



An Ensemble Prediction System Based on Artificial Neural Networks and Deep Learning Methods for Deterministic and Probabilistic Carbon Price Forecasting

Yi Yang¹, Honggang Guo^{2*}, Yu Jin² and Aiyi Song²

¹School of Information Science and Engineering, Lanzhou University, Lanzhou, China, ²School of Statistics, Dongbei University of Finance and Economics, Dalian, China

OPEN ACCESS

Edited by:

Wendong Yang,
Shandong University of Finance and
Economics, China

Reviewed by:

Yao Dong,
Jiangxi University of Finance and
Economics, China

Jamshid Piri,
Zabol University, Iran

*Correspondence:

Honggang Guo
ghg970612@163.com

Specialty section:

This article was submitted to
Environmental Economics and
Management,
a section of the journal
Frontiers in Environmental Science

Received: 12 July 2021

Accepted: 23 August 2021

Published: 17 September 2021

Citation:

Yang Y, Guo H, Jin Y and Song A
(2021) An Ensemble Prediction System
Based on Artificial Neural Networks
and Deep Learning Methods for
Deterministic and Probabilistic Carbon
Price Forecasting.
Front. Environ. Sci. 9:740093.
doi: 10.3389/fenvs.2021.740093

Carbon price prediction is important for decreasing greenhouse gas emissions and coping with climate change. At present, a variety of models are widely used to predict irregular, nonlinear, and nonstationary carbon price series. However, these models ignore the importance of feature extraction and the inherent defects of using a single model; thus, accurate and stable prediction of carbon prices by relevant industry practitioners and the government is still a huge challenge. This research proposes an ensemble prediction system (EPS) that includes improved data feature extraction technology, three prediction submodels (GBiLSTM, CNN, and ELM), and a multiobjective optimization algorithm weighting strategy. At the same time, based on the best fitting distribution of the prediction error of the EPS, the carbon price prediction interval is constructed as a way to explore its uncertainty. More specifically, EPS integrates the advantages of various submodels and provides more accurate point prediction results; the distribution function based on point prediction error is used to establish the prediction interval of carbon prices and to mine and analyze the volatility characteristics of carbon prices. Numerical simulation of the historical data available for three carbon price markets is also conducted. The experimental results show that the ensemble prediction system can provide more effective and stable carbon price forecasting information and that it can provide valuable suggestions that enterprise managers and governments can use to improve the carbon price market.

Keywords: carbon price forecasting, ensemble prediction system, deep learning methods, error distribution function, multiobjective optimization algorithm

INTRODUCTION

This section describes the research background, provides a literature review, and states the purpose and innovation of this study.

Research Background

With the rapid development of the economy, the environment and the climate will inevitably change. Climate change is clearly a common problem facing all countries. On February 16, 2005, the Kyoto

Protocol went into effect. According to the situation in each country, specific emission reduction plans and schedules were formulated. In January 2005, the EU emissions trading scheme (EU ETS), which was designed to achieve the emission reduction targets stipulated in the Kyoto protocol, was introduced (Arouri et al., 2012). The EU ETS allocates carbon trading quotas to different emission entities according to its regulations, and entities that exceed the quota must purchase emission rights from entities that are lower than the quota through the carbon trading market. This measure of using a market trading mechanism provides valuable experience for solving the problem of global climate change.

As the world's largest carbon dioxide emitter (in 2018, its total carbon dioxide emissions reached 10 billion tons, accounting for approximately 30% of global carbon dioxide emissions), China has successively established eight carbon emission trading markets since 2013. However, this system is still in the construction stage, and the market mechanism is not perfect and needs further improvement. By studying the regular price fluctuation pattern of the EU ETS and China's carbon trading market, analyzing the influencing factors, and forecasting the carbon market price accordingly, we can better understand the fluctuation law of the carbon market and obtain a reference for formulating carbon market policies and mechanisms to improve the ability to regulate this market.

Carbon prices have important implications for governments, companies, and long-term investors. For governments, carbon pricing is one of the mechanisms used to reduce carbon emissions, and it can also be a source of revenue. Companies can use internal carbon pricing to assess the impact of mandatory carbon pricing on their businesses and to identify potential climate risks and revenue opportunities. Long-term investors are using carbon pricing to reevaluate their investment strategies. Therefore, regardless of the point of view, it is necessary to establish an accurate and stable carbon price forecasting system.

Literature Review

Most of the research methods used in carbon price prediction rely on historical data to build models to predict carbon prices. Carbon prices display high volatility and nonlinear structure, and many studies of carbon price prediction based on historical data have been conducted in recent years. The prediction methods can be divided into three categories: 1) statistical measurement methods; 2) artificial intelligence methods; and 3) decomposition integration hybrid forecasting methods.

Statistical Measurement Method

As a classical time series forecasting method, statistical measurement methods, including linear regression models, autoregressive integrated moving averages (ARIMAs), generalized autoregressive conditional heteroscedasticity (GARCH) models, and gray model GM (1, 1) (Chevallier, 2009; Byun and Cho, 2013; Zhu and Wei, 2013) are widely used in carbon trading price prediction and volatility analysis. For example, Benz and Trück (2008) proposed the Markov transition and AR-GARCH model for stochastic modeling and

analyzed the short-term price of the carbon dioxide emission quota of the EU ETS. Through the empirical results obtained, it was demonstrated that the prediction performance of the Markov state transition model is better than that of the GARCH model. Zhu and Wei (2013) combined least squares SVM with the ARIMA model, and the results showed that the developed model was more robust than the single-prediction model. Zhu B et al. (2018) used grey correlation analysis to analyze the carbon price market. The traditional statistical model has high prediction accuracy and wide adaptability in linear and stable time series. However, because carbon prices show strong volatility, nonlinearity, and instability, traditional statistical measurement methods cannot capture internal structural characteristic data (Lu et al., 2019). Therefore, accurate forecasting of carbon prices requires the use of a method with a strong nonlinear feature extraction ability that enables it to take into account potential nonlinear characteristics. In addition, the traditional statistical measurement method is more suitable for the long-term prediction of time series, and its short-term carbon price prediction performance is poor (Cheng and Wang, 2020).

Owing to the shortcomings of statistical models, *artificial intelligence methods* (AI) have gradually become widely used in time series prediction; these methods are suitable for nonlinear prediction without any assumption of data distribution (Wang et al., 2020). Increasing evidence shows that the performance of AI in nonlinear time series is better than that of other models (Zhang et al., 2017). AI, including back-propagation neural networks (BPs), multilayer perceptual neural networks (MLPs), least squares support vector regression (LSSVR), and hybrid prediction methods combined with optimization algorithms, have also been widely used in carbon price forecasting. Atsalakis (2016) combined a hybrid fuzzy controller called PATSOS with an adaptive neuro fuzzy inference system (ANFIS). The research shows that this method can produce accurate and timely prediction results. Fan et al. (2015) studied the chaotic characteristics of the EU ETS, used the neural network model of MLP to predict carbon prices, and found that the forecasting accuracy of the model was significantly improved. Tian and Hao (2020) used phase space reconstruction technology and the ELM under the multiobjective grasshopper optimization algorithm (MOGOA-ELM) to predict the trend of the EU ETS and China's carbon prices. The empirical results show that this method can be used effectively to predict carbon prices.

In recent years, with the development of deep learning theory (DL) in image detection, audio detection, and other fields, DL has become the focus of many scholars (Liu et al., 2021). The unique storage unit structure of deep learning allows it to retain past historical data and has significant advantages for processing time series data that feature long processing intervals and delays (Zhang B et al., 2018). Niu et al. (2020) combined LSTM and GRU to establish a deep learning recursive forecasting unit for forecasting multiple financial data. Liu et al. (2020) proposed a new wind speed prediction model based on an error correction strategy and the LSTM algorithm to predict short-term wind speed. The experimental results demonstrated that its performance is better than that of other comparable models.

However, application of deep learning frameworks to carbon price prediction is still very limited.

In addition to the selection and optimization of prediction methods, data preprocessing technology also plays an indispensable role in the prediction accuracy of the prediction model (Wang et al., 2021). **Decomposition and integration methods**, including empirical mode decomposition (EMD), singular spectrum denoising (SSA), and variational mode decomposition (VMD) are widely used in time series data preprocessing. These methods aim to decompose and reconstruct the original time series data and extract the effective features of the time series. Decomposing the original time series into a series of simple patterns that exhibit strong regularity can significantly improve the prediction accuracy of time series. Wei et al. (2018) used wavelet transform and kernel ELM to predict carbon prices. Zhu J et al. (2018) explored an efficient prediction model based on VMD mode reconstruction and optimal combination and thereby greatly improved the prediction accuracy of carbon prices. However, the above decomposition methods still have some shortcomings. For example, in wavelet decomposition and VMD, it is necessary to determine the wavelet basis function and the decomposition level. Although in EMD it is not necessary to determine the number of decomposition levels, mode aliasing and insufficient noise separation cannot be solved (Jin et al., 2020). Therefore, it is very important to extract the nonlinear peculiarities of carbon prices by using appropriate data preprocessing methods.

A single prediction model cannot achieve good performance on every dataset. Therefore, researchers began to focus on combination forecasting models. In essence, combination forecasting models combine different hybrid forecasting methods or single forecasting methods using weighting. In many experimental studies, it is found that the use of a combination of prediction methods produces better predictions than the use of a method that is based on a single-prediction model. The advantage of using combination models is that different time series may have different information sets, information features, and modeling structures, and the use of a combination of prediction methods can result in good performance in the case of such structure mutations. Although the use of a combination forecasting method to forecast time series is very common, use of a combination forecasting model to forecast carbon prices is still in its infancy.

The above analysis indicates that most research on carbon prices is driven by single or hybrid forecasting models and that it tends to emphasize prediction strategies that are based on certainty and to largely ignore the importance of uncertainty analysis of carbon prices. Regardless of the type of prediction model used, there are inherent and irreducible uncertainties in each prediction that will greatly increase the possibility of miscalculation (Du et al., 2020). Therefore, quantification of the uncertainty of carbon price prediction plays an indispensable role in exploring the complexity of the carbon price market and strengthening the ability to conduct effective market anti-risk management.

Objectives and Contributions

To supplement the existing research on carbon price prediction, an ensemble prediction system (EPS) based on the ICEEMDAN data preprocessing method, the deep learning algorithm (DL), the extreme learning machine (ELM), and the multiobjective dragonfly optimization algorithm (MODA) is developed and used to analyze the certainty and uncertainty of carbon prices. Specifically, ICEEMDAN is employed to decompose and reconstruct the original carbon price data and extract the effective features of the data, and the results are transferred into the submodels of EPS as training data (the submodels are ICEEMDAN-GBiLSTM, ICEEMDAN-CNN, and ICEEMDAN-ELM). Using the MODA, the final carbon price point forecast results are then obtained through a weighted combination of the submodel prediction results. For interval prediction, the upper and lower bounds of the prediction interval are constructed based on the prediction value of ESP and the best fit distribution function of error, namely, the T location-scale (TLS) distribution. The main innovations presented in this study are as follows:

- 1) **An effective ensemble prediction system of carbon prices is developed.** Two hybrid prediction models based on a deep learning algorithm (ICEEMDAN-GBiLSTM and ICEEMDAN-CNN) and a feedforward neural network (ICEEMDAN-ELM) are combined to **overcome the inherent defects of a single hybrid prediction model.**
- 2) A deep learning recurrent neural network, GBiLSTM, is first proposed as a prediction submodel of the EPS. **GBiLSTM combines two recursive deep learning algorithms; it can effectively deal with time series with long memory and increase the accuracy of carbon price forecasting.**
- 3) The MODA is employed as an effective method of weighting the ensemble prediction system. It optimizes the weight coefficient of **the ensemble model from the perspective of prediction accuracy and prediction stability**, thereby overcoming the obvious defect that single objective optimization can only select one objective function.
- 4) **To overcome the nonlinearity and strong volatility of the original carbon price data, an effective time series preprocessing technique is developed.** ICEEMDAN sequence decomposition technology is employed to decompose and reconstruct the original carbon price data, extract the salient features of the data, and improve the prediction accuracy of the EPS.
- 5) **By fitting the best error distribution, the uncertainty of the carbon price is mined.** In the past, the error distribution of a prediction was usually assumed to be a Gaussian distribution. In this study, **five types of parameter distribution functions are used to fit the prediction error**, the best error distribution function is found, and the ranges of carbon price interval prediction are constructed.

The remainder of the study is organized as follows. In Section *Model Theory and Related Work* and Section *Ensemble Prediction System and its Interval Forecasting Framework*, we introduce the theoretical method and the framework used in the proposed EPS.

Section *Experiment and Analysis* describes the experimental data and the prediction performance evaluation index. The point prediction and interval prediction of the carbon price are then simulated. Section *Discussion* is a further discussion of EPS, and a summary of the study is presented in Section *Conclusion*.

MODEL THEORY AND RELATED WORK

This section introduces the corresponding theories and describes the functions of the data preprocessing module, the combination prediction module, and the uncertainty mining module of the EPS prediction system.

Data Preprocessing

The data processing module includes the data feature extraction method, which is based on improved complex ensemble empirical mode decomposition with adaptive noise (ICEEMDAN), and the data feature selection method, which is based on the partial autocorrelation function (PACF).

Data Feature Extraction

To improve the problem of mode aliasing in the traditional noise reduction method EMD and the slight residual noise in CEEMDAN, the ICEEMDAN technology is improved. The CEEMDAN method adds Gaussian white noise during the decomposition process, while the ICEEMDAN method adds a special type of white noise, $E_k(\omega^{(i)})$, that is, the k -th IMF component of the Gaussian white noise (M.E. Torres et al., 2011; M.A. Colominas et al., 2014). The local mean value of the added noise is calculated for each modal component, and the IMF is defined as the difference between the residual signal and the local mean.

- 1) The definition operator $E_k(\cdot)$ represents the k -th IMF after EMD decomposition, and $M(\cdot)$ represents the local mean value of the signal. There is $E_1(x) = x - M(x)$. Operator means taking the mean value, and x represents the original data of the study, and then the local average value is calculated by EMD:

$$x^{(i)} = x + \alpha_0 E_1(\omega^{(i)}) \tag{1}$$

where $\omega^{(i)}$ is the i th white noise added and α_0 is the standard deviation of the noise. The first residual component $r_1 = \langle M(x^{(i)}) \rangle$ is obtained by taking the local mean value. The first intrinsic mode function IMF_1 value $\tilde{d}_1 = x - r_1$ is calculated.

- 2) The value \tilde{d}_2 of the second mode component IMF_2 is calculated:

$$\tilde{d}_2 = r_1 - r_2 = r_1 - \langle M(r_1 + \alpha_1 E_2(\omega^{(i)})) \rangle \tag{2}$$

- 3) The k -th residual is calculated:

$$r_k = \langle M(r_{k-1} + \alpha_{k-1} E_k(\omega^{(i)})) \rangle, k = 2, 3, \dots, N \tag{3}$$

- 4) The value of the k -th mode component $IMF_k, \tilde{d}_k = r_{k-1} - r_k$, is calculated, and Eq. 3 is repeated until the residual satisfies the iteration termination condition, which is **Cauchy** convergence. The standard deviation between two adjacent **IMF** components $\chi = \|\tilde{d}_k - \tilde{d}_{k-1}\|_2 / \|\tilde{d}_{k-1}\|_2$ is less than a specified value.

In this study, ICEEMDAN is used to decompose the original carbon price data into several intrinsic mode functions (IMFs). The IMF with the highest frequency is removed, and the remaining IMFs are included. Through this method of deconstruction and reorganization, the problem of strong volatility and randomness of the original data is solved. The data features are effectively extracted, and the prediction veracity of the model is increased.

Data Feature Selection

The partial autocorrelation function (PACF) is an effective method for distinguishing the structural features of sequences (Jiang et al., 2020). It can be used to calculate the partial correlation between the time series and its lag term. If Φ_{kj} is employed to represent the j -th regression coefficient in the k -order autoregressive equation, the model can be expressed as follows:

$$x_t = \Phi_{k1}x_{t-1} + \Phi_{k2}x_{t-2} + \dots + \Phi_{kk}x_{t-k} + \mu_t \tag{4}$$

where x_t is the time series and Φ_{kk} is the last coefficient. If Φ_{kk} is defined as a function of lag time k , then $\Phi_{kk}, k = 1, 2, \dots$ is named partial autocorrelation function.

In this study, PACF is used to find the lag terms that have the strongest correlation with the time series; these are then used as the input characteristics of the forecast model.

Ensemble Prediction Module

The prediction value calculated by the ensemble prediction system is obtained using an ensemble of the prediction results of different single-prediction components through the weighting strategy. In this section, the three submodes of the proposed EPS and the MODA weighting optimization strategy are introduced.

Convolutional Neural Network

A CNN is an incompletely connected DL network structure that is composed of two special neural networks: a convolution layer and a down sampling layer (Wang, 2020). The neurons in each layer of the CNN are locally connected, enabling them to realize hierarchical feature extraction and transformation of the input. Neurons with the same connection weight are connected to different regions of the upper neural network; in this way, a translation-invariant neural network is obtained (Wang, 2018).

- 1) **The training of the Convolution Layer.** The CNN is connected to the local region of the feature surface by a convolution kernel. The output characteristic surface size of each convolution layer must meet the following requirement:

$$oMapN = \left(\frac{(iMapn - CWindow)}{CInterval} + 1 \right) \tag{5}$$

In Eq. 5, $oMapN$ is the number of output feature surfaces of each convolution layer, $iMapN$ is the number of input feature surfaces, $CWindow$ is the size of the convolution kernel, and $CInterval$ is the sliding size of the convolution kernel.

In general, to ensure integral division in the above formula, it is necessary to train the number of parameters for each convolution layer of the CNN so as to satisfy the following condition:

$$CParams = (iMap \times CWindow + 1) \times oMap \quad (6)$$

where $CParams$ is the number of parameters, $iMap$ is the input feature surface, and $oMap$ is the output feature surface.

The output value x_{nk}^{out} can be obtained by the convolution layer; the formula is as follows:

$$x_{nk}^{out} = f_{cov} \left(x_{1h}^{in} \times w_{1(h)n(k)} + x_{1(h+1)}^{in} \times w_{1(h+1)n(k)} + x_{1(h+2)}^{in} \times w_{1(h+2)n(k)} + \dots + b_n \right) \quad (7)$$

where x_{nk}^{in} is the input value, b_n is the offset value of the output characteristic surface n , and $f_{cov}(\cdot)$ is the excitation function. The excitation function is usually the ReLU function, and the formula for its calculation is as follows:

$$f_{cov}(x) = MAX(0, x) \quad (8)$$

2) **The output of the Pooling Layer.** The pooling layer is also composed of several feature surfaces, and the number of feature surfaces does not change. The output value of the pooling layer is

$$t_{nl}^{out} = f_{sub} \left(t_{nq}^{in}, t_{n(q+1)}^{in} \right) \quad (9)$$

where t_{nq}^{in} is the output value of the q -th neuron on the n -th input characteristic surface of the pooling layer and $f_{sub}(\cdot)$ is a function that takes either the maximum value or the mean value. The size $DoMapN$ of each output feature surface of the pooling layer is

$$DoPapN = \left(\frac{oMapN}{DWindow} \right) \quad (10)$$

3) Full connection layer output. In the CNN structure, one or more fully connected layers are connected after the multiple convolution layers and the pooling layers are obtained. The ReLU function is also used in the excitation function of the whole connected layer.

In this study, CNN, as a component of the combined forecasting system, forecasts the carbon price.

Deep Learning Recursive Network Structure (GBiLSTM)

In this study, we developed a deep learning recurrent network structure, which is a hybrid of BiLSTM and GRU. The structure diagram is shown in Figure 1.

Bidirectional Long Short Term Memory Neural Network

BiLSTM is an improved network of LSTM. LSTM cannot capture information from back to front; however, BiLSTM can solve this

problem. When bidirectional sequence information is captured, the time series can be predicted more accurately (Hochreiter and Schmidhuber, 1997).

1) The LSTM mechanism consists of three memory gates: an input gate (i_t), a forgetting gate (f_t), and an output gate (O_t). The specific expression is as follows:

$$\begin{cases} i_t = \sigma(\omega_i X_t + R_i H_{t-1} + B_i) \\ f_t = \sigma(\omega_f X_t + R_f H_{t-1} + B_f) \\ o_t = \sigma(\omega_o X_t + R_o H_{t-1} + B_o) \\ \tilde{c} = f_t \otimes c_{t-1} \\ H_t = o_t \otimes \tanh(c_t) \end{cases} \quad (11)$$

where x_t , σ , and c_t represent the input sample, the **sigmoid** activation function, and the storage unit of time t , respectively, and (B_f, B_i, B_o) and $(\omega_f, \omega_i, \omega_o)$ represent the deviation and the weight matrix, respectively, of each gate. The symbol \otimes represents the corresponding multiplication of elements. First, H_{t-1} , c_{t-1} , and X_t transmit input information to the LSTM unit. The LSTM gate then interacts with X_t . After a new cell state c_{t-1} is established. In this stage, f_t determines which information needs to be stored or deleted and then updates the cell status.

2) Because BiLSTM transmits time series data to LSTM from both the forward and backward directions, it has two output layers: a forward layer $H_t^f = o_t^f \otimes \tanh(c_t^f)$ and a backward layer $H_t^b = o_t^b \otimes \tanh(c_t^b)$.

3) The final predicted output value \hat{y}_t is obtained by integrating the forward layer and the backward layer; in Eq. 12, α and β are numerical factors that satisfy the equation $\alpha + \beta = 1$ (Shi et al., 2015).

$$\begin{cases} H_t = \alpha H_t^f + \beta H_t^b \\ \hat{y}_t = \sigma(H_t) \end{cases} \quad (12)$$

Gated Recurrent Unit

GRU is an effective variant of LSTM; its structure is simpler than that of LSTM, and it can well capture the nonlinear relationship between sequence data, thereby effectively alleviating the problems of traditional RNN gradient disappearance (Chung et al., 2014).

1) The GRU model has two gating units: an update gate z_t and a reset gate r_t . An update gate is employed to equilibrate the historical information. The smaller the value of the update gate is, the more concentrated the output of the model on the information of the previous hidden layer h_t is

$$Z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (13)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (14)$$

In Eqs 13, 14, W is the model weight, and σ is the activation function.

2) By resetting the gate r_t , the candidate vector $\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$ can be calculated. Taking the

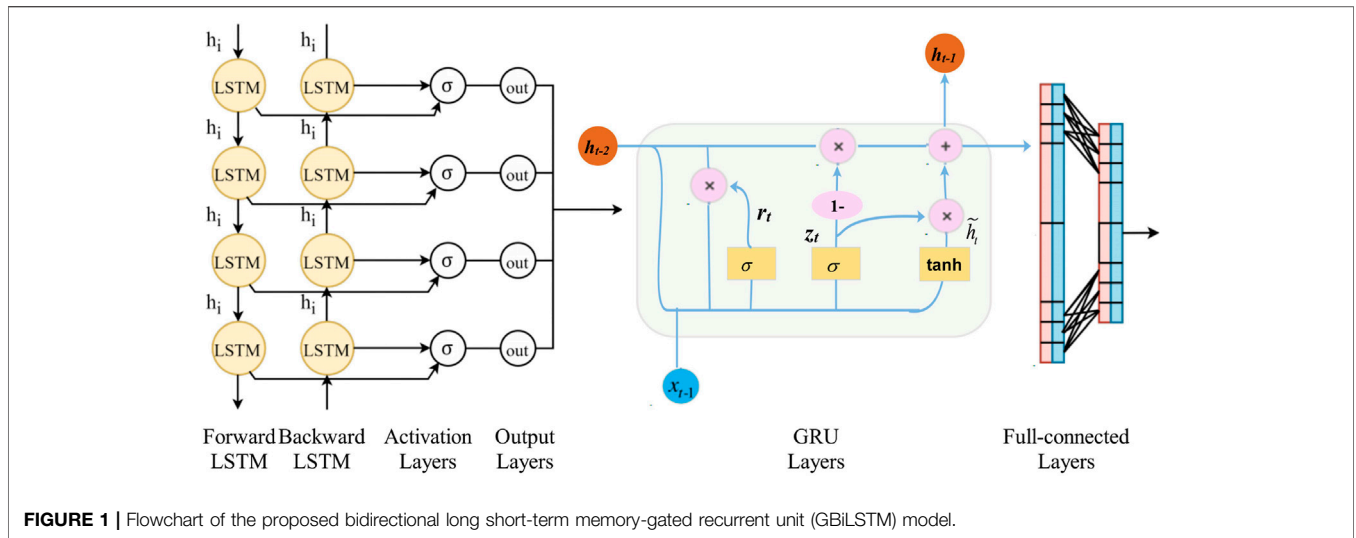


FIGURE 1 | Flowchart of the proposed bidirectional long short-term memory-gated recurrent unit (GBiLSTM) model.

value of the update gate as the weight, \tilde{h}_t to be added and the state of the hidden layer at the previous time step are recorded as the output of the GRU network at time step t , as follows:

$$h_t = (1 - Z_t) * h_{t-1} + Z_t * \tilde{h}_t \quad (15)$$

- 4) A set of training samples is input into the GRU; finally, the final output o is obtained by adding the fully connected layer after the GRU layer.

$$\begin{cases} h_{end} = GRU(x_{t-1}, x_{t-2}, \dots, x_{t-w}) \\ o = V \cdot h_{end} \end{cases} \quad (17)$$

In this study, a deep learning recursive network structure (GBiLSTM) based on BiLSTM and GRU is constructed. To reduce fitting error, the time series are trained by the BiLSTM layer and then transferred to the GRU network. Through this double deep learning layer network structure, we can better fit the carbon price data and reduce the prediction error.

Extreme Learning Machine

ELM is a type of feedforward neural network. On the premise of randomly selecting the input layer weight and the hidden layer neuron threshold, the output weight of the ELM can be obtained through a one-step calculation. ELM has the advantages of higher network generalization ability and strong nonlinear fitting ability (G B Huang et al., 2006; Jiang et al., 2021a; Jiang et al., 2021b).

- 1) For N different inputs (x_i, t_i) and $x_i \in R^p, t_i \in R^p, i = 1, 2, \dots, N$, the ELM with L nodes and the excitation function $f(x)$ can be expressed as

$$\sum_{i=1}^N \beta_i f(w_i \cdot x_j + b_i) = y_j \quad i = 1, \dots, N \quad (19)$$

where $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ is the weight connecting the i -th hidden layer node and the input node, β_i is the connection weight vector, and y_j is the output value of the j -th node. The training of

the network is equivalent to approximating N training samples with zero error; that is, $\hat{w}_i, \hat{b}_i, \hat{\beta}_i$ make

$$\sum_{i=1}^L \hat{\beta}_i f(\hat{w}_i \cdot x_j + \hat{b}_i) = t_j, j = 1, \dots, N \quad (20)$$

$$H\beta = T; \quad (21)$$

$$H = \begin{bmatrix} f(w_1 x_1 + b_1) & \dots & f(w_L x_1 + b_L) \\ \vdots & & \vdots \\ f(w_1 x_N + b_1) & \dots & f(w_L x_N + b_L) \end{bmatrix}_{N \times L} \quad (22)$$

In Eq. 20 through Eq. 22, $\beta = [\beta_1^T, \dots, \beta_L^T]_{L \times m}$; $T = [t_1^T, \dots, t_N^T]_{N \times m}$, and the i -th column of H represents the output vector of the i -th hidden layer node corresponding to the i -th hidden layer neuron of the input x_1, x_2, \dots, x_N .

- 2) The input connection weight W and the hidden layer node bias b can be randomly selected at the beginning of training, and the output connection weight β can be solved by solving the linear Eq. 23.

$$\min_{\beta} \|H\beta - T\| \quad (23)$$

- 3) The solution is $\hat{\beta} = H^\dagger T$; H^\dagger is the Moore-Penrose generalized inverse of the hidden layer output matrix H .

In this study, ELM is used as an excellent traditional neural network prediction component of the combined forecasting system to predict carbon prices.

Combination Strategy

It is generally believed that no single prediction model can achieve the best prediction performance for all datasets. Combining the values predicted by different prediction models usually reduces the overall risk of incorrect model selection. It is hoped that the diversity of models can help improve the final prediction results. However, the previously developed average weighting and weighted weighting methods cannot guarantee the

global optimality of the results (Wang, Y et al., 2018), and it is necessary to find an adaptive variable weight combination strategy.

In this study, the MODA algorithm is used to weigh the three prediction components. For the weighting strategy, we formulated the MODA algorithm as a linear programming (LP) problem to minimize the loss function. These theories are introduced in detail below:

Ensemble Method

Owing to the different weights given to each individual component of the ensemble prediction system, the formula used in the ensemble forecasting method is as follows:

$$f(\hat{x}_{c,t}) = \sum_{j=1}^m \omega_j f(\hat{x}_{j,t}), \quad t = 1, 2, \dots \quad (24)$$

where $f(\hat{x}_{c,t})$ is the final output, $f(\hat{x}_{j,t})$ is the prediction component of the EPS, m is the number of submodels, and ω_j is the weight of the component models. The experimental results demonstrate that the ensemble model can obtain ideal results when these weights are in the range of $[-2, 2]$.

Multiobjective Dragonfly Optimization algorithm

The dragonfly algorithm is a population-based heuristic intelligent algorithm that is easy to understand and implement (Mirjalili, 2016). The dragonfly algorithm is inspired by the static and dynamic group behaviors of dragonflies. In the static group behavior, the group preys; in the dynamic group behavior, the group migrates. These two behaviors are very similar to the two important stages in heuristic optimization algorithms: exploration and development. In this research, the MODA is applied to increase the accuracy and stability of the prediction system (Song and Li, 2017).

The mathematical expression methods are as follows:

- 1) The degree of separation refers to avoiding collisions between dragonflies and adjacent individuals.

$$S_i = \sum_{j=1}^N X - X_j \quad (25)$$

- 2) Alignment means that the trends in movement speed are the same in adjacent individuals.

$$A_i = \frac{\sum_{j=1}^N V_j}{N} \quad (26)$$

- 3) Cohesion refers to the tendency of dragonflies to gather near the center of adjacent individuals.

$$C_i = \frac{\sum_{j=1}^N X_j}{N} - X \quad (27)$$

- 4) Food attraction is the degree of attraction of dragonflies to food.

$$F_i = X^+ - X \quad (28)$$

- 5) The repulsive force of natural enemies refers to the repulsion of the group to natural enemies when dragonflies encounter natural enemies.

$$E_i = X^- + X \quad (29)$$

In Eq. 25 through Eq. 29, X is the position of the current dragonfly individual, X_j represents the position of the j -th adjacent dragonfly, V_j represents the speed of the j -th adjacent dragonfly, N represents the number of individuals adjacent to the i -th dragonfly individual, X^+ indicates the location of the food source, and X^- indicates the position of the natural enemy.

Based on the above five behaviors, the step length and the position of the next generation of dragonflies are calculated as follows:

$$\Delta X_{t+1} = (sS_i + aA_i + cC_i + fF_i + eE_i) + \omega \Delta X_t \quad (30)$$

$$X_{t+1} = X_t + \Delta X_{t+1} \quad (31)$$

Whether dragonflies are adjacent to each other can be judged by the Euclidean distance, which is similar to a circle with a radius of r around each dragonfly, and all individuals in the circle are adjacent. To speed up the convergence, the radius r should gradually increase during the iterative process and should finally include the entire search space (Sun et al., 2018). At the beginning of the iteration, the radius r is very small, and some individuals may have no adjacent individuals. To enhance the search power of the algorithm, the random walk is adopted to replace the step update formula, as shown below.

$$Levy(x) = 0.01 \times \frac{r_1 \times \sigma}{|r_2|^{\frac{1}{\beta}}} \quad (32)$$

In Eq. 32, r_1 and r_2 represent random numbers between $[0,1]$, β is a constant (here, $3/2$), and σ is calculated as follows:

$$\sigma = \left[\frac{\Gamma(1 + \beta) \sin(\frac{\pi\beta}{2})}{\Gamma(\frac{1+\beta}{2}) \beta 2^{\frac{\beta-1}{2}}} \right]^{\frac{1}{\beta}} \quad (33)$$

The corresponding position update formula can be derived as shown in the following formula:

$$X_{t+1} = X_t + Levy(d) \times X_t \quad (34)$$

In Eq. 34, d represents the dimension of the position vector. In MODA, the nondominated Pareto optimal solution that is obtained in the optimization process is stored and retrieved through the storage unit of the external archive. More importantly, to improve the distribution of solutions in the document and maintain the diversity of Pareto solution sets, the algorithm uses a roulette method with probability $P_i = c/N_i$ to keep the nondominated solution sets well distributed. N_i represents the number of solutions near the i -th solution, and c is a constant.

Objective Function of MODA

Generally, the multiobjective optimization problem can be regarded as the solution of the constraint problem. The

constraint problem with J inequalities and K equations can be written as follows:

$$\text{Min } F(x) = (\text{obf}_1(x), \text{obf}_2(x), \dots, \text{obf}_s(x))^T \quad (35)$$

$$\text{s.t. } \mathbf{g}_j(x) \geq 0, j = 1, 2, \dots, J, \quad x \in \Omega \quad (36)$$

$$\mathbf{h}_k(x) = 0, k = 1, 2, \dots, K,$$

where $(\text{obf}_1(x), \text{obf}_2(x), \dots, \text{obf}_s(x))^T$ is the decision vector.

In this study, the objective of the optimization algorithm is to determine the weight of each single-prediction component ω_j to minimize the error between the final combined forecast value $f(\hat{x}_{c,t})$ and the real value of the carbon price Y. The optimization algorithm can be expressed as follows:

$$\text{Min} \begin{cases} \text{obf}_1 = \text{std}(f(\hat{x}_{c,t}), Y) = \text{std}\left(\sum_{j=1}^m \omega_j f(\hat{x}_{j,t}) - Y_j\right) \\ \text{obf}_2 = \text{MAPE}(f(\hat{x}_{c,t}), Y) = \text{MAPE}\left(\sum_{j=1}^m \omega_j f(\hat{x}_{j,t}) - Y_j\right) \end{cases} \quad (37)$$

$$\text{s.t. } -2 \leq \sum_{j=1}^J \omega_j \leq 2$$

Therefore, we can solve the component weight ω_j :

$$\omega_j = \text{argMin} \begin{cases} \text{Std}(f(\hat{x}_{c,t}), Y) \\ \text{MAPE}(f(\hat{x}_{c,t}), Y) \end{cases} \quad (38)$$

$$\text{s.t. } -2 \leq \sum_{j=1}^J \omega_j \leq 2$$

Through continuous iteration of the MODA optimization algorithm, the weight vector $\omega = (\omega_1, \omega_2, \dots, \omega_j)$ that minimizes the error between the combination forecast value $f(\hat{x}_{c,t})$ and the real value of carbon price Y is obtained. In this study, $m = 3$.

Uncertainty Mining Module

The uncertainty information of point prediction results can be used to more deeply analyze the characteristics of carbon prices. In this article, an innovative interval prediction scheme based on prediction error distribution modeling in the training stage is proposed. Unlike previous research in which it is assumed that the prediction error follows a Gaussian distribution, this article uses maximum likelihood estimation (MLE) to conduct statistical research on carbon price error data and to explore its distribution. Among the five distribution functions developed, the function that best fits the distribution of carbon price prediction error is found. Based on its probability distribution function (PDF), the upper and lower bounds of the carbon price prediction interval are constructed. The details of the five distribution functions and interval prediction methods are given below.

Distribution Function

The probability distribution function plays a very important role in resource evaluation and interval prediction. This study attempts to use different DFSs to fit the distribution function

of prediction error, hoping to analyze the time series in a new way and to mine its uncertainty characteristics. In this section, five types of model prediction error distribution functions (stable, extreme value, normal, logistic, and t location-scale (TLS) functions) are introduced. The relevant probability density functions are shown in **Table 1**.

Interval Prediction Theory

Under the significance level α , for the limit of the model prediction error interval (I_{min} and I_{max}), the probability formula of the prediction model error value \hat{y}_{err} and the prediction error true value Y_{err} can be expressed as follows (Song et al., 2015):

$$P(I_{min} \leq Y_{err} \leq I_{max}) = 1 - 2\alpha \quad (39)$$

Since the error value of the prediction model is a random variable, Eq. 50 can also be written as follows:

$$P\{I_{min} \leq Y_{err} \leq I_{max} | E(Y_{err}) = \hat{y}_{err}\} = 1 - 2\alpha \quad (40)$$

In addition, we assume that the prediction error of the future prediction model has the same distribution function as that of the historical prediction model. Therefore, the probability distribution function (PDF) based on the historical error data of the prediction system can be regarded as a distribution function of future prediction error (Chen and Liu, 2021). Thus, the upper and lower bounds of the function at a certain confidence level can be calculated.

$$\left\{ (I_{min}, I_{max}) | I_{min} \leq Y_{err} \leq I_{max}, \int_{I_{min}}^{I_{max}} f(x|\hat{\Theta}) dx = 1 - 2\alpha \right\} \quad (41)$$

The above equation can also be written as

$$[\hat{I}_{min}, \hat{I}_{max}] = [\hat{I}_{min}, \hat{y}_{err}] [\hat{y}_{err}, \hat{I}_{max}] \quad (42)$$

$$\int_{\hat{I}_{min}}^{\hat{y}_{err}} f(x|\hat{\Theta}) dx = F(\hat{y}_{err}) - \alpha \quad (43)$$

$$\int_{\hat{y}_{err}}^{\hat{I}_{max}} f(x|\hat{\Theta}) dx = 1 - F(\hat{y}_{err}) - \alpha \quad (44)$$

After the optimal statistical distribution of the prediction error is determined, the upper \hat{U} and lower \hat{L} bounds of the carbon price prediction interval can be constructed.

$$[\hat{L}, \hat{U}] = [\hat{y}_{forecast} + \hat{I}_{min}, \hat{y}_{forecast} + \hat{I}_{max}] \quad (45)$$

In Eq. 45, $\hat{y}_{forecast}$ is the carbon price predicted by the carbon price prediction model.

ENSEMBLE PREDICTION SYSTEM AND ITS INTERVAL FORECASTING FRAMEWORK

This section introduces in detail the specific process used in this study. A brief overview of EPS and its uncertainty exploration is shown in **Figure 2**.

TABLE 1 | Probability distribution function (PDF) of the five distribution functions used in the study.

Distribution functions	PDF	Parameters
Extreme value	$f(x; \mu; \sigma) = \sigma^{-1} \exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right) \exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right), x > 0$	$\mu > 0$ location parameter $\sigma > 0$ scale parameter
Logistic	$f(x; \mu; \sigma) = \frac{\exp\left(\frac{x-\mu}{\sigma}\right)}{\sigma \left(1 + \exp\left[\frac{(x-\mu)}{\sigma}\right]\right)^2}$	$\mu > 0$ location parameter $\sigma > 0$ scale parameter
Normal	$f(x; \mu; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$	$\mu > 0$ location parameter $\sigma > 0$ scale parameter
Stable	$f(x; \gamma; \alpha; \beta; \delta) = \exp(-\gamma^{\alpha} x ^{1+\delta})$	$0 < \alpha < 2; -1 \leq \beta \leq 1$ shape parameter $0 < \gamma < \infty; -\infty < \delta < \infty$ scale parameter
T Location-Scale	$f(x; \mu, \sigma, \nu) = \Gamma \frac{\left(\frac{x-\mu}{\sigma}\right)^{\nu-1}}{\sigma\sqrt{\nu\pi} \left(\frac{x-\mu}{\sigma}\right)^2 + \frac{\nu-1}{2}}$	$\nu > 0$ shape parameter $\mu > 0$ location parameter $\sigma > 0$ scale parameter

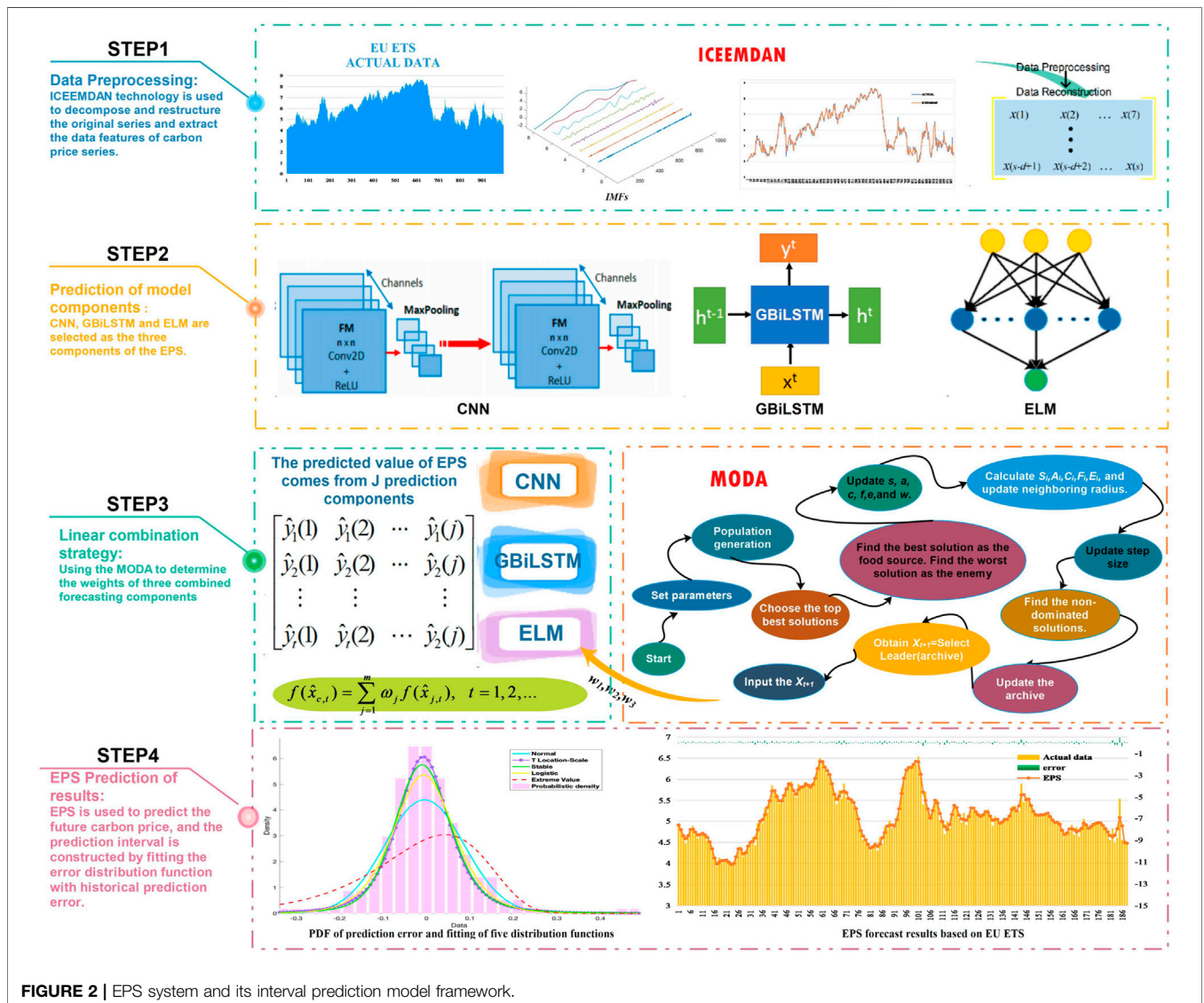


FIGURE 2 | EPS system and its interval prediction model framework.

Step 1: Data Preprocessing and Feature Selection Module

In this article, ICEEMDAN technology is employed to decompose and reconstruct the original carbon price data. ICEEMDAN decomposes the original carbon price into several IMFs and residual terms. The IMF with the highest frequency is then eliminated, and the remaining IMFs are reorganized to extract the effective features of the data. For multivariate time series, effective feature selection is also very important. In this study, partial autocorrelation analysis (PACF) is employed to determine the input feature length of carbon price prediction to achieve feature selection.

Step 2: EPS of Model Components

Owing to the high randomness, volatility, and instability of carbon price data, it is not easy to find its rules of motion, and the single hybrid prediction model has inherent defects. Therefore, the use of a combination of hybrid forecasting models is an effective means of obtaining satisfactory prediction performance and improving prediction accuracy. In this study, two deep learning hybrid models (ICEEMDAN-GBiLSTM and ICEEMDAN-CNN) and a feedforward neural network (ICEEMDAN-ELM) are used as the prediction components of the EPS. They have high prediction accuracy and good learning ability for time series.

Step 3: Component Ensemble Strategy

Given the advantages and disadvantages of different hybrid models, it is very important to select a weight combination strategy with strong adaptability and good fusion effect to compensate for the defects of the individual hybrid models and improve the performance and accuracy of carbon price prediction. Therefore, the MODA is selected to determine the fusion weight among the three prediction model components.

Step 4: Exploring Uncertainty

Quantifying the uncertainty associated with carbon price prediction is a considerable challenge. In this study, a new interval prediction scheme based on forecasting error distribution modeling in the model training stage is proposed. Unlike previous research based on the assumption that the prediction error follows a Gaussian distribution, this article uses MLE to conduct statistical research on carbon price error data and to explore its distribution. Among the five DFs developed, the function that best fits the distribution of carbon price prediction error is found. After confirming that the optimal fit to the distribution of forecast error is provided by the t location-scale, the upper and lower bounds of the carbon price prediction interval are constructed based on its PDF.

EXPERIMENT AND ANALYSIS

This section will introduce the experimental setup and analysis in detail, including the simulation experiment dataset and three

different groups of empirical experiments that are used to verify the prediction performance of EPS.

Data Selection and Analysis

In this article, three datasets based on the carbon price market (the EU Emission Trading System (EU ETS), the Shenzhen (SZ), and the Beijing (BJ) datasets) are used as experimental data. The dataset can be downloaded from the wind database (<http://www.wind.com.cn/>). The first 80% of each dataset is used as the training set, and the last 20% is used as the test set. Specifically, for the EU emission trading system dataset, a total of 1,000 daily quota settlement prices from July 10, 2013 to May 3, 2017 are selected. For the Shenzhen and Beijing datasets, this study used daily spot carbon price data collected from January 14, 2014 to February 7, 2017, including 800 data points. Detailed statistical descriptions of the three datasets are given in **Table 2**. In addition, in constructing the model input vector, we adopted a rolling acquisition mechanism.

Model Parameter Setting

The model parameters determine the performance of the prediction system to a large extent. The different parameters of the EPS proposed in this study are obtained by referring to the literature and to the results of the experiments conducted in this study. The parameter settings for each component of the ensemble system are listed in **Table 3**; this information is valuable and useful because it provides a reference for future research.

Evaluation Index System

To quantify the performance of the developed system, this study constructs an evaluation system using a variety of error evaluation criteria. The system is evaluated and analyzed based on the deterministic point estimation evaluation index and the probabilistic interval estimation evaluation index (Wang R. et al., 2018; Jiang et al., 2021). In the deterministic prediction part, four evaluation indicators, MAE, RMSE, MAPE, and IA, are used. MAE can better express the prediction error under actual conditions. RMSE reflects the deviation between the prediction value and the true value. MAPE expresses the accuracy of prediction using the ratio of error to true value. IA is applied to measure the concordance between the predicted value and the actual value. During interval prediction and evaluation, the three general indicators FICP, FINAW, and AWD are employed to evaluate the quality of the prediction interval. FICP reflects the possibility that the original value falls within the forecast period. FINAW measures the width of the prediction interval. AWD represents the degree of deviation between the observed value and the prediction interval. For FICP, unlike other indicators, a larger value indicates better performance of the model. **Table 4** lists the details of the above evaluation indicators.

Experiment 1: Comparison of Different Data Processing Methods

In this experiment, the original carbon price series and the data based on ICEEMDAN, EEMD, and singular spectrum

TABLE 2 | Statistical description of the carbon prices reported at three sites.

Statistical Indicators		Number	Max	Min	Mean	Std
Equation		—	—	—	Mean = $\frac{\sum_{i=1}^N x_i}{N}$	S = $\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$
EU ETS	Total	1,000	8.67	3.93	6.01	1.24
	Training	800	8.67	4.02	6.24	1.26
	Testing	200	6.54	3.93	5.08	0.57
BJ	Total	800	77	30.63	49.56	7.08
	Training	640	77	30.63	48.96	7.65
	Testing	160	69	39.45	51.98	3.02
SZ	Total	800	88.45	17.83	43.13	16.05
	Training	640	88.45	18.98	46.92	15.58
	Testing	160	42.81	17.83	28.01	5.39

TABLE 3 | Model parameters.

Model	Parameters	Default value
MODA	Maximum number of iterations	50
	Maximum number of archives	500
	Dragonfly number	30
	Upper and lower limits of the weight coefficient	[-2,2]
	Objective functions	$\text{Min} \begin{cases} obf_1 = Std(f(\tilde{x}_{c,t}), Y) \\ obf_2 = MAPE(f(\tilde{x}_{c,t}), Y) \end{cases}$
ICEEMDAN	Noise standard deviation	0.05
	Number of realizations	50
	Maximum number of sifting iterations	500
ELM	Input nodes number	Based on PAC
	Output nodes number	1
	Hidden nodes number	5
	Learning rate	0.001
	Iterations number	200
GBiLSTM	Number of inputs	Based on PAC
	Number of hidden units	200
	Number of outputs	1
	Maximum iteration	250
	Initial learning rate	0.01
	Training algorithm	Adam
CNN	Number of inputs	Based on PAC
	Number of outputs	1
	The kernel sizes	2 × 2
	The activation function	Relu
	The max pooling layer size	2
	Training algorithm	Adam

decomposition (SSA) are used as the training input for different prediction models. The purpose of this experiment is to explore the effect of using different signal decomposition techniques on the prediction accuracy of prediction models. **Table 6** compares the results obtained using the corresponding models.

Feature Selection Analysis

The prediction performance of both machine learning and deep learning methods is closely related to the input variables. The PACF method is used to select appropriate features as the best input of the prediction model. The best input characteristics

obtained from the PACF results of each subsequence are shown in **Table 5**. (In the follow-up experiments, the input units of each prediction model were obtained according to the PACF results.)

Prediction Results Obtained Using the Different Data Preprocessing Methods

To verify the effectiveness of the ICEEMDAN data preprocessing method in data feature extraction, in this experiment the performance of ICEEMDAN is compared with that of the classical feature extraction methods EEMD and SSA. The detailed results are described below.

TABLE 4 | Basic evaluation metrics.

Metric	Definition	Equation
MAE	The mean absolute error	$MAE = \frac{1}{N} \sum_{i=1}^N Y_{PRE(i)} - Y_{ACT(i)} $
RMSE	Root mean squared error	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_{PRE(i)} - Y_{ACT(i)})^2}$
MAPE	The mean absolute percentage error	$MAPE = \frac{1}{N} \sum_{i=1}^N \left \frac{Y_{PRE(i)} - Y_{ACT(i)}}{Y_{ACT(i)}} \right \times 100\%$
IA	Concordance index	$IA = 1 - \frac{\sum_{i=1}^N (Y_{PRE(i)} - Y_{ACT(i)})^2}{\sum_{i=1}^N (Y_{PRE(i)} - \bar{Y}_{ACT} + Y_{ACT(i)} - \bar{Y}_{ACT})^2}$
FICP	Forecast interval coverage probability	$FICP = \sum_{i=1}^N C_i \times 100\% / N$
FINAW	Forecast interval normalized average width	$FINAW = \sum_{i=1}^N (U_i - L_i) \times 100\%$
AWD	Accumulated width deviation of testing dataset	$AWD = \sum_{i=1}^N AWD_i / N$ $AWD_i = \begin{cases} L_i - Y_{ACT(i)} / U_i - L_i, & Y_{ACT(i)} < L_i \\ 0, & L_i \leq Y_{ACT(i)} \leq U_i \\ Y_{ACT(i)} - U_i / U_i - L_i, & Y_{ACT(i)} > U_i \end{cases}$

Note: This table lists the full names and calculation methods of the evaluation indices included in the evaluation system. N is the size of the test sample, \bar{y} is the average value of y , $Y_{ACT(i)}$ is the i -th actual value, and $Y_{PRE(i)}$ is the i -th forecast value. U_i and L_i represent the upper and lower limits, respectively, of the prediction interval. C_i represents the number of true values contained in the construction interval $[U_i, L_i]$ and is the i -th prediction interval.

TABLE 5 | Optimal input characteristics based on the PACF.

Site	Input combination
EU ETS	$X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}, X_{t-5}, X_{t-6}, X_{t-7}, X_{t-8}, X_{t-9}$
SZ	$X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}, X_{t-5}, X_{t-6}, X_{t-7}, X_{t-8}, X_{t-9}, X_{t-10}$
BJ	$X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}, X_{t-5}, X_{t-6}, X_{t-7}, X_{t-8}, X_{t-9}, X_{t-10}$

From the results in **Table 6**, it can be seen that the use of data preprocessing technology can effectively ameliorate the prediction ability of the prediction model. MAPE, MAE, RMSE, and IA were adopted to evaluate the prediction ability of the model based on the prediction accuracy and the fitting situation. For the data on the three carbon trading markets, the MAPE, the MAE, and the RMSE of the prediction model are significantly lower than those of the model that was directly trained on the original dataset, regardless of which data preprocessing technology is adopted. When the original dataset is used directly, the $MAPE_{actual}$ of the prediction model under the EU ETS, SZ, and BJ datasets is between 3 and 14%, while the $MAPE_{ICEE}$ of the prediction model is reduced to 1–5% when ICEEMDAN noise reduction technology is used. This is sufficient to indicate the necessity for data preprocessing.

In addition, for the three components of EPS, different data preprocessing methods are used as model inputs. The experimental results show that ICEEMDAN is more effective than the other methods. For the EU ETS dataset, the prediction system using ICEEMDAN noise reduction technology has the highest prediction accuracy, and the average RMSE value of the three component models, $RMSE_{ICEE}^{EUETS} = 0.1081$, is the best; the prediction performance of the model based on EEMD is the worst: $RMSE_{EEMD}^{EUETS} = 0.1313$ in the SZ dataset, $MAE_{ICEE}^{SZ} = 1.2518$, $RMSE_{ICEE}^{SZ} = 1.9671$, and $MAPE_{ICEE}^{SZ} = 4.8261\%$. In the BJ dataset, the average values of these three indicators are 0.5573, 1.2169, and 1.0573%, obviously better than those obtained using other data processing technologies.

IA is an effective index that can be used to measure the correlation and consistency between the predicted values and

the original data. The higher the index value is, the better the fitting effect of the model is. The ICEEMDAN feature extraction technology proposed in this article achieves the highest IA of all the models tested under the three carbon price datasets. In the SZ dataset, the IA values are $IA_{ICEE-ELM}^{SZ} = 0.9644$, $IA_{ICEE-CNN}^{SZ} = 0.9529$, and $IA_{ICEE-BiLSTM}^{SZ} = 0.9696$. These values are 0.0138, 0.0025, and 0.0173 units higher than $IA_{SSA-ELM}^{SZ}$, $IA_{SSA-CNN}^{SZ}$, and $IA_{SSA-GBiLSTM}^{SZ}$, respectively, and 0.0628, 0.0504 and 0.0657 units higher than $IA_{EEMD-ELM}^{SZ}$, $IA_{EEMD-CNN}^{SZ}$ and $IA_{EEMD-GBiLSTM}^{SZ}$, respectively. In summary, compared with other data preprocessing technologies, ICEEMDAN data preprocessing is more effective for data feature extraction and has incomparable advantages in improving the performance of the prediction model.

Key Finding: Compared with the original carbon price series and other classical data preprocessing techniques, ICEEMDAN preprocessing technology can extract the data characteristics of carbon prices more effectively, significantly enhances the prediction accuracy of the prediction model, and is a more reliable data preprocessing tool.

Experiment 2: Point Forecasting: Comparison of the EPS With Reference Models

To verify the effectiveness of EPS in carbon price prediction, the traditional single forecast model and the classical hybrid prediction model are compared with EPS. These models include the traditional statistical models ARIMA and ICEEMDAN-ARIMA, the traditional single neural network models BP, ELM, GRNN, the deep learning models LSTM, CNN, and the classical hybrid prediction models GWO-BP and ICEEMDAN-GWO-BP. In addition, to explore the expansibility of the model, the experimental content of multistep point prediction is included in the experiment. In the multistep prediction, rolling prediction is adopted. The specific method used to perform multistep prediction is shown in **Figure 3**. The experimental results are shown in

TABLE 6 | Comparison of the performances of prediction models based on different data feature extraction techniques.

Model	EU ETS				SZ				BJ			
	MAPE	RMSE	MAE	IA	MAPE	RMSE	MAE	IA	MAPE	RMSE	MAE	IA
ELM	3.1190	0.2148	0.1583	0.9574	13.3955	4.4118	3.5126	0.7965	3.9351	4.0281	2.5107	0.7225
CNN	3.1581	0.1659	0.1659	0.9544	12.5333	4.4030	3.3666	0.8007	3.7483	3.8029	2.4362	0.7236
GBiLSTM	3.3159	0.2241	0.1744	0.9523	13.1742	4.3402	3.4707	0.7955	3.7406	3.7846	2.4355	0.7239
EEMD-ELM	1.8531	0.1235	0.0942	0.9857	7.8973	2.8796	2.2638	0.9016	2.9211	2.2281	1.5135	0.8443
EEMD-CNN	1.8892	0.1249	0.0962	0.9880	7.8426	2.9683	2.2237	0.9025	2.9058	2.2111	1.5042	0.8483
EEMD-GBiLSTM	2.2015	0.1456	0.1117	0.9829	7.6835	2.5719	2.0024	0.9029	2.8966	2.2051	1.4991	0.8495
SSA-ELM	1.4597	0.1021	0.0734	0.9922	5.8985	2.2233	1.6648	0.9506	1.5922	1.4362	0.8198	0.9318
SSA-CNN	1.5684	0.1100	0.0795	0.9905	5.9112	2.2256	1.6727	0.9504	1.5435	1.4568	0.8079	0.9139
SSA-GBiLSTM	2.0231	0.1293	0.1009	0.9861	5.8622	2.2085	1.6519	0.9513	1.5479	1.4577	0.8085	0.9137
ICEE-ELM	1.2806	0.0915	0.0647	0.9934	4.5032	1.7141	1.2424	0.9644	1.0571	1.2013	0.5561	0.9397
ICEE-CNN	1.5162	0.1075	0.0771	0.9910	5.6835	2.5719	1.3624	0.9529	1.0624	1.2485	0.5596	0.9359
ICEE-GBiLSTM	1.8746	0.1253	0.0941	0.9878	4.2916	1.6153	1.1508	0.9686	1.0525	1.2008	0.5561	0.9399
EPS	1.2657	0.0904	0.0640	0.9936	4.0156	1.6096	1.1372	0.9687	1.0064	1.2049	0.5312	0.9402

Note: The best indicator values are shown in bold type.

Table 7. The detailed experimental analysis is described below.

- 1) In comparison with the traditional single prediction model, we find that EPS displays incomparable advantages in the four indicators in both one-step and multiple-step prediction. This shows that the EPS developed by us is effective in predicting carbon prices. In addition, the MAPE values of GBiLSTM in the three datasets are $MAPE_{GBiLSTM}^{EUETS} = 3.3159$, $MAPE_{GBiLSTM}^{SZ} = 13.1740$, and $MAPE_{GBiLSTM}^{BJ} = 3.7406$; these values are better than those obtained using a single LSTM, proving the effectiveness of the GBiLSTM. The other two prediction components, ELM and CNN, have outstanding prediction performance in all single prediction models, so it is reasonable to choose them as the submodes of the EPS. In addition, we can see that for the traditional statistical model ARIMA, the average value of MAPE of the three stations is $\overline{MAPE}_{ARIMA} = 11.0072$ in one-step prediction; this is not as high as the prediction accuracy achieved using other prediction models, indicating that the traditional linear statistical model is not suitable for the prediction of carbon price series with high volatility and complexity.
- 2) We can observe that under different datasets, different prediction models have different prediction performances. Under the EU ETS dataset, the neural network ELM performs best, yielding $RMSE_{ELM}^{EUETS} = 0.2148$; this is better than the RMSE values of the deep learning algorithms CNN and GBiLSTM, which are $RMSE_{CNN}^{EUETS} = 0.2218$ and $RMSE_{BiLSTM}^{EUETS} = 0.2241$, respectively. However, under the SZ and BJ datasets, the deep learning algorithms CNN and GBiLSTM achieve better prediction results than ELM. The same phenomenon occurs in multistep forecasting, and there the forecast advantage of the deep learning framework is more obvious. However, the prediction accuracy of EPS remains the highest under any of the tested datasets. The RMSE values in the one-step forecast are $RMSE_{EPS}^{EUETS} = 0.0904$, $RMSE_{EPS}^{SZ} = 1.6096$, and $RMSE_{EPS}^{BJ} = 1.2046$. The RMSE values in the two-step forecast are $RMSE_{EPS}^{EUETS} = 0.1829$, $RMSE_{EPS}^{SZ} = 3.3244$, and $RMSE_{EPS}^{BJ} = 1.7567$. This shows that the combination strategy

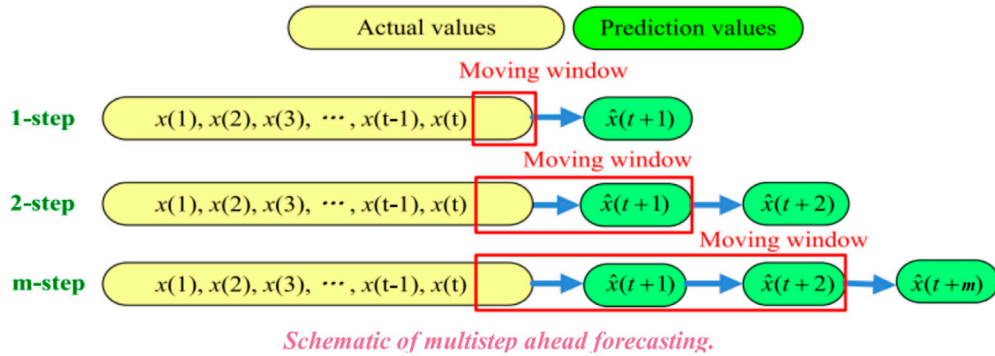
retains the forecasting advantages obtained by using different forecasting components and that it compensates for each model's defects; as a result, EPS has strong robustness and wide adaptability.

In comparison with the classic hybrid forecasting models ICEE-GWO-BP, SSA-GRNN, and GWO-BP, several sets of hybrid forecasting methods have achieved good forecasting performance; however, because the index values used in these models are very similar, it is not easy to intuitively present the predictive ability of the model. In this case, we measure the percentage of improvement in the evaluation index to make the analysis more intuitive. The percentage of improvement in the evaluation index is a measure of the degree of improvement achieved by EPS compared with the index value of the comparison model; it can be expressed as follows:

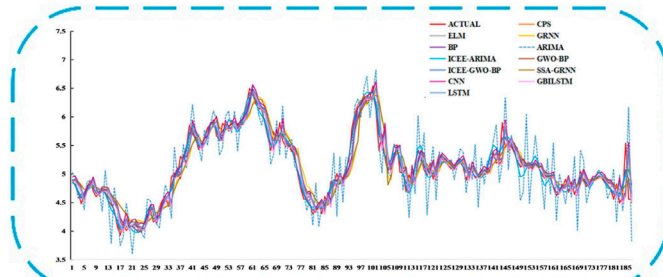
$$P_{Index}^{model} = \left| \frac{Index_{model} - Index_{EPS}}{Index_{model}} \right| \times 100\% \quad (46)$$

where P_{Index}^{model} is the percentage of improvement indicators, $Index_{model}$ stands for the index value of the comparison model, and $Index_{EPS}$ is the index value of the EPS.

In the EU ETS dataset, the improved MAPE values for one-step prediction of the three hybrid models are $P_{MAPE}^{ICEE-GWO-BP} = 1.9673\%$, $P_{MAPE}^{SSA-GRNN} = 60.4950\%$, and $P_{MAPE}^{GWO-BP} = 59.5377\%$. The improved MAPE values of the two-step prediction are $P_{MAPE}^{ICEE-GWO-BP} = 9.2410\%$, $P_{MAPE}^{SSA-GRNN} = 30.7285\%$ and $P_{MAPE}^{GWO-BP} = 51.8867\%$. Under the BJ dataset, the improved IA values predicted for the three hybrid models using the one-step method are $P_{IA}^{ICEE-GWO-BP} = 0.9064\%$, $P_{IA}^{SSA-GRNN} = 48.1563\%$, and $P_{IA}^{GWO-BP} = 33.3896\%$. The improved IA values obtained by two-step prediction are $P_{MAPE}^{ICEE-GWO-BP} = 1.4313\%$, $P_{MAPE}^{SSA-GRNN} = 97.2385\%$, and $P_{MAPE}^{GWO-BP} = 104.3018\%$. The index improvement percentage more intuitively shows the improvement in prediction performance obtained using EPS. Compared with the classical hybrid prediction model, the EPS shows significant improvement in both prediction accuracy and fitting consistency.



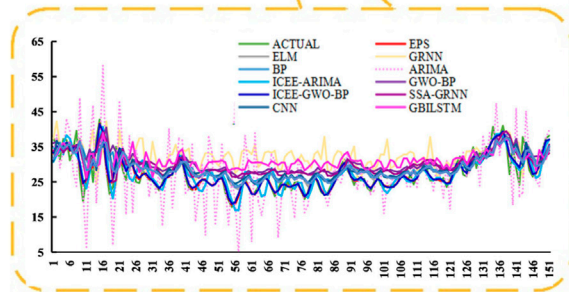
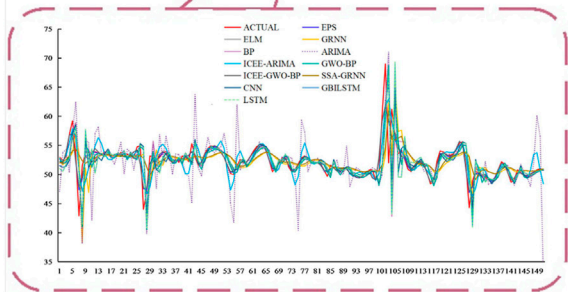
The results of one-step prediction



BJ

EU ETS

SZ



Comparison of two-step prediction index values.

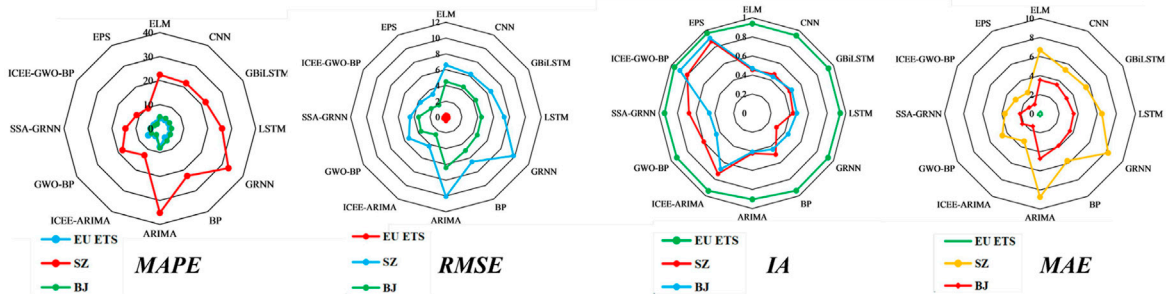


FIGURE 3 | Prediction results obtained using EPS and comparison models under different prediction steps.

Figure 3 shows a comparison of the prediction results obtained using EPS and the comparison model when different numbers of prediction steps are used.

Key finding: The difference in the prediction results between the EPS system and other prediction models is significant. Specifically, under each dataset and for each prediction step,

TABLE 7 | Comparison of the prediction ability of the proposed system with those of some traditional benchmark models and classic hybrid models.

Dataset	Model	ONE-STEP				TWO-STEP			
		MAPE	RMSE	MAE	IA	MAPE	RMSE	MAE	IA
EU ETS	ELM	3.119	0.2148	0.1583	0.9574	4.2962	0.2842	0.2169	0.9386
	CNN	3.1581	0.2218	0.1659	0.9544	4.1304	0.2805	0.2112	0.9392
	GBiLSTM	3.3159	0.2241	0.1744	0.9523	4.1862	0.2794	0.2133	0.9390
	LSTM	3.5642	0.2470	0.1815	0.9407	4.2575	0.2827	0.2164	0.9388
	GRNN	3.5405	0.2432	0.1799	0.9418	4.7743	0.3232	0.2432	0.9341
	BP	3.2941	0.2235	0.1704	0.9516	4.3052	0.2903	0.2193	0.9380
	ARIMA	5.8692	0.3789	0.2947	0.9327	7.6365	0.4814	0.3011	0.9036
	ICEE-ARIMA	1.9758	0.1323	0.1000	0.9862	3.6109	0.2401	0.1839	0.9403
	GWO-BP	3.1281	0.2149	0.1584	0.9571	5.5972	0.3500	0.2828	0.9322
	SSA-GRNN	3.2039	0.2170	0.1621	0.9548	3.8876	0.2632	0.1986	0.9385
	ICEE-GWO-BP	1.2911	0.0891	0.0644	0.9862	2.9672	0.1988	0.1506	0.9617
	EPS	1.2657	0.0904	0.0640	0.9936	2.6930	0.1829	0.1378	0.9674
	SZ	ELM	13.3955	4.4118	3.5126	0.7965	22.4148	6.6057	6.6813
CNN		12.5333	4.4030	3.3666	0.8007	21.8746	6.2643	5.3190	0.4686
GBiLSTM		13.1740	4.3402	3.4707	0.7955	22.034	6.5272	5.5683	0.4617
LSTM		15.1983	6.8455	4.5637	0.7514	25.9514	7.3299	6.4834	0.4352
GRNN		13.7152	5.1715	3.8576	0.7654	33.0502	9.8132	8.2422	0.2973
BP		14.1428	5.8032	4.2144	0.7608	22.8251	6.5453	5.7117	0.5004
ARIMA		21.8403	7.8476	5.9355	0.6595	35.1742	10.0714	8.7415	0.4211
ICEE-ARIMA		8.9814	3.2795	2.5983	0.8943	12.8818	4.2559	3.3362	0.7351
GWO-BP		12.3461	4.1508	3.2907	0.8017	18.0382	5.3858	4.5591	0.5993
SSA-GRNN		5.9875	2.2372	1.6791	0.9494	14.3684	4.5285	3.6886	0.6773
ICEE-GWO-BP		4.6025	1.8286	1.3027	0.9631	11.1825	3.7085	2.9526	0.8039
EPS		4.0156	1.6096	1.1372	0.9687	9.7600	3.3244	2.5621	0.8701
BJ		ELM	3.9351	4.0281	2.5107	0.7225	4.8978	4.4902	3.5380
	CNN	3.7483	3.8029	2.4362	0.7236	4.8724	4.4065	3.5240	0.4410
	GBiLSTM	3.7406	3.7846	2.4355	0.7239	4.6629	4.3132	3.2137	0.4852
	LSTM	3.8462	3.9033	2.4871	0.7231	4.8787	4.4753	3.5231	0.4749
	GRNN	4.1345	4.3247	2.9737	0.6273	5.0379	4.5516	3.6191	0.4416
	BP	4.0350	4.2179	2.8450	0.6355	5.9108	4.9031	3.8741	0.4379
	ARIMA	5.3120	4.6312	3.7397	0.5627	8.1714	6.4412	4.7140	0.4058
	ICEE-ARIMA	2.4705	2.0481	1.2635	0.9045	2.9320	2.5669	1.5213	0.6763
	GWO-BP	2.9202	3.0563	1.5062	0.7113	4.0738	3.6183	2.1437	0.4440
	SSA-GRNN	3.0901	2.6153	1.5929	0.6346	3.9989	3.4553	2.0979	0.4599
	ICEE-GWO-BP	1.1544	1.2383	0.5966	0.9488	2.4931	2.1482	1.3095	0.8943
	EPS	1.0064	1.2049	0.5312	0.9402	2.1558	1.7567	1.1272	0.9071

Note: ICEE is an abbreviation for ICEEMDAN. The best indicator values are shown in bold type.

the EPS has better prediction performance. Therefore, it is concluded that the advanced ensemble prediction system has better carbon price forecasting ability and potential than the traditional single models and classical hybrid models.

Experiment 3: Interval Forecasting: Uncertainty Analysis of Carbon Price

In Experiment 2, the accuracy and stability of the prediction system were discussed through the prediction method of certainty point estimation. However, the point prediction results do not reflect the uncertainty in the dataset. To further prove that the ESP system has a wider range of adaptability than other predictive models, this section uses the interval prediction method to mine the uncertainty of carbon prices. Unlike point prediction, interval prediction can provide the upper and lower bounds of the observed value, making it possible to construct the prediction interval under a given significance level. It can provide additional

information for carbon price market policymakers and can help them analyze the carbon price market.

Distribution Function of Prediction Error

In previous studies, most of the prediction errors of the prediction model defaulted to obey the normal distribution. However, the normal distribution function does not effectively reflect the distribution of forecast model errors. Therefore, this research develops five fitting distribution functions and uses the MLE method to conduct an in-depth investigation of the prediction error to obtain a distribution function (DF) with better fitting performance. The most suitable probability distribution for further interval prediction is selected.

In this section, five DFs, namely, extreme value, normal, logistic, stable, and t location-scale, are used to represent the distribution of carbon price prediction errors. **Table 1** shows the relative PDF of these DFs. **Table 8** lists the five DF parameters estimated by the MLE method. These parameters can be used to

TABLE 8 | Parameter values of the different distribution functions determined by MLE.

Datasets	Extreme value		Logistic		Normal		Stable				T Location-scale		
	μ	σ	μ	σ	μ	σ	α_1	α_2	β	δ	μ	σ	ν
EU ETS	0.0197	0.0342	0.0028	0.0189	0.0030	0.0334	1.9347	-0.0264	0.0227	0.0031	0.0030	0.0314	19.0078
SZ	0.2878	0.7713	-0.0660	0.3496	-0.0542	0.6631	1.7226	0.2460	0.3768	-0.092	-0.0691	0.5031	4.7444
BJ	0.1150	0.4266	-0.0642	0.1978	-0.0839	0.4106	1.2249	-0.1238	0.1469	-0.047	-0.0531	0.1665	1.5469

TABLE 9 | R^2 and RMSE values of the five distribution functions by MLE.

Datasets	EU ETS		SZ		BJ	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
Extreme value	0.9086	1.1897	0.8115	0.0941	0.6103	0.2913
Logistic	0.9546	0.8382	0.9685	0.0384	0.8780	0.1630
Normal	0.9760	0.6095	0.9512	0.0479	0.7207	0.2466
Stable	0.9771	0.5956	0.9507	0.0481	0.9668	0.0851
T location-scale	0.9877	0.4355	0.9791	0.0313	0.9671	0.0846

describe the scale and location of these DFs. In addition, the coefficient of determination ($0 \leq R^2 \leq 1$) and RMSE are used to determine the degree of fit of these DFs. The larger the R^2 value is, the lower the RMSE value is, and the better is the fitting ability of the DFs. The index values reflecting the fitting abilities of the five DFs are shown in **Table 9** and **Figure 4**.

Table 9 shows that the t location-scale function fits the EPS prediction error best. Its R^2 value is higher than 0.96, and its RMSE value is the lowest, indicating that it can provide better estimates in all cases, followed by stable distribution, normal distribution, logistic distribution, and extreme value distribution. In addition, although the stable distribution has a slightly worse fitting effect than the t location-scale distribution, it is still better than the normal distribution that the previous prediction error hypothesis obeys; this further proves the necessity of fitting the distribution of the prediction error. In addition, the motivation for estimating the distribution function of the carbon price dataset in this section is to prepare for further research on the establishment of carbon price interval predictions, as discussed in Section *Interval Prediction of Carbon Price*.

Interval Prediction of Carbon Price

Unlike the deterministic information given by the point forecast, the interval forecast can provide the forecast range, a confidence level, and other uncertain information on future values; this information is helpful to decision-makers who are attempting to analyze and supervise the reasonable operation of the carbon price market. Owing to the generalization ability of the forecasting model, the complex patterns of carbon price series fluctuations and other factors inevitably produce forecast errors, and the ability to effectively transform the uncertainty caused by forecast errors into measurable features is of great significance. Therefore, in this study, a new interval prediction scheme based on modeling of the prediction error distribution in the model training phase is proposed.

According to the point prediction results of the proposed EPS system, the t location-scale distribution function, which

determines the prediction error in Section *Distribution Function of Prediction Error*, and the interval prediction method introduced in Section *Interval Prediction Theory*, the prediction interval is constructed under the given significance level α . To verify that the prediction interval constructed by the t location-scale model has the best fit, it is compared with the other four error distribution functions.

In addition, three evaluation indicators, FINAW, PICP, and AWD, listed in **Table 4**, are introduced in this section to present the effect of interval prediction. It is worth mentioning that the optimal interval prediction should satisfy the following conditions: under a given significance level α , the larger the PICP value is ($0 \leq \text{PICP} \leq 1$), the smaller the FINAW value is, and the better is the prediction performance of interval prediction. However, it is obvious that there is a contradiction between FINAW and PICP. When the PICP value increases, the FINAW of the response average bandwidth will certainly increase. Therefore, the AWD index is introduced as a supplement to the evaluation index system. **Table 10** shows the prediction intervals of the EPS system as evaluated based on the three carbon price markets using five different error distributions.

In theory, when the PICP is greater than the significance level, the constructed prediction interval is effective. As seen from **Table 10**, the models satisfying this condition are EPS-TLS and EPS-Extreme value; these models are effective at the 95, 90, and 80% significance levels. However, if the value of the PICP is the only goal, the result will become meaningless, as increased PICP inevitably leads to a larger FINAW. Based on different α conditions, the value of $FINAW_{EPS-Extreme\ value}$ is significantly higher than the FINAW value obtained through modeling of other distributions. At the same time, considering that in Section *Distribution Function of Prediction Error*, the fitting of extreme value distribution to EPS prediction error is very bad, it can be considered that the prediction interval constructed by EPS-Extreme value is not reasonable.

The PICP of EPS-TLS in the data from the three carbon trading markets is higher than the significance level α . In addition, under different α , the FINAW values of the three markets are $FINAW_{EPS-TLS}^{EUETS} = 0.0542$, $FINAW_{EPS-TLS}^{SZ} = 0.1225$, and $FINAW_{EPS-TLS}^{BJ} = 0.1397$. The FINAW value is not optimal in any of the five interval prediction models, but with only a small increase in the FINAW value, the other index values achieve better results. All things considered, this can be accepted.

For the prediction interval constructed using a normal distribution, in the BJ dataset, the PICP values of EPS-Normal

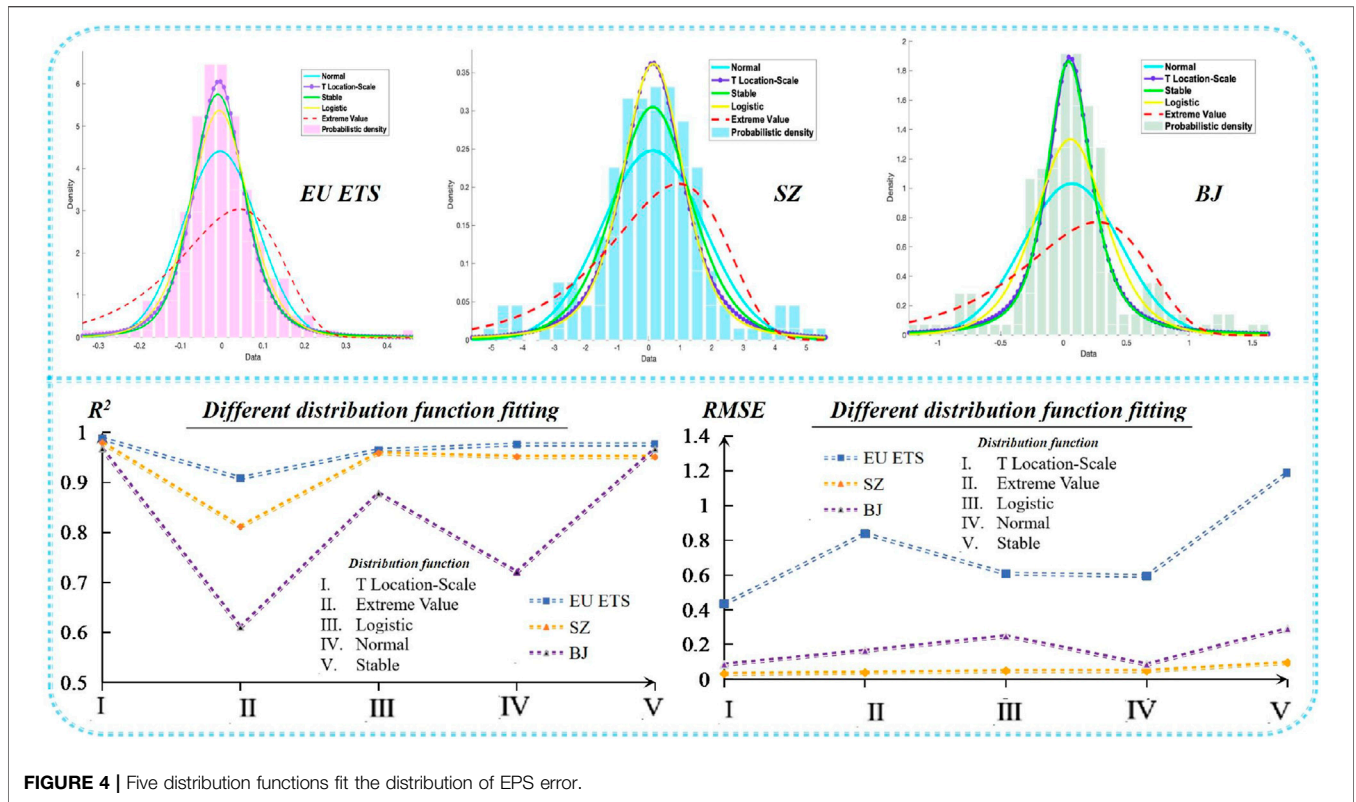


FIGURE 4 | Five distribution functions fit the distribution of EPS error.

are $PICP_{EPS-Normal}^{BJ} = 92.0530$ and $PICP_{EPS-Normal}^{BJ} = 88.0795$ under significance levels of 0.95 and 0.90, respectively; they fail to meet the condition of a level of α that is greater than significance. This also reflects the necessity of detailed research on error distribution. For AWD indicators, although $AWD_{EPS-TLS}$ is not all better than the benchmark model, there is little difference. Considering the comprehensive performance of the three indicators of the proposed scenario, EPS-TLS still has obvious advantages over the four benchmark models in constructing the prediction interval.

In addition, the carbon price prediction intervals generated by the three proposed schemes of the three carbon trading markets are shown in Figure 5. It can be observed that EPS-TLS has a smaller bandwidth and is surrounded by these constructed prediction intervals in most target values. The constructed confidence interval is very effective.

DISCUSSION

In this section, we will discuss the robustness, application, and further development of EPS in the carbon price market.

Robustness Discussion

Because the results of both deep learning and metaheuristic optimization algorithms are always accompanied by randomness and probability mechanisms, the results of each experiment will still have deviations even when the parameters are set exactly the same. At the same time, in the actual forecast, the actual values of the future

carbon price cannot be predicted in advance; thus, it is impossible to use the evaluation index to verify the future value in advance. Therefore, the stability of the EPS is also an important factor that affects the prediction.

The standard deviation is an effective measure of system stability. It can be indicated as $SD_{(M)} = \sqrt{\sum_k^n (M_k - \bar{M})^2 / n}$, where n is the number of training iterations, M_k is the predicted value of the k -th training result, and \bar{M} is the average of the N -th results (Xiao et al., 2017). The smaller the value of SD , the higher the stability of the model.

To evaluate the stability of the different models, the $SD_{(M)}$ values of four evaluation indices were calculated in 30 prediction experiments using three carbon price datasets.

Table 11 shows a comparison of the stability test results of different prediction systems based on ICEEMDAN processing. In the EU ETS dataset, ICEE-ELM has good stability ($SD_{(MAPE)}^{ICEE-ELM} = 0.0105$), but the stability is still slightly lower than that of the EPS prediction system ($SD_{(MAPE)}^{EPS} = 0.0101$). In the BJ and SZ datasets, CNN has good prediction accuracy in previous experiments, but its robustness is not good, and the prediction results fluctuate greatly. In contrast, EPS obtains a smaller SD value regardless of which dataset is used. This further shows that different single prediction systems have different robustness when used with different datasets and indicates that EPS with a combination weighting strategy can be considered the prediction method that obtains the best prediction results.

It is worth mentioning that the average prediction stability of GBiLSTM in the three prediction datasets is better than that of the traditional LSTM model; that is, $SD_{(RMSE)}^{GBiLSTM} = 0.0496$, but $SD_{(RMSE)}^{LSTM} = 0.0566$. The results show that the proposed

TABLE 10 | Carbon price range prediction results based on EU ETS, SZ, and BJ under different significance levels.

Site	PINC	Distribution	PICP	FINAW	AWD
EU ETS	95%	EPS-Extreme value	97.3822	0.0701	0.0027
		EPS-Logistic	83.2461	0.0387	0.0354
		EPS-Normal	95.2880	0.0516	0.0083
		EPS-Stable	95.8115	0.05361	0.0073
		EPS-TLS	95.8115	0.0542	0.0070
	90%	EPS-Extreme value	96.8586	0.0573	0.0062
		EPS-Logistic	74.8691	0.0316	0.0584
		EPS-Normal	90.0109	0.0412	0.0194
		EPS-Stable	90.5759	0.0443	0.0162
		EPS-TLS	90.0524	0.0426	0.0188
	80%	EPS-Extreme value	90.5759	0.0434	0.0155
		EPS-Logistic	62.3037	0.0240	0.1170
		EPS-Normal	80.5340	0.0331	0.0460
		EPS-Stable	79.5812	0.0342	0.0419
		EPS-TLS	81.675	0.0352	0.0378
Site	PINC	Distribution	PICP	FINAW	AWD
SZ	95%	EPS-Extreme value	98.0132	0.1811	0.0029
		EPS-Logistic	94.7020	0.1191	0.0115
		EPS-Normal	95.3642	0.1240	0.0105
		EPS-Stable	95.3642	0.1207	0.0112
		EPS-TLS	95.3642	0.1225	0.0100
	90%	EPS-Extreme value	95.3642	0.1477	0.0051
		EPS-Logistic	92.0530	0.0970	0.0224
		EPS-Normal	90.0514	0.1028	0.0183
		EPS-Stable	92.0530	0.0927	0.0242
		EPS-TLS	92.0530	0.0967	0.0222
	80%	EPS-Extreme value	90.7285	0.1121	0.0158
		EPS-Logistic	81.4570	0.0724	0.0507
		EPS-Normal	81.4570	0.0706	0.0547
		EPS-Stable	81.4570	0.0682	0.0585
		EPS-TLS	86.7550	0.0801	0.0375
Site	PINC	Distribution	PICP	FINAW	AWD
BJ	95%	EPS-Extreme value	92.0530	0.1546	0.0110
		EPS-Logistic	89.4040	0.1054	0.0199
		EPS-Normal	92.0530	0.1177	0.0137
		EPS-Stable	96.6887	0.1550	0.0054
		EPS-TLS	96.0265	0.1397	0.0065
	90%	EPS-Extreme value	92.0530	0.1263	0.0175
		EPS-Logistic	85.4305	0.0848	0.0401
		EPS-Normal	88.0795	0.0987	0.0258
		EPS-Stable	86.0927	0.0900	0.0354
		EPS-TLS	91.7951	0.1017	0.0224
	80%	EPS-Extreme value	88.7417	0.0958	0.0332
		EPS-Logistic	84.1060	0.0632	0.0793
		EPS-Normal	85.4305	0.0769	0.0524
		EPS-Stable	75.4967	0.0520	0.1179
		EPS-TLS	80.7947	0.0522	0.1134

Note: In the table above, PICP, FINAW, and AWD are selected to verify the prediction performance of different models, where $FICP = \sum_{i=1}^N C_i \times 100\% / N$, $FINAW = \sum_{i=1}^N (U_i - L_i) \times 100\%$, and $AWD = \sum_{i=1}^N AWD_i / N$.

GBILSTM not only has better prediction accuracy than LSTM but also that its robustness is significantly improved.

Application of the Proposed Ensemble Prediction System

1) A stable carbon price prediction system plays a prominent role in the initial allocation of the carbon quota, in transaction pricing and in effective monitoring of market risk. The

proposed EPS system not only shows accurate point prediction performance but also reasonably analyzes and mines the potential uncertainty of carbon prices by constructing the carbon price prediction interval based on error distribution fitting.

- 2) The proposed EPS system combines a deep learning framework with a traditional neural network and thereby provides a new idea for carbon price prediction and an effective reference tool that policymakers can use to research the volatility of the carbon market.
- 3) Comparing the EU ETS market data with the carbon price markets in Shenzhen and Beijing, it is helpful for China to analyze the evolution of the mature carbon trading market price in the EU; this will help the regulatory authorities adjust the policy and ensure the steady development of China's carbon market.
- 4) EPS has high practical value and strong expansibility and can easily fit highly volatile nonlinear time series. It thus provides a new intelligent supervision scheme for building a sound global carbon trading market in the future. At the same time, the use of a deep learning integrated forecasting system with high accuracy and strong stability is expected to become a new direction of energy and financial markets in the future.

Suggestions on Further Improvement of Carbon Price Market

More accurate prediction of carbon prices can provide some effective suggestions through which governments and enterprises can build and improve the carbon price market in the future. These are outlined below.

Improvement of the Initial Allocation Mechanism of Carbon Emission Rights

In the initial allocation of carbon quotas, we should pay attention to the fairness of allocation. First, the government should formulate incentive policies to encourage regional governments and local enterprises to reduce emissions and should give appropriate incentives or policy support to the regions and enterprises that use emission reduction technologies. Second, the initial allocation of carbon emission rights requires effective operation and an effective regulatory system; both of these components directly affect the efficiency and fairness of carbon quota allocation. Strengthening the construction of a carbon emission rights regulatory system will help achieve efficiency and fairness of resource allocation as China's total emission reduction target is being met.

Rationalization of the Carbon Trading Pricing System

Owing to the imperfect development of the carbon trading market and the carbon trading pricing system, the carbon trading price is easily affected by monopoly enterprises. At present, there is a certain monopoly phenomenon in the carbon trading market in some regions of China. Some small buyer enterprises can only passively accept the carbon price, and this allows monopoly enterprises to disproportionately influence the supply and demand of the market and reduces total social welfare. The establishment of a reasonable pricing system that

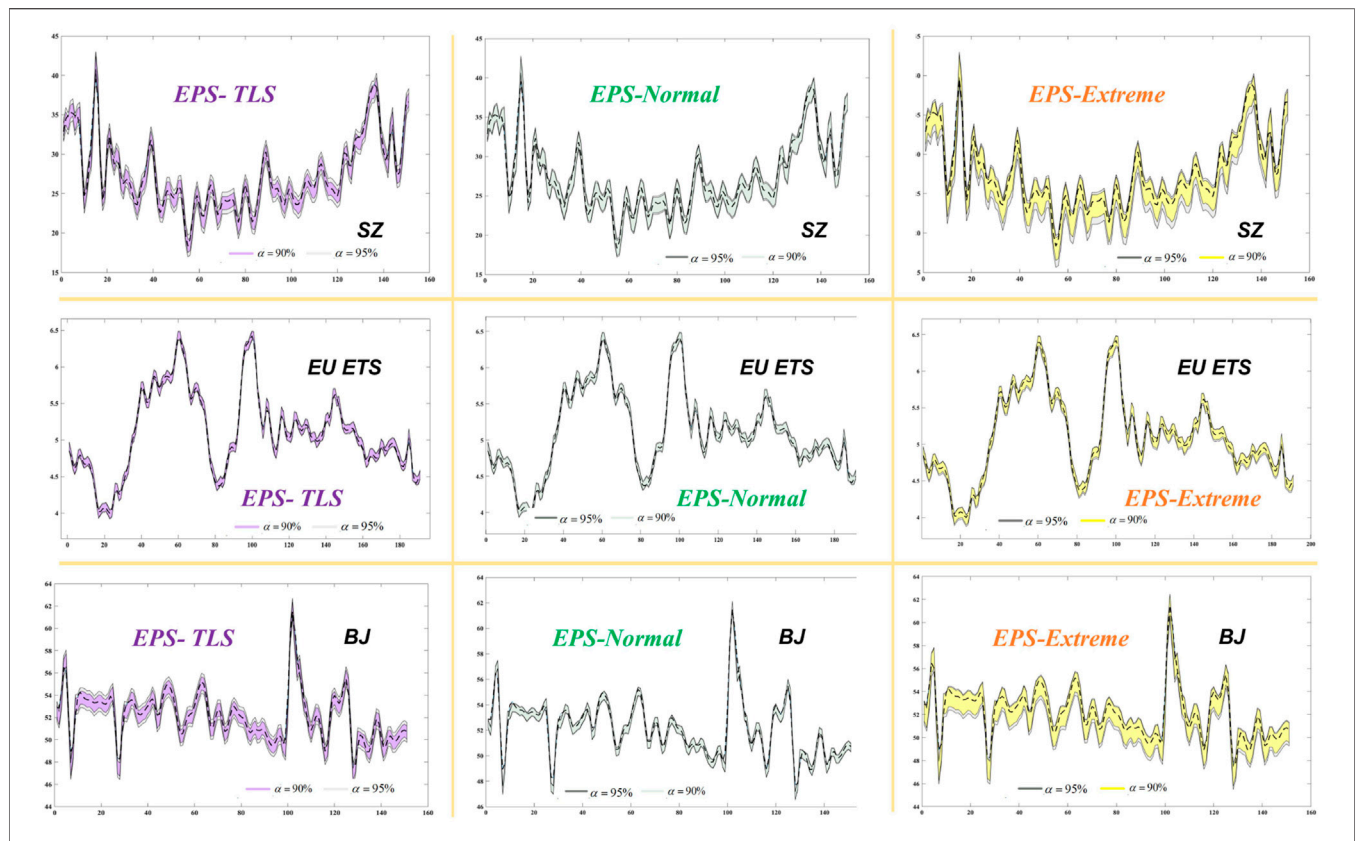


FIGURE 5 | Carbon price prediction intervals generated by the three proposed schemes.

TABLE 11 | Certainty of different forecasting methods.

Model	ICEE-ELM	ICEE-GBiLSTM	ICEE-CNN	EPS	ICEE-LSTM	ICEE-GWO-BP
EU ETS						
SD _(MAPE)	0.0105	0.0243	0.0902	0.0101	0.3277	0.0583
SD _(RMSE)	0.0004	0.0007	0.0041	0.0003	0.0042	0.0007
SD _(MAE)	0.0005	0.0008	0.0044	0.0005	0.0033	0.0005
SD _(IA)	0.0001	0.0001	0.0007	0.0001	0.0007	0.0001
BJ						
SD _(MAPE)	0.0255	0.0204	0.0757	0.0147	0.0438	0.0171
SD _(RMSE)	0.0091	0.0117	0.0224	0.0106	0.0212	0.0107
SD _(MAE)	0.0129	0.0123	0.0379	0.0114	0.0321	0.0819
SD _(IA)	0.0011	0.0011	0.0019	0.0012	0.0033	0.0038
SZ						
SD _(MAPE)	0.1845	0.1130	0.7846	0.1124	0.3831	0.1695
SD _(RMSE)	0.0344	0.1363	0.1745	0.1352	0.1443	0.2029
SD _(MAE)	0.0448	0.0062	0.1934	0.0054	0.0179	0.0348
SD _(IA)	0.0021	0.0061	0.0076	0.0018	0.0067	0.0015

Note: ICEE is an abbreviation for ICEEMDAN. The best indicator values are shown in bold type.

avoids monopoly price manipulation is conducive to the return of carbon prices to real value levels and to the optimal allocation of resources.

Improvement of the Carbon Market Risk Management and Control System

In the process of price fluctuation risk control, an accurate and effective price forecasting model can be used to monitor price

fluctuations. Using the relevant data, such a model can be used to predict long-term and short-term carbon trading prices, predict future fluctuation trends, and establish an effective carbon trading price risk early warning index system to effectively monitor the volatility risk caused by market price fluctuations. Through prediction of the carbon trading market price, we can grasp the fluctuations in carbon prices and take measures in advance to exercise macro control and reduce the level of risk when large market price fluctuations occur.

CONCLUSION

The availability of a reliable carbon price forecasting system is significant in the emission trading market because it can help decision-makers evaluate climate policies and adjust the emission ceiling to effectively maintain the reliable operation of the market. In this study, the EPS system, which adopts advanced data feature extraction and selection methods, combines the three optimal submodels through a multiobjective dragonfly optimization algorithm and explores the deterministic and uncertainty prediction of carbon price series. This study has several important implications: 1) ICEEMDAN is better than traditional signal decomposition at extracting data features. This can improve the accuracy of the prediction system. 2) The deep learning algorithm has a better ability than other algorithms to forecast carbon price series. The developed GBiLSTM model has better predicted performance and stability than the traditional LSTM. 3) Unlike previous studies in which it was assumed that the prediction error obeys a Gaussian distribution, this study explores five fitting distribution functions of prediction error, finds a more accurate error distribution function, and constructs a more reasonable carbon price prediction interval. The experimental results indicate that the EPS prediction system achieves the best prediction performance, with MAPE values of 1.2657, 4.0156, and 1.0064% for the three datasets. In addition, according to the optimal distribution fitting function of EPS prediction error, the carbon price prediction interval is constructed to mine the uncertainty of carbon price fluctuation. At various significance levels, the constructed prediction interval contains most of the observations, showing that the interval prediction has good performance. Therefore, the system is an

effective supplement to the existing carbon price prediction research framework and contributes to the ability of the government to reduce market risk and stabilize the market.

Although the combined prediction system proposed in this article achieves good prediction performance, there are still some limitations due to practical factors. Future research will analyze the carbon price trend from two perspectives, historical carbon price series and external factors, to obtain more accurate and stable prediction results.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

YY: Conceptualization, Methodology. HG: Writing-Reviewing and Editing, Data curation. YJ: Formal analysis, Supervision. AS: Software, Validation and Visualization, Investigation.

FUNDING

This work was supported by the Major Project of the National Social Science Fund (Grant No. 17ZDA093).

REFERENCES

- Arouri, M. E. H., Jawadi, F., and Nguyen, D. K. (2012). Nonlinearities in Carbon Spot-Futures price Relationships during Phase II of the EU ETS. *Econ. Model.* 29 (3), 884–892. doi:10.1016/j.econmod.2011.11.003
- Atsalakis, G. S. (2016). Using Computational Intelligence to Forecast Carbon Prices. *Appl. Soft Comput.* 43, 107–116. doi:10.1016/j.asoc.2016.02.029
- Benz, E. A., and Trück, S. (2009). Modeling the price Dynamics of CO₂ Emission Allowances. *Energ. Econ.* 31, 4–15. doi:10.1016/j.eneco.2008.07.003
- Byun, S. J., and Cho, H. (2013). Forecasting Carbon Futures Volatility Using GARCH Models with Energy Volatilities. *Energ. Econ.* 40, 207–221. doi:10.1016/j.eneco.2013.06.017
- Chen, C., and Liu, H. (2021). Dynamic Ensemble Wind Speed Prediction Model Based on Hybrid Deep Reinforcement Learning. *Adv. Eng. Inform.* 48, 101290. doi:10.1016/j.aei.2021.101290
- Cheng, Z., and Wang, J. (2020). A New Combined Model Based on Multi-Objective Salp Swarm Optimization for Wind Speed Forecasting. *Appl. Soft Comput.* 92, 106294. doi:10.1016/j.asoc.2020.106294
- Chevallier, J. (2009). Carbon Futures and Macroeconomic Risk Factors: A View from the EU ETS. *Energ. Econ.* 31 (4), 614–625. doi:10.1016/j.eneco.2009.02.008
- Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,” in *NIPS 2014 Workshop on Deep Learning*. ArXiv e-prints Available at: <https://arxiv.org/abs/1412.3555>.
- Colominas, M. A., Schlotthauer, G., and Torres, M. E. (2014). Improved Complete Ensemble EMD: a Suitable Tool for Biomedical Signal Processing. *Biomed. Signal Process. Control.* 14, 19–29. doi:10.1016/j.bspc.2014.06.009
- Du, P., Wang, J., Yang, W., and Niu, T. (2020). Point and Interval Forecasting for Metal Prices Based on Variational Mode Decomposition and an Optimized Outlier-Robust Extreme Learning Machine. *Resour. Pol.* 69, 101881. doi:10.1016/J.RESOURPOL.2020.101881
- Fan, X., Li, S., and Tian, L. (2015). Chaotic Characteristic Identification for Carbon price and an Multi-Layer Perceptron Network Prediction Model. *Expert Syst. Appl.* 42 (8), 3945–3952. doi:10.1016/j.eswa.2014.12.047
- 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., Zhu, Q-Y., and Siew, C-K. (2006). Extreme Learning Machine: Theory and Applications. *Neurocomputing* 1. doi:10.1016/j.neucom.2005.12.126
- Jiang, H., Tao, C., Dong, Y., and Xiong, R. (2020). Robust Low-Rank Multiple Kernel Learning with Compound Regularization. *Eur. J. Oper. Res.* 295, 634–647. doi:10.1016/J.EJOR.2020.12.024
- Jiang, H., Luo, S., and Dong, Y. (2021a). Simultaneous Feature Selection and Clustering Based on Square Root Optimization. *Eur. J. Oper. Res.* 289, 214–231. doi:10.1016/j.ejor.2020.06.045
- Jiang, H., Zheng, W., and Dong, Y. (2021b). Sparse and Robust Estimation with ridge Minimax Concave Penalty. *Inf. Sci.* 571, 154–174. doi:10.1016/J.IINS.2021.04.047
- Jiang, P., Liu, Z., Niu, X., and Zhang, L. (2021). A Combined Forecasting System Based on Statistical Method, Artificial Neural Networks, and Deep Learning Methods for Short-Term Wind Speed Forecasting. *Energy* 217, 119361. doi:10.1016/J.ENERGY.2020.119361
- Jin, Y., Guo, H., Wang, J., and Song, A. (2020). A Hybrid System Based on LSTM for Short-Term Power Load Forecasting. *Energies* 13 (23), 6241. doi:10.3390/en13236241
- Liu, Z., Jiang, P., Zhang, L., and Niu, X. (2020). A Combined Forecasting Model for Time Series: Application to Short-Term Wind Speed Forecasting. *Appl. Energ.* 259, 114137. doi:10.1016/j.apenergy.2019.114137
- Liu, Z., Jiang, P., Wang, J., and Zhang, L. (2021). Ensemble Forecasting System for Short-Term Wind Speed Forecasting Based on Optimal Sub-model Selection

- and Multi-Objective Version of Mayfly Optimization Algorithm. *Expert Syst. Appl.* 177, 114974. doi:10.1016/J.ESWA.2021.114974
- Lu, H., Ma, X., Huang, K., and Azimi, M. (2020). Carbon Trading Volume and price Forecasting in China Using Multiple Machine Learning Models. *J. Clean. Prod.* 249, 119386. doi:10.1016/j.jclepro.2019.119386
- Mirjalili, S. (2016). Dragonfly Algorithm: a New Meta-Heuristic Optimization Technique for Solving Single-Objective, Discrete, and Multi-Objective Problems. *Neural Comput. Applic* 27, 1053–1073. doi:10.1007/s00521-015-1920-1
- Niu, T., Wang, J., Lu, H., Yang, W., and Du, P. (2020). Developing a Deep Learning Framework with Two-Stage Feature Selection for Multivariate Financial Time Series Forecasting. *Expert Syst. Appl.* 148, 113237. doi:10.1016/j.eswa.2020.113237
- Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., and Woo, W. C. (2015). “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting,” in *Advances in Neural Information Processing Systems (NIPS 2015)*. ArXiv e-prints Available at: <https://arxiv.org/abs/1506.04214>.
- Song, J., and Li, S. (2017). Elite Opposition Learning and Exponential Function Steps-Based Dragonfly Algorithm for Global Optimization. *IEEE Int. Conf. Inf. Autom. ICIA*. doi:10.1109/ICInfA.2017.8079080
- Song, Y., Qin, S., Qu, J., and Liu, F. (2015). The Forecasting Research of Early Warning Systems for Atmospheric Pollutants: A Case in Yangtze River Delta Region. *Atmos. Environ.* 118, 58–69. doi:10.1016/j.atmosenv.2015.06.032
- Sun, Y., Wang, X., Chen, Y., and Liu, Z. (2018). A Modified Whale Optimization Algorithm for Large-Scale Global Optimization Problems. *Expert Syst. Appl.* doi:10.1016/j.eswa.2018.08.027
- Tian, C., and Hao, Y. (2020). Point and Interval Forecasting for Carbon price Based on an Improved Analysis-Forecast System. *Appl. Math. Modelling(C)*. doi:10.1016/j.apm.2019.10.022
- Torres, M. E., Colominas, M. A., Schlotthauer, G., and Flandrin, P. (2011). “A Complete Ensemble Empirical Mode Decomposition with Adaptive Noise,” in *ICASSP, IEEE International Conference on Acoustics, Speech and signal processing- Proceedings*, 4144e7. doi:10.1109/ICASSP.2011.5947265
- Wang, J., Niu, T., Du, P., and Yang, W. (2020). Ensemble Probabilistic Prediction Approach for Modeling Uncertainty in Crude Oil price. *Appl. Soft Comput.* 95, 106509. doi:10.1016/j.asoc.2020.106509
- Wang, J., Niu, X., Zhang, L., and Lv, M. (2021). Point and Interval Prediction for Non-ferrous Metals Based on a Hybrid Prediction Framework. *Resour. Pol.* 73, 102222. doi:10.1016/J.RESOURPOL.2021.102222
- Wang, R., Li, J., Wang, J., and Gao, C. (2018). Research and Application of a Hybrid Wind Energy Forecasting System Based on Data Processing and an Optimized Extreme Learning Machine [J]. *Energies* 11, 1712. doi:10.3390/en11071712
- Wang, Y., Zhang, N., Tan, Y., Hong, T., Kirschen, D. S., and Kang, C. (2019). Combining Probabilistic Load Forecasts. *IEEE Trans. Smart Grid* 10, 3664–3674. doi:10.1109/TSG.2018.2833869
- Wang, G. (2018). Design of Damage Identification Algorithm for Mechanical Structures Based on Convolutional Neural Network. *Concurrency Computat Pract. Exper* 30, e4891. doi:10.1002/cpe.4891
- Wang, J. (2020). OCT Image Recognition of Cardiovascular Vulnerable Plaque Based on CNN. *IEEE Access* 8, 140767–140776. doi:10.1109/access.2020.3007599
- Wei, S., Chongchong, Z., and Cuiping, S. (2018). Carbon Pricing Prediction Based on Wavelet Transform and K-ELM Optimized by Bat Optimization Algorithm in China ETS: the Case of Shanghai and Hubei Carbon Markets. *Carbon Manage.* 9 (6), 605–617. doi:10.1080/17583004.2018.1522095
- Xiao, L., Qian, F., and Shao, W. (2017). Multi-step Wind Speed Forecasting Based on a Hybrid Forecasting Architecture and an Improved Bat Algorithm. *Energ. Convers. Manage.* 143, 410–430. doi:10.1016/j.enconman.2017.04.012
- Zhang, X., Wang, J., and Zhang, K. (2017). Short-term Electric Load Forecasting Based on Singular Spectrum Analysis and Support Vector Machine Optimized by Cuckoo Search Algorithm. *Electric Power Syst. Res.* 146, 270–285. doi:10.1016/j.epsr.2017.01.035
- Zhang, X., and Wang, J. (2018). A Novel Decomposition-ensemble Model for Forecasting Short-term Load-time Series with Multiple Seasonal Patterns. *Soft Comput.* 65, 478–494. doi:10.1016/j.asoc.2018.01.017
- Zhu, B., and Wei, Y. (2013). Carbon price Forecasting with a Novel Hybrid ARIMA and Least Squares Support Vector Machines Methodology. *Omega* 41, 517–524. doi:10.1016/j.omega.2012.06.005
- Zhu, B., Yuan, L., and Ye, S. (2019). Examining the Multi-Timescales of European Carbon Market with Grey Relational Analysis and Empirical Mode Decomposition. *Physica A: Stat. Mech. its Appl.* 517, 392–399. doi:10.1016/j.physa.2018.11.016
- Zhu, J., Wu, P., Chen, H., Liu, J., and Zhou, L. (2019). Carbon price Forecasting with Variational Mode Decomposition and Optimal Combined Model. *Physica A: Stat. Mech. its Appl.* 519, 140–158. doi:10.1016/j.physa.2018.12.017

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.

Copyright © 2021 Yang, Guo, Jin and Song. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.

APPENDIX

APPENDIX A1 | The list of abbreviations in this study.

List of terminologies (method and indices)

EMD	Empirical model decomposition	EEMD	Ensemble empirical mode decomposition
ICEEMDAN	Improved complementary ensemble empirical mode decomposition with adaptive noise		
GRU	Gated recurrent unit	CNN	Convolutional neural networks
ELM	Extreme learning machine	BP	Back propagation neural network
CDF	Cumulative density function	ARIMA	Autoregressive interval moving average model
GWO	Grey wolf optimization algorithm	MODA	Multiobjective dragonfly optimization algorithm
GWO-BP	BP after GWO algorithm	GBiLSTM	Bidirectional long short-term memory-gated recurrent unit
BiLSTM	Bidirectional long short-term memory	SSA	Singular spectrum analysis
FINAW	Forecast interval normalized average width	FICP	Forecast interval coverage probability
AWD	Accumulated width deviation of testing dataset	IMFs	Intrinsic mode functions
DF	Distribution function	CDF	Cumulative distribution function
MAPE	Mean absolute percentage error	MAE	Mean absolute error
RMSE	Root mean square error	IA	Concordance index
TLS	T location-scale function	GRNN	Generalized regression neural network
R2	Coefficient of determination	DL	Deep learning probability density function
		PDF	
ANNs	Artificial neural networks	LSTM	Long short-term memory