniques have been widely applied to optimize the parameters of ensemble models to improve experimental results. Yang et al. used multi-objective Ant Lion Optimizer (MOALO) to optimize their prediction results (Yang H. et al., 2022). Meanwhile, Fallahi et al. also used the Multi-Objective Grey Wolf Optimizer (MOGWO) to improve the accuracy of carbon emission reduction prediction results (Fallahi et al., 2023). In addition to multi-objective optimization ensemble models, Liang Dua et al. successfully synthesized ten candidate models using a Bayesian dynamic ensemble method, which has higher adaptability and better generalization performance (Du et al., 2022). Ensemble models have also successfully addressed the non-linear features of time series prediction. For example, the digital currency prediction hybrid model proposed by Aytaç Altan et al. can capture the non-linear characteristics of digital currency time series (Altan et al., 2019). The prediction model proposed by Seçkin Karasu et al. can also

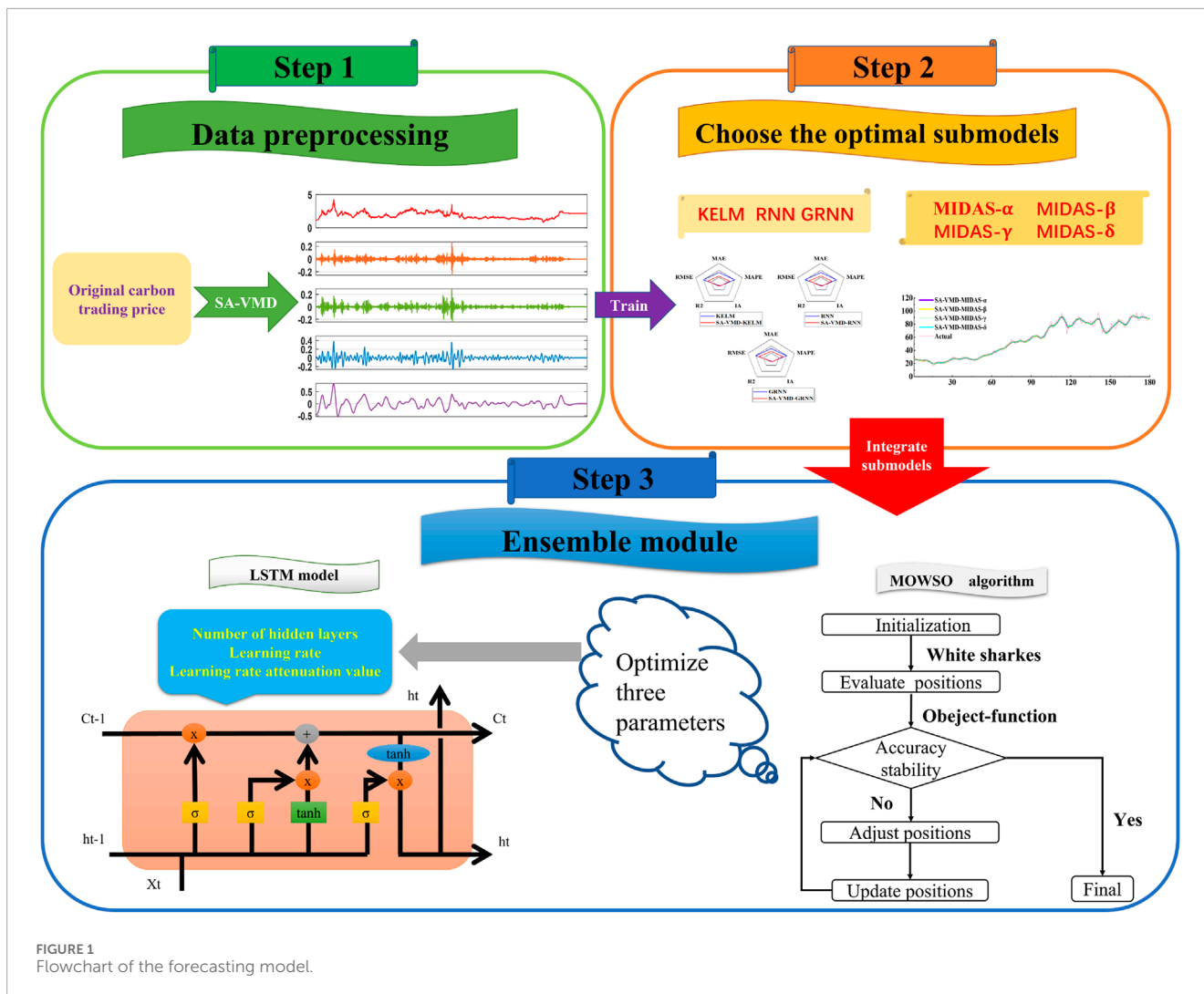


FIGURE 1 Flowchart of the forecasting model.

capture the non-linear characteristics of crude oil time series, with better prediction performance in terms of accuracy and volatility than existing prediction models (Karasu et al., 2020; Karasu and Altan, 2022). Therefore, ensemble models can better integrate the strengths of different models to improve the accuracy of carbon trading price prediction while also playing an important role in capturing non-linear features.

Although the current carbon trading price forecasting system is developing rapidly, some issues still need to be considered in future development. Early research mainly focused on common frequency research (Niu et al., 2022), with less use of mixed frequency techniques. To better handle the impact of high-frequency data on complex datasets with low-frequency data, Lahiri et al. improved their predictive ability by using mixed frequency techniques and a single model to predict New York State tax revenue (Lahiri and Yang, 2022). Chang et al. used a single mixed frequency model for carbon emission prediction (Chang et al., 2023). Jiang et al. used a single MIDAS model to explore the relationship between carbon emissions and economic growth, fully leveraging the advantages of mixed frequency techniques (Jiang and Yu, 2023). In the era of big data, researchers can obtain time series data at different frequencies, where high-frequency information has the advantages of timeliness

and stability relative to low-frequency information. In carbon trading price prediction, high-frequency data usually has a higher update frequency and can reflect the dynamic changes in market prices more timely. High-frequency data is essential for decision-makers who need to respond quickly because high-frequency data can provide more timely predictions and decision support. In addition, high-frequency data is relatively stable as it better captures the microstructure and behavior of the market, which means that high-frequency data can provide more stable prediction results, reducing the volatility and uncertainty of predictions. Therefore, mixed frequency prediction has yet to be considered in the current carbon trading price prediction field, or the impact of ensemble models and mixed frequency techniques needs to be considered.

Therefore, Some issues exist in the current carbon trading price forecasting field: Lack of technical updates in data mining, such as disregard for mixed-frequency prediction and indifference towards ensemble models utilizing mixed-frequency techniques. Specifically, most carbon trading price prediction studies use time series analysis, machine learning, and other methods to mine historical price data to predict future price trends. However, this approach often needs to pay more attention to the interactions and impacts between different frequency data. Therefore, when

TABLE 1 Parameter values utilized in the study’s methodologies.

	Parameter settings	Value	Range for variables
LSTM	MaxEpochs	50	~
	GradientThreshold	1	~
	InitialLearnRate	0.005	~
	Numer of hidden layer	60	[1,100]
	LearnRateDropPeriod	0.5	[0.1,1]
	LearnRateDropFactor	0.8	[1,200]
	MOWSO	Number of variables	3
Number of objective functions		2	~
Population size		10	~
Repository size		200	~
Maximum number of generations		50	~
Number of grids in each dimension		30	~
Maxmium vel in percentage		5	~
Maximum frequency of the wavy motion		0.75	~
Minimum frequency of the wavy motion		0.07	~
tau		4.11	~
pmin		0.5	~
pmax		1.5	~
a0		6.25	~
a1		100	~
a2		0.0005	~

building an ensemble forecasting model, it is necessary to consider the influence of both high-frequency and low-frequency data to predict carbon trading prices more accurately. In addition, the current mainstream data preprocessing method, VMD, has an issue of uncertain decomposition mode, which reduces the predictive power of the ensemble model. If the number of modes decomposed by VMD is too small, important price fluctuation features may be missed. These unconsidered features may significantly impact the prediction results, leading to a decrease in prediction accuracy. On the other hand, if the number of modes decomposed by VMD is too large, mode overlap may occur. Mode overlap means that

TABLE 2 Index of valuation.

Index	Equation
MAE	$MAE = \sum_{i=1}^L  A_i - \hat{A}_i  / L$
RMSE	$RMSE = \sqrt{\sum_{i=1}^L (A_i - \hat{A}_i)^2} / L$
MAPE	$MAPE = \sum_{i=1}^L  (A_i - \hat{A}_i) / A_i  / L \times 100\%$
IA	$IA = 1 - \sum_{i=1}^L (A_i - \hat{A}_i)^2 / \sum_{i=1}^L ( A_i - \hat{A}_i  +  \hat{A}_i - A_i )^2$
R2	$R2 = 1 - \sum_{i=1}^L (A_i - \hat{A}_i)^2 / \sum_{i=1}^L (\hat{A}_i - \bar{A}_i)^2$
U1	$U1 = \sqrt{\sum_{i=1}^L (A_i - \hat{A}_i)^2} / L / \left( \sqrt{\sum_{i=1}^L A_i^2} / L + \sqrt{\sum_{i=1}^L \hat{A}_i^2} / L \right)$
U2	$U2 = \sqrt{\sum_{i=1}^L ((A_{i+1} - \hat{A}_{i+1}) / A_i)^2} / L / \left( \sqrt{\sum_{i=1}^L ((A_{i+1} - \hat{A}_i) / A_i)^2} / L \right)$

some modes may overlap or interact, making the prediction model complex and challenging to interpret and reducing the model’s predictive power and robustness. In conclusion, there are the following issues:

- (1) Neglecting mixed-frequency prediction and having an indifferent attitude towards integrated models utilizing mixed-frequency techniques, i.e., giving less consideration to the temporal distribution characteristics of carbon trading prices.
- (2) The current mainstream data preprocessing method VMD has the problem of uncertain decomposition patterns. Additionally, both the sub-models and ensemble models suffer from parameter randomness, leading to the issue of parameter uncertainty in carbon trading price prediction.

In order to address these issues, this research aims to consider and resolve the temporal distribution characteristics and uncertainties of carbon trading prices better. Specifically, this study develops a novel ensemble forecasting method comprised of three modules to predict carbon trading prices. Firstly, the data preprocessing module employs effective methods to address the prediction uncertainties in carbon trading prices. As high-frequency data may contain more noise and fluctuations, the VMD decomposition technique has an advantage over EMD and EWT in decomposing complex and nonlinear data because it can better overcome the mode mixing problem of EMD and EWT by using a constrained variational problem, resulting in more accurate decomposition results (Lian et al., 2018; Huang et al., 2021). However, the original VMD technique may need more accurate forecasting due to randomly chosen decomposition mode numbers. Recently, self-attention mechanisms have been shown to effectively mitigate this issue by comparing the similarity between the vectors of the ground truth and predicted values (Chen et al., 2023), thereby determining the number of modal decompositions. Therefore, this study proposes a new data preprocessing method called SA-VMD to improve the drawback of uncertain mode decomposition numbers in VMD. Additionally, the current research framework mainly focuses on the influence of traditional energy sources on carbon trading prices, as mentioned by Song et al. regarding the significant impact of natural gas, fuel oil, and crude oil prices on carbon trading prices (Song et al., 2022).

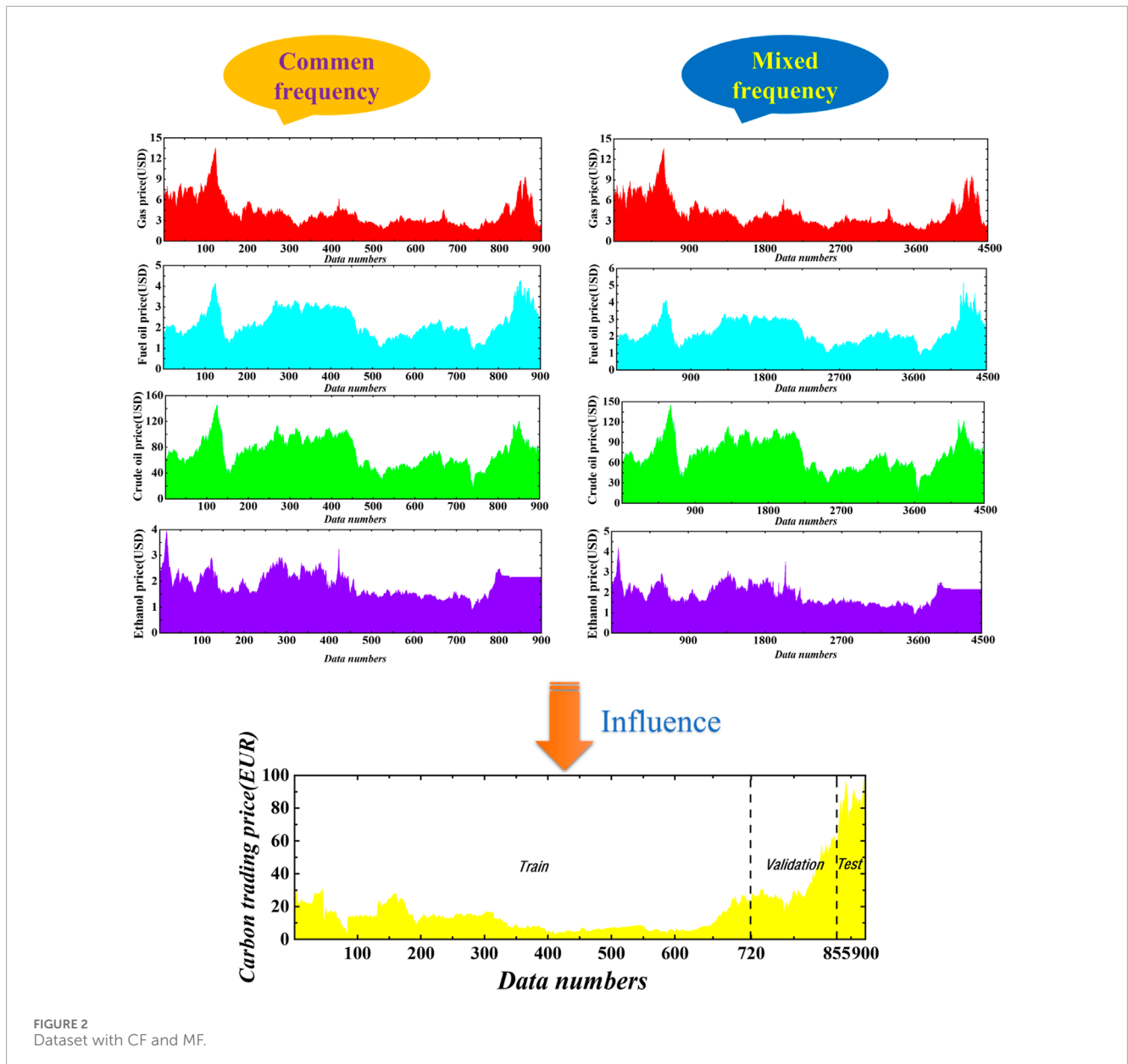


FIGURE 2  
Dataset with CF and MF.

However, the research overlooks clean energy sources (Hao and Tian, 2020). Wang et al. mention the correlation between ethanol and carbon trading price predictions (Wang et al., 2023). Therefore, this study incorporates ethanol data and considers four multivariate factors: natural gas, fuel oil, crude oil, and ethanol. Secondly, in the mixed-frequency and machine learning module, this study utilizes the Mixed Data Sampling (MIDAS) model to transform the high-frequency data of the four factors into low-frequency data for predicting carbon trading prices. Simultaneously, three low-frequency machine learning models are employed for comparative experiments in mixed-frequency prediction, which better accounts for the temporal distribution characteristics in carbon trading price prediction. Furthermore, a newly proposed evaluation criterion can better determine the models' lag order and other parameter information, thus avoiding errors caused by parameter uncertainties. Finally, LSTM is chosen as the deep

learning tool in the ensemble module due to its powerful learning capabilities. The ensemble model can fully leverage the advantages of each sub-model and capture the nonlinear features of carbon trading prices. To address the issue of parameter uncertainties in carbon trading price prediction, this study optimizes the parameters of the LSTM model based on Multi-Objective Whale Optimization (MOWSO), reducing the uncertainty and improving the accuracy and stability of price predictions. This study constructs a new framework for carbon price prediction by integrating the effects of three machine learning models based on four multivariate factors and four mixed-frequency models. This framework fully considers the temporal distribution characteristics of carbon trading price prediction by harnessing the advantages of integrating high-frequency and low-frequency data using LSTM. It aims to address current research gaps, fill in the research void, and ultimately provide a new framework for carbon price prediction.

TABLE 3 Statistical characteristics of the dataset.

Data	Carbon price	Gas price		Fuel oil price		Crude oil price		Ethanol price	
Frequency	weekly	weckly	daily	weekly	daily	weekly	daily	woekly	daily
Number	900	900	4,500	900	4,500	900	4,500	900	4,500
Median	13.1900	3.6510	3.6180	2.0734	2.0737	70.9200	70.5600	1.8605	1.8530
Average	20.5446	4.1910	4.1788	2.2452	2.2481	72.3843	72.1334	1.8998	1.8997
Minimum	0.0100	1.4950	1.4820	0.7328	0.6104	16.9400	-37.6300	0.8450	0.8200
Mlaximum	98.0100	13.5770	13.5770	4.2582	5.1354	145.2900	145.2900	3.9400	4.2300
Variance	509.8659	4.1421	4.1257	0.4814	0.5058	511.3039	510.9963	0.2211	0.2224
Standard deviation	22.5802	2.0352	2.0312	0.6939	0.7112	22.6120	22.6052	0.4702	0.4716
Skewncss	1.9444	1.4784	1.4871	0.4170	0.4751	0.2534	0.2139	0.5455	0.5581
Kurtosis	5.9113	5.3474	5.3252	2.4382	2.6832	2.5017	2.5550	3.1780	3.2846
Kolmogorov-Smirnov test	-	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>

Bold values indicate the optimal values.

### 1.2.5 Contributions and innovations

The data preprocessing modules proposed, including the mixed frequency and machine learning module and ensemble forecasting module, were used to develop a new deep learning model ensemble system. In summary, the proposed new ensemble model can effectively improve the forecasting performance of carbon trading prices and lay the foundation for better use of mixed frequency technology prediction methods.

The contributions and innovations can be summarized as follows:

- (1) Efficient data preprocessing methods. Data preprocessing technology can improve data quality and help enhance subsequent prediction processes' accuracy and performance. Previous research employed VMD data preprocessing algorithms that contained an uncertainty in the number of modes to be decomposed, leading to decreased prediction accuracy. This paper uses self-attention-variational mode decomposition (SA-VMD) to obtain better prediction results than VMD techniques to eliminate this uncertainty in the number of modes to be decomposed in VMD. The experiment shows that after data preprocessing, the sub-model improves prediction performance and stability compared to the corresponding sub-model.
- (2) Capable mixed frequency technology. To better consider the impact of different periodic factors on the prediction experiment, this article considers the effect of mixed frequencies on carbon price prediction. The introduction of mixing technology not only compensates for the previous studies that did not consider the impact of different cycle factors on the experiment but also lays a solid foundation for proposing novel ensemble prediction modules. Experiments have shown that excellent sub-models can be created using novel mixed frequency and common frequency techniques in the proposed modeling module.

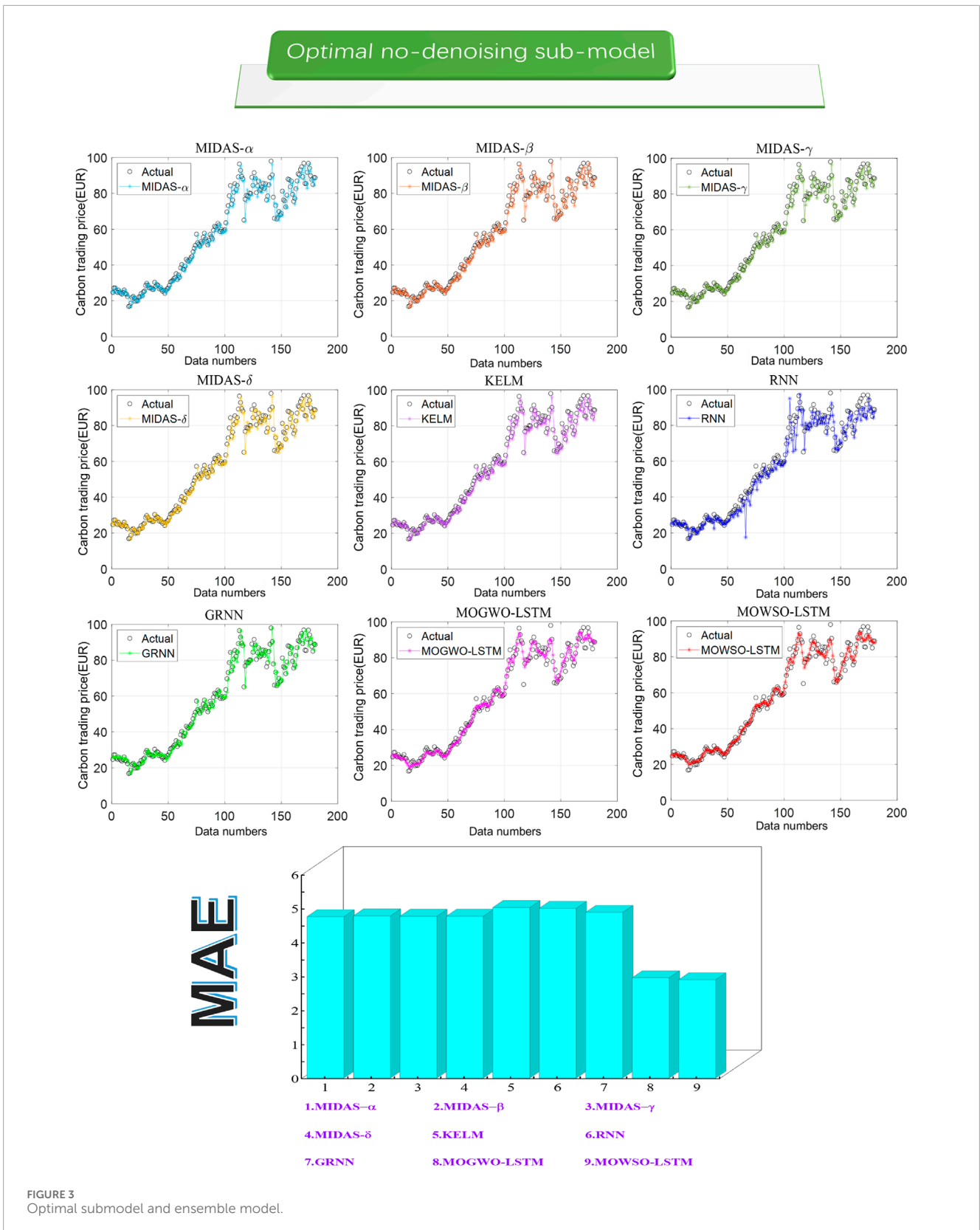
- (3) Excellent multi-objective optimizer and ensemble model. In the current carbon trading price forecasting model, only some studies simultaneously consider the effects of high-frequency and low-frequency data. To better integrate the strengths of common frequency, mixed frequency, and deep learning ensemble strategies, this article considers using the LSTM model to integrate the mixed frequency and the common frequency models. At the same time, a new recently proposed multiobjective algorithm is introduced to optimize the three parameters of the ensemble model and predict the performance based on the carbon trading price. The successful experiment proves that the new ensemble system effectively improves the accuracy and stability of predicting carbon trading prices.

## 2 Framework of the novel ensemble forecasting system for the carbon trading price

The ensemble forecasting model for carbon emissions trading developed in this paper is shown in Figure 1, and the framework is listed in detail below:

Stage 1: Since data preprocessing makes the data reliable and interpretable and enables effective investigation, SA-VMD is used to process the original data in this work. SA-VMD has high robustness to sampling and noise and can effectively avoid modal aliasing. Through the data preprocessing method, this paper effectively removes the noise and preserves the main features of the original data.

Stage 2: Measure the machine learning and mixed frequency models using new evaluation criteria to obtain the optimal partial model in each case. The evaluation criteria were successfully used to obtain the optimal sub-models of



four mixed frequencies and three machine learning models to take the best advantage of mixed frequencies, which provide a basis for the ensemble prediction of carbon trading prices.

Stage 3: Considering the powerful learning ability of machine learning, this paper selects the LSTM model in the machine learning model and uses an innovative multi-objective algorithm to find the optimal solution for the hidden LSTM layer, learning

rate, and learning rate attenuation value to obtain the final ensemble model.

### 3 Novel ensemble predicting model modules

The novel ensemble prediction model consists of the data preparation module, the mixed frequency module, the machine learning module, and the ensemble modeling module.

#### 3.1 Data preparation module

##### 3.1.1 Variational mode decomposition

VMD is a powerful signal processing technique that gained widespread acceptance in the last few years (Dragomiretskiy and Zosso, 2013). The basic idea behind VMD is to decompose a given signal into a series of elementary modes or components, each with frequency and amplitude characteristics. These components are combined to reconstruct the original signal, each component making up a specific part of the overall signal.

The mathematical basis of VMD can be complex. However, at its core, it finds a set of optimal functions (intrinsic mode functions) that can effectively represent the input signal. These mode functions are determined by solving an optimization problem that minimizes the difference between the signal and its decomposition.

Suppose  $u_p = \{u_1, u_2, \dots, u_p\}$  is the  $K$ th Intrinsic Mode Functions obtained from decomposition,  $\omega_p = \{\omega_1, \omega_2, \dots, \omega_p\}$  means the center frequency of each component.  $\delta_t$ ,  $f(t)$ , and  $*$  represent impulse function, original signal, and convolutional operator, respectively. Optimization problem mentioned above, which is also a variational question, lists as follows:

$$\begin{cases} \min_{u_p, \omega_p} \left\{ \sum_{p=1}^P \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \cdot u_p(t) \right] e^{-j\omega_p t} \right\|_2^2 \right\} \\ \text{s.t. } \sum_{p=1}^P u_p = f(t) \end{cases} \quad (1)$$

To solve this problem, the above problem can be transformed into a variational problem without constraint by an augmented Lagrangian function by introducing the penalty factor  $\alpha$  and the Lagrange multiplier  $\gamma$ .

$$\begin{aligned} L(u_p, \omega_p, \lambda) = & \alpha \sum_{p=1}^P \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \cdot u_p(t) \right] e^{-j\omega_p t} \right\|_2^2 \\ & + \left\| f(t) - \sum_{p=1}^P u_p(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{p=1}^P u_p(t) \right\rangle \end{aligned} \quad (2)$$

The problem can be solved using the alternating direction multiplier method (ADMM). The idea is to fix the other two variables and update one of them as follows:

$$\begin{aligned} u_p^{n+1} &= \arg \min_{u_p} L(\{u_{i < p}^{n+1}\}, \{\omega_p^n\}, \lambda^n) \\ \omega_p^{n+1} &= \arg \min_{\omega_k} L(\{u_p^n\}, \{\omega_{i < p}^{n+1}\}, \lambda^n) \\ \lambda^{n+1} &= \lambda^n + \rho \left( f(t) - \sum_p u_p^{n+1}(t) \right) \end{aligned} \quad (3)$$

Taking the derivative of the above equation in the frequency domain, we can conclude that:

$$\hat{u}_p^{n+1} = \frac{\hat{f}(\omega) - \sum_{i \neq p} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_p)^2} \quad (4)$$

Similarly, it can be concluded that:

$$\hat{\omega}_p^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_p(\omega)| d\omega}{\int_0^\infty |\hat{u}_p(\omega)| d\omega} \quad (5)$$

$$\hat{\lambda}^{n+1}(w) = \hat{\lambda}^n(w) + \tau \left[ \hat{f}(w) - \sum_m \hat{u}_p^{n+1}(w) \right] \quad (6)$$

##### 3.1.2 Self attention-variational mode decomposition

The self-attention (SA) mechanism is an effective method for obtaining information that has emerged in machine learning in recent years. Yuan et al. improved the EMD algorithm based on the SA mechanism (Yuan and Yang, 2023). Usually, VMD decomposition achieves the effect of data preprocessing by randomly setting the number of decomposition modes and removing the high-frequency part of the data. However, this method may have the problem of excessive error (Gao et al., 2023). In order to avoid randomly selecting the number of decomposed modes as much as possible, this article will improve the VMD noise reduction method using the SA method.

In traditional sequence models, such as recurrent neural networks (RNNs), the representation of each position can only be obtained through sequential processing to get contextual information. The self-attention mechanism allows each part to focus directly on other elements in the input sequence to receive more comprehensive contextual details. At the same time, the self-attention mechanism obtains weights by calculating the correlation between each position and other positions. Then, it applies these weights to the weighted sum of all parts in the input sequence. In this way, the representation of each location can contain information from other places.

The advantage of the self-attention mechanism is that it can calculate the correlation between various positions in parallel without relying on sequential processing like traditional recurrent neural networks. To better utilize the advantages of the self-attention mechanism to improve VMD, it is considered to determine the most suitable number of patterns by calculating the cosine angle value corresponding to the original data after data preprocessing. The smaller the absolute value of the included angle, the higher the correlation between the two. The relationship between the original data and the target can be successfully measured by calculating cosine similarity to select an appropriate number of decomposition modes.

$$\text{sim}(\widehat{Y}_m, Y_m) = \left| \frac{\widehat{Y}_m^T \cdot Y_m}{\|\widehat{Y}_m\| * \|Y_m\|} \right| \quad (7)$$

Where  $\widehat{Y}_m$  and  $Y_m$  represent the component obtained after VMD decomposition and the component of original data, respectively, while  $m = 1, 2, \dots, M$  denotes the number of decomposition modes.



### 3.2 Mixed frequency and machine learning modules

To better leverage the role of data preprocessing, it is crucial to propose effective artificial intelligence models. Mixed frequency and machine learning modules are introduced to increase the impact of different cycles on the results.

#### 3.2.1 Mixed frequency model

MIDAS (Mixed Data Sampling) regression is a method for predicting time series data that combines high-frequency and low-frequency data (Ghysels and Valkanov, 2004). This technique has been applied in various fields, including finance, economics, and environmental studies.

The basic idea of MIDAS regression is to use high-frequency data as independent variables and low-frequency data as dependent variables (Xu et al., 2021; Wichitaksorn, 2022). The approach involves fitting a regression model in which the independent variables are lagged high-frequency data, and the dependent variable is low-frequency data. The mathematical formula for the MIDAS regression model is as follows:

$$Y_t = \beta_0 + \beta_1 B(L^{1/m}; \theta) X_t^{(m)} + \varepsilon_t \tag{8}$$

where  $Y_t, X_t$  represent the low-frequency data and the high-frequency data, respectively, at time  $v$ .  $m$  is the value obtained by dividing the high frequency by the low frequency and will be 7 in this article.  $\varepsilon_t$  is the residual term.  $L^{k/m}$  lags the value of  $X_t$  backward by  $k$ , and their specific mathematical expressions are listed as follows:

$$L^{k/m} X_t^m = X_{t-k/m}^{(m)} \tag{9}$$

$$B(L^{1/m}; \theta) = \sum_{k=0}^K \omega(k; \theta) L^{k/m}$$

where  $K$  denotes the maximum delay order of  $Y_t$ , and  $\omega(k; \theta)$  is a polynomial function that includes some common functions such as the beta polynomial, etc.

A variant of the MIDAS model is the autoregressive MIDAS (AR-MIDAS) model, which extends the basic MIDAS model with autoregressive terms for the dependent variable. This allows modeling of time-varying coefficients and dynamics in the relationship between variables. The AR-MIDAS model is represented as follows:

$$Y_t = \beta_0 + \sum_{l=1}^L \eta_l Y_{t-l} + \beta_1 B(L^{1/m}; \theta) X_{t/m}^{(m)} + \varepsilon_t \tag{10}$$

Similar to the above lag of the low-frequency variables, introducing the step size  $h$  to lag the high-frequency variables also improves the result of the prediction.

$$Y_t = \beta_0 + \sum_{l=1}^L \eta_l Y_{t-l} + \beta_1 B(L^{1/m}; \theta) X_{t-h/m}^{(m)} + \varepsilon_t \tag{11}$$

For the convenience of abbreviation, the following notation will be used later. The impact of gas on carbon emissions is captured as MIDAS- $\alpha$ , while the impact of fuel oil, crude oil, and ethanol on carbon emissions are referred to as MIDAS- $\beta$ , MIDAS- $\gamma$ , and MIDAS- $\delta$ , respectively.

#### 3.2.2 Machine learning model

Artificial neural network (ANN) prediction is a type of machine learning that uses algorithms to model and predict complex relationships between inputs and outputs. The ANN consists of multiple layers of interconnected nodes or artificial neurons, which make predictions based on patterns in the data.

##### 3.2.2.1 Kernel extreme learning machine

The Extreme Learning Machine (ELM) is a powerful model that has gained popularity in recent years due to its ability to make fast and accurate predictions (Huang et al., 2006). The ELM model is a type of feedforward neural network that can be trained quickly and efficiently, making it ideal for large data sets.

The connecting weight  $\beta$  is between the hidden layer and the output layer. The output of hidden layer nodes and output are denoted as  $H, T$  respectively. ELM model can be expressed as:

$$H\beta = T' \tag{12}$$

While the minimum of the loss function can get the value of  $\beta$ .

$$\beta = (H^T H)^{-1} H^T T \tag{13}$$

The Kernel Extreme Learning Machine (KELM) is an extended version of the ELM model that includes kernel features to improve accuracy and performance. Like the ELM model, the KELM is a feedforward neural network that uses a random weight initialization process to speed up training, but instead of a single hidden layer with linear activation functions, it incorporates kernel functions to handle non-linear relationships between input and hidden layers (Huang et al., 2011). This allows the model to capture more complex relationships in the data, resulting in higher accuracy and better performance (He et al., 2020).

$$\beta = H^T \left( H^T H + \frac{I}{C} \right)^{-1} T \tag{14}$$

##### 3.2.2.2 Recurrent neural network

Recurrent neural network (RNN) are a powerful type of artificial neural network that are particularly well suited for predicting sequential data (Medsker and Jain, 2001). RNN consist of a series of interconnected nodes or “cells” that are capable of both receiving and transmitting information (Park et al., 2020).

One of the main advantages of RNN is their ability to learn long-term dependencies between inputs. In this way, the network can detect patterns in the input data that occur over long periods of time, such as seasonal trends or periodic fluctuations (Son et al., 2019).

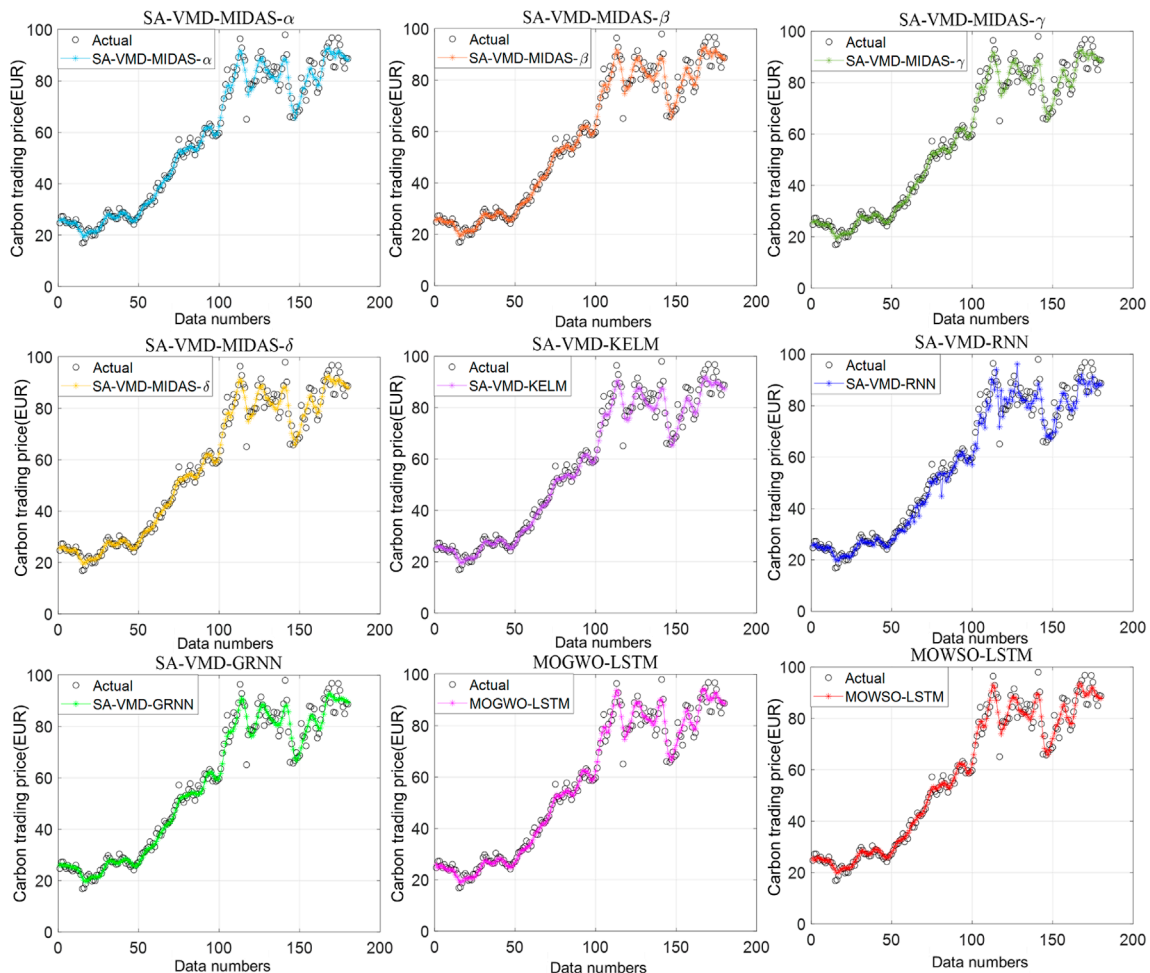
The RNN has only two inputs:  $x_t$  and  $h_{t-1}$ . These two inputs are the sample input at the current time and the hidden layer state input at the previous time. We fix  $h_{t-1}$ , so we only change  $x_t$ . Two outputs in the RNN are  $O_t$  and  $h_t$ . These two outputs are the output of the current time step and the output of the hidden layer of the current time step. The current transition of the hidden layer is used as input for the calculation of the next time step.

Assume  $\sigma, W, c$  are activation function, weight and bias, respectively. The calculation formulas from input to output  $O_t$  and  $h_t$  are listed as follows:

$$h_t = \sigma(U * x_t + V * h_{t-1} + b) \tag{15}$$

$$O_t = W * h_t + c$$

Optimal SA-VMD-sub-model



MAE



FIGURE 4 Optimal submodel with data preprocessing and ensemble model.

3.2.2.3 General regression neural network

GRNN is a network-based model which can be used for tasks such as classification, regression, and clustering

(Specht et al., 1991). The full name of GRNN is General Regression Neural Network and was proposed by American scientist Donald F. Specht.

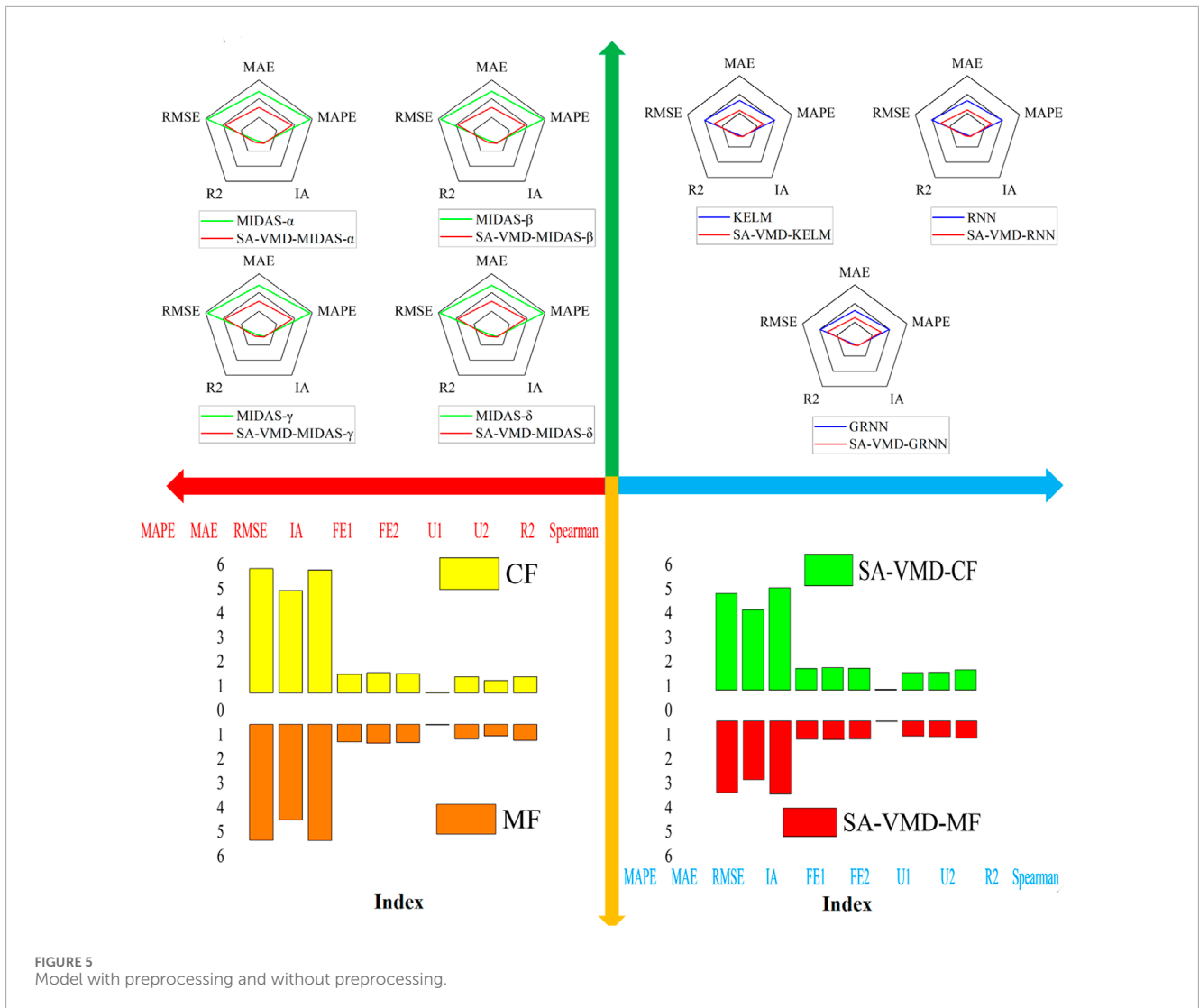


FIGURE 5 Model with preprocessing and without preprocessing.

The advantage of GRNN is that it does not require too much parameter adjustment and training time, and it works well even with small data sets. In addition, GRNN is also very resistant to noise and high interpretability, which makes it widely applicable in practise.

The theoretical basis of GRNN is nonlinear regression analysis, in which the analysis of the nonindependent variable  $Y$  with respect to the independent variable  $u$  actually computes  $v$  with the maximum probability value. Let the observed value of  $u$  be known to be  $u$ , then let the regression of  $v$  be with respect to  $X$ . Where  $X_i, Y_i$  are the sample observations of the random variables  $u, v$ .  $\sigma$  is the width coefficient of the Gaussian function. Computing a net output  $\hat{Y}(X)$  as:

$$\hat{Y}(X) = \frac{\sum_{i=1}^n Y_i e^{-\frac{(X-X_i)^T(X-X_i)}{2\sigma^2}}}{\sum_{i=1}^n e^{-\frac{(X-X_i)^T(X-X_i)}{2\sigma^2}}} \quad (16)$$

### 3.2.2.4 Long short-term memory network

LSTM is a variant of RNN to solve the long-term memory problem (Hochreiter and Schmidhuber, 1997).

LSTM has certain advantage in sequence modeling and has a long-term memory function which leads to implementation easily (Gallardo-Antolín and Montero, 2021). Meanwhile, LSTM solves the problem of gradient vanishing and gradient explosion during long sequence training (Hao et al., 2022). As a deep learning machine learning model, LSTM has the advantage of significantly improving prediction performance and has been increasingly noticed and used by scholars. Meanwhile, LSTM is a network commonly used for sequence model prediction, widely used in fields such as time series analysis (Akhtar and Feng, 2022). LSTM has strong anti-noise performance. When processing sequence data in practical applications, there are often situations where noise and outliers exist in the data (Alsulami et al., 2022).

The gate structure of LSTM (3 in total) is listed as follows:

- o Forget Gate: based on the current input and the previous state of the hidden layer, the sigmoid activation function is used to determine which information to forget.

$$f_t = \sigma(W_f \cdot [h_{t-1}, i_t] + b_f) \quad (17)$$

TABLE 4 Result of MIDAS model SA-VMD-MIDAS model.

Model	MIDAS- $\alpha$		MIDAS- $\beta$		MIDAS- $\gamma$		MIDAS- $\delta$		P-M
	N	S	N	S	N	S	N	S	S
MAPE	5.7765%	<b>3.7497%</b>	5.8036%	<b>3.7421%</b>	5.7912%	<b>3.7397%</b>	5.7901%	<b>3.7462%</b>	<b>3.3596%</b>
MAE	4.7597	<b>3.0673</b>	4.7843	<b>3.0637</b>	4.7710	<b>3.0610</b>	4.7723	<b>3.0672</b>	<b>2.7438</b>
RMSE	5.8297	<b>3.8045</b>	5.7449	<b>3.8271</b>	5.7839	<b>3.8246</b>	5.8459	<b>3.8272</b>	<b>3.5397</b>
IA	0.8786	<b>0.9431</b>	0.8814	<b>0.9424</b>	0.8808	<b>0.9425</b>	0.8777	<b>0.9424</b>	<b>0.9536</b>
FE1	0.9422	<b>0.9625</b>	0.9420	<b>0.9626</b>	0.9421	<b>0.9626</b>	0.9421	<b>0.9625</b>	<b>0.9664</b>
FE2	0.9028	<b>0.9353</b>	0.9047	<b>0.9349</b>	0.9037	<b>0.9350</b>	0.9025	<b>0.9349</b>	<b>0.9392</b>
U1	0.0351	<b>0.0228</b>	0.0346	<b>0.0229</b>	0.0348	<b>0.0229</b>	0.0352	<b>0.0229</b>	<b>0.0212</b>
U2	0.7294	<b>0.7811</b>	0.7151	<b>0.7910</b>	0.7170	<b>0.7907</b>	0.7299	<b>0.7909</b>	<b>0.7659</b>
VAR	33.4206	<b>14.7997</b>	32.5579	<b>14.9692</b>	32.8771	<b>14.9514</b>	33.4837	<b>14.9693</b>	<b>12.8063</b>
R2	0.5628	<b>0.8138</b>	0.5754	<b>0.8116</b>	0.5696	<b>0.8118</b>	0.5603	<b>0.8116</b>	<b>0.8388</b>
Spearman	0.7889	<b>0.8885</b>	0.7876	<b>0.8921</b>	0.7929	<b>0.8888</b>	0.7883	<b>0.8900</b>	<b>0.9200</b>
MDM	<b>5.2017</b>	-	<b>5.0898</b>	-	<b>5.1680</b>	-	<b>5.1966</b>	-	-
$p\_value$	<b>0.0000</b>	-	<b>0.0000</b>	-	<b>0.0000</b>	-	<b>0.0000</b>	-	-

The bold values mean the optimal values. N represents No-denoising, S represents SA-VMD, and P-M represents the proposed model.

- Input Gate: after some information is forgotten, it is necessary to add some useful information based on the last state, and at this point an input gate is used to determine what information is newly added.

$$\begin{aligned}
 m_t &= \sigma(W_m \cdot [h_{t-1}, i_t] + b_m) \\
 \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, i_t] + b_C) \\
 C_t &= f_t * C_{t-1} + m_t * \tilde{C}_t
 \end{aligned}
 \tag{18}$$

- Output Gate: the output gate determines the output at the current time based on the input at the current time, the state of the hidden layer at the previous time, and the last state.

$$\begin{aligned}
 o_t &= \sigma(W_o [h_{t-1}, i_t] + b_o) \\
 h_t &= o_t * \tanh(C_t)
 \end{aligned}
 \tag{19}$$

Where  $i_t$ ,  $f_t$ ,  $C_t$ ,  $o_t$  and  $\tilde{C}_t$  are the input value, forget gate, input gate, and input gate, the unit state, respectively.  $h_{t-1}$  is output at the previous time step.  $\sigma$  is the sigmoid function,  $W$  and  $b$  are weights and biases parameters.

In conclusion, LSTM networks offer several advantages and unique features that make them powerful tools for integrating models with common frequency and mixed frequency. Although LSTM is a widely used deep learning model, the number of hidden layers and the value of the learning rate are uncertain, which leads to its instability. The multiobjective algorithm is used to find the optimal number of hidden layers, learning rate, and learning rate attenuation value.

### 3.3 Ensemble modeling module

To enhance the results of prediction, an ensemble predictor based on multi-objective algorithms is developed as an ensemble prediction component in this paper.

#### 3.3.1 Multi-objective optimization techniques

White Shark Optimizer (WSO) is a new algorithm proposed by Malik Braik, Abdelaziz Hammuri et al., inspired by the behavior of the great white shark during hunting (Braik et al., 2022). Improved multi-objective WSO algorithm (MOWSO) was proposed to solve multi-objective optimization problems (Xing et al., 2023). The details are described below:

WSO needs to be initialized before the iteration starts, and the initial solution is represented by the following equation:

$$\omega = \begin{bmatrix} W_{11} & W_{12} & \dots & W_{1n} \\ W_{21} & W_{21} & \dots & W_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ W_{d1} & W_{d2} & \dots & W_{dn} \end{bmatrix}
 \tag{20}$$

The dimensional of the space is  $n$ , the population of  $d$  great white sharks (i.e., the population size) and the location of each great white shark can be represented as a  $d$ -dimensional matrix to represent the proposed solutions to a problem.

$$W_{ji} = L_i + r \times (U_i - L_i)
 \tag{21}$$

$W_{ji}$  is the initial vector of the  $j$ th white shark in the  $i$ th dimension.  $U_i$  and  $L_i$  represent the lower and upper bounds of

TABLE 5 Result of ANN model SA-VMD-ANN model.

Model	KELM		RNN		GRNN		P-M
	N	S	N	S	N	S	S
MAPE	6.0841%	<b>4.1554%</b>	6.0658%	<b>4.2071%</b>	5.9619%	<b>4.5739%</b>	<b>3.3596%</b>
MAE	5.0294	<b>3.4535</b>	5.0076	<b>3.5245</b>	4.8879	<b>3.7442</b>	<b>2.7438</b>
RMSE	6.0285	<b>4.3557</b>	6.1787	<b>4.5715</b>	5.9380	<b>4.7150</b>	<b>3.5397</b>
IA	0.8701	<b>0.9242</b>	0.8314	<b>0.9124</b>	0.8744	<b>0.9090</b>	<b>0.9536</b>
FE1	0.9392	<b>0.9584</b>	0.9393	<b>0.9579</b>	0.9404	<b>0.9543</b>	<b>0.9664</b>
FE2	0.9008	<b>0.9264</b>	0.8966	<b>0.9256</b>	0.8998	<b>0.9174</b>	<b>0.9392</b>
U1	0.0365	<b>0.0263</b>	0.0374	<b>0.0276</b>	0.0355	<b>0.0282</b>	<b>0.0212</b>
U2	0.7366	<b>0.7525</b>	0.8056	<b>0.7628</b>	0.7715	<b>0.7545</b>	<b>0.7659</b>
VAR	33.8879	<b>18.0991</b>	36.4603	<b>19.4217</b>	36.0214	<b>22.7236</b>	<b>12.8063</b>
R2	0.5325	<b>0.7559</b>	0.5089	<b>0.7311</b>	0.5464	<b>0.7140</b>	<b>0.8388</b>
Spearman	0.7784	<b>0.8813</b>	0.6983	<b>0.8686</b>	0.7577	<b>0.8389</b>	<b>0.9200</b>
MDM	<b>5.1632</b>	-	<b>3.1289</b>	-	<b>4.2100</b>	-	-
p_value	<b>0.0000</b>	-	<b>0.0020</b>	-	<b>0.0000</b>	-	-

The bold values mean the optimal values. N represents No-denoising, S represents SA-VMD, and P-M represents the proposed model.

TABLE 6 Result in the experiment 3.

Frequency	CF	MF	SA-VMD-CF	SA-VMD-MF	P-M
MAPE	5.8121%	<b>5.7783%</b>	4.1547%	<b>3.7444%</b>	<b>3.3596%</b>
MAE	4.7878	<b>4.7629</b>	3.4490	<b>3.0648</b>	<b>2.7438</b>
RMSE	5.7461	<b>5.7953</b>	4.3865	<b>3.8203</b>	<b>3.5397</b>
IA	0.8723	<b>0.8798</b>	0.9206	<b>0.9426</b>	<b>0.9536</b>
FE1	0.9419	<b>0.9422</b>	0.9585	<b>0.9626</b>	<b>0.9664</b>
FE2	0.9037	<b>0.9035</b>	0.9258	<b>0.9350</b>	<b>0.9392</b>
U1	0.0346	<b>0.0349</b>	0.0264	<b>0.0229</b>	<b>0.0212</b>
U2	0.7549	<b>0.7225</b>	0.7451	<b>0.7884</b>	<b>0.7659</b>
VAR	32.6157	<b>33.0158</b>	18.8969	<b>14.9187</b>	<b>12.8063</b>
R2	0.5752	<b>0.5679</b>	0.7525	<b>0.8122</b>	<b>0.8388</b>
Spearman	0.7610	<b>0.7908</b>	0.8708	<b>0.8910</b>	<b>0.9200</b>
MDM	5.1922	5.1602	3.4281	-	-
p_value	0.0000	0.0000	0.0008	-	-

The bold values mean the optimal values. P-M represents the proposed model.

TABLE 7 The comparison of ensemble models.

Model	MOALO	MOGWO	MOWSO
MAPE	3.6975%	3.6434%	<b>3.3596%</b>
MAE	3.0096	2.9608	<b>2.7438</b>
RMSE	3.7234	3.6838	<b>3.5397</b>
IA	0.9463	0.9486	<b>0.9536</b>
FE1	0.9630	0.9636	<b>0.9664</b>
FE2	0.9363	0.9367	<b>0.9392</b>
U1	0.0223	0.0220	<b>0.0212</b>
U2	0.8096	0.8067	<b>0.7659</b>
VAR	14.1117	13.6289	<b>12.8063</b>
R2	0.8216	0.8254	<b>0.8388</b>
Spearman	0.9069	0.9070	<b>0.9200</b>

The bold values mean the optimal values.

the space in the  $i$ th dimension, respectively.  $r$  is a random number created in the interval  $[0, 1]$ . And  $k$  and  $K$  stand for the current and maximum number of iterations.

The speed to the prey:

$$V_{k+1}^j = \mu \left[ V_k^j + P_1 (W_{gb} - W_k^j) \times C_1 + P_2 (W_b - W_k^j) \times C_2 \right] \quad (22)$$

Where  $j = 1, 2, \dots, d$  is the great white shark of the  $n$  large population,  $V_{(k+1)}^j$  represents the new velocity vector at step  $(k + 1)$ ,  $V_k^j$  defines the current velocity vector at step  $k$ ,  $W_{gb}$  represents the global optimal position vector achieved by each great white shark in the  $k$ th iteration so far,  $W_k^j$  is the current position vector of the first great white shark in step  $k$ ,  $W_b$  is the best known position vector of the population, and  $C_1, C_2$  are two random numbers generated uniformly in the range of  $[0, 1]$ .  $\mu$  refers to the constriction factor proposed in WSO.  $P_1$  and  $P_2$  represent the forces of the white sharks that control the effect of  $W_{gb}$  and  $W_b$  on  $V_k^j$ .

Approaching the best prey:

$$W_{k+1}^j = \begin{cases} W_k^j \times \Lambda \oplus (u \times t_1) + (u \times t_2) & \text{rand} < mv \\ W_k^j + (W_k^j / f) & \text{rand} \geq mv \end{cases} \quad (23)$$

Among them,  $t_1 = \text{sgn}(W_k^j - u) > 0$  and  $t_2 = \text{sgn}(W_k^j - l) < 0$  represent one-dimensional binary vectors determined by the formula,  $f$  denotes the frequency parameters  $mv$  of the waves,  $\Lambda$  is a negation operator,  $\oplus$  is a bit-wise xor operation, which are supposed to express the auditory and olfactory intensity of the great white sharks, increasing as an iterative function.  $rand$  defines a random number created in the range from 0 to 1, and  $u$  denotes the lower and upper limits of the search space. A smaller value  $mv$  leads to a local search, while a larger value  $mv$  leads to a global search, where

$mv$  is represented by the following equation:

$$mv = \frac{1}{\left( a_0 + e^{\frac{K/2-K}{a_1}} \right)} \quad (24)$$

Where  $a_0$  and  $a_1$  are two positive constants employed to manage exploration and exploitation behaviors.

Optimal positional motion (convergence):

$$W_{k+1}^{j'} = W_{gb} + p_1 D_w \text{sgn}(p_2 - 0.5) \quad p_3 < S \quad (25)$$

$W_{k+1}^{j'}$  is the updated location of the  $j$ th great white shark relative to its prey location,  $\text{sgn}(p_2 - 0.5)$  gives 1 or  $-1$  for the change in search direction.  $p_1, p_2$  and  $p_3$  are random numbers in the range  $[0, 1]$ ,  $D_w$  is the distance between the prey and the great white shark, and  $S$  is the odor and sight intensity of the proposed parameter great white shark when following other great white sharks near the best prey.

### 3.3.2 Ensemble forecasting module

In this section, the new ensemble strategy integrates the submodels to acquire the ultimate carbon trading price forecasting results. Considering the advantages of Deep Learning in machine learning and price forecasting (Lin et al., 2022), this paper attempts to develop a novel ensemble predicting model based on the LSTM model as an ensemble forecasting model. Although LSTM is a widely used deep learning model, the number of hidden layers and the learning rate's value are uncertain, leading to its instability. To overcome this problem, this paper considers introducing an advanced MOWSO algorithm to solve this problem. The multiobjective algorithm finds the optimal number of hidden layers, learning rate, and learning rate attenuation value. Among them, accuracy and stability are represented by  $f_1$  and  $f_2$  respectively, as shown below:

$$\min \begin{cases} f_1 = \frac{1}{L} \sum_{l=1}^L (\widehat{A}_l - A_l)^2 \\ f_2 = \text{Std}(\widehat{A}_l - A_l) \end{cases} \quad (26)$$

Where  $A_l$  and  $\widehat{A}_l$  represent the actual and predicted values, respectively.  $t = 1, 2, \dots, L$ ,  $L$  is the test sample size.

By combining the advantages of LSTM and MOWSO, a MOWSO-LSTM model is built to obtain the optimal parameter solution and objective function value. Finally, based on the MOWSO-LSTM model, a novel ensemble forecasting model is proposed, which provides an optimal prediction of carbon trading prices by integrating the advantages of each sub-model. Table 1 presents the parameter values employed in this study for the MOWSO optimization algorithm and the LSTM deep learning model.

## 4 Experiments and analysis

The evaluation index, data descriptions, and experimental design are constructed in this section.

TABLE 8 The results of ensemble model and other optimal submodels.

Model	KELM	RNN	GRNN	MIDAS- $\alpha$	MIDAS- $\beta$	MIDAS- $\gamma$	MIDAS- $\delta$	P-M
	S	S	S	S	S	S	S	S
MAPE	4.1554%	4.2071%	4.5739%	3.7497%	3.7421%	3.7397%	3.7462%	<b>3.3596%</b>
MAE	3.4535	3.5245	3.7442	3.0673	3.0637	3.0610	3.0672	<b>2.7438</b>
RMSE	4.3557	4.5715	4.7150	3.8045	3.8271	3.8246	3.8272	<b>3.5397</b>
IA	0.9242	0.9124	0.9090	0.9431	0.9424	0.9425	0.9424	<b>0.9536</b>
FE1	0.9584	0.9579	0.9543	0.9625	0.9626	0.9626	0.9625	<b>0.9664</b>
FE2	0.9264	0.9256	0.9174	0.9353	0.9349	0.9350	0.9349	<b>0.9392</b>
U1	0.0263	0.0276	0.0282	0.0228	0.0229	0.0229	0.0229	<b>0.0212</b>
U2	0.7525	0.7628	0.7545	0.7811	0.7910	0.7907	0.7909	<b>0.7659</b>
VAR	18.0991	19.4217	22.7236	14.7997	14.9692	14.9514	14.9693	<b>12.8063</b>
R2	0.7559	0.7311	0.7140	0.8138	0.8116	0.8118	0.8116	<b>0.8388</b>
Spearman	0.8813	0.8686	0.8389	0.8885	0.8921	0.8888	0.8900	<b>0.9200</b>

The bold values mean the optimal values. N represents No-denoising, S represents SA-VMD, and P-M represents the proposed model.

## 4.1 Evaluation index

To evaluate submodels and the overall model, the evaluation module of this article includes two parts: the new comprehensive evaluation index and other evaluation indicators.

### 4.1.1 Novel comprehensive evaluation indicators

This article evaluates the accuracy of carbon trading price forecasting by introducing five classical valuation indicators:

Mean Absolute Percent Error (MAPE) is a measure of relative error that uses absolute values to avoid canceling positive and negative errors. Mean absolute error (MAE) describes prediction errors, while root mean square error (RMSE) indicates the coefficient of determination of prediction errors. R2 is an indicator for assessing the goodness of fit of regression models and the index of agreement (IA) (Wang et al., 2022c).

There are various evaluation criteria but no universal method for measuring the quality of the various models. Following previous literature, this article proposes a new evaluation index to assess forecasting models' accuracy comprehensively. Comprehensive evaluation indicators (CEI) were defined based on the five typical indicators used in this article for forecast evaluation. Considering the importance of each indicator, this article considers the method of assigning weights to determine the importance of different indicators. In this study, all indicators have the same importance, so each weight of CEI is assumed to be 0.2. The detailed formula is as follows:

$$CEI = \omega_1 \cdot MAE + \omega_2 \cdot RMSE + \omega_3 \cdot MAPE + \omega_4 \cdot (1 - IA) + \omega_5 \cdot (1 - R2) \quad (27)$$

### 4.1.2 Other evaluation indicators

In order to provide a more comprehensive and accurate evaluation of the predictive effect of carbon trading prices, this paper adds some more indicators as supplements. The typical point prediction evaluation indicators Theil U-statistic 1 (U1) and Theil U-statistic 2 (U2) measure the degree of model superiority. Some evaluation indices are listed as follows in Table 2, where  $A_t$  and  $\hat{A}_t$  are the actual values and are the predicted values,  $L$  is the number of test samples, and  $\bar{A}_L$  and  $\bar{\hat{A}}$  are the average value of the actual data and the average value of the predicted values, respectively.

#### 4.1.2.1 Correlation strength

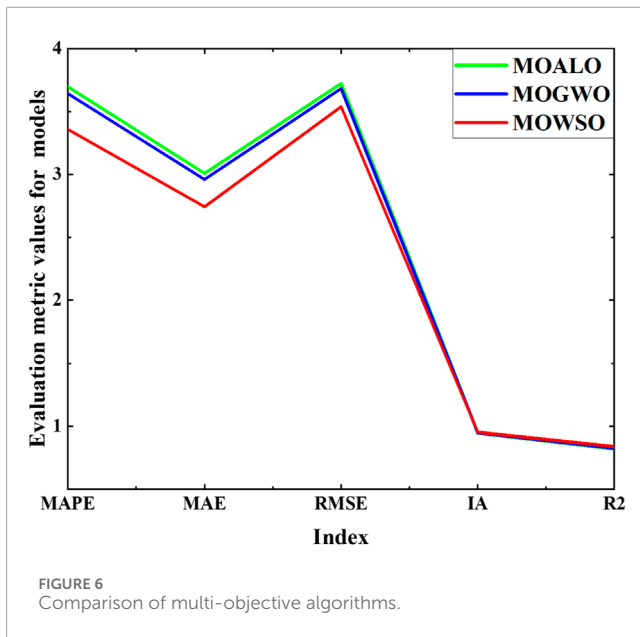
The Spearman correlation coefficient (Spearman), also called Spearman's rank correlation coefficient, is used to measure the correlation between two variables. Spearman proposed a different method, using the order of the data size instead of the numerical value itself. The correlation strength indicators can show the superiority of the innovative ensemble prediction model.

#### 4.1.2.2 Forecasting effectiveness

Prediction efficiency is an essential means of measuring prediction performance. Generally, the evaluation criteria for prediction efficiency are first calculated as an absolute percentage error. Then, the mean and variance of absolute percentage error are used to obtain first-order forecasting effectiveness (FE1) and second-order forecasting effectiveness (FE2).

#### 4.1.2.3 Forecasting stability

Besides prediction efficiency, prediction stability is a significant criterion for evaluating model performance and can be used



to evaluate ensemble models. The stability of the prediction is generally measured by the variance of the prediction error (VAR). The smaller the value VAR, the more stable the prediction.

#### 4.1.2.4 Statistical significance

The modified form of the Diebold-Mariano test (MDM) is used to determine whether there are significant differences between different data sets and whether the results of the ensemble prediction model are significantly different from those of other models.

## 4.2 Data description

To make accurate predictions about carbon emission prices, publicly available carbon trading data from the U.S. Intercontinental Exchange on the investing website were selected for this article. Looking at the classification of external factors affecting carbon emissions, it is clear that energy is the most important factor affecting the carbon trading price selected in this article. Therefore, gas, fuel oil, crude oil, and ethanol data from the same investment website were selected for this article. The impact of gas on carbon emissions is captured as MIDAS- $\alpha$ , while the impact of fuel oil, crude oil, and ethanol on carbon emissions are referred to as MIDAS- $\beta$ , MIDAS- $\gamma$ , and MIDAS- $\delta$ , respectively.

All five datasets contain high-frequency (daily) and low-frequency (weekly) price data, with weekly carbon trading prices being the target of the prediction in this article. To improve the experiments of the prediction in this article, a total of 900 weekly carbon emission price data are divided into three parts: The first part is a training set to train each sub-model to obtain prediction results, with a total of 720 data; the second and third parts contain a total of 180 data; the second part is the validation set to determine the optimal sub-model, with a total of 135 data points; the third part is

the test set to test the error size between the actual value and the test, with a total of 45 data.

Because the weekly dataset is fixed in time, this article considers controlling the start and end times of the training, validation, and testing datasets to ensure that each mixing factor consistently affects the carbon trading price. A total of 4,500 data were obtained by filling in the missing data. To increase the accuracy of the data, 900 data were obtained by processing and completing some outlier data of weekly carbon trading prices. In this paper, a mixed-frequency (MF) example of weekly carbon dioxide emissions trading prices and daily heating oil prices and a common-frequency (CF) example of weekly carbon dioxide emissions trading prices and weekly natural gas prices are selected as typical examples, as shown in Figure 2 and Table 3. The statistical characteristics tables and the two-sample Kolmogorov-Smirnov test table for all datasets are shown below and indicate that the  $p$ -value is 0.0000. Therefore, it can be assumed that a significant difference exists between the different factors, and using them as factors affecting carbon prices to achieve good results is reasonable.

## 4.3 Experimental design

Four experiments were conducted in this article to demonstrate the rationality and effectiveness of carbon price prediction. In experiment 1, the lag order is set from 1 to 6, and the CEI index is used to evaluate the optimal lag order of the different submodels and select the optimal submodel in the CF models. The MF models require the simultaneous discussion of the lag order of  $x$  and  $y$ . In the second experiment, the preprocessed submodels were compared with the corresponding submodels, confirming the superiority and effectiveness of SA-VMD decomposition. In the third experiment, the prediction effects and errors of the submodels CF and MF and the effects and errors of the preprocessed submodels CF and MF were compared. The comparison of the ensemble model with the previous submodels shows the effectiveness and superiority of the ensemble method, as shown in Experiment 4.

### 4.4 Experiment 1: establishment of optimal submodel

In carbon price prediction, it is crucial to select the optimal submodel to obtain a satisfactory ensemble model in the future. Therefore, this article considers comparisons based on data preprocessing techniques and determining the optimal lag order by CEI indicators to determine the optimal submodel. Specifically, this article selects the optimal index by creating an MF-MIDAS model with lag orders  $y$  from 1 to 6, lag orders  $x$  from 7 to 42, and intervals of 7. Moreover, the three machine learning models with the CF all have lag orders from 1 to 8, and the specific prediction results are shown in the table. In addition, the optimal submodel is the one that has the lowest CEI value. The CEI results corresponding to the different lag orders are shown in Figures 3, 4 below. Due to the consideration that the lag order table is too large, the table has been moved to the Supplementary.



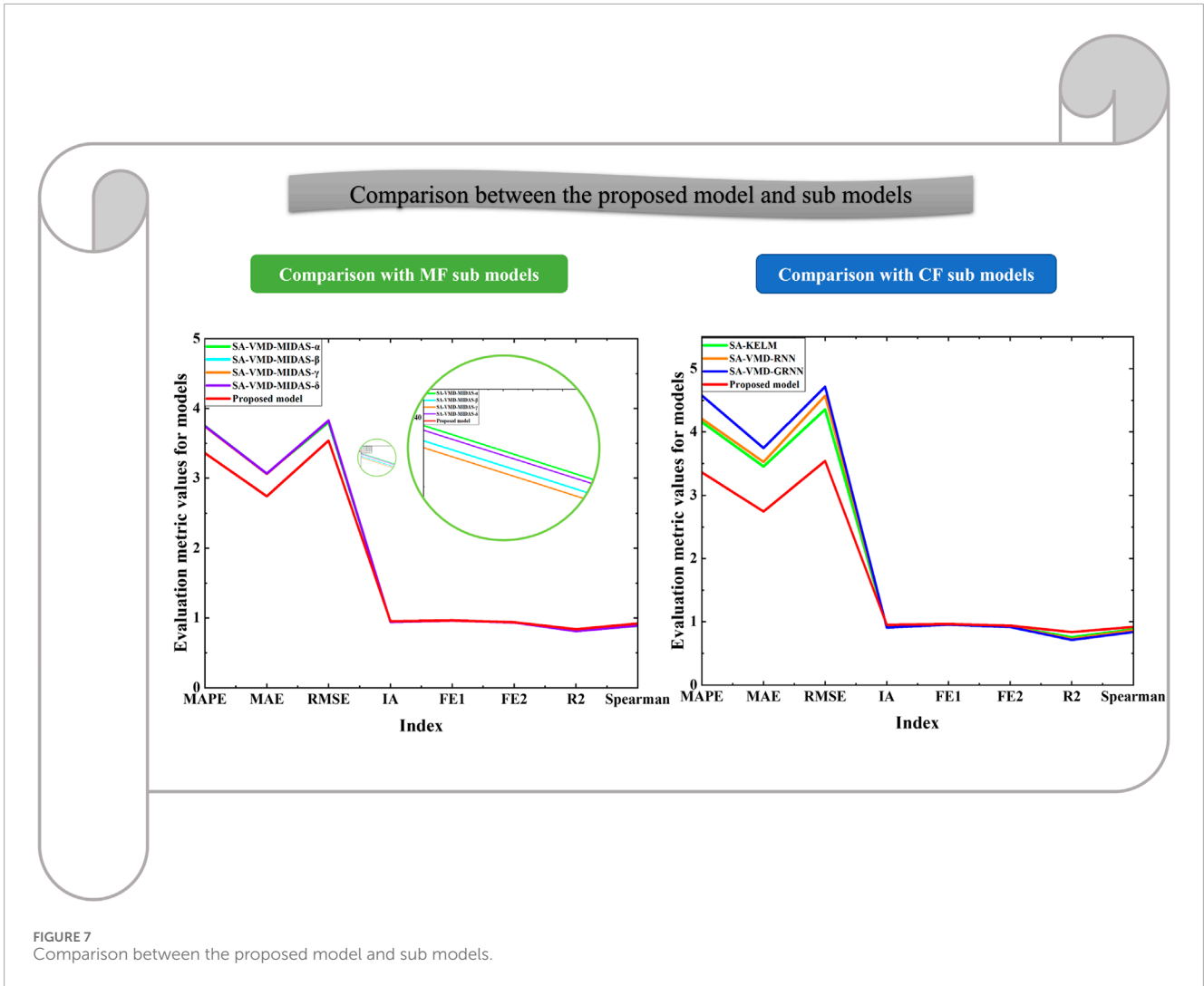


FIGURE 7 Comparison between the proposed model and sub models.

- (1) For the mixed frequency MIDAS model, as shown in the table in the appendix, the optimal AR-MIDAS model is determined by taking values from 1 to 6 and 7 to 42 for different lag orders  $y$  and  $x$ , respectively, using the lowest CEI value and polynomial comparison. For example, in MIDAS- $\alpha$  without data preprocessing, it can be found that when  $y_{lag} = 5$ ,  $x_{lag} = 35$ , and Legendre polynomials is selected as the polynomial, CEI achieves the lowest value of 3.3849, which is superior to other lag orders and polynomial selection. Therefore, this model is the optimal MIDAS- $\alpha$  sub-model at this time. Similarly, it can be seen that MIDAS- $\beta$ , MIDAS- $\gamma$ , and MIDAS- $\delta$  are the optimal sub-models when  $y_{lag} = 5$  and  $x_{lag} = 7$ , and the polynomials are Legendre polynomials. For the AR-MIDAS model after data preprocessing:  $y_{lag} = 3$ ,  $x_{lag} = 35$  when Legendre polynomials is selected as the polynomial, it is the optimal SA-VMD-MIDAS- $\alpha$  submodel;  $y_{lag} = 6$ ,  $x_{lag} = 35$ , when the polynomial is Legendre polynomials, it is the optimal SA-VMD-MIDAS- $\beta$  submodel;  $y_{lag} = 6$ ,  $x_{lag} = 35$ , when the polynomial is Legendre polynomials, it is the optimal SA-VMD-MIDAS- $\gamma$  submodel; When  $y_{lag} = 6$ ,  $x_{lag} = 35$ , and the polynomial is a Beta polynomial, it is the optimal SA-VMD-MIDAS- $\delta$  submodel.
- (2) For artificial intelligence models of the same frequency, as shown in the table in the appendix, three optimal SA-VMD-CF models (SA-VMD-KELM, SA-VMD-RNN, SA-VMD-GRNN) and three optimal CF models (KELM, RNN, GRNN) were obtained by comparing the order of 1–6 lags. For example, by comparing the CEI values corresponding to different lag orders, it is easy to see that when the lag order is 2, the SA-VMD-KELM submodel with data preprocessing has the lowest CEI value of 2.6602, which is better than the results of others. This means that a lag order of 2 is the optimal SA-VMD-KELM submodel. Similarly, the optimal lag orders for KELM, RNN, and GRNN are 1, 1, and 5, respectively. Similarly, SA-VMD-RNN and SA-VMD-GRNN, which have undergone data preprocessing, obtain the optimal submodels when the lag orders are 6 and 6, respectively.

Remark: Experiment 1 selects the optimal MIDAS and machine learning models by comparing the CEI index, laying a solid foundation for future ensemble prediction. The experimental results indicate the optimal lag order of the optimal submodel.

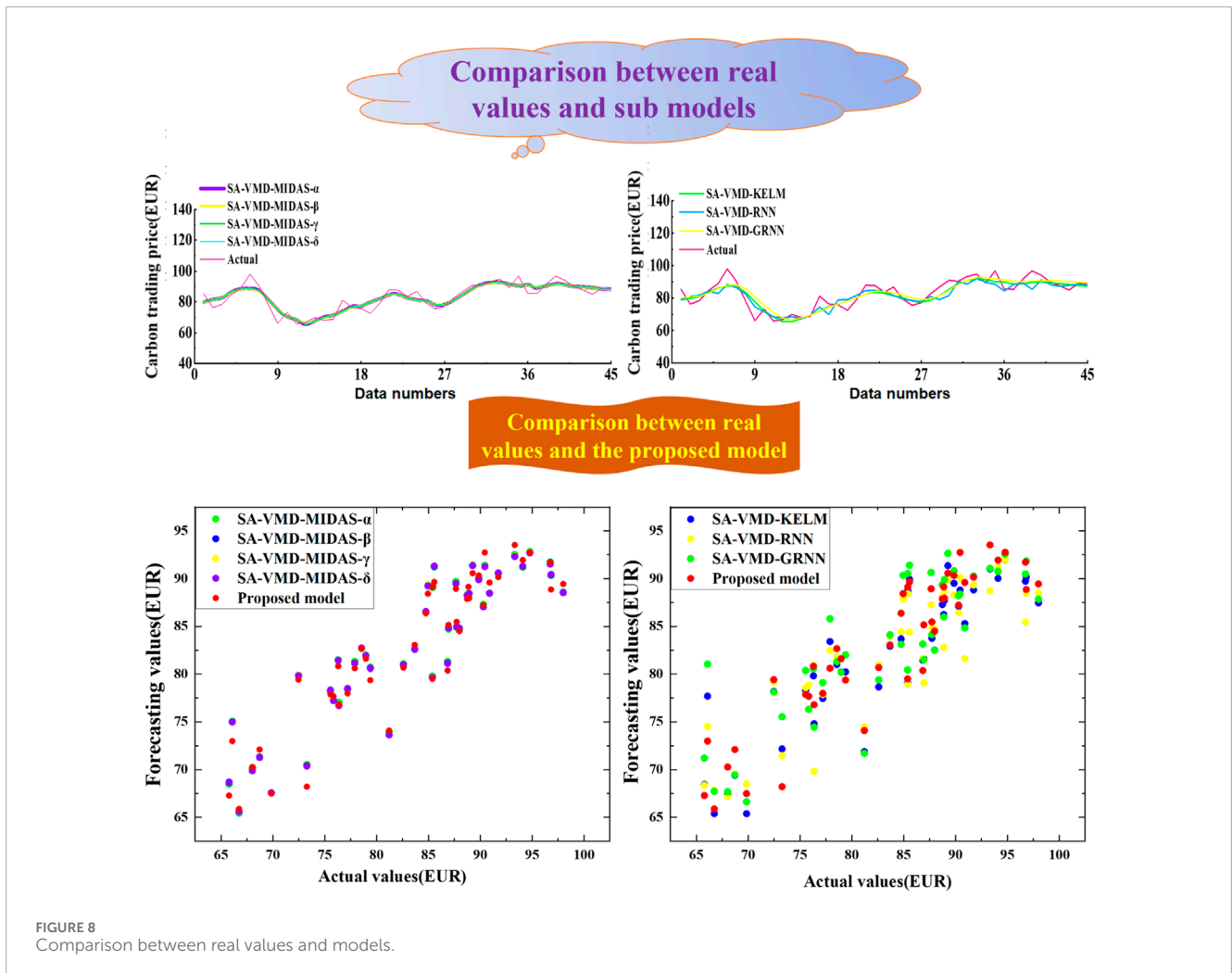


FIGURE 8 Comparison between real values and models.

### 4.5 Experiment 2: comparison of sub-models with SA-VMD and sub-models

This experiment verifies the superiority of data preprocessing and the advantages of SA-VMD decomposition by contrasting the submodels with corresponding submodels after data preprocessing. We compared the various indicators of the optimal submodel obtained in Experiment 1 and compared the indicators of the submodel with or without data preprocessing as shown in Figure 5 and Tables 4, 5 below:

- (1) The first comparison between the submodels and the corresponding submodels after data preprocessing shows that data preprocessing plays a crucial role in predicting the mixing of carbon emission prices. As shown in the table, all the blending submodels without data preprocessing have lower indicators than those decomposed with SA-VMD, indicating that the submodels decomposed with SA-VMD outperform the corresponding submodels in various aspects of prediction. For example, the MAPE indicator of MIDAS- $\alpha$  is 5.7765, while the MAPE indicator of the SA-VMD-MIDAS- $\alpha$  model is only 3.7497; After data preprocessing, the MAE decreased from 4.7597 to 3.0673, and the MDM test  $p$ -value of 0.0000 also fully proves the superiority of SA-VMD-MIDAS- $\alpha$  compared

to MIDAS- $\alpha$ . Similarly, it can be found that SA-VMD-MIDAS- $\beta$ , SA-VMD-MIDAS- $\gamma$ , and SA-VMD-MIDAS- $\delta$  all outperform their corresponding sub-models in terms of predictive indicators.

- (2) In the second comparison, the superiority and importance of SA-VMD decomposition in the CF artificial intelligence deep learning submodels can be established by comparing the corresponding submodels without SA-VMD decomposition and SA-VMD decomposition. As shown in the table, the submodels with SA-VMD technology outperform the corresponding submodels in predicting indicators in all aspects. This indicates that the submodels decomposed by SA-VMD are superior to the corresponding submodels. For example, the MAPE indicator of the KELM model is 6.0841, while the MAPE indicator of SA-VMD-KELM is only 4.1554. The MAE and RMSE decreased from 5.0294 to 6.0285 to 3.4535 and 4.3557, respectively. The result of the MDM test also proves the superiority of SA-VMD-KELM over KELM. Similarly, both SA-VMD-RNN and SA-VMD-GRNN can outperform their corresponding submodels in predicting indicators.

Remark: Experiment 2 compared the evaluation indicators of the submodels after data preprocessing with the corresponding

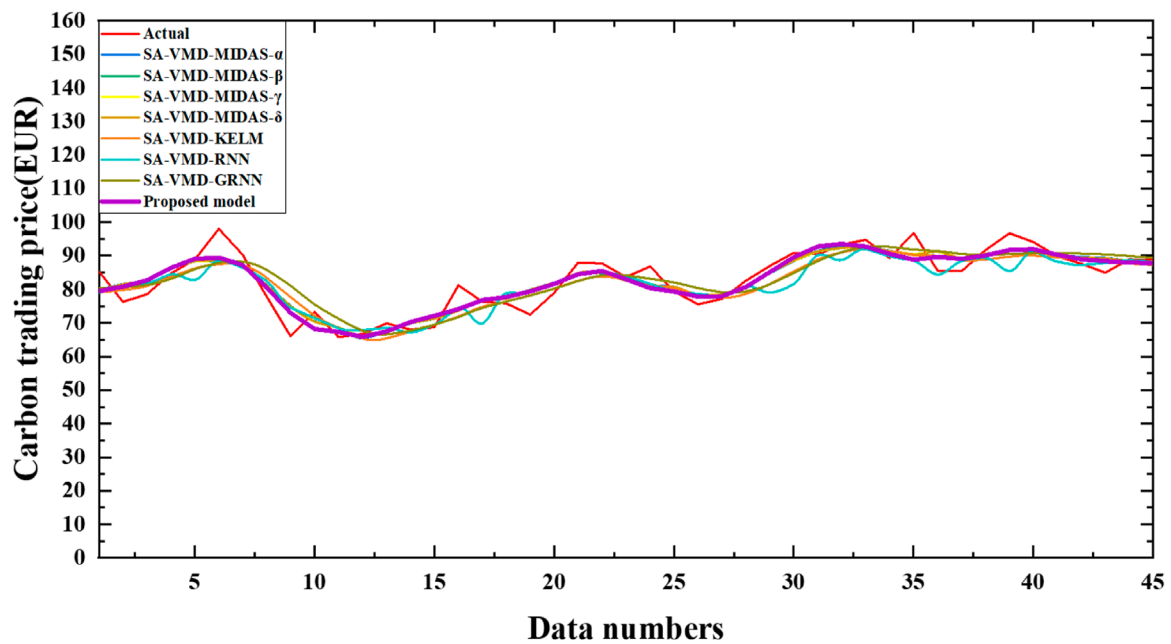


FIGURE 9  
Comparison between models.

submodels, and the results showed that the submodels after SA-VMD was significantly better in prediction than the corresponding sub-model.

#### 4.6 Experiment 3: comparison of the MF and CF models

This experiment demonstrates the important role and irreplaceability of MF in carbon emission price prediction through comparison. Also, it demonstrates the rationality and necessity of using MF sub-models in prediction. The result is listed in Table 6.

- (1) In the first comparison, several MF and CF submodels were averaged to obtain the forecasting results, as shown in Table 6 below. From the comparison of the CF model and the MF model, almost all the evaluation indices of the mixture model are better than the initial frequency, proving the mixing and common frequency prediction's effectiveness and necessity. For example, the MAPE of the common frequency model is 5.8121, while it is reduced to 5.7783 for the mixed frequency model, and the MAE is also reduced from 4.7878 to 4.7629. The results show that adding a mixed model can improve prediction ability and accuracy.
- (2) In the second comparison, it can be seen that the preprocessed SA-VMD-MF is significantly better than SA-VMD-CF, and all indicators of the mixing model after VMD decomposition are significantly better than the same frequency model after SA-VMD decomposition. For example, the MAPE of the SA-VMD-CF model has decreased from 4.1547 to 3.7444 of the SA-VMD-MF model, while the MAE has also decreased from 3.4490 to 3.0648. The comparison

results prove the effectiveness of the mixing model and once again prove the progressiveness and rationality of the preprocessing.

Remark: Experiment 3 compared the prediction performance between common frequency and mixed frequency submodels and found that the mixed frequency model outperformed the common frequency model, demonstrating the necessity and importance of incorporating a mixed frequency model in carbon trading price prediction.

#### 4.7 Experiment 4: comparison of ensemble models

To verify the superiority of the ensemble technology and fully exploit the efficiency of Deep Learning, a new ensemble strategy MOWSO-LSTM model is developed in this experiment. As a generalization of the basic RNN model, the LSTM can effectively overcome the RNN gradient explosion problem and is robust to noise. This paper measures the importance of the new integration methods for carbon price prediction by comparing the ensemble effects of MOALO-LSTM, MOGWO-LSTM, and MOWSO-LSTM. At the same time, the comparison of each optimal sub-model and the ensemble model also shows the progressiveness and effectiveness of the ensemble strategy. The comparison between the proposed model and other ensemble models and the comparison results of the optimal submodels and the proposed model are shown in Tables 7, 8. The result is listed in Figures 6–9.

- (1) The first comparison shows that each indicator of the new ensemble strategy model is superior to other ensemble models

by comparing the errors of the three ensemble models, which demonstrates the advantages of the new ensemble strategy. For example, the MAPE of MOALO-LSTM is 3.6975, while the MAPE of MOWSO-LSTM is only 3.3596. At the same time, MAE decreased from 3.0096 to 2.7438, and RMSE fell from 3.7234 to 3.5397. MOWSO-LSTM also outperforms other ensemble strategies MOGWO-LSTM in every indicator. The experiment fully proves that the ensemble strategy MOWSO-LSTM is necessary and can enhance the prediction of carbon trading prices.

- (2) The second comparison compares the optimal artificial intelligence submodels and MF with the new ensemble model after preprocessing the data. It can be seen that each indicator of MOWSO-LSTM is superior to all the individual sub-models, which once again proves the effectiveness and superiority of the ensemble strategy. For example, the MAPE of the optimal submodel of KELM after data preprocessing is 4.1554, the MAPE of MIDAS- $\gamma$  after data preprocessing is 3.7397, and the MAPE of MOWSO-LSTM is only 3.3596. The MAPE index of the proposed model improved by 28.3526% compared to machine learning models. The MAE of MOWSO-LSTM is 2.7438, and the IA is 0.9536, both of which are superior to the optimal submodels. The MAE index of the proposed model improved by 30.2597% compared to machine learning models. At the same time, the other indicators of the new ensemble model are also prominent to a single optimal submodel. The experiment once again clearly illustrates the progressiveness of the new ensemble strategy and provides an effective method to improve the price forecasting ability of carbon trading.

Remark: In Experiment 4, by comparing the LSTM model based on the deep learning strategy using different multi-objective optimization algorithms, the proposed model is superior to other multi-objective ensemble models and any optimal sub-model.

## 5 Conclusion

The current prediction of carbon trading prices faces complexities in terms of temporal distribution characteristics and uncertainties, making it challenging to achieve high accuracy. This study introduces the concept of mixed frequencies into the carbon trading price prediction field to consider the temporal distribution characteristics of carbon trading prices better to overcome these challenges. A novel integrated prediction method is developed to reduce uncertainties and improve prediction results. Specifically, this research includes a data preparation module, a mixed-frequency modeling module based on attention mechanism, and a machine learning module. The mixed-frequency and machine learning module consists of three CF (continuous frequency) models and 4 MF (mixed frequency) models, and the optimal sub-models are determined based on proposed evaluation criteria to consider the temporal distribution characteristics of carbon trading prices sufficiently. Lastly, an ensemble module based on the LSTM model is developed, taking advantage of the strengths of deep learning to reduce parameter uncertainties in carbon trading price prediction.

A multi-objective algorithm is used to find the optimal solution for LSTM parameters, resulting in the proposed model. The proposed model outperforms all individual sub-models in various indicators. For example, the MAPE of the best sub-models, SA-VMD RNN and SA-VMD MIDAS2, are 4.2071% and 3.7421%, respectively, while the MAPE of MOWSO-LSTM is only 3.5327%. The FE1 and FE2 of the proposed model are 0.9664 and 0.9392, which are better than each best sub-model.

Some specific innovations and conclusions can be listed as follows:

- (a) By successfully incorporating mixed frequencies, the prediction of carbon trading prices is improved, and the predictive performance is enhanced. Experimental results show that adequately utilizing mixed-frequency data helps reduce price prediction errors and provides a more comprehensive consideration of the temporal distribution characteristics than a single machine learning model.
- (b) The best sub-models are preprocessed using SA-VMD. The optimal solution for the varying patterns is successfully searched through iterations, and various mode functions and central frequencies are continuously updated to obtain multiple mode functions with specific bandwidths. Comparing the data preprocessing sub-models with their corresponding sub-models demonstrates the necessity of data preprocessing in price prediction, avoiding potential parameter uncertainties in data preprocessing.
- (c) New evaluation criteria based on machine learning and mixed-frequency modules are proposed, which can fully utilize various evaluation indicators to obtain the optimal sub-models. These evaluation criteria measure the performance among different models and help select the optimal sub-models, avoiding potential parameter uncertainties during the sub-model prediction process.
- (d) Based on the optimal sub-models, a deep learning-based multi-objective optimization ensemble module is proposed to enhance the results of carbon trading price prediction. It is found that MOWSO optimizes three parameters of the LSTM ensemble forecaster, avoiding parameter uncertainties that may occur in ensemble prediction.

In conclusion, the proposed novel ensemble prediction model can be used to investigate better the temporal distribution characteristics and uncertainties of carbon trading prices. It considers and overcomes the issues of data preprocessing decomposition techniques and parameter uncertainties in ensemble models while fully utilizing the temporal distribution characteristics of carbon trading prices at different frequencies. The results also demonstrate that the model has more potential and value compared to existing models, and these methods can also be applied to other prediction problems. However, there are still some limitations and prospects in this study. Specifically, although the static ensemble model performs better than each sub-model, a dynamic ensemble based on Bayesian methods may yield even better results. Furthermore, the current iteration of dynamic ensemble weights is mainly based on single accuracy evaluation indicators. Further research can explore multi-objective optimization methods

to update weights, simultaneously considering dynamic ensemble weights' accuracy and stability. In addition, regarding data preprocessing, this study only considers the closeness in the direction of vector values when determining the number of decomposition modes using the self-attention mechanism of VMD without considering the approximation in magnitude. There is still room for further optimization, and other data preprocessing methods, such as EMD and EWT, may also be applicable depending on the data's frequency and temporal distribution characteristics. Different data preprocessing methods for data with different frequencies and temporal distribution characteristics may yield better results.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://cn.investing.com/>.

## Author contributions

ZL: Investigation, Writing—original draft, Writing—review and editing. JL: Writing—review and editing. LL: Writing—review and editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work

## References

- Akhtar, M. S., and Feng, T. (2022). Detection of malware by deep learning as cnn-lstm machine learning techniques in real time. *Symmetry* 14, 2308. doi:10.3390/sym14112308
- Alhakami, H., Kamal, M., Sulaiman, M., Alhakami, W., and Baz, A. (2022). A machine learning strategy for the quantitative analysis of the global warming impact on marine ecosystems. *Symmetry* 14, 2023. doi:10.3390/sym14102023
- Alsulami, A. A., Abu Al-Hajja, Q., Alqahtani, A., and Alsini, R. (2022). Symmetrical simulation scheme for anomaly detection in autonomous vehicles based on lstm model. *Symmetry* 14, 1450. doi:10.3390/sym14071450
- Altan, A., Karasu, S., and Bekiros, S. (2019). Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. *Chaos, Solit. Fractals* 126, 325–336. doi:10.1016/j.chaos.2019.07.011
- Borup, D., Rapach, D. E., and Schütte, E. C. M. (2023). Mixed-frequency machine learning: nowcasting and backcasting weekly initial claims with daily internet search volume data. *Int. J. Forecast.* 39, 1122–1144. doi:10.1016/j.ijforecast.2022.05.005
- Braik, M., Hammouri, A., Atwan, J., Al-Betar, M. A., and Awadallah, M. A. (2022). White shark optimizer: a novel bio-inspired meta-heuristic algorithm for global optimization problems. *Knowledge-Based Syst.* 243, 108457. doi:10.1016/j.knosys.2022.108457
- Chai, F., Li, Y., Zhang, X., and Chen, Z. (2023). Daily semiparametric garch model estimation using intraday high-frequency data. *Symmetry* 15, 908. doi:10.3390/sym15040908
- Chang, T., Hsu, C.-M., Chen, S.-T., Wang, M.-C., and Wu, C.-F. (2023). Revisiting economic growth and co2 emissions nexus in taiwan using a mixed-frequency var model. *Econ. Analysis Policy* 79, 319–342. doi:10.1016/j.eap.2023.05.022
- Chen, Y., Dodwell, T., Chuaqui, T., and Butler, R. (2023). Full-field prediction of stress and fracture patterns in composites using deep learning and self-attention. *Eng. Fract. Mech.* 286, 109314. doi:10.1016/j.engfractmech.2023.109314
- Chullamonthon, P., and Tangamchit, P. (2023). Ensemble of supervised and unsupervised deep neural networks for stock price manipulation detection. *Expert Syst. Appl.* 220, 119698. doi:10.1016/j.eswa.2023.119698
- Dong, Y., Wang, J., Wang, R., and Jiang, H. (2023). Accurate combination forecasting of wave energy based on multiobjective optimization and fuzzy information granulation. *J. Clean. Prod.* 386, 135772. doi:10.1016/j.jclepro.2022.135772
- Dragomiretskiy, K., and Zosso, D. (2013). Variational mode decomposition. *IEEE Trans. Signal Process.* 62, 531–544. doi:10.1109/tsp.2013.2288675
- Du, L., Gao, R., Suganthan, P. N., and Wang, D. Z. (2022). Bayesian optimization based dynamic ensemble for time series forecasting. *Inf. Sci.* 591, 155–175. doi:10.1016/j.ins.2022.01.010
- Emam, W., and Tashkandy, Y. (2023). Modeling the amount of carbon dioxide emissions application: new modified alpha power weibull-x family of distributions. *Symmetry* 15, 366. doi:10.3390/sym15020366
- Fallahi, A., Shahidi-Zadeh, B., and Niaki, S. T. A. (2023). Unrelated parallel batch processing machine scheduling for production systems under carbon reduction policies: nsga-ii and mogwo metaheuristics. *Soft Comput.* 27, 17063–17091. doi:10.1007/s00500-023-08754-0
- Gallardo-Antolin, A., and Montero, J. M. (2021). An auditory saliency pooling-based lstm model for speech intelligibility classification. *Symmetry* 13, 1728. doi:10.3390/sym13091728
- Gao, C., Zhang, N., Li, Y., Lin, Y., and Wan, H. (2023). Adversarial self-attentive time-variant neural networks for multi-step time series forecasting. *Expert Syst. Appl.* 231, 120722. doi:10.1016/j.eswa.2023.120722
- Gao, R., Du, L., Duru, O., and Yuen, K. F. (2021). Time series forecasting based on echo state network and empirical wavelet transformation. *Appl. Soft Comput.* 102, 107111. doi:10.1016/j.asoc.2021.107111

was funded by the Guangxi Science and Technology Project (No. GuikeAD21220082), Guangxi Key Laboratory of Big Data in Finance and Economics (No. FED2202), Guangxi First-class Discipline Statistics Construction Project (Nos. 2022SXYP08, TJYLXKDSJ 2022A02), Guangxi Natural Science Foundation (Nos. 2020GXNSFAA159014, 2021GXNSFAA220114).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fenrg.2023.1341881/full#supplementary-material>

- Gao, X., Wang, J., and Yang, L. (2022). An explainable machine learning framework for forecasting crude oil price during the covid-19 pandemic. *Axioms* 11, 374. doi:10.3390/axioms11080374
- Ghysels, E., and Valkanov, R. (2004). The midas touch: mixed data sampling regression models. *Cirano Work. Pap.* 5, 512–517.
- Guan, S., and Zhao, A. (2017). A two-factor autoregressive moving average model based on fuzzy fluctuation logical relationships. *Symmetry* 9, 207. doi:10.3390/sym9100207
- Guo, Y., Guo, J., Sun, B., Bai, J., and Chen, Y. (2022). A new decomposition ensemble model for stock price forecasting based on system clustering and particle swarm optimization. *Appl. Soft Comput.* 130, 109726. doi:10.1016/j.asoc.2022.109726
- Han, M., Ding, L., Zhao, X., and Kang, W. (2019). Forecasting carbon prices in the shenzhen market, China: the role of mixed-frequency factors. *Energy* 171, 69–76. doi:10.1016/j.energy.2019.01.009
- Hao, X., Liu, Y., Pei, L., Li, W., and Du, Y. (2022). Atmospheric temperature prediction based on a bilstm-attention model. *Symmetry* 14, 2470. doi:10.3390/sym14112470
- Hao, Y., and Tian, C. (2020). A hybrid framework for carbon trading price forecasting: the role of multiple influence factor. *J. Clean. Prod.* 262, 120378. doi:10.1016/j.jclepro.2020.120378
- He, W., Xie, Y., Lu, H., Wang, M., and Chen, H. (2020). Predicting coronary atherosclerotic heart disease: an extreme learning machine with improved salp swarm algorithm. *Symmetry* 12, 1651. doi:10.3390/sym12101651
- Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735
- Huang, G.-B., Zhou, H., Ding, X., and Zhang, R. (2011). Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man, Cybern. Part B Cybern.* 42, 513–529. doi:10.1109/TSMCB.2011.2168604
- Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2006). Extreme learning machine: theory and applications. *Neurocomputing* 70, 489–501. doi:10.1016/j.neucom.2005.12.126
- Huang, Y., Dai, X., Wang, Q., and Zhou, D. (2021). A hybrid model for carbon price forecasting using garch and long short-term memory network. *Appl. Energy* 285, 116485. doi:10.1016/j.apenergy.2021.116485
- James, N., and Menzies, M. (2022). Global and regional changes in carbon dioxide emissions: 1970–2019. *Phys. A Stat. Mech. Its Appl.* 608, 128302. doi:10.1016/j.physa.2022.128302
- Jiang, W., and Yu, Q. (2023). Carbon emissions and economic growth in China: based on mixed frequency var analysis. *Renew. Sustain. Energy Rev.* 183, 113500. doi:10.1016/j.rser.2023.113500
- Karasu, S., and Altan, A. (2022). Crude oil time series prediction model based on lstm network with chaotic henry gas solubility optimization. *Energy* 242, 122964. doi:10.1016/j.energy.2021.122964
- Karasu, S., Altan, A., Bekiros, S., and Ahmad, W. (2020). A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. *Energy* 212, 118750. doi:10.1016/j.energy.2020.118750
- Lahiri, K., and Yang, C. (2022). Boosting tax revenues with mixed-frequency data in the aftermath of covid-19: the case of New York. *Int. J. Forecast.* 38, 545–566. doi:10.1016/j.ijforecast.2021.10.005
- Li, X., Zhang, X., and Li, Y. (2022). High-dimensional conditional covariance matrices estimation using a factor-garch model. *Symmetry* 14, 158. doi:10.3390/sym14010158
- Lian, J., Liu, Z., Wang, H., and Dong, X. (2018). Adaptive variational mode decomposition method for signal processing based on mode characteristic. *Mech. Syst. Signal Process.* 107, 53–77. doi:10.1016/j.ymssp.2018.01.019
- Lin, Y., Liao, Q., Lin, Z., Tan, B., and Yu, Y. (2022). A novel hybrid model integrating modified ensemble empirical mode decomposition and lstm neural network for multi-step precious metal prices prediction. *Resour. Policy* 78, 102884. doi:10.1016/j.resourpol.2022.102884
- Medsker, L. R., and Jain, L. (2001). Recurrent neural networks. *Des. Appl.* 5, 64–67. doi:10.1007/978-1-4842-6513-0\_8
- Niu, X., Wang, J., Wei, D., and Zhang, L. (2022). A combined forecasting framework including point prediction and interval prediction for carbon emission trading prices. *Renew. Energy* 201, 46–59. doi:10.1016/j.renene.2022.10.027
- Park, J., Yi, D., and Ji, S. (2020). Analysis of recurrent neural network and predictions. *Symmetry* 12, 615. doi:10.3390/sym12040615
- Son, G., Kwon, S., and Park, N. (2019). Gender classification based on the non-lexical cues of emergency calls with recurrent neural networks (rnn). *Symmetry* 11, 525. doi:10.3390/sym11040525
- Song, X., Wang, D., Zhang, X., He, Y., and Wang, Y. (2022). A comparison of the operation of China's carbon trading market and energy market and their spillover effects. *Renew. Sustain. Energy Rev.* 168, 112864. doi:10.1016/j.rser.2022.112864
- Specht, D. F. (1991). A general regression neural network. *IEEE Trans. Neural Netw.* 2, 568–576. doi:10.1109/72.97934
- Sun, W., and Huang, C. (2020). A novel carbon price prediction model combines the secondary decomposition algorithm and the long short-term memory network. *Energy* 207, 118294. doi:10.1016/j.energy.2020.118294
- Taborski, V. M., and Pacca, S. A. (2022). Carbon dioxide emission reduction potential for low income housing units based on photovoltaic systems in distinct climatic regions. *Renew. Energy* 198, 1440–1447. doi:10.1016/j.renene.2022.08.091
- Wang, F., Jiang, J., and Shu, J. (2022a). Carbon trading price forecasting: based on improved deep learning method. *Procedia Comput. Sci.* 214, 845–850. doi:10.1016/j.procs.2022.11.250
- Wang, H., Shi, W., He, W., Xue, H., and Zeng, W. (2023). Simulation of urban transport carbon dioxide emission reduction environment economic policy in China: an integrated approach using agent-based modelling and system dynamics. *J. Clean. Prod.* 392, 136221. doi:10.1016/j.jclepro.2023.136221
- Wang, J., Cui, Q., and He, M. (2022b). Hybrid intelligent framework for carbon price prediction using improved variational mode decomposition and optimal extreme learning machine. *Chaos, Solit. Fractals* 156, 111783. doi:10.1016/j.chaos.2021.111783
- Wang, J., Zhang, H., Li, Q., and Ji, A. (2022c). Design and research of hybrid forecasting system for wind speed point forecasting and fuzzy interval forecasting. *Expert Syst. Appl.* 209, 118384. doi:10.1016/j.eswa.2022.118384
- Wang, X., Gao, X., Wang, Z., Ma, C., and Song, Z. (2021). A combined model based on eobl-ssvm for power load forecasting. *Symmetry* 13, 1579. doi:10.3390/sym13091579
- Wang, Y., Li, X., and Liu, R. (2022d). The capture and transformation of carbon dioxide in concrete: a review. *Symmetry* 14, 2615. doi:10.3390/sym14122615
- Wichitaksorn, N. (2022). Analyzing and forecasting Thai macroeconomic data using mixed-frequency approach. *J. Asian Econ.* 78, 101421. doi:10.1016/j.asieco.2021.101421
- Xia, Y., Wang, J., Zhang, Z., Wei, D., and Yin, L. (2023). Short-term pv power forecasting based on time series expansion and high-order fuzzy cognitive maps. *Appl. Soft Comput.* 135, 110037. doi:10.1016/j.asoc.2023.110037
- Xing, Q., Wang, J., Jiang, H., and Wang, K. (2023). Research of a novel combined deterministic and probabilistic forecasting system for air pollutant concentration. *Expert Syst. Appl.* 228, 120117. doi:10.1016/j.eswa.2023.120117
- Xu, Q., Liu, S., Jiang, C., and Zhuo, X. (2021). Qrnn-midas: a novel quantile regression neural network for mixed sampling frequency data. *Neurocomputing* 457, 84–105. doi:10.1016/j.neucom.2021.06.006
- Yang, H., Li, P., and Li, H. (2022a). An oil imports dependence forecasting system based on fuzzy time series and multi-objective optimization algorithm: case for China. *Knowledge-Based Syst.* 246, 108687. doi:10.1016/j.knosys.2022.108687
- Yang, R., Liu, H., and Li, Y. (2023). An ensemble self-learning framework combined with dynamic model selection and divide-conquer strategies for carbon emissions trading price forecasting. *Chaos, Solit. Fractals* 173, 113692. doi:10.1016/j.chaos.2023.113692
- Yang, W., Tian, Z., and Hao, Y. (2022b). A novel ensemble model based on artificial intelligence and mixed-frequency techniques for wind speed forecasting. *Energy Convers. Manag.* 252, 115086. doi:10.1016/j.enconman.2021.115086
- Yuan, E., and Yang, G. (2023). Sa-emd-lstm: a novel hybrid method for long-term prediction of classroom pm2.5 concentration. *Expert Syst. Appl.* 230, 120670. doi:10.1016/j.eswa.2023.120670
- Zhao, J., Kou, L., Jiang, Z., Lu, N., Wang, B., and Li, Q. (2022). A novel evaluation model for carbon dioxide emission in the slurry shield tunnelling. *Tunn. Undergr. Space Technol.* 130, 104757. doi:10.1016/j.tust.2022.104757