



OPEN

## Carbon price prediction based on decomposition technique and extreme gradient boosting optimized by the grey wolf optimizer algorithm

Mengdan Feng<sup>1✉</sup>, Yonghui Duan<sup>1</sup>, Xiang Wang<sup>2</sup>, Jingyi Zhang<sup>1</sup> & Lanlan Ma<sup>1</sup>

It is essential to predict carbon prices precisely in order to reduce CO<sub>2</sub> emissions and mitigate global warming. As a solution to the limitations of a single machine learning model that has insufficient forecasting capability in the carbon price prediction problem, a carbon price prediction model (GWO–XGBOOST–CEEMDAN) based on the combination of grey wolf optimizer (GWO), extreme gradient boosting (XGBOOST), and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) is put forward in this paper. First, a random forest (RF) method is employed to screen the primary carbon price indicators and determine the main influencing factors. Second, the GWO–XGBOOST model is established, and the GWO algorithm is utilized to optimize the XGBOOST model parameters. Finally, the residual series of the GWO–XGBOOST model are decomposed and corrected using the CEEMDAN method to produce the GWO–XGBOOST–CEEMDAN model. Three carbon emission trading markets, Guangdong, Hubei, and Fujian, were experimentally predicted to verify the model's validity. Based on the experimental results, it has been demonstrated that the proposed hybrid model has enhanced prediction precision compared to the comparison model, providing an effective experimental method for the prediction of future carbon prices.

Climate change has evolved into a formidable menace to the survival of humanity in the twenty-first century. Greenhouse gases are considered a major factor contributing to global warming<sup>1</sup>. To cope with the global warming crisis, the international community has actively reduced carbon emissions by formulating climate policies and other measures. Among them, the European Emissions Trading System (EU-ETS) was implemented in 2005, reducing carbon emissions and energy consumption<sup>2</sup>. Furthermore, China plays a significant role in international climate protection as one of the top carbon emitters worldwide. China has implemented eight carbon trading pilots in various regions, namely Beijing (2013), Shanghai (2013), Guangdong (2013), Tianjin (2013), Shenzhen (2013), Chongqing (2014), Hubei (2014), and Fujian (2016), in order to reduce global emissions<sup>3</sup>.

Carbon trading has emerged as an emerging financial industry. A carbon price reflects fluctuations in supply and demand for carbon energy within the carbon emissions market, where carbon energy can be traded as a commodity<sup>4</sup>. Because of the uncertainty of the internal mechanism and external factors, carbon prices demonstrate nonlinear and non-stationary features<sup>5,6</sup>. The risks associated with carbon trading are greater than those associated with traditional financial products. Accurate carbon price forecasting not only helps governments grasp the changes in market conditions and make reliable decisions, but also helps enterprises and investors grasp the characteristics of carbon prices. This will make sensible resource allocations and realize the value-added of carbon assets. As a result, it is crucial to establish a system that is stable and effective for the research of carbon prices.

In accordance with the previous literature review, carbon price research can be categorized into two classifications: models based on historical data<sup>7–9</sup> and models based on influencing factors<sup>10–13</sup>.

<sup>1</sup>Department of Civil Engineering, Henan University of Technology, No. 100, Lianhua Street, Gaoxin District, Zhengzhou 450001, China. <sup>2</sup>Department of Civil Engineering, Zhengzhou University of Aeronautics, No. 15, Wenyuan West Road, Zhengdong New District, Zhengzhou 450015, China. ✉email: mengdanfeng139499@foxmail.com

Grounded on historical data, carbon price forecasting methods can be classified into three categories: statistical and econometric methods, artificial intelligence (AI), and integration methods.

In the past, statistical and econometric methods were extensively employed for forecasting carbon prices as classical time series forecasting methods. Main statistical methods are the autoregressive integrated moving average model (ARIMA)<sup>14</sup>, generalized autoregressive conditional heteroskedasticity model (GARCH)<sup>15</sup>, gray model (GM)<sup>16</sup>, etc. For instance, Carolina et al. (2013) employed an ARIMA model in order to forecast carbon prices, ultimately achieving more accurate predictive outcomes<sup>17</sup>. According to Dutta (2018), an exponential GARCH model was used for forecasting carbon price volatility, and outliers were processed to improve accuracy<sup>18</sup>. Under the assumption of linearity, the statistical and econometric methods perform well for short-term forecasting, but when forecasting nonlinear, non-stationary time series of carbon prices, the prediction accuracy is not satisfactory<sup>19</sup>.

As a result, AI that does not require linear assumptions is broadly utilized across different sectors. For example, credit risk prediction<sup>20</sup>, disease treatment<sup>21</sup>, and traffic congestion<sup>22,23</sup>. For carbon price prediction, least squares support vector machines (LSSVM) and artificial neural networks (ANN) are commonly used. Using 1074 daily carbon price results, Atsalakis (2016) developed a neural network (NN) model to predict time series. ANN was found to be the most effective method for predicting carbon prices based on the final results<sup>24</sup>. Zhu et al. (2016) introduced an adaptive multiscale integrated learning approach grounded in LSSVM to effectively capture the non-stationary and non-linear attributes of carbon prices. The findings demonstrated that their proposed model surpassed the performance of the ARIMA and GARCH models<sup>25</sup>. Despite the fact that AI exceeds traditional statistical models in forecasting non-linear and non-stationary data, a single AI model fails to possess sufficient forecasting stability and does not meet researchers' expectations for accurate carbon price predictions across different markets<sup>26</sup>.

Given the constraints of conventional statistical approaches in handling non-stationary feature data and the shortcomings of a single AI model, experts have started to focus on researching integrated methods to boost data analysis and forecasting precision. A number of decomposition methods have been proposed based on different theoretical foundations, including the wavelet transform (WT)<sup>27</sup>, variational mode decomposition (VMD)<sup>28</sup>, and ensemble empirical mode decomposition (EEMD)<sup>29</sup>. E et al. (2019) realized that carbon valence has nonlinear and nonstationary properties. To address this issue, they combined VMD with a gated recurrent unit (GRU) to predict carbon prices' future trends. Experimental results confirmed its validity and reliability<sup>30</sup>. Jinpei Liu et al. (2019) employed empirical mode decomposition (EMD) and a reconstruction algorithm to transform the original data into three subseries of varying frequencies. Subsequently, they individually analyzed these three types of data using ARIMA, partial least squares (PLS), and NN methods. The findings demonstrated the superior predictive performance of the model<sup>31</sup>. Using EEMD to preprocess the data, Zhou et al. (2018) constructed different combinations of models to identify different frequencies. The existing hybrid models, although they enhance carbon price prediction accuracy, have drawbacks<sup>32</sup>. For example, existing hybrid models usually have model subseries obtained from decomposition without considering noise. This can reduce prediction accuracy and efficiency<sup>29</sup>.

Carbon prices are impacted by a combination of historical data and external factors. The existing literature primarily utilizes carbon price time series data for the modeling process. However, the dynamics of carbon trading prices are influenced by various factors, including energy factors, macroeconomic factors, and industry structures<sup>11</sup>. In general, since external factors can be analyzed, carbon price forecasting built upon multiple influencing factors is important for carbon market research. Therefore, carbon price prediction models for influencing factors are favored by scholars. Using oil, coal, and natural gas prices as the basis, Tsai and Kuo (2013) devised an ant-based radial basis function network (ARBFN) model for carbon price prediction. The inclusion of multiple influencing factors in carbon price forecasting models can indeed pose challenges due to the potential for error accumulation. When considering multiple factors, the complexity of the model increases, and uncertainties associated with each factor can accumulate throughout the forecasting process<sup>33</sup>.

Reviewing previous studies, we identify potential research gaps in the prediction of carbon prices. One is that most carbon price forecasting models rely only on past carbon price data series. They ignore the impact of external factors on the carbon market. This limitation may result in models that do not adequately take into account the full range of market conditions when forecasting carbon prices. Second, most current carbon price forecasting models fail to fully explore and utilize other useful information. Useful information means that after the model prediction, there are still a large number of nonlinear residual sequences, which are not random walks<sup>34</sup> and still contain carbon price information. Ignoring the residual series leads to the potential problem of incomplete information in predicting carbon prices. To this end, it is of particular importance that the above issues are addressed and a new perspective on carbon price forecasting is proposed.

In order to bridge these gaps, this study first established the index system of influencing factors of carbon price and selected indicators by the random forest method to find out the main factors affecting carbon price, so as to improve the prediction accuracy of the model. Secondly, XGBOOST is used to establish the carbon price prediction model. Meanwhile, with the objective of avoiding the prediction error caused by the parameter setting of the XGBOOST model, GWO is used to find the optimization of the model parameters. To increase the precision of model predictions, the residual series of XGBOOST predictions is corrected using the CEEMDAN method, and a combined GWO–XGBOOST–CEEMDAN model is derived. The contributions can be summarized as follows:

- (1) In the majority of prior studies, carbon price forecasts relied on historical time series data on carbon prices. This ignores the effects of multiple factors when predicting carbon prices, so there are limits to the information that can be provided and the extent to which carbon markets can be managed. In this study, multiple influencing factors are considered in conducting carbon price forecasts with the aim of addressing the problem of carbon price forecasting. In order to develop a richer indicator system that is more appropriate

- to China's national conditions, the carbon price time series data as well as the various influencing factors are treated as candidate input features for carbon price modeling.
- (2) In this study, Partial Autocorrelation Function (PACF) and Random Forest (RF) are introduced as feature selection methods to build carbon price prediction models in accordance with numerous influencing factors and reduce the influence of redundant information between features. A significant improvement has been made in the model's prediction performance.
  - (3) Most previous hybrid models first decompose the data and then perform carbon price prediction studies. However, this study adopts a different approach by first predicting carbon prices and then decomposing the residual series. After the carbon price information is predicted by the strong master model, the useful information of the residual sequence is difficult to obtain, so the CEEMDAN algorithm is used to further process the residual information and decompose it into modal information that is easy to extract and a sequence that is more difficult to extract. This is to dig deeper into residual effective information. According to the experiment, carbon price prediction is more accurate and practical than most previous studies. The method of prediction and then decomposition offers innovative thought for carbon price prediction research, and it will serve as a strong reference in the future.

## Algorithm introduction

### Feature election

#### Random forest

After the initial selection of 11 metrics, feature screening is performed next. It can enhance the model's ability to generalize, reduce the risk of overfitting, reduce the computational complexity of the model, etc. Common feature selection methods are Gray correlation, Pearson correlation coefficient, and random forests (RF). Gray correlation and Pearson correlation coefficient are both linear relationship-based methods, while RF can handle more complex nonlinear relationships. This means RF can select features in a wide range of situations. As a result, in this paper, the RF method is used for screening carbon price primary indicator systems.

Based on the results of the RF method, the primary features are ranked in terms of importance and then selected. Consider a sample size of  $A$  and a feature dimension of  $m$ . Provide a set of training samples  $\{(x_1, y_1), \dots, (x_N, y_N)\}$  and create a self-help sample set  $C_t$  of size  $A$ ;  $K_t$  is obtained by classification and regression tree (CART) on  $C_t$ ; Taking a random sample of  $m_{try} = \sqrt{m}$  features from each tree and selecting the most significant  $m_{try}$  features for node splitting; Analyzing whether  $t$  satisfies  $t \leq ntree$  until the loop is not exited, and then generating  $G = Uniform(\{K_t\})$ .

In the calculation of feature importance, the Gini Index is used as a segmentation function to calculate "Gini Importance" as the degree of importance of a feature. This can be expressed as follows:

$$Gini(C) = 1 - \sum_{i=1}^{|E|} [F_i]^2 \quad (1)$$

$C$  represents the sample set;  $F_i$  represents the probability of belonging to the  $i$ th class in the sample set  $C$ ; There are a number of sample classes in  $E$ . The Gini index of the sample set  $C$  is defined when feature  $G$  is known.

$$Giniindex(C, G) = \sum_{H=1}^H \frac{|C^H|}{C} Gini(C^H) \quad (2)$$

$H$  represents the number of features  $G$  values, i.e.,  $C$  is divided into  $H$  subsets according to the feature  $G$  values  $\{C^1, C^2, \dots, C^H\}$ , and the samples within each subset are of the same feature  $G$  value.  $G$  feature that has the smallest Gini index after division is considered to be the optimal feature in the selection process.

#### Partial auto-correlation function

PACF is a statistical tool for time series analysis that helps determine the relationship between each observation in a time series and its lag values. Its function is to recognize the order of the AR (Autoregressive) model in a time series, i.e., how many lags need to be considered in that model. The PACF model actually adjusts the autocorrelation function (ACF) by eliminating the part already explained by the previous lags so that the remaining part more accurately reflects the relationship between the observations and the lags at the current moment.  $(X_t, X_{t+v}|X_{t+1}, \dots, X_{t+v-1})$  represents the conditional correlation between  $X_t$  and  $X_{t+v}$  after removing the effects of the intervening variables  $X_{t+1}, \dots, X_{t+v-1}$ , i.e., the partial autocorrelation between  $X_t$  and  $X_{t+v}$ .

### CEEMDAN model

Empirical mode decomposition (EMD) is to decompose the nonlinear and non-stationary raw data into inherent mode functions (IMFs) with various fluctuation scales. However, due to the intermission of the raw data, mode confusion is easy to occur. This will affect the decomposition effect. Wu<sup>35</sup> proposed an ensemble empirical mode decomposition (EEMD) method by adding a certain degree of Gaussian white noise to the original data for repeated decomposition. Although the mode overlap phenomenon can be effectively solved, residual white noise still exists in the component of the eigenmode function derived by this method, resulting in low reconstruction accuracy. Building upon this, Torres<sup>36</sup> moved to the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) method, which addresses the issue of significant reconstruction errors in the EEMD method by introducing adaptive white noise at each stage. Therefore, in this essay, the CEEMDAN method is used to forecast each component of the eigenmode function and the trend term separately.

CEEMDAN can be broken down as follows:

*Step 1* As a result of adding a Gaussian white noise sequence to the residual sequence, an updated sequence with noise is obtained:

$$\bar{y}_i(t) = y(t) + \sigma n_i(t), i = 1, 2, \dots, N \quad (3)$$

where  $y(t)$  is the residual sequence, and  $\bar{y}_i(t)$  is the new sequence with the addition of Gaussian white noise;  $n_i(t)$  denotes the white noise added to the residual data;  $\sigma$  is the adaptive coefficient.

*Step 2* EMD decomposition is performed on the new sequence with white noise added to obtain  $N$  modal components, and the first modal component of CEEMDAN is obtained by the overall averaging of the  $N$  modal components as follows:

$$imf_1(t) = \frac{1}{N} \sum_{i=1}^N imf_{1i}(t) \quad (4)$$

At this point,  $R_1(t)$  is the residual component.

$$R_1(t) = \bar{y}_i(t) - imf_1'(t) \quad (5)$$

*Step 3* The adaptive white noise sequence  $\sigma n_i(t)$  is added to  $R_1(t)$  to form a new sequence  $R_1(t) + \sigma E_1(n_i(t))$  with noise, where  $E_j(\cdot)$  is the  $j$ th eigenmodal component obtained after EMD decomposition. At this point, the EMD decomposition is performed on the new sequence and averaged to obtain the second modal component and the residual component as follows:

$$imf_2(t) = \frac{1}{N} \sum_{i=1}^N E_1(R_1(t) + \sigma_1 E_1(n_i(t))) \quad (6)$$

$$R_2(t) = R_1(t) - imf_2'(t) \quad (7)$$

*Step 4* Repeat the above three steps to obtain the  $(j + 1)$ th modal component and the  $j$ th residual component:

$$imf_{j+1}(t) = \frac{1}{N} \sum_{i=1}^N E_j(R_j(t) + \sigma_j E_j(n_i(t))) \quad (8)$$

$$R_j(t) = R_{j-1}(t) - imf_j'(t) \quad (9)$$

*Step 5* Repeat the above steps until the CEEMDAN can no longer be decomposed by EMD. Finally, the original sequence  $y(t)$  is decomposed into multiple eigenmodal components and a trend component.

$$y(t) = imf(t) + R_{es}(t) \quad (10)$$

After CEEMDAN has decomposed the residual series, the GWO-XGBOOST model is applied to each eigenfunction component. The final residual forecast is derived by linearly combining the results of each component.

### XGBOOST model

Extreme gradient boosting (XGBOOST) was developed by Chen et al.<sup>37</sup> in 2016, which integrates a linear scale solver with a categorical regression tree learning algorithm. The model combines models with low prediction accuracy through certain strategies. The purpose of this is to construct an integrated model that is more accurate in terms of prediction. During the model training process, XGBOOST optimizes the boosting process. Each iteration generates an updated decision tree to fit the residuals generated in the previous iteration. XGBOOST can continuously improve its prediction accuracy and generalization capacity through iterative optimization. While traditional gradient boosting decision tree (GBDT) methods utilize only first-order derivatives, XGBOOST does a second-order Taylor expansion of the loss function, controls model complexity by introducing regularization terms to avoid overfitting problems, and employs a more refined evaluation approach when splitting nodes to better capture the nonlinear relationships between features. In recent years, the XGBOOST model has shown superior performance in financial risk control, medical health, natural language processing, and other fields. This model is based on the following mathematical principles:

An integration model for the definition tree can be described as follows:

$$\hat{y}_i = \sum_{m=1}^M f_m(x_i), f_m \in F \quad (11)$$

where  $\hat{y}_i$  is the prediction value;  $M$  is the number of decision trees;  $F$  is the tree selection space;  $x_i$  is the first  $i$  input feature.

XGBOOST's loss function is as follows:

$$Q = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{m=1}^M \theta(f_m) \quad (12)$$

The first part of the function is the prediction error between the predicted value and the real training value of the XGBOOST model, and the second part represents the complexity of the tree, which is mainly used to control the regularization of the model complexity:

$$\theta(f_m) = \gamma T + \frac{1}{2} \tau \|\omega\|^2 \quad (13)$$

where  $\gamma$  and  $\tau$  are penalty factors.

By adding an incremental function  $f_t(x_i)$  to Eq. (13), the value of the loss function is minimized. Then the objective function of the  $t$  th time is

$$Q_{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{m=1}^M \theta(f_m) = \sum_{i=1}^n l(y_i, \hat{y}_i^{t-1} + f_t(x_i)) + \theta(f_t) \quad (14)$$

The second-order Taylor expansion of Eq. (15) is used to approximate the objective function, and the set of samples in each child of the  $j$  tree is defined as  $I_j = \{i | q(x_i = j)\}$ . At this point the  $Q_{(t)}$  can be approximated as

$$Q_{(t)} \cong \sum_{j=1}^T \left[ \left( \sum_{i \in I_j} g_i \right) \omega_j + (1/2) \left( \sum_{i \in I_j} h_i + \tau \right) \omega_j^2 \right] + \gamma T \quad (15)$$

where  $g_i = \partial_{\hat{y}_i^{t-1}} l(y_i, \hat{y}_i^{t-1})$  is the first order derivative of the loss function;  $h_i = \partial_{\hat{y}_i^{t-1}}^2 l(y_i, \hat{y}_i^{t-1})$  is the second order derivative of the loss function. Defining  $G_i = \sum_{i \in I_j} g_i$ ,  $H_i = \sum_{i \in I_j} h_i$  then we have:

$$Q_{(t)} \cong \sum_{j=1}^T [G_j \omega_j + (1/2)(H_j + \tau) \omega_j^2] + \gamma T \quad (16)$$

The partial derivative of  $\omega$  yields

$$\omega_j = -G_j / (H_j + \tau) \quad (17)$$

By incorporating weights into the objective function, we get

$$Q_{(t)} \cong -(1/2) \sum_{j=1}^T G_j^2 / (H_j + \tau) + \gamma T \quad (18)$$

A large portion of the model's performance is determined by parameter selection during the training process of the XGBOOST model. There are 23 hyperparameters in the XGBOOST algorithm, mainly divided into general parameters for macroscopic function control, booster parameters for booster detail control, and learning target parameters for training target control. The GWO-XGBOOST combinatorial model combines the three hyperparameters that have a significant impact on the performance of XGBOOST (learning\_rate, n\_estimators, and max\_depth) as the position vector of the head wolf  $\alpha$  in the GWO algorithm and continuously updates them through the iterations of the GWO algorithm to continuously find the optimal position until the global optimal position is output as the final parameter of the XGBOOST model.

### GWO model

A pack intelligence optimization algorithm, the grey wolf optimizer (GWO), based on the predatory behavior of grey wolves, was proposed by Mirjalili et al.<sup>38</sup> in 2014, inspired by the predatory behavior of grey wolves. The optimization process of the GWO algorithm can be analogized to the hunting behavior of the gray wolf pack. Among them,  $\alpha$ ,  $\beta$ , and  $\delta$  wolves with the highest social level in each generation of the population act as the leaders of the gray wolf pack. A predator searches, encircles, and attacks prey to achieve its optimization goal. GWO has strong global convergence ability, robustness, and fewer parameters to adjust, and is now used in many fields for optimization problems.

Firstly, the mathematical definition of how a wolf pack searches for and surrounds its prey is as follows:

$$A = |B \cdot F_p(t) - F(t)| \quad (19)$$

$$F(t+1) = F_p(t) - C \cdot A \quad (20)$$

$$c = 2 - 2D/E \quad (21)$$

$$C = 2c \cdot r_1 - c \quad (22)$$

$$B = 2 \cdot r_2 \quad (23)$$

where  $F(t)$  is the position of the prey after the  $t$  th iteration;  $F_p(t)$  is the position of the gray wolf at the  $t$  iteration;  $A$  is the distance between the gray wolf and the prey;  $F(t+1)$  is the update of the position of the gray wolf;  $C$  and  $B$  are the coefficient vectors;  $c$  is the convergence factor whose value decreases linearly from 2 to 0 with the number of iterations,  $D$  is the number of previous iterations, and  $E$  is the maximum number of iterations;  $r_1$  and  $r_2$  are the random numbers between [0,1].

Secondly, the prey is finally determined by constantly updating the positions of the three optimal wolves  $\alpha$ ,  $\beta$ , and  $\delta$ . The mathematical definition of the hunting process of the gray wolf pack is

$$A_\alpha = |B_1 \cdot F_\alpha(t) - F(t)| \quad (24)$$

$$A_{\beta} = |B_2 \cdot F_{\beta}(t) - F(t)| \quad (25)$$

$$A_{\delta} = |B_3 \cdot F_{\delta}(t) - F(t)| \quad (26)$$

$$F_1(t+1) = F_{\alpha}(t) - C_1 \cdot A_{\alpha} \quad (27)$$

$$F_2(t+1) = F_{\beta}(t) - C_2 \cdot A_{\beta} \quad (28)$$

$$F_3(t+1) = F_{\delta}(t) - C_3 \cdot A_{\delta} \quad (29)$$

$$F(t+1) = (F_1(t+1) + F_2(t+1) + F_3(t+1))/3 \quad (30)$$

where  $F_{\alpha}(t)$ ,  $F_{\beta}(t)$  and  $F_{\delta}(t)$  are the positions of  $\alpha$ ,  $\beta$  and  $\delta$  wolves when the population is iterated to generation  $t$ ;  $F(t)$  is the position of individual gray wolves in generation  $t$ ;  $C_1$  and  $B_1$ ,  $C_2$  and  $B_2$ ,  $C_3$  and  $B_3$  are the coefficient vectors of  $\alpha$ ,  $\beta$  and  $\delta$  wolves, respectively;  $F_1(t+1)$ ,  $F_2(t+1)$  and  $F_3(t+1)$  are the positions of  $\alpha$ ,  $\beta$  and  $\delta$  wolves after  $(t+1)$  iterations, respectively;  $F(t+1)$  is the position of the next generation of gray wolves.

### GWO–XGBOOST–CEEMDAN model

To improve carbon price prediction, we propose to combine the CEEMDAN, XGBOOST, and GWO models to build the GWO–XGBOOST–CEEMDAN model. The general idea is as follows: First, the GWO–XGBOOST model is established, and the GWO algorithm is used for optimizing the parameters of the XGBOOST model. Secondly, the CEEMDAN method is applied to decompose the residual series of the GWO–XGBOOST model to establish the GWO–XGBOOST–CEEMDAN hybrid model. Finally, the predicted values and the accumulated values of the residual predictions are summed up to get the final prediction results of the model. Figure 1 illustrates the specific process.

## Data description

### Data source

Accurate carbon price forecasts smooth investment decisions and maintain carbon market stability. There are big differences between China's carbon trading pilots. The Hubei carbon trading market is the only carbon trading market in central China<sup>12</sup>. In addition, the Guangdong carbon market was officially launched in 2013, setting five first places in China's carbon market trading<sup>39</sup>. Fujian is the first ecological civilization demonstration zone in China. The carbon market is aligned with the overall idea of the national carbon market, and it is the first pilot to adopt carbon verification standards and guidelines issued by the state. In particular, the data direct reporting system is completely consistent with the national system under construction standards, and the construction starting point is high<sup>40,41</sup>. To sum up, this paper chooses Guangdong, Hubei, and Fujian carbon trading markets as research objects. In this paper, we collect data on the three carbon markets from the Choice financial terminal and the Wind database. The selected carbon prices take into account public holidays, differences in trading hours, and missing values of variables at home and abroad. In the above data, the Bohai Sea Power Coal Price Index and Natural Gas Market Quotation are weekly and ten-day data, and Eviews software is used to convert them into daily data. A hybrid model is evaluated by using 80% of the data for training and 20% for testing. The carbon price information for the three trading markets is shown in Table 1, and Table 2 presents descriptive statistics for each indicator.

### ADF inspection

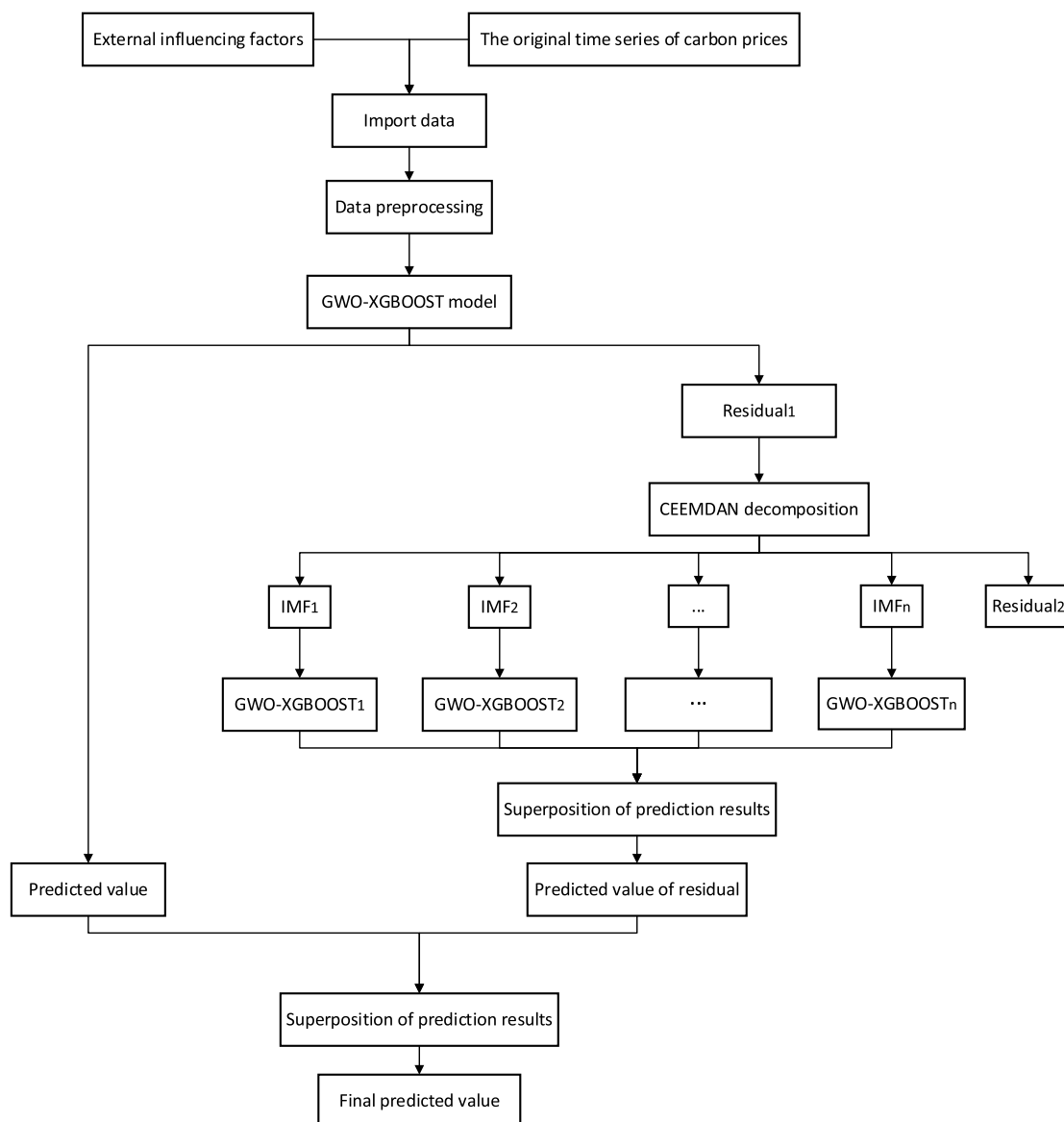
The ADF (Augmented Dickey–Fuller) test was proposed by economists David Dickey and Wayne Fuller in 1979. The test is a statistical method used to determine whether the time series data has a unit root (or the root of the series), i.e., to verify whether the data has smoothness. The ADF test gives a Guangdong  $p$ -value of 0.912439, a Hubei  $p$ -value of 0.638039, and a Fujian  $p$ -value of 0.988874, which are greater than the usually chosen significance level (e.g., 0.05 or 0.01). Therefore, the original hypothesis cannot be rejected; that is, the historical carbon price data of the three carbon markets is not stationary. In short, it is not possible to use traditional econometric methods for experiments, and an integrated learning approach is used for the prediction study of non-stationary time series of carbon prices.

### Data pre-processing

The factors often have different magnitudes and units of magnitude. It is crucial to pretreat the data to be limited to  $[0, 1]$  to remove the adverse effects caused by odd sample data and make the data comparable.

$$Z^* = \frac{Z - Z_{min}}{Z_{max} - Z_{min}} \quad (31)$$

where  $Z^*$  represents the normalized value of the data;  $Z$  is the input data, and  $Z_{min}$  and  $Z_{max}$  represent the minimum and maximum values of the input data, respectively.



**Figure 1.** GWO-XGBOOST-CEEMDAN model prediction process.

Market	Dataset	Training set	Test set	Training date	Test date	Time nodes
Guangdong	703	563	140	2020/4/3–2022/7/14	2022/7/15–2023/2/3	2020/4/3–2023/2/3
Hubei	750	601	149	2020/1/2–2022/8/24	2022/8/25–2023/4/10	2020/1/2–2023/4/10
Fujian	699	560	139	2018/1/2–2022/9/16	2022/9/19–2023/5/5	2018/1/2–2023/5/5

**Table 1.** Carbon trading market data information.

#### Four aspects that affect the price of carbon

Carbon prices are impacted by a number of factors. This paper builds primary indicators of carbon price influencing factors from four aspects: macroeconomics, energy prices, international carbon markets, and weather conditions. The detailed classification and secondary quantification of each indicator level are shown in Table 3.

##### Macroeconomics

The macroeconomy directly determines the boom in the carbon market<sup>42</sup>. The macroeconomic situation, specifically the advancement of the industrial economy, is the most representative of CO<sub>2</sub> emissions, which will affect the price of carbon trading<sup>43</sup>. Guo Fuchun says that when macroeconomic conditions are favorable, production and business activities become active, and the carbon trading price will enter a relatively stable operation. In contrast, when the economy slows down, the carbon trading price will fluctuate sharply<sup>44</sup>. Meanwhile, carbon

Index name	Mean	Std	Var	Min	25%	50%	75%	Max
Research object I: Guangdong carbon market—related variables								
Y	50.95700	21.67600	469.86900	27.04000	28.48000	42.61000	77.02000	95.26000
X1.1	3105.00500	283.22600	80,216.69900	2247.32000	2856.65000	3019.20000	3229.08000	3544.69000
X1.2	31,804.61600	3344.51100	11,185,756.23000	21,052.53000	29,883.79000	32,861.34000	34,451.23000	36,799.65000
X1.3	3938.82600	489.80400	239,908.31400	2488.65000	3655.04000	3963.51000	4348.87000	4793.54000
X1.4	4579.15800	513.71300	263,900.86800	3508.70000	4098.71000	4712.31000	4964.77000	5807.72000
X2.2	662.55000	88.60400	7850.71000	525.35000	573.14000	683.08000	735.00000	850.94000
X2.3	5188.03000	1821.86100	3,319,179.05200	2490.57000	3491.55000	5445.17000	6740.67000	8615.91000
X3.1	56.53500	23.80500	566.68300	17.86000	30.81000	57.70000	79.72000	97.59000
X4.3	62.03100	29.04700	843.73200	13.00000	41.00000	53.00000	76.00000	201.00000
Research object II: Hubei carbon market—related variables								
Y	37.91793	9.21166	84.85500	24.49000	28.96750	38.26000	47.44000	61.48000
X1.2	31,662.85910	3400.11127	11,560,756.66000	18,591.93000	29,869.74000	32,630.30500	34,293.65000	36,799.65000
X1.3	3910.49480	496.30847	246,322.10100	2237.40000	3653.15250	3944.67500	4275.00000	4793.54000
X1.4	4527.37730	518.30981	268,645.06100	3508.70000	4046.36750	4628.22500	4939.75750	5807.72000
X2.1	72.60910	23.53668	553.97500	19.50000	51.43750	74.10000	87.70250	129.47000
X2.2	662.95280	85.60991	7328.91900	508.96000	576.48740	673.27550	735.00000	863.67000
X2.3	5135.90040	1746.30901	3,049,595.17500	2488.45000	3614.98900	5246.80680	6604.83080	8676.97000
X3.1	57.30340	24.86076	618.05800	15.45000	29.74500	58.40500	80.95500	97.67000
X4.3	63.57550	31.84088	1013.84200	13.00000	42.00000	63.23500	73.25000	239.00000
Research object III: Fujian carbon market—related variables								
Y	20.74330	6.73573	2,712,776.15600	2542.36000	16.25000	19.19000	25.28000	38.12000
X1.1	2774.70790	356.36149	126,993.51400	1975.77000	2481.34000	2835.87000	3032.92000	3544.69000
X1.2	29,673.45940	3707.38569	13,744,708.68000	22,445.37000	26,112.53000	29,673.46000	33,156.41000	36,432.22000
X1.3	3535.71570	627.95655	394,329.42300	2416.62000	2884.43000	3535.72000	4017.82000	4791.19000
X2.1	76.34070	19.25121	370.60900	39.54000	64.40000	74.78000	87.23000	124.37000
X2.2	650.75130	86.31098	7449.58600	531.04000	575.71000	595.10000	734.00000	884.03000
X2.3	5113.61820	1647.05074	2,712,776.15600	2542.36000	3678.41000	4684.76000	6602.58000	8557.84000
X3.1	49.74140	30.12951	907.78700	7.62000	24.90000	32.02000	80.84000	97.67000
X4.3	41.86680	14.44742	208.72800	13.00000	31.00000	41.87000	49.00000	126.00000

**Table 2.** Descriptive statistics of carbon price data.

Primary indicators	Secondary indicators	Factor symbols	Serial number	Literature sources
Macroeconomics	Shanghai Stock Exchange Industrial Index	SHZQ	X1.1	10,64
	Dow Jones Industrial Average	DQS	X1.2	10
	S&P 500 Index	BP500	X1.3	10,64
	CSI300 Index	CSI300	X1.4	65,10
Energy prices	Brent Crude Oil CFD	CFD	X2.1	66,47,10
	Bohai Sea Power Coal Price Index	HBH	X2.2	10
	Natural Gas Market Offer	TRQ	X2.3	10,47
International carbon markets	EU Carbon Emission Allowances	EUA	X3.1	66,11
Weather conditions	Maximum Temperature	ZGQW	X4.1	47
	Minimum Temperature	ZDQW	X4.2	47
	Air Quality Index	AQI	X4.3	47,10,13

**Table 3.** Primary indicators of carbon price impact factors.

prices also reflect a country's economic growth to a certain extent<sup>45</sup>. In addition, China's carbon trading market is developing and maturing. It covers a broad spectrum of industries, including chemicals, construction materials, steel, non-ferrous metals, and various other sectors. Therefore, the rapid economic development and extension of China's carbon market highlight the significance of carbon pricing mechanisms<sup>46</sup>.



### Energy prices

It is often considered that energy is the factor that has the greatest impact on carbon prices<sup>47</sup>. With this in mind, researchers have been empirically searching for the drivers of carbon prices and understanding their future value through adequate predictive analysis. Early studies have identified energy prices as one of the indicators of the main influencing factors of carbon prices<sup>26,48–50</sup>. For example, crude oil<sup>51</sup> and natural gas prices<sup>52</sup>. The majority of these studies have found that energy prices have a significant impact on carbon prices. However, investment strategies need to focus not only on the impact of external factors on carbon prices but also on the predictability of future returns. Consequently, numerous recent studies have concentrated on the predictability of carbon prices' future values<sup>53,54</sup>. In conclusion, it is valuable to study the influence of energy prices on carbon prices both now and in the future.

### International carbon markets

Foreign carbon markets influence China's carbon prices<sup>11</sup>. On the one hand, Chinese carbon markets are still in the development stage. In contrast, foreign carbon markets have been established for a longer period of time and are relatively well-established. China's carbon market will, to a certain extent, refer to foreign carbon markets when setting carbon emission quotas. For instance, the EUA price acts as the primary reference point in the global carbon trading market, significantly shaping carbon emission allowances and, consequently, carbon prices<sup>55</sup>. Specifically, EUA, certified emission reductions (CER), and other similar products are used to reference the fulfillment of carbon reduction obligations<sup>29</sup>. On the other hand, the disparity in economic development between China and other nations can result in variations in carbon market pricing. If China's carbon market is priced low, transnational companies will speculate heavily in the Chinese carbon market to buy a large number of carbon emission rights, thus adding to the demand for carbon emission rights in the Chinese market and driving up the Chinese carbon price until it reaches parity with the international carbon price. Furthermore, at the macro level, an increase in carbon emissions reduces foreign direct investment, which affects the trading of carbon allowances and indirectly causes price volatility<sup>10</sup>. As of now, China's carbon market is not yet in line with international standards. Consequently, the investor base remains relatively modest in size. Nevertheless, once the two are connected, the issue of speculation is expected to escalate. Hence, foreign carbon prices will have a dual effect: they will inform the establishment of carbon prices in China and potentially drive up the carbon price in the country through speculative activities<sup>11</sup>.

### Weather conditions

Global warming is becoming more severe and the primary cause of this issue can be attributed to greenhouse gas emissions, especially CO<sub>2</sub><sup>56</sup>. Climate change can affect carbon price volatility through multiple channels. Earlier studies have shown that climate change can alter fossil energy consumption and thus affect carbon price fluctuations<sup>57–60</sup>. In the past few years, researchers have mainly addressed the significance of climate change on carbon prices from different perspectives. From a production standpoint, when temperatures become excessively high or low, residents resort to cooling or heating equipment, resulting in a temporary upswing in energy consumption and subsequent CO<sub>2</sub> emissions. In addition, from a business perspective, extreme weather and catastrophic events are exposing new energy companies to a significant physical risk, leading to changes in the energy mix and having a considerable impact on carbon prices<sup>61,62</sup>. To be more precise, the generation of environmentally friendly energy sources like wind, solar, and hydropower is strongly influenced by various weather factors, including temperature, precipitation, and humidity<sup>63</sup>. Therefore, it is very important to consider climate change when predicting carbon prices<sup>46</sup>.

## Experimental results and discussion

### Evaluation indicators

In this study, five common metrics are used, as shown in Table 4. The larger the R<sup>2</sup> and the smaller the remaining indicators, the better the predictive performance of the model.

Indicators	Definition	Formula
MSE	Mean square error	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
MAE	Mean absolute error	$MAE = \frac{1}{n} \sum_{i=1}^n  y_i - \hat{y}_i $
RMSPE	Root mean square percentage error	$RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left  \frac{y_i - \hat{y}_i}{y_i} \right ^2} * 100\%$
MAPE	Mean absolute percentage error	$MAPE = \frac{1}{n} \sum_{i=1}^n \left  \frac{y_i - \hat{y}_i}{y_i} \right  * 100\%$
R <sup>2</sup>	Coefficient of determination	$R^2 = 1 - \frac{SS_{residual}}{SS_{total}}$

**Table 4.** Evaluation indicators.

## Experiment one: comparison of this paper's model with different benchmark models

### Algorithm table

To demonstrate the superiority of the prediction performance of the proposed GWO–XGBOOST–CEEMDAN model in practical applications, four other different benchmark models are first set up for comparison in this paper, namely GBDT, XGBOOST, GWO–XGBOOST, and GWO–XGBOOST–EEMD. As indicated in Table 5.

### Model parameter setting

To validate the prediction accuracy of the GWO–XGBOOST–CEEMDAN model, various comparison algorithms are utilized to evaluate forecasting performance. For the prediction of the base model, the GBDT and XGBOOST models are selected to compare and analyze the prediction effect of GWO–XGBOOST. For the prediction of the combined residual correction model, the EEMD and CEEMDAN methods are used to compare and decompose the residual sequences generated by GWO–XGBOOST. The parameter settings of each model are shown in Table 6. According to the actual situation of the three carbon trading markets, the parameters of each model are adjusted as shown in Table 7, and the remaining parameters are set by default in Python.

### Screening analysis of carbon price influencing factors

This study considers both the selection of historical carbon price variables and the identification of external influences in three carbon trading markets. More specifically, for the purpose of examining the correlation between historical carbon price variables and the carbon price data, we use PACF in order to identify the relevant input data characteristics for forecasting. Figure 2 shows the PACF results for Guangdong, Hubei, and Fujian. This analysis reveals a notable fourth-order autocorrelation in the carbon price data of Guangdong, whereas both Hubei and Fujian demonstrate a third-order autocorrelation.  $X_i$  is the output feature;  $\{X_{i-1}, X_{i-2}, X_{i-5}, X_{i-7}\}$  are the input historical variables for the Guangdong dataset;  $\{X_{i-1}, X_{i-3}, X_{i-5}\}$  are the input historical variables for the Hubei dataset; and  $\{X_{i-1}, X_{i-2}, X_{i-3}\}$  are the input history variables for the Fujian dataset.

Furthermore, with the intention of identifying the main external influences on carbon prices, this paper selects 11 primary carbon price indicators and uses random forest to rank the importance of the indicators for the purpose of indicator screening, as shown in Fig. 3. The screened carbon price indicators are shown in Table 8. The input variables for the Guangdong data set are  $\{X_{i-1}, X_{i-2}, X_{i-5}, X_{i-7}, X_{1.1}, X_{1.2}, X_{1.3}, X_{1.4}, X_{2.2}, X_{2.3}, X_{3.1}, X_{4.3}\}$ , and the input variables for the Hubei data set are  $\{X_{i-1}, X_{i-3}, X_{i-5}, X_{1.2}, X_{1.3}, X_{1.4}, X_{2.1}, X_{2.2}, X_{2.3}, X_{3.1}, X_{4.3}\}$ , and the input variables for the Fujian dataset are  $\{X_{i-1}, X_{i-2}, X_{i-3}, X_{1.1}, X_{1.2}, X_{1.3}, X_{2.1}, X_{2.2}, X_{2.3}, X_{3.1}, X_{4.3}\}$ .

Models	Decomposition technology	Parameter optimization
GBDT		
XGBOOST		
GWO–XGBOOST		√
GWO–XGBOOST–EEMD	√	√
PROPOSED MODEL	√	√

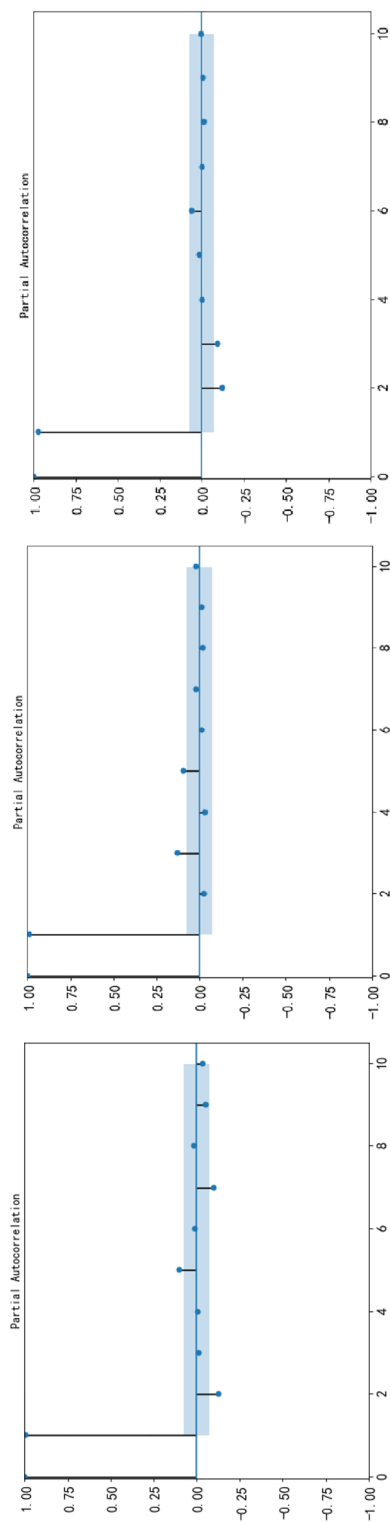
**Table 5.** Comparison of different models.

Models	Adjustment parameters
GBDT	n_estimators; min_samples_leaf; learning_rate; max_depth; min_samples_split
XGBOOST	n_estimators; colsample_bytree; learning_rate; max_depth
GWO–XGBOOST	n_estimators; colsample_bytree; learning_rate; max_depth; subsample

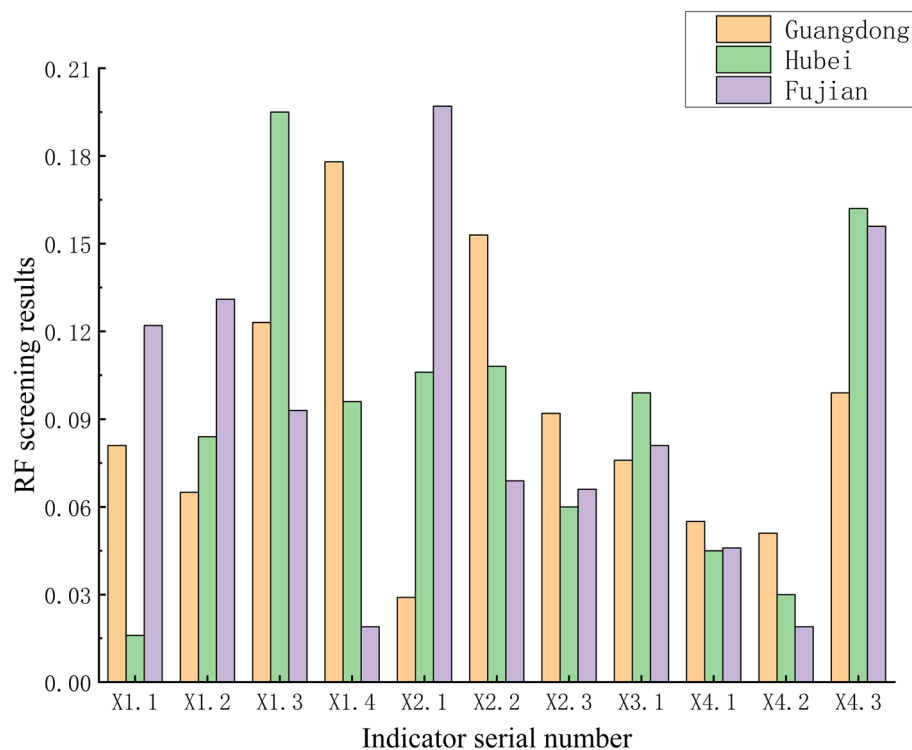
**Table 6.** Parameters of each model.

Models	Guangdong parameters	Hubei parameters	Fujian parameters
GBDT	(100,2,0.1,8,4)	(30,9,0.1,20,20)	(15,8,0.1,20,20)
XGBOOST	(100,0.5,0.1,10)	(100,0.2,0.1,18)	(100,0.3,0.1,20)
GWO–XGBOOST	(37.5037,0.3176,0.2457, 56.2732, 0.5770)	(31.1208, 0.4177, 0.0441, 40.7127, 0.1817)	(11.3251, 0.1224, 0.0683, 19.2370, 0.5652)

**Table 7.** Model parameter settings.



**Figure 2.** PACF results for carbon price data.



**Figure 3.** Random Forest screening results.

	Guangdong	Hubei	Fujian
Macroeconomics	Shanghai Stock Exchange Industrial Index (X1.1)	Dow Jones Industrial Average (X1.2)	Shanghai Stock Exchange Industrial Index (X1.1)
	Dow Jones Industrial Average (X1.2)	S&P 500 Index (X1.3)	Dow Jones Industrial Average (X1.2)
	S&P 500 Index (X1.3)	CSI300 Index (X1.4)	S&P 500 Index (X1.3)
	CSI300 Index (X1.4)		
Energy prices	Bohai Sea Power Coal Price Index (X2.2)	Brent Crude Oil CFD (X2.1)	Brent Crude Oil CFD (X2.1)
	Natural Gas Market Offer (X2.3)	Bohai Sea Power Coal Price Index (X2.2)	Bohai Sea Power Coal Price Index (X2.2)
		Natural Gas Market Quotes (X2.3)	Natural Gas Market Quotes (X2.3)
International carbon markets	EU Carbon Emission Allowances (X3.1)	EU Carbon Emission Allowances (X3.1)	EU Carbon Emission Allowances (X3.1)
Weather conditions	Air Quality Index (X4.3)	Air Quality Index (X4.3)	Air Quality Index (X4.3)

**Table 8.** Carbon price finalization indicators.

*Carbon price forecast results I*

Table 9 shows the effect of the fitted curves on the carbon price predictions for each model in the three test sets. Furthermore, to facilitate a visual comparison of the prediction outcomes between the proposed model and other comparative models, Figs. 4, 5, and 6 depict the prediction results for the Guangdong, Hubei, and Fujian datasets, respectively.

**Experiment two: comparison of this paper’s model with different input feature models**

*Algorithm table*

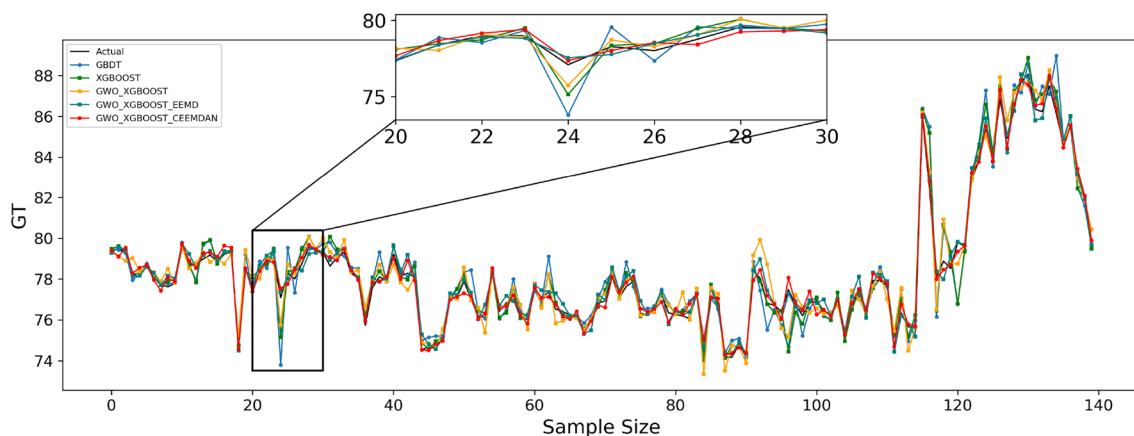
To further confirm the effectiveness of the feature selection algorithm, GWO–XGBOOST–CEEMDAN\*, ALL-VARIABLE-GWO–XGBOOST–CEEMDAN, and the proposed GWO–XGBOOST–CEEMDAN model are compared in this paper, and the algorithm experiment table is shown in Table 10.

*Carbon price forecast results II*

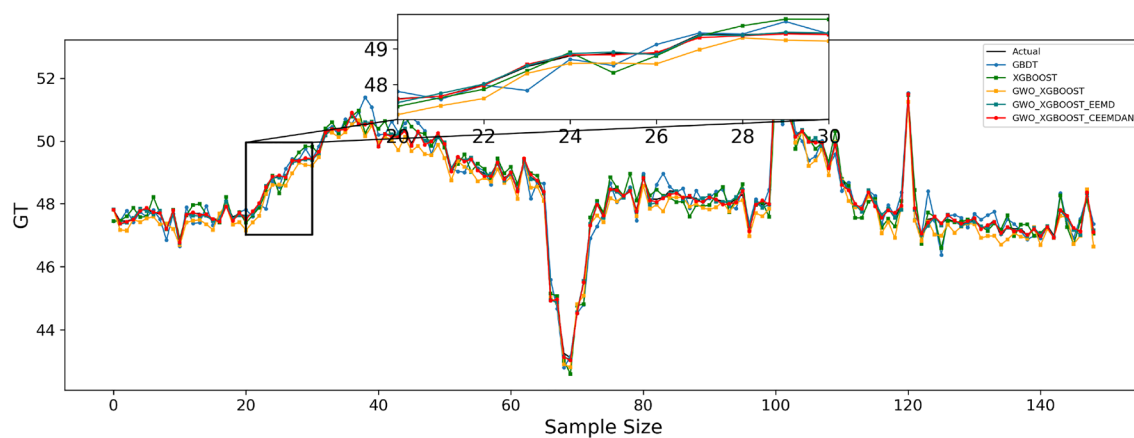
The predicted effect graphs are shown in Figs. 7, 8, and 9, and the algorithm experimental results table is shown in Table 11.

Forecast range	Evaluation indicators	GBDT	XGBOOST	GWO-XGBOOST	GWO-XGBOOST-EEMD	GWO-XGBOOST-CEEMDAN
Guangdong	MSE	0.650193	0.410348	0.368359	0.119977	<b>0.078120</b>
	MAE	0.508363	0.391618	0.436538	0.275044	<b>0.223989</b>
	RMSPE	1.013249	0.806493	0.775963	0.441638	<b>0.356462</b>
	MAPE	0.642882	0.495045	0.556786	0.003506	<b>0.002856</b>
	R <sup>2</sup>	0.931976	0.957069	0.961462	0.9874478	<b>0.991826</b>
Hubei	MSE	0.101325	0.081960	0.075187	0.005043	<b>0.004184</b>
	MAE	0.240302	0.216567	0.232021	0.056902	<b>0.050459</b>
	RMSPE	0.657276	0.589355	0.567327	0.147217	<b>0.134632</b>
	MAPE	0.496693	0.447754	0.480307	0.001180	<b>0.001047</b>
	R <sup>2</sup>	0.949081	0.958812	0.962216	0.997466	<b>0.997898</b>
Fujian	MSE	0.299317	0.170274	0.075187	0.005043	<b>0.004184</b>
	MAE	0.407047	0.332075	0.232021	0.056902	<b>0.050459</b>
	RMSPE	1.802013	1.347809	0.567327	0.147217	<b>0.134632</b>
	MAPE	1.328623	1.080592	0.480307	0.001180	<b>0.001047</b>
	R <sup>2</sup>	0.945945	0.958812	0.962216	0.997466	<b>0.997898</b>

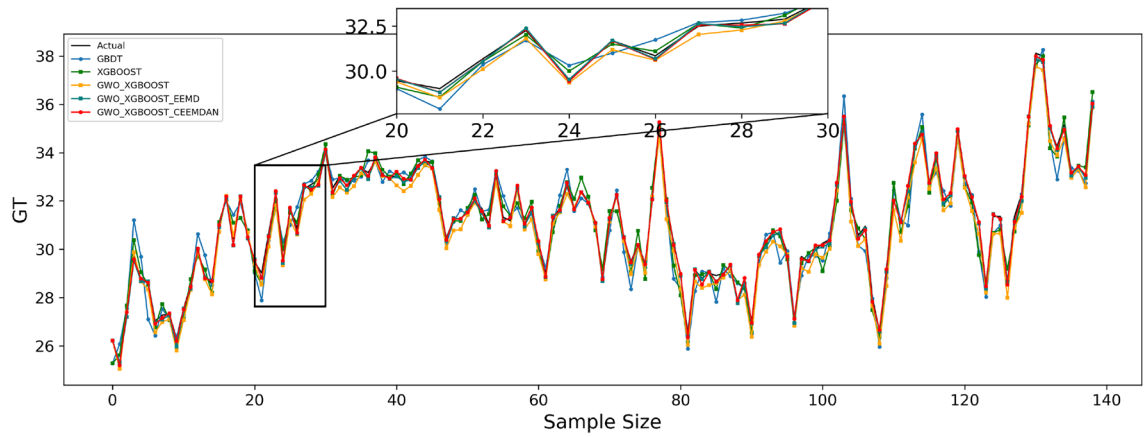
**Table 9.** Algorithm experiment results table. Significant values are given in Bold.



**Figure 4.** Guangdong model prediction effect.



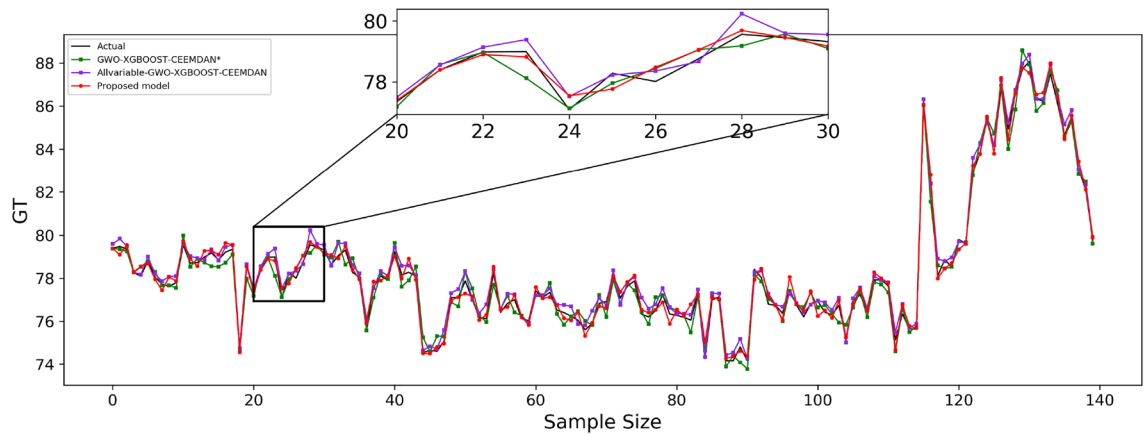
**Figure 5.** Hubei model prediction effect.



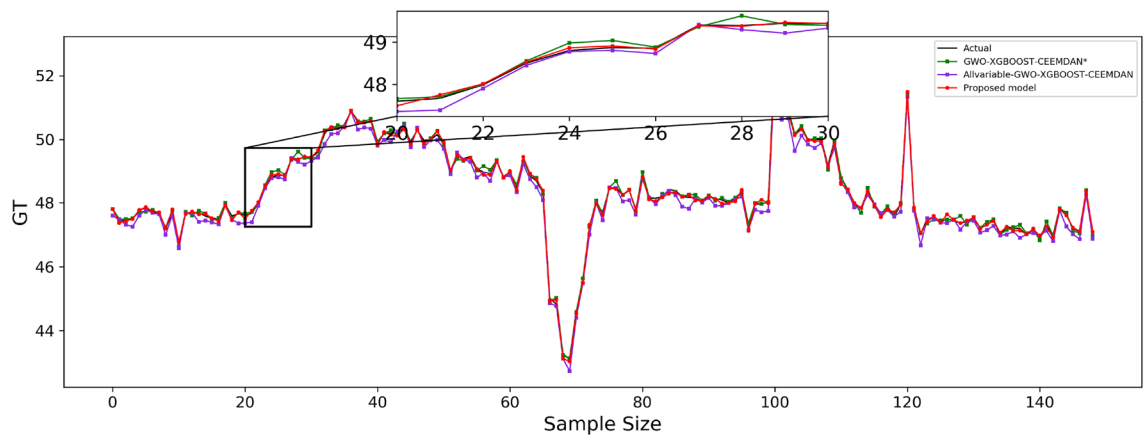
**Figure 6.** Fujian model prediction effect.

Models	External influencing factors	Historical carbon price
GWO-XGBOOST-CEEMDAN*		√
ALL-VARIABLE- GWO-XGBOOST-CEEMDAN	√	√
PROPOSED MODEL	√	√

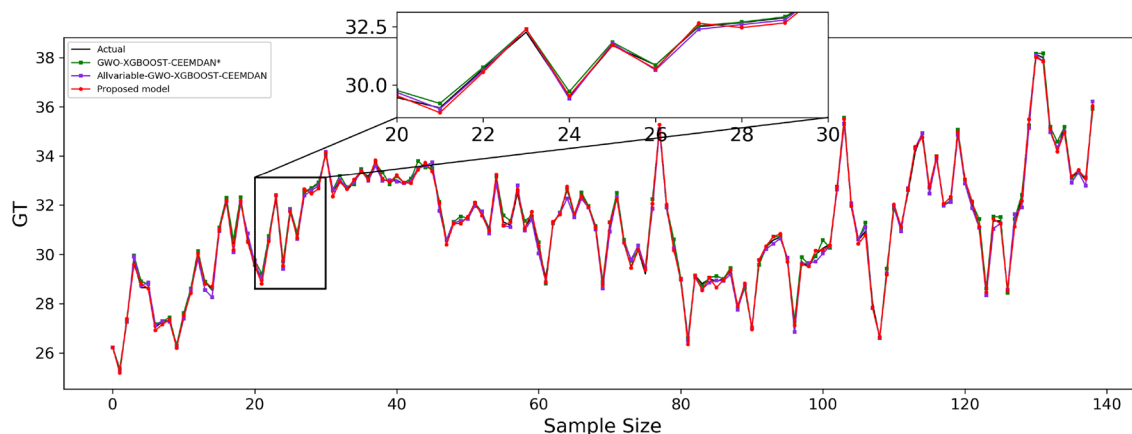
**Table 10.** Algorithm experiment table.



**Figure 7.** Guangdong model prediction effect.



**Figure 8.** Hubei model prediction effect.



**Figure 9.** Fujian model prediction effect.

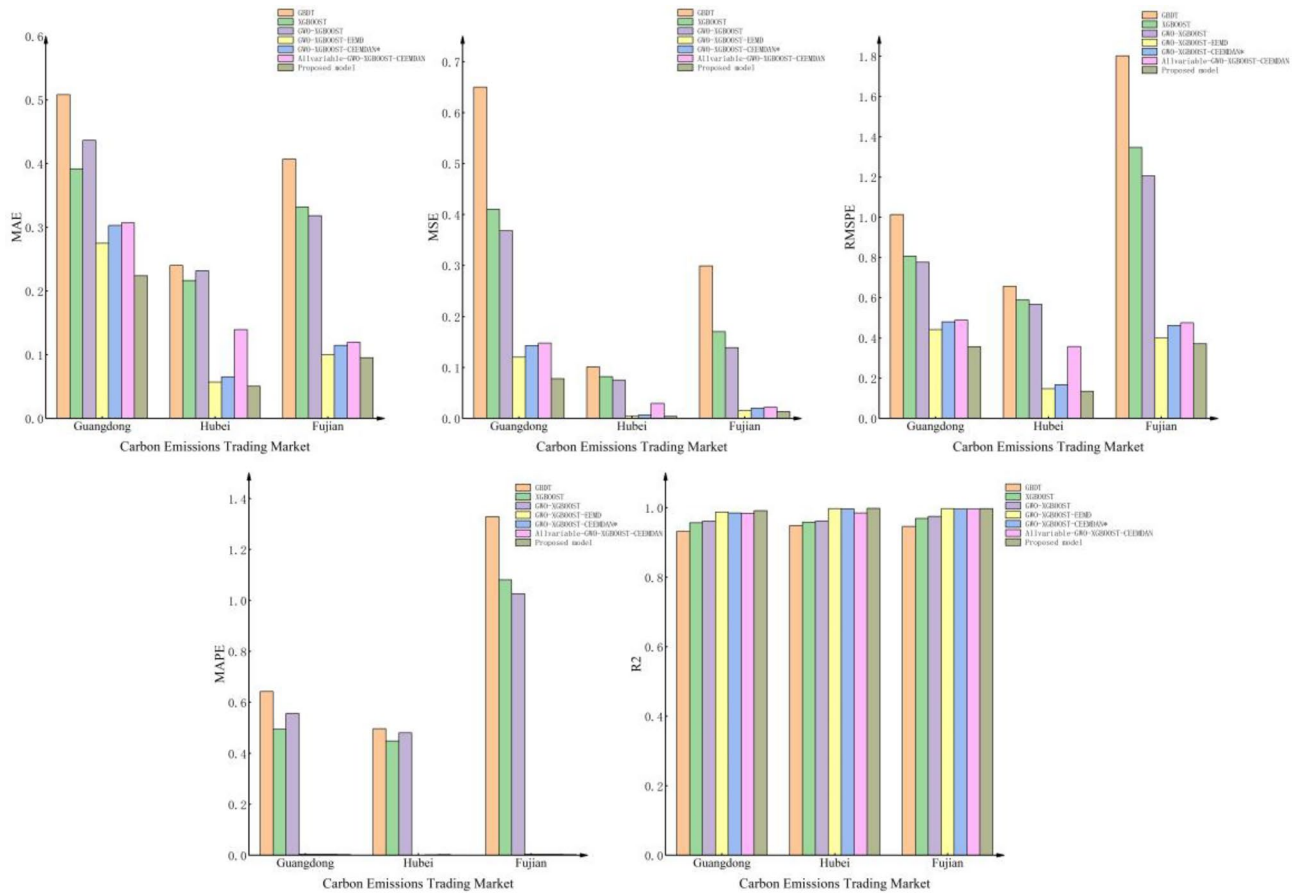
Forecast range	Evaluation indicators	GWO-XGBOOST-CEEMDAN*	ALL-VARIABLE-GWO-XGBOOST-CEEMDAN	PROPOSED MODEL
Guangdong	MSE	0.142878	0.147749	<b>0.078120</b>
	MAE	0.302748	0.307266	<b>0.223989</b>
	RMSPE	0.479748	0.488676	<b>0.356462</b>
	MAPE	0.003858	0.003911	<b>0.002856</b>
	R <sup>2</sup>	0.985052	0.984542	<b>0.991826</b>
Hubei	MSE	0.006618	0.029784	<b>0.004184</b>
	MAE	0.065209	0.139503	<b>0.050459</b>
	RMSPE	0.168078	0.357332	<b>0.134632</b>
	MAPE	0.001348	0.002888	<b>0.001047</b>
	R <sup>2</sup>	0.996674	0.985033	<b>0.997898</b>
Fujian	MSE	0.020320	0.021822	<b>0.013272</b>
	MAE	0.114830	0.119552	<b>0.095404</b>
	RMSPE	0.462370	0.475852	<b>0.371209</b>
	MAPE	0.003709	0.003846	<b>0.003070</b>
	R <sup>2</sup>	0.996330	0.996059	<b>0.997603</b>

**Table 11.** Algorithm experiment results table. Significant values are given in Bold.

### Analysis of experimental results

Figure 10 shows the predicted results of the evaluation indexes of the above two groups of experiments. By comparing and analyzing the above two sets of experiments, we can draw the following conclusions:

- (1) In single-model prediction, the GWO-XGBOOST model has the best prediction effect, which is mainly attributed to the following reasons: First, GBDT builds an integrated model by training a series of decision trees; each tree is trained on the residuals of the previous tree, so when the number of trees is large, the model may be over-fitted on the training set, resulting in large model prediction errors. Second, XGBOOST is an integrated learning method that enhances and optimizes on the basis of GBDT and improves generalization ability but has weaker performance in dealing with the category imbalance problem. Finally, GWO, as an optimization algorithm for searching for the global optimal solution, can not only tune the hyper-parameters in XGBOOST, such as learning rate, tree depth, subsample ratio, etc., but also fuse multiple XGBOOST models, adjusting the weights and parameters of different models to achieve the combination and integration of models. This can, to some degree, enhance the model's stability and generalization ability, thus increasing its overall capability.
- (2) For the combination algorithm, all combination models outperform comparison models in relation to predictive accuracy. The Guangdong dataset is used as an example, and the Hubei and Fujian datasets are consistent with this conclusion. First, to verify the effectiveness of the decomposition method proposed in this paper, GWO-XGBOOST-EEMD is contrasted with the model in this paper, and it is found that the prediction accuracy of the proposed model in this paper is improved by 34.888%, 18.562%, 19.286%, 18.540%, and 0.443% for MSE, MAE, RMSPE, MAPE, and R<sup>2</sup>, respectively. The results indicate that the CEEMDAN approach proposed in this study offers an additional enhancement to the prediction accuracy of the GWO-XGBOOST model when compared to EEMD. Secondly, to further demonstrate the superior-



**Figure 10.** Results of index evaluation of each model.

ity of the feature selection algorithm, the GWO–XGBOOST–CEEMDAN\* model only considers carbon price historical data, and the All-VARIABLE-GWO–XGBOOST–CEEMDAN model takes carbon price historical data and all external influences as input variables. The results of these three models indicate that the feature selection algorithm is helpful in improving the prediction performance of the hybrid model.

### Discussion

#### Validation on other data sets

In order to verify the generalization ability of the model as well as its strong robustness, we selected a Q1 partitioned article published in the journal ENVIRONMENTAL SCIENCE AND POLLUTION RESEARCH<sup>29</sup>. This paper selects the daily carbon prices in Beijing, Hubei, and Shanghai as sample data, all of which come from the China Carbon Trading Network, and the experiment uses 75% of the data as the training set and 25% as the test set as a way to verify the accuracy of the hybrid model. We use the raw data from this published paper to do a comparative analysis between our proposed model and the model prediction results from the published paper, which not only verifies again that our own model has strong stability and prediction accuracy but also makes the study richer and more convincing. The information on the three carbon trading markets is shown in Table 12.

The effect of the fitting curves of our carbon price prediction for three carbon markets using the proposed GWO–XGBOOST–CEEMDAN model is shown in Fig. 11. In addition, Table 13 is a table of the experimental results of the algorithms of the two models for the three carbon markets, which compares more intuitively the prediction accuracy capability of the proposed model with the models of the published papers.

Comparing the prediction accuracy of our own proposed model with the published paper model for the same dataset, we can draw the following conclusions:

Market	Dataset	Training set	Test set	Training date	Test date	Time nodes
Beijing	1577	1183	394	2013.11.28–2018/9/7	2018/9/10–2020.4.27	2013.11.28–2020.4.27
Hubei	1525	1144	381	2014.4.3–2018/9/27	2018/9/28–2020.4.27	2014.4.3–2020.4.27
Shanghai	1561	1171	390	2013.12.19–2017/4/19	2017/4/20–2020.4.27	2013.12.19–2020.4.27

**Table 12.** Carbon trading market data information.



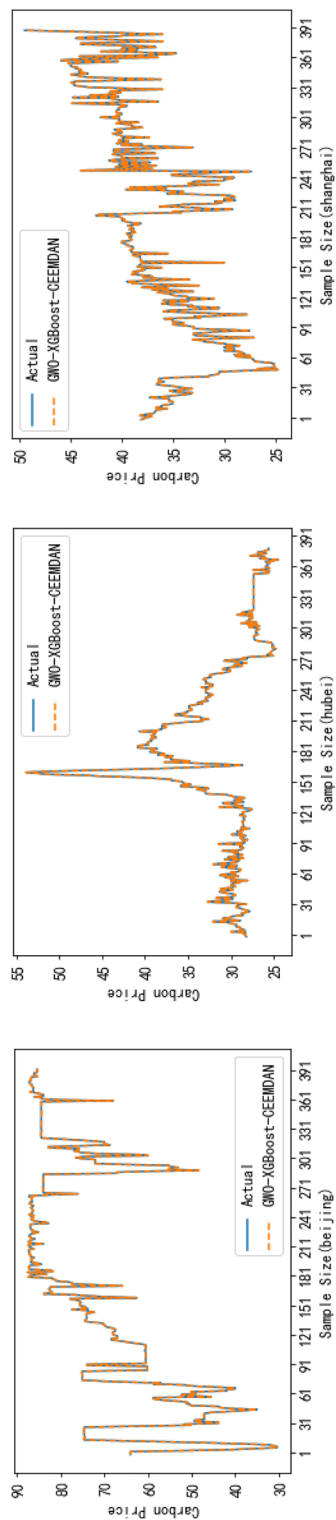


Figure 11. Model prediction effect.

Forecast range	Model	MAE	MAPE	RMSE
Beijing	Published paper model	2.543	0.232	3.363
	Proposed model	<b>0.0847</b>	<b>0.0012</b>	<b>0.7177</b>
Hubei	Published paper model	0.654	0.144	1.048
	Proposed model	<b>0.0265</b>	<b>0.0009</b>	<b>0.0328</b>
Shanghai	Published paper model	1.212	0.143	1.673
	Proposed model	<b>0.0212</b>	<b>0.0006</b>	<b>0.0271</b>

**Table 13.** Algorithm experiment results table. Significant values are given in Bold.

- (1) Most of the previous research focused on the analysis of historical carbon price prediction, and the selected published paper is a typical representative of the previous research, which is different from one of the innovations of my own thesis: considering the historical carbon price and the influencing factors at the same time so as to create a rich indicator system. Therefore, this published paper also validates, to a certain extent, the prediction accuracy of the proposed model when only considering the historical carbon price.
- (2) The research objects of the published papers are selected as Beijing, Hubei, and Shanghai, which are different from the objects of our own paper. Firstly, it verifies that the proposed model is not limited to the carbon market in Guangdong, Hubei, and Fujian but can also be used in carbon price prediction studies in other regions, which means that the results can be extended to other regions, such as Beijing and Shanghai. Secondly, the research object in the published paper still includes the Hubei carbon market, which verifies that the proposed model still has high accuracy in the case of another dataset that only considers the historical carbon price. In conclusion, by comparing the carbon price prediction with the published paper for the same dataset, it shows that the proposed model has stronger generalization ability and robustness.
- (3) The GWO–XGBOOST–CEEMDAN model is more suitable than the VMD–SE–DRNN–GRU model used in the published paper to deal with the problem of forecasting time series data. Firstly, CEEMDAN automatically determines the number of modes to be generated based on the dataset and generates intrinsic modal functions (IMFs), while VMD, although it can also perform modal decomposition, needs to pre-specify the number of modes to be decomposed into, which requires some domain knowledge or experiments to determine. If the number of modes chosen is inappropriate, it may lead to inaccurate decomposition results, thus CEEMDAN has more adaptive and flexible compared to VMD. Secondly, the GWO–XGBOOST–CEEMDAN model has lower complexity compared to deep learning models, thus it is easier to train with limited data and does not require a large amount of computational resources, which can make it more practical in some applications, such as the field of carbon price prediction. Finally, XGBOOST models are usually very interpretable and can provide feature importance rankings, whereas deep learning models such as DRNN and GRU are usually more difficult to interpret, especially in highly complex network structures. In conclusion, both VMD–SE–DRNN–GRU and GWO–XGBOOST–CEEMDAN are sophisticated carbon price prediction methods, and GWO–XGBOOST–CEEMDAN may be a better choice in cases of complex or non-stationary data.

#### DM test

To further examine the prediction performance between the proposed GWO–XGBOOST–CEEMDAN integrated combination model and the comparison models, this section uses the DM test to analyze statistical errors from the perspective of statistical errors. The bold values in the table indicate that the  $p$ -value is below the significance threshold of 0.05. To visually assess the predictive performance of the GWO–XGBOOST–CEEMDAN model and other models, we analyze their predictive ability using the coverage ratio based on the DM results. The coverage ratio is expressed as the ratio of the number of DM results rejecting the original hypothesis to the total number of DM results. When the models exhibit comparable predictive capabilities, a lower number of DM test results have a  $p$ -value less than 0.05, resulting in a coverage rate below 50%. When the models demonstrate significantly superior predictive capabilities compared to the benchmark model, a higher number of DM test results exhibit a  $p$ -value below 0.05, leading to a coverage rate exceeding 50%.

This further analysis of the DM test results in Table 14 revealed that DM coverage for the GWO–XGBOOST–CEEMDAN model was 83.3% in all three datasets, demonstrating that the GWO–XGBOOST–CEEMDAN model outperformed the proposed benchmark model in the majority of instances, and thus the proposed hybrid model was statistically significant.

#### Limitations of the current study and future work

Although the constructed hybrid forecasting framework showcases superior performance in carbon trading price prediction and fills the current research gap in carbon price prediction, there are still a few shortcomings that need further improvement and development. Following are the main limitations of this study:

- (1) Due to data availability limitations, the hybrid prediction framework we developed only considers eight influencing factors.
- (2) This study provides information for related scholars. Firstly, this paper performs carbon price prediction first and then decomposes the residual series. As a result, data can be explored and utilized more effectively.

Models	Guangdong		Hubei		Fujian	
	MAE	MSE	MAE	MSE	MAE	MSE
GBDT	1.31480e-07	0.000140	0.000000	2.31306e-10	0.000000	3.09520e-10
XGBOOST	0.000100	0.002870	0.000000	3.12004e-10	0.000000	4.68075e-11
GWO-XGBOOST	3.02359e-09	7.92154e-06	0.000000	0.000000	0.000000	0.000000
GWO-XGBOOST-EEMD	0.015140	0.009720	0.130250	0.215960	0.464508	0.254400
GWO-XGBOOST-CEEMDAN*	0.000610	0.000350	0.003640	0.007930	0.033014	0.009580
All-VARIABLE-GWO-XGBOOST-CEEMDAN	0.411980	0.632980	0.000000	1.04716e-12	0.006089	0.001100

**Table 14.** DM test results. a  $p$ -value of 0.0 does not mean that the data is 0, it just means that the  $p$ -value is small and the probability tends to be 0.

tively, and prediction accuracy and reliability can be improved. Compared with the traditional method of decomposition followed by prediction, this paper provides an alternative way of thinking and method for carbon price prediction, which scholars can refer to further to explore and analyze the intrinsic mechanism of carbon price, expand the research field, and deepen theoretical understanding. Second, in the modeling process, in addition to the factors already considered, other factors that may affect carbon prices can be further considered. This will reveal more factors affecting carbon price change and further improve the prediction effect. In addition to better adapting to changes in data characteristics, we need to improve the hybrid forecasting framework. Specifically, we can implement automatic key parameter settings and build a smart carbon price prediction framework. This framework automatically adapts to data changes, improving prediction accuracy and stability. Finally, future research can also extend carbon price prediction to other fields, for instance, energy market prediction and climate change risk management, which will further prove the value of our research results.

#### *Impact on sustainability*

This study examines the effects of macroeconomics, energy prices, international carbon markets, and weather conditions on improving carbon trading price forecasts, which have crucial ramifications for sustainable development. In particular, risk managers can incorporate multiple factors, such as energy factors and global carbon prices, into carbon market management. In addition, investors can grasp carbon market dynamics based on influencing factors and improve market participants' flexibility and motivation. This paper conducts research related to carbon price forecasting, which is helpful for the government and enterprises to grasp the characteristics of carbon prices and helps carbon market management and investment decisions, especially the solution of the carbon price prediction problem, which is linked to whether the double carbon target can be achieved on time or in advance. Therefore, this study aims to provide a reasonable forecast of carbon prices in order to facilitate carbon market participants in achieving their goals and help real producers reduce emissions efficiently through market mechanisms. In conclusion, the forecasting framework and the associated research findings we have developed hold significant implications for the advancement of sustainable development.

#### *Feature importance analysis*

To identify the key determinants in carbon price prediction, this paper uses XGBOOST and GBDT models for feature importance analysis, respectively. The results and statistical plots of feature importance indices for each model are shown in Table 13 and Fig. 12. Through preliminary observation, it is evident that the feature ranking results in the two models for the three data sets vary to some extent. A more in-depth examination of the feature rankings in Table 15 uncovers that the historical carbon price, natural gas market offers, and Bohai Ring Power Coal Price Index in energy prices, the S&P 500 and Dow Jones Industrial Index in macroeconomics, and the EU carbon emission allowances in the international carbon market rank ahead of the two models for the three data sets XGBOOST and GBDT as the key factors for carbon price prediction.

For governments, our findings suggest that historical carbon prices, natural gas market quotes, the Bohai Ring Power Coal Index, the S&P 500 Index, the Dow Jones Industrial Average, and EU carbon emission allowances can be effective ways to improve the predictive power of carbon prices in regional carbon trading markets, and policymakers can refer to our findings to make decisions about carbon market policies. First, for historical carbon price data, when the carbon price rises, the government makes stricter carbon reduction policies. This is to stimulate emission reduction measures. And when carbon prices fall, the government may reduce subsidies and support for carbon abatement to prevent a burden on the Treasury. Second, changes in natural gas supply and the Bohai Ring Power Coal Index can affect energy security. When these prices rise, the government can increase domestic production and reserves to guarantee energy supply stability. When prices fall, the government should promote the energy market by increasing subsidies and controlling imports. Moreover, the S&P 500 and the Dow Jones Industrial Average reflect the macroeconomic environment. When prices fall, the government takes stimulus measures, such as cutting taxes or increasing spending, to promote economic growth. When prices rise, the government can take restraining measures, such as strengthening regulation or controlling capital inflows, to prevent overheating. Finally, EU carbon emission allowances reflect international carbon markets. When quotas rise, the government should increase carbon emission quotas to ease enterprises' economic burden. This will avoid excessive carbon prices that lose them competitiveness. When the quota decreases, the government should

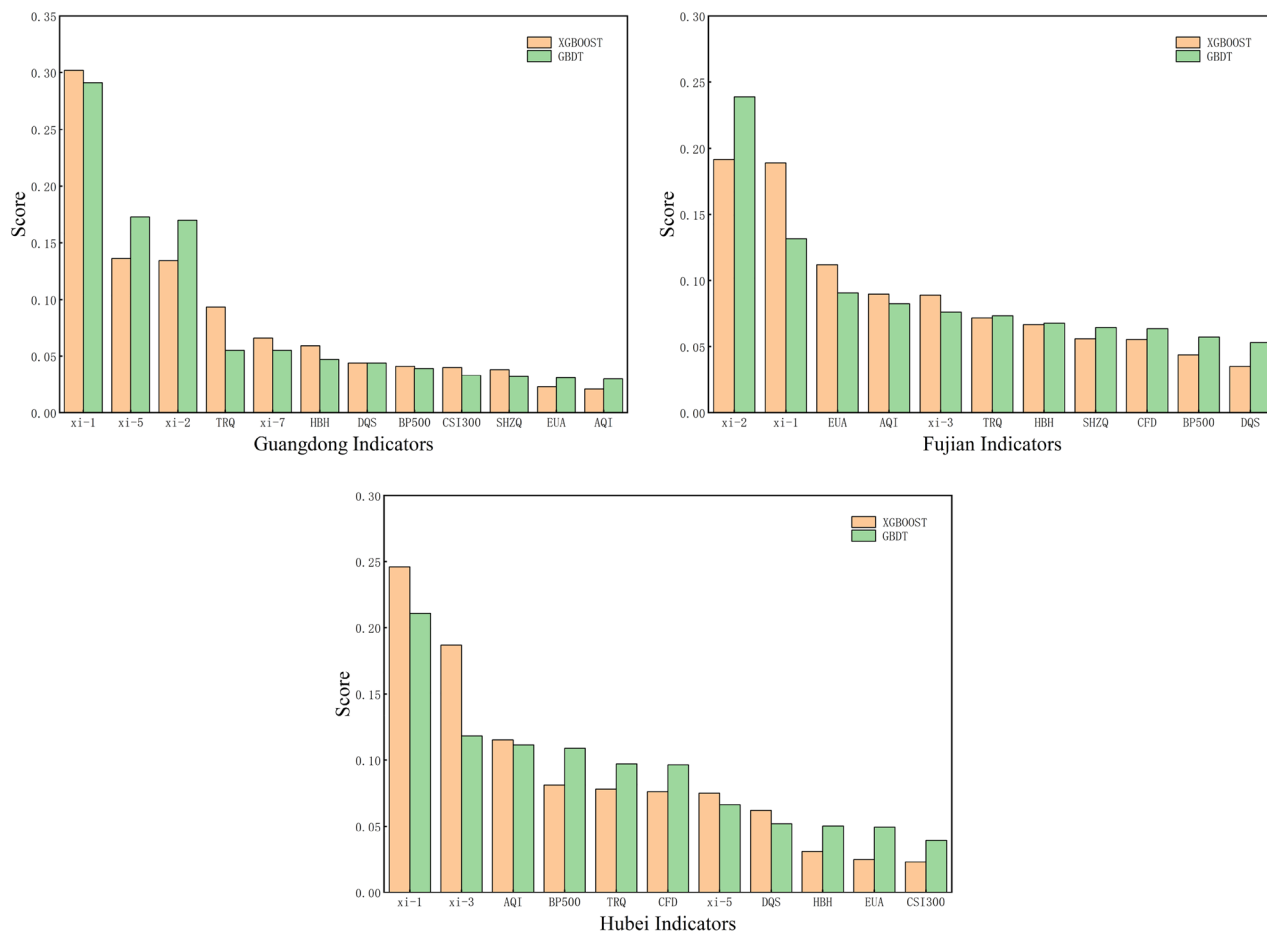


Figure 12. Feature importance diagram.

Guangdong				Hubei				Fujian			
XGBOOST		GBDT		XGBOOST		GBDT		XGBOOST		GBDT	
Features	Score	Features	Score	Features	Score	Features	Score	Features	Score	Features	Score
xi-1	0.302	xi-1	0.291	xi-1	0.246	xi-1	0.211	xi-2	0.192	xi-1	0.239
xi-5	0.136	xi-5	0.173	xi-3	0.187	xi-3	0.118	xi-1	0.189	xi-3	0.132
xi-2	0.134	xi-2	0.170	AQI	0.115	xi-5	0.111	EUA	0.112	DQS	0.091
TRQ	0.093	xi-7	0.055	BP500	0.081	TRQ	0.109	AQI	0.090	BP500	0.083
xi-7	0.066	EUA	0.055	TRQ	0.078	CSI300	0.097	xi-3	0.089	EUA	0.076
HBH	0.059	BP500	0.047	CFD	0.076	EUA	0.096	TRQ	0.072	SHZQ	0.073
DQS	0.044	DQS	0.044	xi-5	0.075	CFD	0.066	HBH	0.067	xi-2	0.068
BP500	0.041	TRQ	0.039	DQS	0.062	HBH	0.052	SHZQ	0.056	CFD	0.065
CSI300	0.040	HBH	0.033	HBH	0.031	DQS	0.050	CFD	0.056	HBH	0.064
SHZQ	0.038	AQI	0.032	EUA	0.025	AQI	0.050	BP500	0.044	TRQ	0.057
EUA	0.023	CSI300	0.031	CSI300	0.023	BP500	0.039	DQS	0.035	AQI	0.053
AQI	0.021	SHZQ	0.030								

Table 15. Carbon price characteristics in order of importance.

reduce carbon emission quotas or support low-carbon technologies to reduce carbon emissions. In addition, the government should strengthen regulation and management to reduce fraud and misconduct on the carbon market, which will improve market transparency and stability.

In conclusion, the carbon price forecasting study in this paper can act as a point of reference for policymakers to consider various factors. This will ensure policy sustainability and effectiveness.

## Conclusions

This paper uses sophisticated data decomposition methods and efficient feature subset selection algorithms to propose a comprehensive consideration of multiple influencing factors in a carbon price forecasting framework, thus achieving the expected forecasting results. To validate the effectiveness of the designed hybrid forecasting framework, we conducted an empirical study on three regional carbon emission trading markets in Guangdong Province, Hubei Province, and Fujian Province, China, and evaluated five performance evaluation indicators, six benchmark models, three case analyses, and four discussions to systematically and holistically examine the developed hybrid forecasting framework. The prediction results clearly demonstrate the superior performance of the developed prediction framework over all benchmark models. Therefore, we contend that the proposed carbon price forecasting framework offers a valuable and effective approach to predicting carbon prices. Specifically, this study's findings can be briefly outlined as follows:

- (1) The XGBOOST model based on the boosting integrated learning framework has high prediction accuracy and strong generalization ability. However, the model's predictive performance is sensitive to parameter settings, and the grey wolf optimization algorithm facilitates rapid determination of the optimal parameters for the XGBOOST model. Among the single-model carbon price prediction approaches, the proposed GWO–XGBOOST model exhibits the highest level of accuracy.
- (2) The nonlinear residual series generated by an individual machine learning model for carbon price prediction still contains valid information. In this paper, CEEMDAN is utilized to decompose the residual series generated by the GWO–XGBOOST model into sub-series with different frequencies, so that each sub-series is predicted separately, and finally, the prediction results of each sub-series are superimposed to achieve the overall prediction result. In comparison to the traditional single model, the combined GWO–XGBOOST–CEEMDAN model presented in this paper has the most accurate prediction effect.
- (3) Two machine learning models, XGBOOST and GBDT, are used to model carbon prices separately and perform feature importance analyses. It is found that the historical carbon price, the natural gas market price, and the Bohai Sea Power Coal Price Index in the energy price, the macroeconomic S&P 500 and Dow Jones Industrial Average in the macroeconomy, and EU carbon emission allowances in the international carbon market are the main influencing factors for carbon price prediction.
- (4) This study constructs a carbon price prediction model in accordance with multiple influencing factors and introduces a feature selection method. By selecting the features with the most predictive power from many possible input features and mitigating the adverse impacts of redundant information and noise among features, the model's predictive performance and generalization capability are enhanced.
- (5) In this paper, external influences and historical carbon price data are jointly used as input features of the model, such as macroeconomics, energy prices, and international carbon markets. Compared with previous studies that only considered historical carbon price data and all external influences and historical carbon price data concurrently, the accuracy of carbon price prediction has significantly improved. For one thing, it shows the significance of feature selection, and for another, it indicates that the combination of external factors and historical carbon price data can forecast the carbon price trend more precisely.

## Data availability

The data used to support the findings of this study are available from the corresponding author upon request.

Received: 29 June 2023; Accepted: 20 October 2023

Published online: 27 October 2023

## References

1. Hashim, H. *et al.* An integrated carbon accounting and mitigation framework for greening the industry. *Energy Procedia* **75**, 2993–2998 (2015).
2. An, K. *et al.* Socioeconomic impacts of household participation in emission trading scheme: A computable general equilibrium-based case study. *Appl. Energy* **288**, 116647. <https://doi.org/10.1016/j.apenergy.2021.116647> (2021).
3. Song, Y., Liang, D., Liu, T. & Song, X. How China's current carbon trading policy affects carbon price? An investigation of the Shanghai Emission Trading Scheme pilot. *J. Clean. Prod.* **181**, 374–384 (2018).
4. Sun, W. & Zhang, C. Analysis and forecasting of the carbon price using multi—Resolution singular value decomposition and extreme learning machine optimized by adaptive whale optimization algorithm. *Appl. Energy* **231**, 1354–1371 (2018).
5. Zhu, B., Ye, S., He, K., Chevallier, J. & Xie, R. Measuring the risk of European carbon market: An empirical mode decomposition-based value at risk approach. *Ann. Oper. Res.* **281**, 373–395 (2019).
6. Zhu, J., Wu, P., Chen, H., Liu, J. & Zhou, L. Carbon price forecasting with variational mode decomposition and optimal combined model. *Phys. A* **519**, 140–158 (2019).
7. Yun, P., Zhang, C., Wu, Y. & Yang, Y. Forecasting carbon dioxide price using a time-varying high-order moment hybrid model of NAGARCHSK and gated recurrent unit network. *Int. J. Environ. Res. Public Health* **19**, 899. <https://doi.org/10.3390/ijerph19020899> (2022).
8. Niu, X., Wang, J. & Zhang, L. Carbon price forecasting system based on error correction and divide-conquer strategies. *Appl. Soft Comput.* **118**, 107935. <https://doi.org/10.1016/j.asoc.2021.107935> (2022).
9. Yun, P., Huang, X., Wu, Y. & Yang, X. Forecasting carbon dioxide emission price using a novel mode decomposition machine learning hybrid model of CEEMDAN-LSTM. *Energy Sci. Eng.* **11**, 79–96 (2023).
10. Li, Y. & Song, J. Research on the application of GA-ELM model in carbon trading price—An example of Beijing. *Pol. J. Environ. Stud.* **31**, 149–161 (2022).
11. Du, Y., Chen, K., Chen, S. & Yin, K. Prediction of carbon emissions trading price in Fujian province: Based on BP neural network model. *Front. Energy Res.* **10**, 939602. <https://doi.org/10.3389/fenrg.2022.939602> (2022).

12. Li, H. *et al.* Forecasting carbon price in China: A multimodel comparison. *Int. J. Environ. Res. Public Health* **19**, 6217. <https://doi.org/10.3390/ijerph19106217> (2022).
13. Hao, Y. & Tian, C. A hybrid framework for carbon trading price forecasting: The role of multiple influence factor. *J. Clean. Prod.* **262**, 120378. <https://doi.org/10.1016/j.jclepro.2020.120378> (2020).
14. Zhu, B. & Wei, Y. Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. *Omega* **41**, 517–524 (2013).
15. Paoletta, M. S. & Taschini, L. An econometric analysis of emission allowance prices. *J. Bank. Financ.* **32**, 2022–2032 (2008).
16. Wang, Z. & Ye, D. Forecasting Chinese carbon emissions from fossil energy consumption using non-linear grey multivariable models. *J. Clean. Prod.* **142**, 600–612 (2017).
17. García-Martos, C., Rodríguez, J. & Sánchez, M. J. Modelling and forecasting fossil fuels, CO<sub>2</sub> and electricity prices and their volatilities. *Appl. Energy* **101**, 363–375 (2013).
18. Dutta, A. Modeling and forecasting the volatility of carbon emission market: The role of outliers, time-varying jumps and oil price risk. *J. Clean. Prod.* **172**, 2773–2781 (2018).
19. Sun, W. & Wang, Y. Factor analysis and carbon price prediction based on empirical mode decomposition and least squares support vector machine optimized by improved particle swarm optimization. *Carbon Manage.* **11**, 315–329 (2020).
20. Chi, G., Uddin, M. S., Abedin, M. Z. & Yuan, K. Hybrid model for credit risk prediction: An application of neural network approaches. *Int. J. Artif. Intell. Tools* **28**, 1950017. <https://doi.org/10.1142/S0218213019500179> (2019).
21. Vo, T. H., Nguyen, N. T. K., Kha, Q. H. & Le, N. Q. K. On the road to explainable AI in drug-drug interactions prediction: A systematic review. *Comput. Struct. Biotechnol. J.* **20**, 2112–2123 (2022).
22. Akhtar, M. & Moridpour, S. A review of traffic congestion prediction using artificial intelligence. *J. Adv. Transp.* **2021**, 1–18 (2021).
23. Popoola, S. I., Adebisi, B., Ande, R., Hammoudeh, M. & Atayero, A. A. Memory-efficient deep learning for botnet attack detection in IoT networks. *Electronics* **10**, 1104. <https://doi.org/10.3390/electronics10091104> (2021).
24. Atsalakis, G. S. Using computational intelligence to forecast carbon prices. *Appl. Soft Comput.* **43**, 107–116 (2016).
25. Zhu, B., Shi, X., Chevallier, J., Wang, P. & Wei, Y. An adaptive multiscale ensemble learning paradigm for nonstationary and nonlinear energy price time series forecasting. *J. Forecast.* **35**, 633–651 (2016).
26. Xu, W., Wang, J., Zhang, Y., Li, J. & Wei, L. An optimized decomposition integration framework for carbon price prediction based on multi-factor two-stage feature dimension reduction. *Ann. Oper. Res.* 1–38 (2022).
27. Momeneh, S. & Nourani, V. Application of a novel technique of the multi-discrete wavelet transforms in hybrid with artificial neural network to forecast the daily and monthly streamflow. *Model. Earth Syst. Environ.* **8**, 4629–4648 (2022).
28. Sun, G. *et al.* A carbon price forecasting model based on variational mode decomposition and spiking neural networks. *Energies* **9**, 54. <https://doi.org/10.3390/en9010054> (2016).
29. Wang, J., Cheng, Q. & Sun, X. Carbon price forecasting using multiscale nonlinear integration model coupled optimal feature reconstruction with biphasic deep learning. *Environ. Sci. Pollut. Res.* **29**, 85988–86004 (2022).
30. E, J., Ye, J., He, L. & Jin, H. Energy price prediction based on independent component analysis and gated recurrent unit neural network. *Energy* **189**, 116278. <https://doi.org/10.1016/j.energy.2019.116278> (2019).
31. Liu, J., Guo, Y., Chen, H., Ren, H. & Tao, Z. Multi-scale combined prediction of carbon prices based on non-structural data and popular learning. *Control Decis.* 279–286 (2019).
32. Zhou, J., Yu, X. & Yuan, X. Predicting the carbon price sequence in the shenzhen emissions exchange using a multiscale ensemble forecasting model based on ensemble empirical mode decomposition. *Energies* **11**, 1907. <https://doi.org/10.3390/en11071907> (2018).
33. Tsai, M. T. & Kuo, Y. T. A forecasting system of carbon price in the carbon trading markets using artificial neural network. *Int. J. Environ. Sci. Dev.* **4**, 163–167 (2013).
34. Wang, X. *et al.* A combined prediction model for hog futures prices based on woa-lightgbm-ceemdan. *Complexity* **2022**, 3216036. <https://doi.org/10.1155/2022/3216036> (2022).
35. Wu, Z. & Huang, N. E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **1**, 1–41 (2009).
36. Torres, M. E., Colominas, M. A., Schlotthauer, G. & Flandrin, P. A complete ensemble empirical mode decomposition with adaptive noise. *ICASSP*. 4144–4147 (2011).
37. Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. *KDD*. 785–794 (2016).
38. Mirjalili, S., Mirjalili, S. M. & Lewis, A. Grey wolf optimizer. *Adv. Eng. Softw.* **69**, 46–61 (2014).
39. Sun, W. & Xu, C. Carbon price prediction based on modified wavelet least square support vector machine. *Sci. Total Environ.* **754**, 142052. <https://doi.org/10.1016/j.scitotenv.2020.142052> (2021).
40. Li, Y. Current situation, problems and suggestions of carbon emission trading pilot in Fujian province. *Macrocon. Manag.* 66–71 (2018).
41. Ding, G., Deng, Y. & Lin, S. A study on the classification of China's provincial carbon emissions trading policy instruments: Taking Fujian province as an example. *Energy Rep.* **5**, 1543–1550 (2019).
42. Peng, X. & Zhong, Y. Factors influencing carbon emission trading prices and strategies. *Price Mon.* **12**, 25–31 (2021).
43. Sun, X., Hao, J. & Li, J. Multi-objective optimization of crude oil-supply portfolio based on interval prediction data. *Ann. Oper. Res.* **309**, 611–639 (2022).
44. Guo, F. & Pan, X. Carbon market: Price fluctuations and risk measurement—An empirical analysis based on EU ETS futures contract prices. *Fin. Trade Econ.* 110–118 (2011).
45. Zhang, Y. & Wei, Y. An overview of current research on EU ETS: Evidence from its operating mechanism and economic effect. *Appl. Energy* **87**, 1804–1814 (2010).
46. Qin, Q., He, H., Li, L. & He, L. Y. A novel Decomposition-Ensemble based carbon price forecasting model integrated with local polynomial prediction. *Comput. Econ.* **55**, 1249–1273 (2020).
47. Xie, Q., Hao, J., Li, J. & Zheng, X. Carbon price prediction considering climate change: A text-based framework. *Econ. Anal. Policy* **74**, 382–401 (2022).
48. Christiansen, A. C., Arvanitakis, A., Tangen, K. & Hasselknippe, H. Price determinants in the EU emissions trading scheme. *Clim. Policy* **5**, 15–30 (2005).
49. Mansanet-Bataller, M., Pardo, A. & Valor, E. CO<sub>2</sub> prices, energy and weather. *Energy J.* **28**, 73–92 (2007).
50. Chevallier, J., Nguyen, D. K. & Reboredo, J. C. A conditional dependence approach to CO<sub>2</sub>-energy price relationships. *Energy Econ.* **81**, 812–821 (2019).
51. Ji, Q., Zhang, D. & Geng, J. Information linkage, dynamic spillovers in prices and volatility between the carbon and energy markets. *J. Clean. Prod.* **198**, 972–978 (2018).
52. Alberola, E., Chevallier, J. & Chèze, B. Price drivers and structural breaks in European carbon prices 2005–2007. *Energy Policy* **36**, 787–797 (2008).
53. Tian, C. & Hao, Y. Point and interval forecasting for carbon price based on an improved analysis-forecast system. *Appl. Math. Model.* **79**, 126–144 (2020).
54. Zhu, B. *et al.* A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting. *Energy Econ.* **70**, 143–157 (2018).

55. Huang, Y. & He, Z. Carbon price forecasting with optimization prediction method based on unstructured combination. *Sci. Total Environ.* **725**, 138350. <https://doi.org/10.1016/j.scitotenv.2020.138350> (2020).
56. Dimitriadis, D., Katrakilidis, C. & Karakotsios, A. Investigating the dynamic linkages among carbon dioxide emissions, economic growth, and renewable and non-renewable energy consumption: Evidence from developing countries. *Environ. Sci. Pollut. Res.* **28**, 40917–40928 (2021).
57. Alberola, E., Chevallier, J. & Chèze, B. Emissions compliances and carbon prices under the EU ETS: A country specific analysis of industrial sectors. *J. Policy Model.* **31**, 446–462 (2009).
58. Considine, T. J. The impacts of weather variations on energy demand and carbon emissions. *Resour. Energy Econ.* **22**, 295–314 (2000).
59. Creti, A., Jouvet, P. A. & Mignon, V. Carbon price drivers: Phase I versus Phase II equilibrium?. *Energy Econ.* **34**, 327–334 (2012).
60. Han, M., Ding, L., Zhao, X. & Kang, W. Forecasting carbon prices in the Shenzhen market, China: The role of mixed-frequency factors. *Energy* **171**, 69–76 (2019).
61. Ji, C., Hu, Y., Tang, B. & Qu, S. Price drivers in the carbon emissions trading scheme: Evidence from Chinese emissions trading scheme pilots. *J. Clean. Prod.* **278**, 123469. <https://doi.org/10.1016/j.jclepro.2020.123469> (2021).
62. Zheng, H., Song, M. & Shen, Z. The evolution of renewable energy and its impact on carbon reduction in China. *Energy* **237**, 121639. <https://doi.org/10.1016/j.energy.2021.121639> (2021).
63. Ji, C., Hu, Y. & Tang, B. Research on carbon market price mechanism and influencing factors: A literature review. *Nat. Hazards* **92**, 761–782 (2018).
64. Sun, W., Sun, C. & Li, Z. A hybrid carbon price forecasting model with external and internal influencing factors considered comprehensively: A case study from China. *Pol. J. Environ. Stud.* **29**, 3305–3316 (2020).
65. Li, J. & Liu, D. Carbon price forecasting based on secondary decomposition and feature screening. *Energy* **278**, 127783. <https://doi.org/10.1016/j.energy.2023.127783> (2023).
66. Zhou, J. & Wang, S. A carbon price prediction model based on the secondary decomposition algorithm and influencing factors. *Energies* **14**, 1328. <https://doi.org/10.3390/en14051328> (2021).

## Acknowledgements

This research was funded by the National Natural Science Foundation of China (Grant no. 81973791).

## Author contributions

All authors contributed to the study's conception and design. M.F.: data analysis, visualization of experimental results, and writing the original draft. Y.D.: materials preparation and financial support. X.W.: data collection and methodological interpretation. J.Z.: manuscript check. L.M.: manuscript revision. All authors have read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to M.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023