



An optimized decomposition integration framework for carbon price prediction based on multi-factor two-stage feature dimension reduction

Wenjie Xu¹ · Jujie Wang¹ · Yue Zhang¹ · Jianping Li² · Lu Wei³

Accepted: 23 June 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The carbon trading market is an effective tool to combat greenhouse gas emissions, and as the core issue of carbon market, carbon price can stimulate the market for technological innovation and industrial transformation. However, the complex characteristics of carbon price such as nonlinearity and nonstationarity bring great challenges to carbon price prediction research. In this study, potential influencing factors of carbon price are introduced into carbon price forecasting, and a novel hybrid carbon price forecasting framework is developed, which contains data decomposition and reconstruction techniques, two-stage feature dimension reduction methods, intelligent and optimized deep learning forecasting with nonlinear integrated models and interval forecasting. Firstly, the carbon price series is decomposed into several simple and smooth subsequences using variational modal decomposition. The stacked autoencoder is then used to extract its effective features and reconstruct them into several new subsequences. A two-stage feature dimension reduction method is utilized for feature selection and extraction of exogenous variables. A bidirectional long and short-term memory model optimized based on the cuckoo search algorithm was used for prediction and nonlinear integration. Finally, Gaussian process regression based on a hybrid kernel function is applied to carbon price interval forecasting. The validity of the model was verified on seven real carbon trading pilot datasets in China. The methodology outperforms all benchmark models in the final simulation results, providing a novel and efficient forecasting method for the carbon trading industry.

Keywords Carbon trading market · Influencing factors · Two-stage feature dimension reduction · Bidirectional long and short-term memory · Cuckoo search algorithm

✉ Jujie Wang
jujiawang@126.com

¹ School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China

² School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China

³ School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100081, China

Abbreviations

| | |
|--------|---|
| ADF | Augmented Dickey-Fuller |
| AE | Auto encoder |
| AI | Artificial intelligence |
| ANN | Artificial neural network |
| ARIMA | Autoregressive integrated moving average |
| BDS | Brock-Decher-Scheikman |
| BPNN | Back propagation neural network |
| BiLSTM | Bi-directional long and short-term memory |
| BP | Back propagation |
| CS | Cuckoo search |
| EMD | Empirical modal decomposition |
| GARCH | Generalized autoregressive conditional heteroskedasticity |
| GPR | Gaussian process regression |
| LSSVR | Least squares support vector regression |
| LSTM | Long short-term memory |
| MAE | Mean absolute error |
| MAPE | Mean absolute percentage error |
| MLP | Muti-layer perceptron |
| RBFNN | Radial basis function neural networks |
| RNN | Recurrent neural networks |
| RF | Random forest |
| RMSE | Root mean square error |
| SAE | Stacked autoencoder |
| SSA | Singular spectrum analysis |
| VMD | Variational mode decomposition |

1 Introduction

Global climate change has created enormous challenges to socioeconomic and environmental systems in recent years. Excessive emissions of greenhouse gases, especially carbon dioxide, are one of the significant contributors to climate change. The European Union Emissions Trading System (EU ETS) has proven to be an effective system for combating greenhouse gas emissions (Wei et al., 2018). Meanwhile, carbon dioxide has developed into a tradable commodity under the trading system of the EU ETS. Researchers believe that carbon pricing is one of the most effective strategies to reduce global greenhouse gas emissions (Farouq et al., 2021). As the largest emitter of carbon dioxide, China established eight pilot carbon emissions trading markets in 2011, including Beijing, Chongqing, Guangdong, Shanghai, Tianjin, Hubei, Shenzhen, and Fujian. Driven by the domestic and international emission reduction situation, China is capable of achieving its emission reduction commitments in the coming decades (Meng et al., 2019).

As an emerging financial industry, the carbon trading market is influenced by a variety of factors, including the trading environment, regional development, and related policies. Due to the uncertainty of internal mechanism and the external environment, carbon price shows complex characteristics such as nonlinear and nonstationary (Zhu et al., 2019a, 2019b). As a result, carbon trading carries more unknown risks than conventional financial products. The

carbon trading market in China is still in the stage of exploration and improvement. Accurate carbon price forecasts not only help policy makers gain insight into the volatility features of the carbon trading market and improve the policies and regulations of the carbon trading market, but also promote the reduction of carbon emissions (Yang et al., 2020). Therefore, establishing an effective and stable carbon price forecasting framework for the carbon trading market is a significant research problem that needs to be solved urgently.

To achieve more accurate and effective carbon price forecasting, researchers have proposed many forecasting models in recent years. The existing forecasting models can be classified into three types: statistical models, artificial intelligence (AI) models, and hybrid models. Statistical models can achieve short-term forecasts of carbon prices. However, the statistical model is based on rigorous statistical assumptions, such as linearity and smoothness. It is difficult for statistical models to obtain ideal prediction accuracy when forecasting nonlinear and nonstationary carbon price time series. In contrast, AI models represented by artificial neural networks (ANN) do not need to satisfy statistical assumptions and can effectively learn the nonlinear relationships in data (Abedin et al., 2021). Nowadays, AI models are widely used in the field of time series forecasting, such as credit risk prediction forecasting (Abedin et al., 2018; Chi et al., 2019), energy supply forecasting (Sun et al., 2022), and wind speed forecasting (Li et al., 2022a, 2022b). Due to carbon prices' nonlinear and nonstationary characteristics, AI models have better prediction performance and adaptability than statistical methods. However, carbon prices in different carbon trading markets have different characteristics, making it impossible for a single AI model to suit the needs of all carbon markets. Therefore, the researchers proposed hybrid models to further improve the accuracy and stability of carbon price prediction. Among numerous hybrid models, the decomposition integration strategy shows excellent prediction performance in nonlinear time series forecasting. The decomposition integration model decomposes the original time series into several components by a signal decomposition algorithm, and then predicts them separately. The prediction results are then integrated to get the ultimate prediction results. The decomposition integration model has proven its effectiveness in various fields, such as financial indices (Li et al., 2021a), electricity prices (Zhang et al., 2022), and wind power generation (Liu et al., 2022).

Although the existing carbon price forecasting models have achieved good forecasting performance, there are still some areas for improvement. First, some of the previous studies have modeled and predicted all the components obtained from the decomposition, which increases the computational complexity and reduces the modeling efficiency. Similarities and invalid information exist in the different components obtained by decomposition, but this is often overlooked. Second, the prediction models for each subsequence are often the same. Different subseries have their own unique characteristics, for which separate prediction models with more appropriate hyperparameters should be built. Third, there is some unreasonableness in the way of linear integration of each subseries prediction result. The nonlinear relationship between the subseries prediction results is ignored, which will degrade the prediction performance to some extent. Fourth, most current studies mainly use historical carbon price data for forecasting, while ignoring economic factors, climate factors and other potential factors that affect carbon price fluctuations. Meanwhile, if all factors are used for forecasting studies, the redundancy and correlation among the influencing factors may lead to the problem of error accumulation in the forecasting model. Fifth, existing studies often give only point prediction results of carbon prices. However, carbon assets in the carbon market always give information on the fluctuation of carbon prices with the data of such intervals as [minimum, maximum]. With merely point predictions of carbon prices, it is difficult to capture the range of carbon price fluctuations in the real carbon market. In contrast, interval forecasting can

provide a new way for carbon price forecasting research, stepping out of the traditional mode of studying uncertainty problems with precise observations. In summary, achieving accurate and stable carbon price forecasts remains a great challenge.

To further address the challenges in carbon price forecasting research, this paper develops a novel hybrid forecasting framework. The framework considers the impact of different factors on carbon price prediction and contains a novel VMD algorithm, stacked autoencoder (SAE), random forest (RF), bidirectional long and short-term memory neural network (BiLSTM), cuckoo optimization algorithm (CS), and gaussian process regression (GPR). First, the original carbon price series is decomposed using the VMD algorithm to obtain several simpler and smoother components. The SAE is then used to extract the effective features from the different decomposition components, remove the noise and reconstruct them into several new components to reduce the complexity of the prediction. Second, a two-stage feature dimension reduction method based on RF and SAE is constructed in this paper to deal with the influencing factors. The factors with a strong influence on carbon price are selected using RF, and the factors with low influence are removed. SAE is then used to extract features from the factors after feature selection. SAE can remove redundant information from the factors and effectively extract the intrinsic features to solve the problems such as error accumulation that may result from the introduction of exogenous variables. Third, a prediction model is constructed for each reconstructed component using BiLSTM. And the CS is used to find the optimal hyperparameters of different BiLSTM to obtain the best prediction results. After getting the prediction results of all reconstruction components, the CS-optimized BiLSTM is used as a nonlinear integrated model to improve the prediction accuracy further and get the point prediction results of the carbon price. Finally, the interval prediction results of carbon prices are obtained using a GPR model based on a hybrid kernel function.

In summary, the main innovations and contributions of this paper are summarized in the following four points:

- (1) Considering the similarity and invalid information existing in different decomposition components, this paper combines VMD and SAE to extract the effective features of the original carbon price series. Meanwhile, this paper considers the impact of different influencing factors on carbon price prediction and proposes a two-stage feature dimension reduction method to process the influencing factor data effectively. It further reduces the computational complexity of the prediction model and improves the prediction accuracy.
- (2) To improve the accuracy of prediction, this paper proposes a two-stage prediction of carbon price based on BiLSTM with CS optimization. The CS-BiLSTM is used to model different components separately for prediction, enabling the prediction model to have higher prediction accuracy and stability. In addition, unlike the previous linear integration, this paper also uses CS-BiLSTM as a nonlinear integration model to improve the predictive performance and robustness of carbon price forecasting.
- (3) Interval forecasting can quantify the uncertainty of carbon price changes, and this paper uses a hybrid kernel function-based GPR model to make interval forecasts of carbon prices. Interval forecasting can provide traders and policy makers with more valuable information to reduce the risks they face in their business investment and decision-making efforts.
- (4) In this paper, an intelligent optimized nonlinear integrated carbon price forecasting framework based on multi-factor and two-stage feature dimension reduction is proposed for the first time, which has excellent forecasting performance and robustness.

The following is the rest of the paper's structure: A literature review is presented in Sect. 2. Section 3 presents the hybrid prediction framework proposed in this paper. The data sources, pre-processing, and prediction performance evaluation measures are discussed in Sect. 4. Section 5 discusses the empirical analysis and comparative study. Finally, Sect. 6 describes the research conclusions and future research outlook.

2 Literature review

Carbon price prediction models can be grouped into three types in the existing literature: statistical models, AI models, and hybrid models. In addition, interval forecasting can effectively quantify the uncertainty in financial markets, while some researchers often neglect interval forecasting of carbon prices. This section provides a detailed review of the carbon price forecasting literature and explores various types of carbon price point forecasting models as well as interval forecasting models.

2.1 Statistical prediction models

Traditional statistical models analyze historical data and infer relationships in the data to obtain better prediction results. Common statistical models include autoregressive integrated moving average (ARIMA) model and generalized autoregressive conditional heteroskedasticity (GARCH) model. Byun and Cho (2013) used the generalized autoregressive conditional heteroskedasticity (GARCH) models to achieve a forecast of the volatility of the next day's carbon price. Çanakoğlu et al. (2018) combine econometric time series, institutional transformation, and vector autoregressive models to analyze potential joint relationships in carbon prices with fuel prices and electricity prices. García-Martos et al. (2013) used a dynamic factor model to extract common features in carbon price volatility and effectively simulated carbon price trends using a multivariate GARCH model. Segnon et al. (2017) verify that GARCH based on two-state Markov-switching can achieve short and long-term forecasts of carbon price fluctuations in the EU. Although statistical models can produce good predictions when the time series satisfy certain classical statistical assumptions, this limits their ability to handle nonlinear nonstationary series (Sun et al., 2021). To overcome the effect of nonlinearity on carbon price prediction, researchers have tried to use multivariate linear models to predict carbon prices, such as MIDAS linear regression (Zhao et al., 2018). However, in the face of complex carbon prices, statistical models struggle to achieve the predictive performance expected by researchers.

2.2 AI prediction models

AI models have powerful nonlinear learning ability, which can effectively fit the nonlinear features hidden in the time series (Shajalal et al., 2021). Typical AI models include ANN, least squares support vector regression (LSSVR), multi-layer perceptron (MLP) neural network, and radial basis function neural networks (RBFNN). Mori and Jiang (2008) proposed an ANN-based carbon price prediction method and successfully applied it to real carbon price data. Fan et al. (2015) analyzed the chaotic characteristics of carbon prices and developed an MLP model to characterize the strong nonlinear characteristics of carbon prices. However, while these typical AI models can effectively handle nonlinear features, they have difficulty learning short and long-term dependent information in time series. The recurrent

neural network (RNN) differs from traditional neural networks in that it not only learns non-linear features in data, but also has the ability to remember information (Şaylı & Yılmaz, 2017). Therefore, the RNN is widely used in research areas such as natural language processing, time series prediction, and especially nonlinear time series. Chen et al. (2013) used RNN to predict flood flows and found that RNN can memorize dependencies in time series and obtain better prediction performance. However, RNN is not perfect and suffers from the defects of gradient disappearance and the inability to handle long-term dependencies. Therefore, the researchers proposed a variant RNN, Long Short Term Memory (LSTM) neural network (Peng et al., 2022). Zhang and Xia (2022) used LSTM to predict the price of EU carbon emission allowances and demonstrated that the prediction performance and robustness of LSTM are stronger than statistical models. To forecast carbon emissions in China, Huang et al. (2019) used back propagation (BP) neural networks, Gaussian process regression (GPR), and LSTM, respectively. The experimental results demonstrate that LSTM can be effectively applied to carbon emission prediction with optimal prediction performance. When dealing with complex and volatile carbon markets, a single AI model does not have sufficient predictive stability to achieve the predictive accuracy that researchers expect for carbon prices in different markets. To overcome the shortcomings of statistical models and single AI models, the researchers introduced hybrid models into the carbon price prediction study.

2.3 Hybrid prediction models

In existing research, there are two main types of hybrid models. One type combines different prediction models (Wang et al., 2022a, 2022b), and the second type combines data pre-processing techniques with prediction models (Li et al., 2021b). Hybrid models based on decomposition integration strategies can explore the complex intrinsic characteristics of carbon prices from different perspectives (Jiang et al., 2022; Zhu et al., 2019a, 2019b). The decomposition integration model uses a signal decomposition algorithm to pre-process complex time series and decompose them into a series of subseries with simpler structure and smoother trend. The decomposed subseries are then predicted separately, and the prediction results are integrated to obtain better prediction performance. Zhu et al. (2015) found that decomposing the carbon price allows the characterization of the carbon price at different scales and thus captures the trend of the carbon price more effectively. Wavelet transform (WT) (Momeneh & Nourani, 2022) and empirical modal decomposition (EMD) (Yu et al., 2008) are two popular signal decomposition algorithms. Sun et al. (2018) used WT to decompose and remove the high-frequency components from the carbon price data, and then used the partial autocorrelation function (PACF) to analyze the correlation between historical carbon price data and predict it. Zhu et al. (2013) used EMD to decompose the carbon price into several more stable and simple components. Then generalized ARCH (GARCH) model and the LSSVR were used for forecasting. The experiments demonstrate that GARCH-LSSVR outperforms statistical models and single artificial intelligence models in terms of predictive performance. Zhu et al. (2017) used the particle swarm algorithm (PSO) optimized LSSVR to predict each decomposition component separately based on EMD decomposition.

Although these decomposition algorithms can effectively decompose carbon price series, they also have unavoidable drawbacks. The decomposition performance of WT depends on the subjective choice of wavelet basis functions and decomposition levels. EMD does not require the choice of basis functions, but it also has inherent problems such as mode mixing and endpoint effects. To address the deficiencies in decomposition algorithms such as WT

and EMD, variational mode decomposition (VMD) was proposed (Dragomiretskiy & Zosso, 2014). Compared with WT and EMD, VMD can achieve adaptive decomposition of signals by constructing and solving constrained variational problems, effectively avoiding problems such as mode mixing and boundary effects, and better performing complex signal decomposition. Guo et al. (2022) applied the VMD algorithm to financial time series forecasting research and found that the forecasting performance of VMD-based ARIMA has a substantial improvement over ARIMA. Liu et al. (2022) used WT, EMD, and VMD algorithms to forecast the carbon price separately. It was demonstrated that the VMD-based forecasting model significantly outperformed WT and EMD. Chai et al. (2021) achieve an effective prediction of the carbon price in China based on VMD and extreme learning machine (ELM) and argue that the carbon price in China will generally increase during the recovery phase of covid-19.

2.4 Interval prediction models

Hybrid models are effective in predicting carbon prices, but most studies have only predicted point-value time series of carbon prices. When applied to the real carbon market, the point forecasts inevitably have varying degrees of bias and do not reflect the uncertainty of the carbon price market (Xiong et al., 2015). The interval forecast results contain more variable information and can quantify the uncertainty of the forecast results and give the interval of variation of the forecast results (Maia and de Carvalho, 2011).

Interval forecasting has been widely used in research areas such as load forecasting (Wang et al., 2022a, 2022b) and wind speed forecasting (Khodayar et al., 2022). The existing interval forecasting methods include three main types: quantile regression method, interval construction method, and probabilistic interval prediction method. The quantile regression method can obtain the interval of change in the point forecast under a given confidence interval by calculating different quantile points of the data. Bremnes (2004) predicted the interval of variation of wind power by the local partial regression. However, quantile regression needs to determine the suitable regression model and quantile points based on the data, and it is difficult to maintain good prediction accuracy in long-term forecasting. The interval construction methods typically use point prediction results to simulate interval results. Souza et al. (2017) constructed two regression models and simulated the upper and lower bounds of the interval by parametric methods. Quan et al. (2014) output the prediction interval for wind power using an ANN based on upper and lower bound estimation (LUBE). Although the interval construction method can obtain effective prediction results, it relies on a large and complete amount of data. Its computational complexity is high and difficult for engineering applications. The probabilistic interval forecasting method is mainly based on Bayesian theory. The distribution and expectation of the forecast values can be derived by constructing a distribution model of the forecast quantities, while obtaining interval forecasting results at any confidence level (Li et al., 2022a, 2022b). Gaussian Process Regression (GPR), a machine learning algorithm based on Bayesian theory, can predict the expected value of a location quantity and its distribution. The interval prediction performance of GPR as a probabilistic prediction technique has been proven in several fields (). Zhang et al. (2016) used an autoregressive model to extract the overall structure of wind speed and then used GPR for prediction. The study proved that the model could obtain satisfactory point and interval predictions. Peng and Bai (2019) used GPR for spatial orbit prediction and generated high performance orbit variation interval prediction results. Although GPR can obtain effective interval prediction results, there is still room for improving the prediction performance and

prediction stability of GPR for specific research areas. Therefore, interval prediction of carbon prices remains a significant challenge that requires additional attention from researchers.

3 The proposed hybrid prediction framework

For carbon price forecasting, an innovative decomposed integrated intelligent optimization framework based on multi-factor and two-stage feature dimension reduction is proposed in this study. This study's hybrid prediction framework is divided into three stages: data pre-processing, prediction and nonlinear integration, and interval prediction. Figure 1 depicts the proposed hybrid prediction framework's flow chart.

3.1 Data pre-processing stage

The data pre-processing stage can be divided into three parts. The first part uses the VMD algorithm to decompose carbon price into several components. The second part uses the SAE algorithm to reconstruct carbon price decomposition components to reduce the computational complexity of the prediction model. The third part proposes a two-stage feature dimension reduction technique for feature selection and feature extraction of carbon price influencing factors.

3.1.1 Data decomposition

As a non-recursive optimal decomposition technique, VMD overcomes the deficiencies of WT and EMD, and has stronger signal decomposition capability. In this paper, the carbon price sequence $Z(t)$ is decomposed into k number of band-limited intrinsic mode functions with each mode $m_t(k)$ of a center frequency $c(k)$ by the VMD algorithm. The VMD's principle is to assume that the center frequency bandwidth of each mode is limited and to minimize the sum of their frequency bandwidths under the condition that the sum of each mode equals the original signal.

The unilateral spectrum of each modal function $m_t(k)$ is found using the Hilbert transform. Then the spectrum of the mode function is corrected using exponential correction and shifted to the baseband region. Next, Gaussian smoothing is used on the signal to obtain the bandwidth of each modal function.

$$\min_{m_t(k), c(k)} \left\{ \sum_{i=1}^K \left\| \partial_t \left[(\delta(t) + \frac{j}{\pi t}) m_t(k) \right] e^{-jc(k)t} \right\|^2 \right\} \quad (1)$$

$$s.t. \sum_{i=1}^K m_t(k) = Z(t) \quad (2)$$

where $Z(t)$ denotes the original carbon price time series, $m_t(k)$ represents the k th decomposition mode, $c(k)$ is the center frequency, t implicates the time script, $\delta(t)$ denotes the unit pulse function, and ∂_t refers to the partial derivative of the function for time t . The above problem can be transformed from the objective function to an unconstrained optimization problem by introducing a quadratic penalty term as well as a Lagrange multiplier. For a more detailed rationale of the VMD algorithm, the reader can refer to the literature (Dragomiretskiy & Zosso, 2014).

Eventually, the carbon price is decomposed into several simpler and smoother subsequences by the VMD, which reduces the complexity and increases the predictability of the

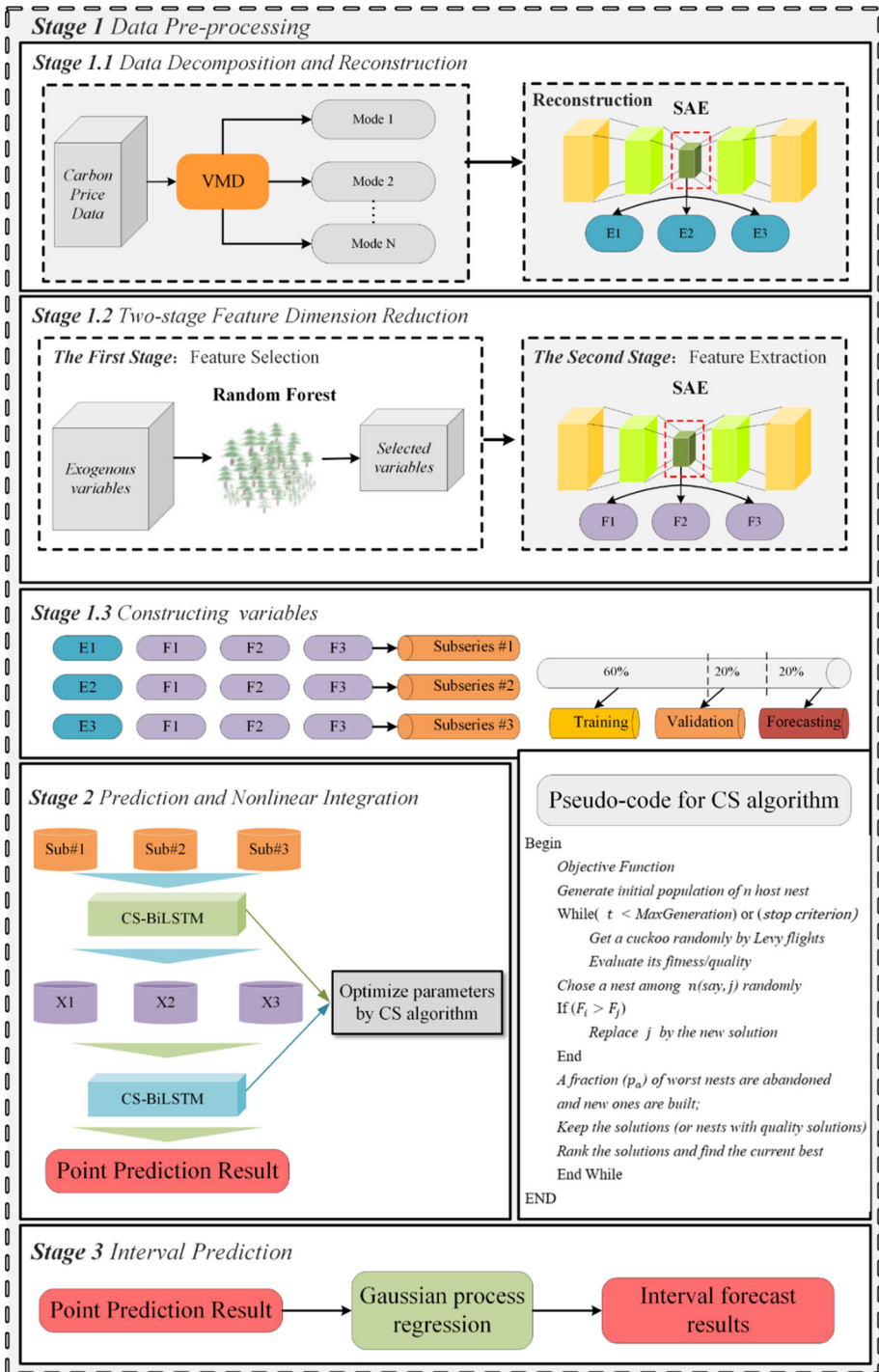


Fig. 1 Flowchart of the proposed model

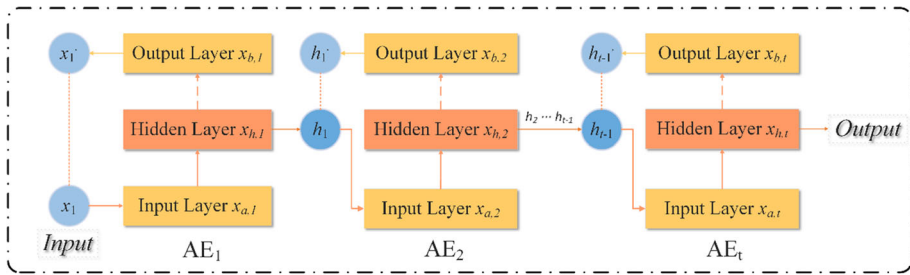


Fig. 2 The structure of SAE

carbon price series. However, there are correlations and invalid information between these decomposition components, so to reduce the complexity of the calculation, they are further processed using SAE in this paper.

3.1.2 Data reconstruction

The traditional autoencoder (AE) model, as a machine learning algorithm for data denoising and dimensionality reduction, usually consists of a three-layer network: input layer x_a , hidden layer h , and output layer x_b . The AE uses the original data as the target output for training, hoping to reconstruct a complete sequence of the original data using fewer data inputs. In practical applications, more attention is paid to the $x_a \rightarrow h$ conversion process, so AE can be used to extract the effective features of the carbon price decomposition components. However, it is difficult to obtain a better data representation with a single AE when dealing with high-dimensional and nonlinear data.

The Stacked autoencoders (SAE) consist of multiple AEs stacked on top of each other, capable of learning multiple representations of the original input data layer by layer, and more suitable for nonlinear data than AE (Xu & Ren, 2022). Figure 2 depicts the SAE structure. In this paper, the SAE model is constructed by layer-by-layer training to learn deeper and more meaningful information in the carbon price decomposition components, while removing the noisy signals. Ultimately, the reconstruction of the carbon price decomposition components is achieved.

3.1.3 Two-stage feature dimension reduction

Multiple potential carbon price influencing factors are introduced into the forecasting analysis in this paper to obtain better carbon price predicting results. Based on this, a two-stage feature dimension reduction method based on random forest (RF) and SAE is developed in this paper. RF is based on an improved bagging ensemble algorithm, which is a typical machine learning method. Unlike other traditional linear regression models, random forests can fit complex nonlinear relationships and measure the importance of variables (Ye et al., 2019). This is because RF constructs decision trees by selecting the best splitting points in a random subspace, and the best attributes are computed according to certain principles for selection, so this process is a feature selection process. This paper first uses RF to calculate the importance magnitude of different potential influencing factors for the first stage of the feature selection process.

Each tree in the random forest is a CART decision tree, and the trees are selected down the feature split nodes based on the corresponding Gini coefficients. The Gini value is denoted by GI, and the importance score of the variables in the random forest is denoted by VIM. Suppose there are n features: $x_1, x_2, x_3, \dots, x_k$. The Gini index score VIM of each feature x_i is the average change of the node-split impurity of the i th feature over all decision trees in the random forest (Wen & Yuan, 2020). If the VIM score is more engaged, it means that the feature is more influential. The Gini index is calculated as follows.

$$GI_n = 1 - \sum_{K=1}^K p_{nk}^2 \quad (3)$$

In this formula k denotes the number of categories and p_{nk} denotes the proportion of category in the node n . And the importance of feature x_i at the node n is:

$$VIM_{in}^{Gini} = GI_n - GI_x - GI_y \quad (4)$$

In the equation, GI_x and GI_y represent the Gini index of the two new nodes after the branch, respectively. Assuming that there are m trees in the random forest, the sum of the importance of all features in the whole forest is calculated to obtain the feature importance score of the i th feature.

$$VIM_i = \frac{VIM_{in}^{Gini}}{\sum_{m=1}^m VIM_{in}^{Gini}} \quad (5)$$

Finally, the importance scores of all features are normalized to obtain the importance score of each feature. Higher feature importance scores indicate that the influence factors have a more significant impact on carbon price fluctuations. To further reduce the complexity of the data, this paper uses SAE to perform the second stage of feature extraction on the feature subset obtained by RF selection. The meaningful information in the feature subset is further extracted by eliminating the noisy signals through SAE. Finally, the optimal features of potential influencing factors of carbon price are obtained after the two-stage feature dimension reduction process.

3.2 Prediction and nonlinear integration stage

In the prediction and nonlinear integration stages, BiLSTM is chosen as the prediction model in this paper, and the hyperparameters of BiLSTM are optimized using the CS algorithm to obtain better prediction performance. The CS-BiLSTM is used to predict each reconstructed component, and the prediction results are nonlinearly integrated using the CS-BiLSTM to obtain the final carbon price point prediction results.

3.2.1 Bi-directional long short-term memory

Hochreiter and Schmidhuber presented the LSTM as an enhanced RNN that addresses the problem of gradient disappearance and long-term memory weakness. The LSTM enhances the RNN with three gate controllers, namely “input gate,” “forget gate,” and “output gate,” allowing it to learn long-term dependent knowledge. The LSTM considers the temporal and nonlinear characteristics of the data, allowing it to effectively predict nonlinear time series. Figure 3 depicts the LSTM’s structure.

$$f_t = \sigma_1(w_{f,1}x_t + w_{f,2}x_{t-1} + b_f) \quad (6)$$

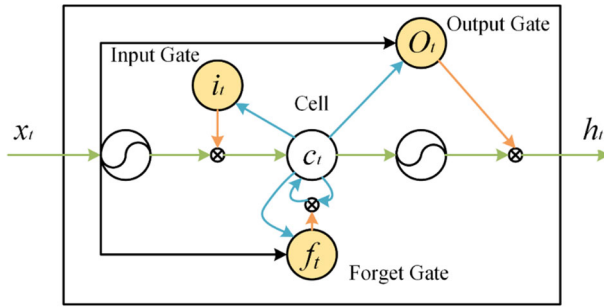


Fig. 3 The structure of LSTM

$$i_t = \sigma_1(w_{i,1}x_t + w_{i,2}h_{t-1} + b_i) \tag{7}$$

$$o_t = \sigma_1(w_{o,1}x_t + w_{o,2}h_{t-1} + b_o) \tag{8}$$

$$\tilde{c}_t = \sigma_2(w_{c,1}x_t + w_{c,2}h_{t-1} + b_c) \tag{9}$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \tag{10}$$

$$h_t = o_t * \sigma_2(c_t) \tag{11}$$

where σ_1 is the sigmoid function and σ_2 is the hyperbolic tangent function. Then, f_t , i_t and o_t represent the forget gate, input gate, and output gate, respectively. c_t and \tilde{c}_t represent the cell state and the intermediate cell state at moment t . x_t and h_t are the input and output values of the hidden layer, respectively. The weights and biases of the forget gates, input gates, output gates, and cell states are represented by w and b .

The BiLSTM consists of two LSTMs superimposed in the forward and reverse directions, and the structure of the BiLSTM model expanded along the time axis at moments $t - 1$, t and $t + 1$ is shown in Fig. 4. Where X represents the model input and Y represents the model output. The forward LSTM layer can be viewed as a forward computation from the

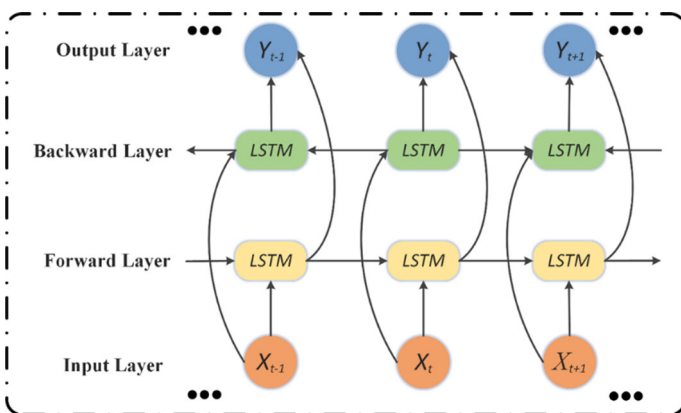


Fig. 4 The structure of BiLSTM

last moment to the end moment, and the backward LSTM layer can be viewed as a reverse computation from the last moment to the start moment. Both LSTM layers are processed in the same way during the computation, and the network weights are updated by the forward and backward propagation of the neurons. It allows BiLSTM to consider past and future data information during the mapping process between input and output sequences, which is more effective than one-way LSTM when dealing with time-series data.

3.2.2 Cuckoo search algorithm

Cuckoo Search (CS), a stochastic global search technique for populations based on simulated bird activity, was proposed by Yang and Deb (2014). CS is based on cuckoo parasitic brood-rearing behavior and can be enhanced by Lévy flights rather than basic isotropic random wandering. It can be widely used in optimization problems such as engineering structures and neural network training. Related studies show that this algorithm may be more effective than genetic algorithms (GA), PSO, and other algorithms (Yildiz, 2013). Several investigations have revealed that many animals and insects exhibit flight behavior similar to Lévy flights with power-law patterns. Lévy flights improve the CS algorithm compared to the basic isotropic random walk technique. To simplify the definition of standard CS, the following three idealized rules are presented in this paper:

- (1) Each cuckoo positions a single egg on a nest, which has been chosen randomly.
- (2) The best nests and eggs are passed along to the following generation.
- (3) The number of available host nests is fixed, and the host discovers the cuckoo's eggs with probability $p_\alpha \in (0, 1)$. The host has the option of either destroying the egg or abandoning the previous nest and constructing a new one.

The algorithm employs a well-balanced mix of a local random walk controlled by the parameter p_α and a global exploratory random walk controlled by the parameter p_α . The following is the local random walk principle:

$$x_i^{t+1} = x_i^t + \alpha s \otimes H(p_\alpha - \varepsilon) \otimes (x_j^t - x_k^t) \quad (12)$$

where x_i^t and x_j^t are two different solutions chosen by random substitution, and $H(\cdot)$ is a Heaviside function. In this formula, ε is a random number drawn from a uniform distribution, s is the step size, and \otimes represents the point product of the two vectors. Furthermore, a global random walk using Lévy flights:

$$x_i^{t+1} = x_i^t + \alpha L(s, \lambda) \quad (13)$$

$$L(s, \lambda) = \frac{\lambda \Gamma(\lambda) \sin(\pi \lambda / 2)}{\pi} \frac{1}{s^{1+\lambda}}, \quad (s \gg s_0 > 0) \quad (14)$$

where $\alpha > 0$ is the step scaling factor because Lévy flights have infinite mean and variance.

3.2.3 BiLSTM optimized by the CS algorithm

As a modified recurrent neural network, BiLSTM fits the model by updating the weights and biases based on the hyperparameters. Therefore, the prediction accuracy of BiLSTM is very sensitive to the setting of hyperparameters. It is shown that increasing the number of neurons in BiLSTM has a direct impact on the enhanced learning ability of the model, while the BiLSTM training time and overfitting risk are increasing. A suitable Batch size can improve

the computational efficiency and convergence accuracy of BiLSTM, too small may lead to long computation time and difficult convergence of the model, too large may fall into local extremes. In addition, BiLSTM may be overfitted when learning too long series, and a part of neurons are eliminated randomly by adding Dropout layer to effectively improve the generalization ability and robustness of BiLSTM. Previous research has relied on repeated trials to determine neural network hyperparameters. Obtaining the ideal hyperparameters for the neural network model is not only time-consuming and labor-intensive, but also complicated.

Related studies have verified that the CS algorithm is more effective than other population algorithms (Yildiz, 2013). The CS algorithm enables the BiLSTM model to determine the optimal hyperparameters quickly and accurately and realize the effective combination of the BiLSTM model network structure and carbon price data features. First, according to literature references and experimental studies, the BiLSTM model with three implicit layers is found to have the best prediction performance for carbon price data. Second, “softsign” is used as the activation function of BiLSTM instead of “tanh” because “softsign” is faster and less prone to saturation. In addition, a Dropout layer is added between each BiLSTM layer, and an early stop method is introduced to prevent overfitting. After determining the basic network structure of BiLSTM, the CS is used to optimize the three BiLSTM hyperparameters: batch size, learning rate, and units. The location information of the nest is randomly initialized according to the range of the hyperparameters.

Next, a BiLSTM model is constructed based on the hyperparameter values corresponding to the bird nest locations. The trained model is predicted using the validation data, and the mean absolute error (MAE) of the predicted model on the validation data set is used as the fitness function. The fitness function f is formulated as follows.

$$f = \frac{1}{n} \sum_{i=1}^n |x(t) - \hat{x}(t)| \quad (15)$$

where n is the number of samples in the validation set, $x(t)$ is the true value of the t -th validation data, and $\hat{x}(t)$ is the predicted value of the t -th validation data.

Then, the position is updated according to the values of nesting fitness in various clusters. The optimal value of the optimization objective is obtained when the termination condition is reached. If not, the population is divided based on the nest position information, and the fitness is calculated and updated until the termination requirement is met. Finally, the CS-BiLSTM model with optimal hyperparameters is obtained.

In this paper, different reconstructed components are modeled and predicted separately using the CS-BiLSTM model. Unlike the previous linear integration, this paper uses CS-BiLSTM for nonlinear integration of the prediction results of the reconstructed components. The final point prediction results of carbon price are obtained. The CS-BiLSTM model’s pseudo-code is provided below.

Algorithm 1 The CS-BiLSTM model**Parameters:**

G_{\max} - maximum number of iterations t - current iteration number
 n - number of nests m - number of dimensions
 x_i - nest i F_i - the fitness function of x_i
 p_α - the probability of finding an egg laid by a cuckoo α - Lévy flight step control quantity
 Lb - lower bounds Ub - upper bounds

1: /* Initialize the CS population $X = \{x_i, i = 1, 2, \dots, n\}$ in a random position. */
2: /* Initialize $t = 0$ and build BiLSTM network structure. */
3: **WHILE** $t < G_{\max}$ **DO**
4: **FOR EACH** $i = 1 : n$ **DO**
5: /* Cuckoos were randomly selected according to Lévy flight. */
6: /* Pass cuckoo nests values to the given BiLSTM network as units, batch size and learning rate. */
7: /* Calculate the corresponding fitness function F_i^* . */
8: /* Find the best nest location x_i^0 and the minimum fitness f_{\min} . */
9: /* Adjust the parameters of the BiLSTM and minimize the error. */
10: **END FOR**
11: **FOR EACH** $i = 1 : n$ **DO**
12: /* Update the current best solution via $X_i^{(t+1)} = X_i^{(t)} + \alpha \otimes L(s, \lambda)$. */
13: /* A fraction p_α of the worst nests are abandoned, and new ones are built. */
14: /* Update the corresponding fitness function F_i^* . */
15: **END FOR**
16: **IF** $f_{\text{best}} < f^*$ **THEN**
17: /* Find the best nest position x_i^j . */
18: **END IF**
19: /* $iter = iter + 1$. */
20: **END WHILE**
21: **RETURN** X_{best}

3.3 Interval prediction stage

In the interval prediction section, this paper uses the GPR model to obtain the interval prediction results of the carbon price. GPR is based on Bayesian theory and is a kernel-based approach to solving regression problems (Jin et al., 2021). Compared to neural networks, GPR has been widely used in areas such as power and renewable energy because of its advantages of rapid deployment and hyperparametric adaptive acquisition of probabilistically meaningful expected outputs (Petelin & Kocijan, 2014). Therefore, this paper chooses to use GPR as an interval prediction model for carbon prices to quantify the uncertainty of the point prediction results and to give probability intervals for carbon price fluctuations. The prediction

performance of building a GPR model mainly depends on the setting of the kernel function. Researchers often use GPR based on a single kernel function for their studies. However, for complex nonlinear carbon price series, it is difficult for a single kernel function to effectively describe its complex fluctuations. Therefore, in this paper, a hybrid kernel function-based GPR is constructed to achieve effective carbon price interval prediction. There are three common kernel functions in GPR as follows.

$$K_{SE}(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^2}{2l^2}\right) \quad (16)$$

$$K_{RQ}(x_i, x_j) = \left(1 + \frac{d(x_i, x_j)^2}{2\alpha l^2}\right)^{-\alpha} \quad (17)$$

$$K_{Matern}(x_i, x_j) = \frac{1}{\Gamma(v)2^{v-1}} \left(\frac{\sqrt{2v}}{l} d(x_i, x_j)\right)^v K_v\left(\frac{\sqrt{2v}}{l} d(x_i, x_j)\right) \quad (18)$$

where K_{SE} is the squared exponential (SE) kernel function, K_{RQ} is the rational quadratic (RQ) kernel function, and K_{Matern} is the Matern kernel function. $d(\cdot, \cdot)$ denotes the euclidean distance. α is the scale mixing parameter representing the shape of the kernel function. l is a hyperparameter representing the length of the kernel. In addition, the modified Bessel and gamma functions are $K_v(\cdot)$ and $\Gamma(\cdot)$, respectively, and v is used as the smoothness of the control function. To compare the effects of different kernel functions on interval predictability, the above three single kernel functions and three hybrid kernel functions are used for analysis in this paper, for a total of six kernel functions. The hybrid kernel functions $K_{hybrid}(x_i, x_j)$ are obtained by combining two single kernel functions and are constructed as shown below, where $K_a(x_i, x_j)$ and $K_b(x_i, x_j)$ are two of the three single kernel functions.

$$K_{hybrid}(x_i, x_j) = K_a(x_i, x_j) + K_b(x_i, x_j) \quad (19)$$

3.4 Evaluation criteria

Researchers have presented various error evaluation metrics to quantify prediction accuracy, and this paper considers the mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) to measure the point prediction accuracy of the proposed model. In addition, the interval prediction results are evaluated using the predicted interval average width (PIAW) and the predicted interval coverage probability (PICP). Under a certain confidence level, if the PIAW is smaller and the PICP is larger, it means that the interval prediction is more valid. The exact formula is represented in Table 1, where k represents the number of test sets in the data, \hat{y} is the final prediction, and y is the true value of the test set, L_n and U_n are the lower and upper value of the forecasting interval, N is the interval prediction's length, and C_n means a Boolean value.

4 Data

4.1 Data collection

China gradually started carbon trading pilot programs in 2011, with a total of eight carbon trading pilots as of December 31, 2021. Among them, the pilot in Fujian Province started

Table 1 Evaluation metrics

| Metric | Definition | Equation |
|--------|---|--|
| MAE | Mean absolute error | $MAE = \frac{1}{k} \sum_{i=1}^k y_i - \hat{y}_i $ |
| RMSE | Root mean square error | $RMSE = \sqrt{\frac{\sum_{i=1}^k (y_i - \hat{y}_i)^2}{k}}$ |
| MAPE | Mean absolute percentage error | $MAPE = \frac{1}{k} \sum_{i=1}^k \left \frac{\hat{y}_i - y_i}{y_i} \right \times 100\%$ |
| PIAW | Predicted interval average width | $PIAW = \frac{1}{N} \sum_{n=1}^N (U_n - L_n)$ |
| PICP | Predicted interval coverage probability | $PICP = \frac{1}{N} \sum_{n=1}^N c_n, \quad c_n = \begin{cases} 1, & A_n \in [L_n, U_n] \\ 0, & A_n \notin [L_n, U_n] \end{cases}$ |

trading in early 2017, and since it started late and only 538 carbon price data are publicly available on trading days, this paper does not consider the data of Fujian pilot. To verify the validity of the proposed hybrid framework, we select daily carbon price data from seven carbon trading pilots for an empirical study. For better analysis, the time windows of the carbon price data of the seven markets are aligned in this paper, and the daily carbon price data from June 19, 2014, to December 31, 2021, are all selected, and the dataset is obtained from the respective exchanges. To better train the prediction model, verify the model's generalization ability and prevent overfitting, in this paper, all carbon trading pilot data are separated into training, validation, and test sets in the ratio of 6:2:2. The total sample size, training set sample size, validation set sample size, and test set sample size of the seven carbon trading pilot data are shown in Table 2.

Figure 5 shows the carbon price trends of three carbon trading markets and their geographical locations. The relevant statistical indicators for the seven market carbon price data are shown in Table 3, containing mean, maximum, minimum, median, and standard deviation. Considering that the different carbon trading markets are still in the pilot stage, the trading dates are not exactly the same as other financial markets, and there are some missing values. Therefore, the missing data for the date when carbon trading was not performed are removed in this paper. In addition, the data are normalized using typical normalization processing methods. The dimensionless data can improve the efficiency of the model operation, and solve the

Table 2 The sample size of data from seven carbon trading pilots in China

| Dataset | Sample size | Training set | Validation set | Test set |
|-----------|-------------|--------------|----------------|----------|
| Beijing | 1702 | 1021 | 340 | 341 |
| Tianjin | 1646 | 987 | 329 | 330 |
| Shanghai | 1693 | 1015 | 339 | 339 |
| Shenzhen | 1817 | 1090 | 363 | 364 |
| Guangdong | 1854 | 1112 | 371 | 371 |
| Hubei | 1833 | 1099 | 367 | 367 |
| Chongqing | 1692 | 1015 | 338 | 339 |

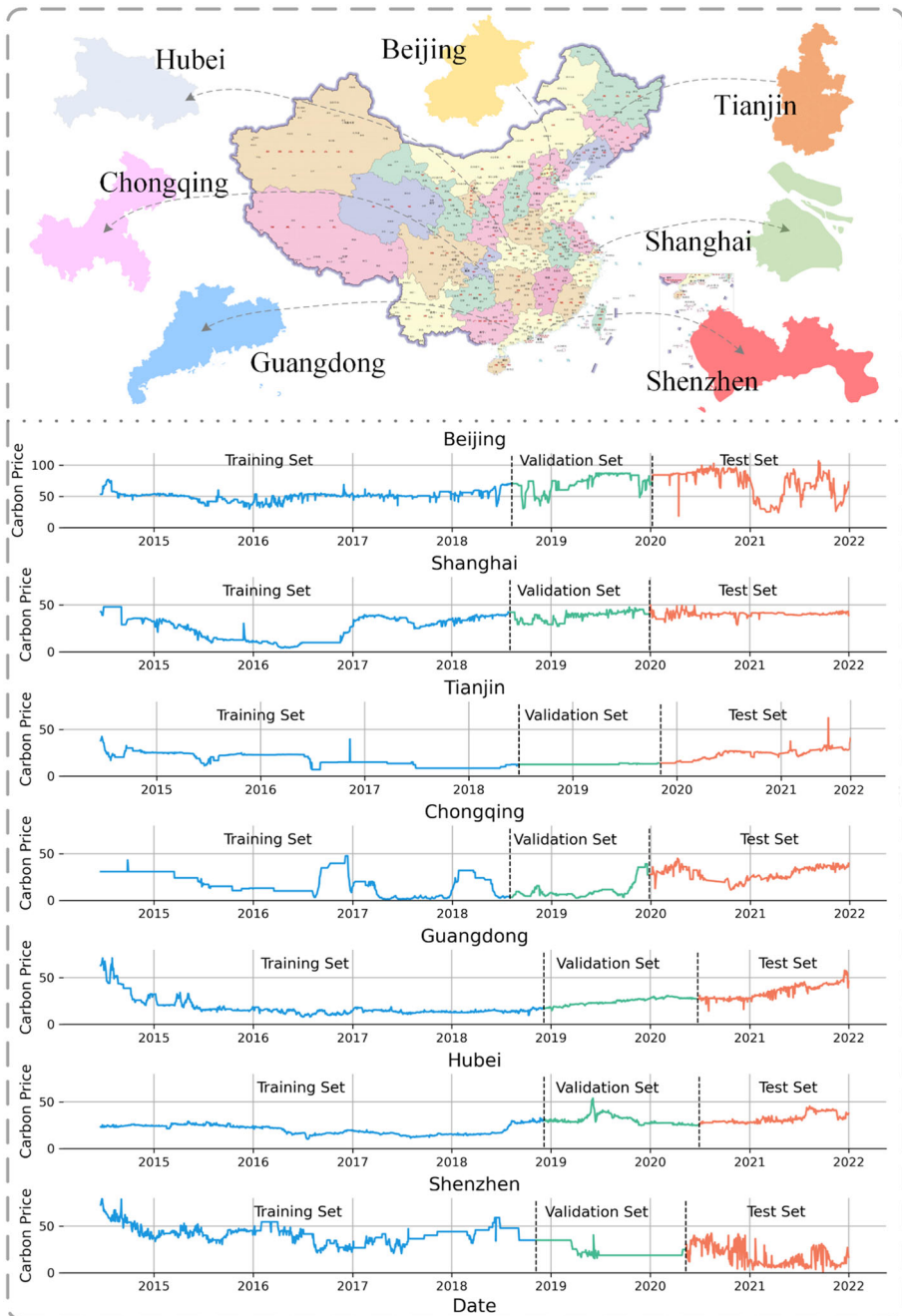


Fig. 5 Carbon price data for the seven carbon trading pilots

Table 3 Statistical indicators of carbon prices

| Dataset | Mean | Max | Min | Median | Std |
|-----------|-------|--------|-------|--------|-------|
| Beijing | 60.00 | 107.26 | 18.63 | 53.40 | 16.54 |
| Tianjin | 17.59 | 62.38 | 7.00 | 15.05 | 6.69 |
| Shanghai | 31.16 | 49.98 | 4.20 | 35.30 | 11.99 |
| Shenzhen | 33.21 | 79.00 | 1.00 | 34.79 | 13.87 |
| Guangdong | 22.89 | 71.09 | 8.10 | 19.53 | 10.82 |
| Hubei | 24.77 | 53.85 | 10.07 | 25.21 | 7.37 |
| Chongqing | 18.59 | 47.52 | 1.00 | 16.13 | 11.86 |

problem of failure to converge that may be caused by odd sample data. The normalization formula is as follows, where x represents the original data and x' is the normalized data.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (20)$$

4.2 Potential influencing factors

Related research showed that variations of carbon prices are impacted by factors like energy and weather (Dutta et al., 2018; Hao & Tian, 2020). When selecting factors influencing carbon price, its data must meet the following characteristics: (1) The data must be real and valid and have a large enough data sample. (2) The factors must be able to influence the fluctuation of the carbon price to some extent. Considering the above issues to the relevant literature, this paper selects the influencing factors in four directions: similar products, energy structure, economic factors, and environmental factors.

(1) Similar products

The international carbon emissions trading market is relatively mature and has important implications for the Chinese carbon market, which is not yet fully open. Similar products such as EU emission allowances (EUA) and certified emission reduction (CER) can be used to fulfill carbon emission reduction obligations. However, since 2013, the EU no longer accepts CER project indicators from emerging countries such as China, and CER data will no longer be publicly available after April 2021. Therefore, this paper chooses EUA to react to the impact of similar products on China's carbon price, using data from the Wind database.

(2) Energy structure

Energy is often considered to be the factor that most directly affects the price of carbon. Furthermore, West Texas Intermediate crude oil futures settlement prices are chosen to reflect crude oil prices, natural gas prices from the New York Mercantile Exchange, coking coal futures, and coke futures settlement prices from the Dalian Commodity Exchange, and power coal futures settlement prices from the Zhengzhou Commodity Exchange. The data used are from the Wind database and the Choice database.

(3) Economic factors

The USD/CNH exchange rate and the China Industrial Index (CHII) have been selected to represent the effect of major economic factors. The data used are from the Wind database and the Choice database.

(4) Environmental factors

Air quality index (AQI), PM2.5, PM10, SO2, CO, and NO2 were selected as environmental factors. The data used were obtained from data.cma.cn.

4.3 Unstructured data based on web search index

To achieve effective forecasting of financial data, researchers have begun to pay more attention to the impact of investor behavior on financial markets. Bank et al. (2011) find that an increase in Google keyword searches correlates with increased stock trading activity and liquidity, and they claim that Google searches minimize asymmetric information costs. Some studies have even concluded that there is a correlation between the volume of carbon-related keyword searches on the Internet and the volatility of carbon prices (Hintermann et al., 2016). Therefore, unstructured data based on web search can respond to some extent to changes in investor behavior, investor attention to the market, and asymmetric information costs.

Given that Google exited the Chinese market in 2010, its impact on the Chinese market is minimal. And Baidu, the search engine with the largest market share in China, is more represented in the country. Therefore, this paper selects the Baidu index as an unstructured data source based on web search index to provide more real-time and responsive information on investors' behavior. In this paper, the Baidu indexes of twenty-four carbon-related keywords were selected from the directions of carbon finance, low-carbon life, and environmental protection, including carbon footprint, carbon tax, and greenhouse effect. All keywords are shown in Table 4, and the Baidu indices of keywords are obtained from the official website of the Baidu index (<https://index.baidu.com/>).

In addition, this paper uses a concise and effective mutual information method to evaluate the degree of association between different keywords and carbon price to determine the weight of each keyword to construct a comprehensive index. Assuming that x_i is the Baidu index of the i th keyword and the target is the carbon price time series y , the mutual information between the two variables is defined as follows.

$$MI(x_i, y) = \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 \left(\frac{P(x_i, y_j)}{P(x_i)P(y_j)} \right) \quad (21)$$

Table 4 Baidu Index keywords

| Carbon-related keywords | | | |
|-------------------------|--------------------------|--------------------------|-------------------------------------|
| Low carbon | Carbon tax | Smog | Carbon dioxide emissions |
| Low-carbon economy | Carbon emissions trading | Greenhouse gases | Pollution discharge |
| Carbon footprint | Carbon sink | Greenhouse gas emissions | Clean energy |
| Emission reduction | Carbon neutralization | Greenhouse effect | Cleaner production |
| Carbon emission | Carbon trading | Air pollution | Low-carbon environmental protection |
| Carbon emissions | Carbon dioxide | Atmospheric pollutant | Low carbon life |

Table 5 The details of the influences on the carbon price in China

| External factors | Factor name | Factor symbol |
|-----------------------|----------------------------------|-------------------|
| Similar products | European union allowance futures | EUA |
| Energy structure | Crude oil price | WTI |
| | Natural gas price | Nature Gas |
| | Coke price | Coke |
| | Steam coal price | Steam coal |
| | Coking coal price | Coking coal |
| Economic factors | USD/CNY exchange rate | USD_CNY |
| | China Industrial Index | CHII |
| Environmental factors | Daily AQI | AQI |
| | Daily PM _{2.5} | PM _{2.5} |
| | Daily PM ₁₀ | PM ₁₀ |
| | Daily SO ₂ | SO ₂ |
| | Daily CO | CO |
| Web search index | Daily NO ₂ | NO ₂ |
| | Comprehensive web index | Baidu |

Then the average mutual information value of all keywords is calculated, and the weight is determined for each keyword based on Eq. (22), where W_i is the weight of the i th keyword. Finally, the comprehensive web index Baidu is obtained.

$$W_i = \frac{MI(x_i, y)}{1/24 \sum_{n=1}^{24} MI(x_i, y)} \quad (22)$$

$$\text{Baidu} = \sum W_i \cdot x_i \quad (23)$$

In summary, a total of fifteen potential influencing factors of the carbon price are selected in this paper, and the information of all influencing factors is displayed in Table 5.

5 Case analysis

5.1 Nonstationary and nonlinear data set tests

In this research, the Augmented Dickey-Fuller (ADF) and Brock-Decher-Scheikman (BDS) are used to assess the original carbon price series to confirm its nonlinearity and nonstationarity. The results of these two tests on the seven pilot datasets are shown in Tables 6 and 7. According to the ADF test results, the carbon price of all pilots are nonstationary within the 10%, 5%, and 1% levels. Moreover, the p-values of the BDS tests on different embedded dimensions are much less than 0.01, which also proves that the carbon price series of all pilots are nonlinear at the 1% level. As shown above, all pilots' original carbon price series are nonstationary and nonlinear.

Table 6 Test results of ADF

| Carbon prices | ADF statistic | 1% | 5% | 10% | Prob |
|---------------|---------------|----------|----------|----------|--------|
| Beijing | - 2.0172 | - 3.4342 | - 2.8632 | - 2.5676 | 0.2527 |
| Tianjin | - 2.2679 | - 3.4343 | - 2.8633 | - 2.5677 | 0.1763 |
| Shanghai | - 1.6093 | - 3.4342 | - 2.8632 | - 2.5676 | 0.4788 |
| Shenzhen | - 2.2555 | - 3.4340 | - 2.8631 | - 2.5676 | 0.1867 |
| Guangdong | - 1.6393 | - 3.4339 | - 2.8631 | - 2.5676 | 0.4626 |
| Hubei | - 1.2216 | - 3.4339 | - 2.8631 | - 2.5676 | 0.6642 |
| Chongqing | - 2.0095 | - 3.4342 | - 2.8632 | - 2.5676 | 0.2542 |

5.2 Data decomposition and reconstruction

The carbon price has strong nonlinearity and non-smoothness, and the VMD technique can weaken the non-smoothness of the series and reduce the prediction difficulty. Referring to related literature studies (Dragomiretskiy & Zosso, 2014), in the first stage, this paper uses the VMD decomposition algorithm to decompose the original carbon price series into several components. Taking the three carbon trading pilots, Beijing, Shanghai, and Tianjin, as an example, the components obtained from the decomposition of the original carbon price, arranged in order from the lowest frequency to the highest frequency, are displayed in Fig. 6. Then, in this paper, the decomposed components are further extracted and reorganized using SAE. This not only removes the noise signal from the decomposed components and extracts the effective features, but also reduces the complexity of the prediction model. Through an empirical study, it is found that the SAE model containing two hidden layers to reconstruct the data works best. In addition, the training of SAE utilizes an unsupervised layer-by-layer greedy method so that the parameters of each hidden layer are locally optimal. Finally, the decomposed modal component sequences are reconstructed into three sub-series, E1, E2, and E3, using SAE. Taking the three carbon trading pilots in Beijing, Shanghai, and Tianjin as examples, the results of the carbon price series using the VMD-SAE method are shown in Fig. 6.

5.3 Two-stage feature dimension reduction

In this paper, a total of fifteen exogenous variables are introduced by considering the impact of potential influences on carbon price forecasting. To solve the problems of correlation, complexity, and redundancy in exogenous variables, a two-stage feature dimension reduction method based on RF-SAE is developed in this paper. First, the feature importance score of each exogenous variable is calculated using RF, and the feature subset is obtained by selecting the top ten features in the feature importance score. Taking the three carbon trading pilots in Beijing, Shanghai, and Tianjin as examples, the histogram of the characteristic importance scores of different influencing factors is shown in the figure.

It is clear that the factors that have a higher degree of influence are not exactly the same in different carbon trading pilots. This is due to the differences in socioeconomic development, industrial structure and other aspects of different carbon trading pilot provinces and cities. To better explore how China's carbon trading market will develop, each carbon trading pilot has

Table 7 Test results of BDS

| Carbon pilots | m-dimensional space | | | | | | | | | | | |
|---------------|---------------------|------|----------|------|----------|------|----------|------|----------|------|------|------|
| | 2 | | 3 | | 4 | | 5 | | 6 | | Prob | BDS |
| | BDS | Prob | BDS | Prob | BDS | Prob | BDS | Prob | BDS | Prob | | |
| Beijing | 4.5814 | 0.00 | 6.3036 | 0.00 | 8.2489 | 0.00 | 10.5668 | 0.00 | 13.5657 | 0.00 | 0.00 | 0.00 |
| Tianjin | 189.1967 | 0.00 | 210.1345 | 0.00 | 237.0415 | 0.00 | 275.2171 | 0.00 | 327.6276 | 0.00 | 0.00 | 0.00 |
| Shanghai | 16.7180 | 0.00 | 6.8294 | 0.00 | 8.9358 | 0.00 | 11.4764 | 0.00 | 14.7692 | 0.00 | 0.00 | 0.00 |
| Shenzhen | 5.3118 | 0.00 | 6.9627 | 0.00 | 8.8965 | 0.00 | 11.2306 | 0.00 | 14.1851 | 0.00 | 0.00 | 0.00 |
| Guangdong | 5.5139 | 0.00 | 6.5366 | 0.00 | 7.7226 | 0.00 | 9.1085 | 0.00 | 10.7723 | 0.00 | 0.00 | 0.00 |
| Hubei | 6.2662 | 0.00 | 8.1124 | 0.00 | 8.1124 | 0.00 | 10.3719 | 0.00 | 16.7180 | 0.00 | 0.00 | 0.00 |
| Chongqing | 4.9200 | 0.00 | 7.0679 | 0.00 | 9.4658 | 0.00 | 12.4795 | 0.00 | 16.4502 | 0.00 | 0.00 | 0.00 |

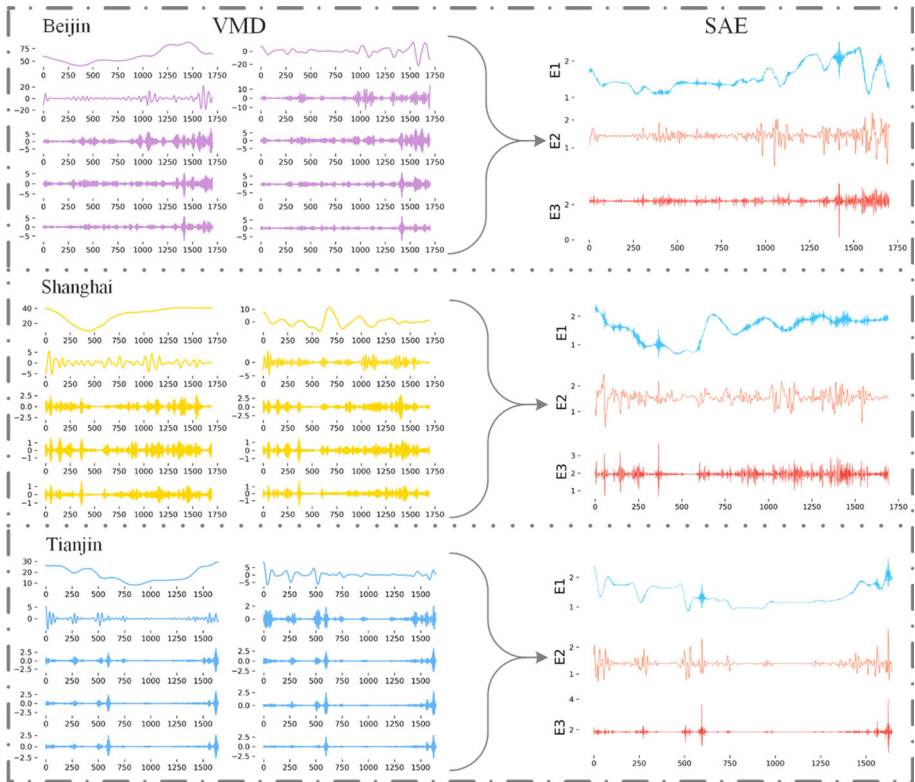


Fig. 6 Results of the VMD-SAE method

set different emission control industries, carbon quota allocation systems, and control thresholds. For example, Beijing and Shanghai include a wide range of emission control industries, including power, construction, and service industries, while Tianjin mainly includes high-emission industries such as power, steel, and chemicals. Second, Beijing's carbon emission allowances consist of year-by-year unpaid allocations and reserved paid allocations, while Shanghai's carbon emission allowances consist of a three-year allocation and occasional paid allocations. At the same time, different markets levy different default penalties on overdue companies, with Beijing being fined three to five times the average market price for default, Shanghai being fined RMB 50,000 to 100,000, and Tianjin only being given a deadline to make corrections. Therefore, although bulk products such as crude oil and coke will directly cause changes in the cost of carbon credits, different carbon trading pilots have different emission control sectors and have different impacts on carbon prices. Influenced by policy differences such as default mechanisms and government intervention, the EU carbon price also has different guiding effects on the carbon price of different carbon trading pilots. Therefore, the impact of different factors on carbon prices has regional variability.

This paper uses RF to select the top ten ranked exogenous variables as a subset of features for the carbon prices of different carbon trading pilots. It ensures that the selected feature subset has sufficient influence on the fluctuation of the carbon price. Then, the SAE model with layer-by-layer greedy training is used to mine the effective features in the feature subset and remove the noisy signals. Finally, the three optimal feature subsets F1, F2, and F3 are obtained

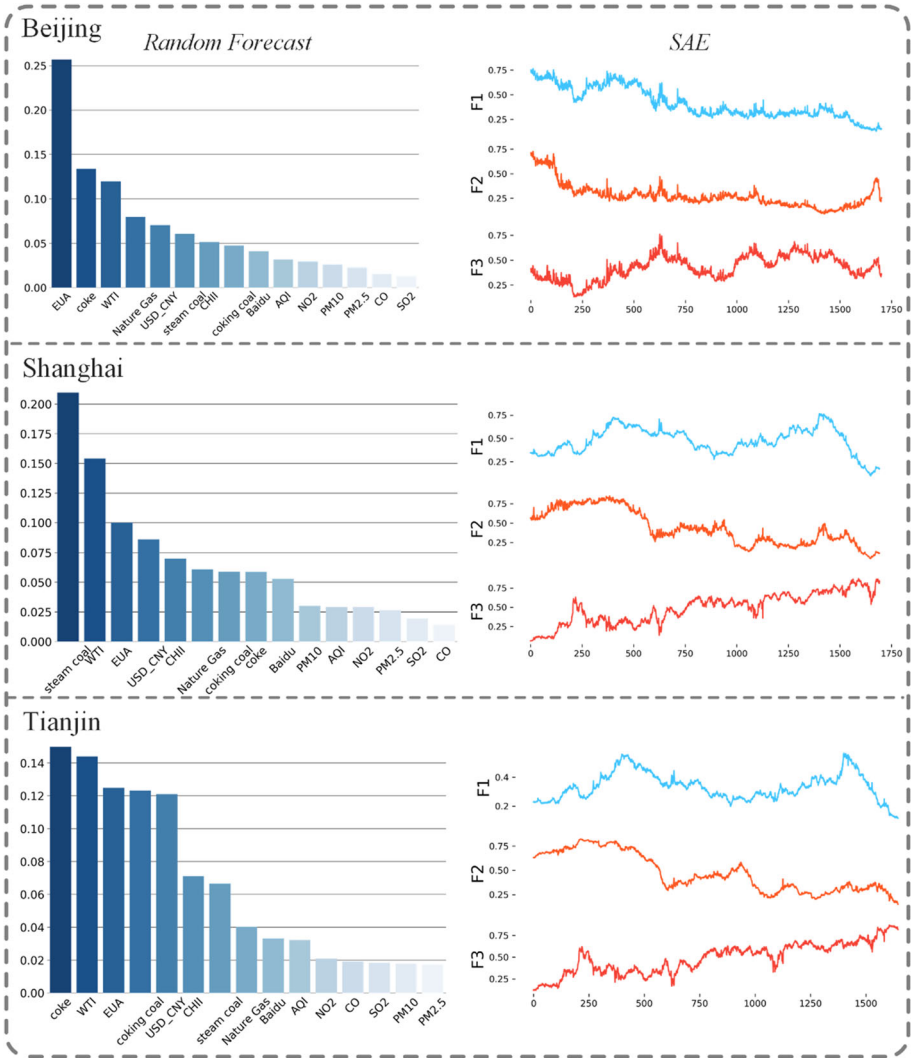


Fig. 7 Two-stage feature dimension reduction results

after processing and incorporated into the final prediction model. Taking Beijing, Shanghai and Tianjin as examples, Fig. 7 shows the three optimal feature subsequences obtained by using the two-stage feature dimension reduction method to process the exogenous variables of the three carbon trading pilots.

5.4 Prediction and nonlinear integration

In the prediction part, BiLSTM is used to build separate prediction models for the different components. Meanwhile, the prediction performance of BiLSTM is very sensitive to the setting of its hyperparameters. In this paper, the CS algorithm is used instead of empirical

tuning, and the hyperparameters of BiLSTM are intelligently sought to obtain the optimal BiLSTM prediction model. Through extensive experiments, the BiLSTM model using three hidden layers was found to have the best accuracy and stability when predicting carbon prices. This paper also uses “softsign” instead of “tanh” as the activation function of BiLSTM. In addition, to prevent overfitting, this study incorporates the Dropout layer into the BiLSTM model. Experimentally, the dropout rate is set to 0.1. Unlike previous empirical tuning, this paper employs the CS algorithm to determine the ideal BiLSTM hyperparameters, such as batch size, learning rate, and units. The optimization range of batch size is set from 2 to 512, the learning rate from 0.001 to 0.0001, and units from 2 to 512. Referring to the relevant literature and practical requirements, the population size of CS calculation is set to 30, the probability of cuckoo eggs being found p_a is taken as 0.25, and the number of iterations is 100 (Yang & Deb, 2014).

Furthermore, the setting of the observation window, i.e., how long in the past time information is referenced for forecasting, is very important for time series forecasting. An observation window that is too short is difficult to contain sufficient time-series information, while too long may introduce early irrelevant information. Therefore, the observation windows are set to 3, 4, 5, and 6 respectively to select the optimal observation window size and improve the point prediction performance of carbon price. Then, this paper makes full use of the nonlinear learning ability and feature extraction ability of CS-BiLSTM to nonlinearly integrate the prediction results of each reconstructed component. The final point prediction value of carbon price is obtained.

Taking the three carbon trading pilots, Beijing, Shanghai and Tianjin, as examples, Fig. 8 shows the point forecast results and the forecast performance evaluation results at an observation window of 4. To further validate the effectiveness and robustness of the proposed hybrid model, this paper also conducts an empirical study using data from four carbon pilot markets, namely Shenzhen, Guangdong, Hubei and Chongqing. The empirical analysis process for these four carbon trading pilots is the same as the one described above for Beijing, Shanghai and Tianjin. Table 8 shows the point prediction results of the seven carbon trading pilots under different observation windows. It can be found that the hybrid model proposed in this paper has the optimal prediction performance and prediction stability when the observation window is set to 4, i.e., the carbon price of the future day is predicted using the data of the first 4 days. In summary, the hybrid model proposed in this paper maintains good point prediction performance in the seven carbon trading pilot data sets, and is an effective tool for point prediction of carbon price.

5.5 Point prediction comparison experiment

5.5.1 Comparison of hyperparametric optimization methods

In order to verify the superiority of the CS-BiLSTM model proposed in this paper, four different common hyperparameter optimization algorithms, including differential evolutionary algorithm (DEA), genetic algorithm (GA) and particle swarm optimization algorithm (PSO), and ant colony algorithm (ACO) are selected for comparison to optimize the hyperparameters of BiLSTM and perform carbon price prediction. The same number of populations and iterations are set for the four different optimization algorithms, and the optimization ranges of the hyperparameters of the BiLSTM are all the same. All algorithms run on Intel Xeon Gold 6139 CPU, NVIDIA Tesla V100 GPU, 86 GB RAM, Linux, and Python 3.8. Taking the Beijing dataset as an example, the results of the optimized BiLSTM model for the four

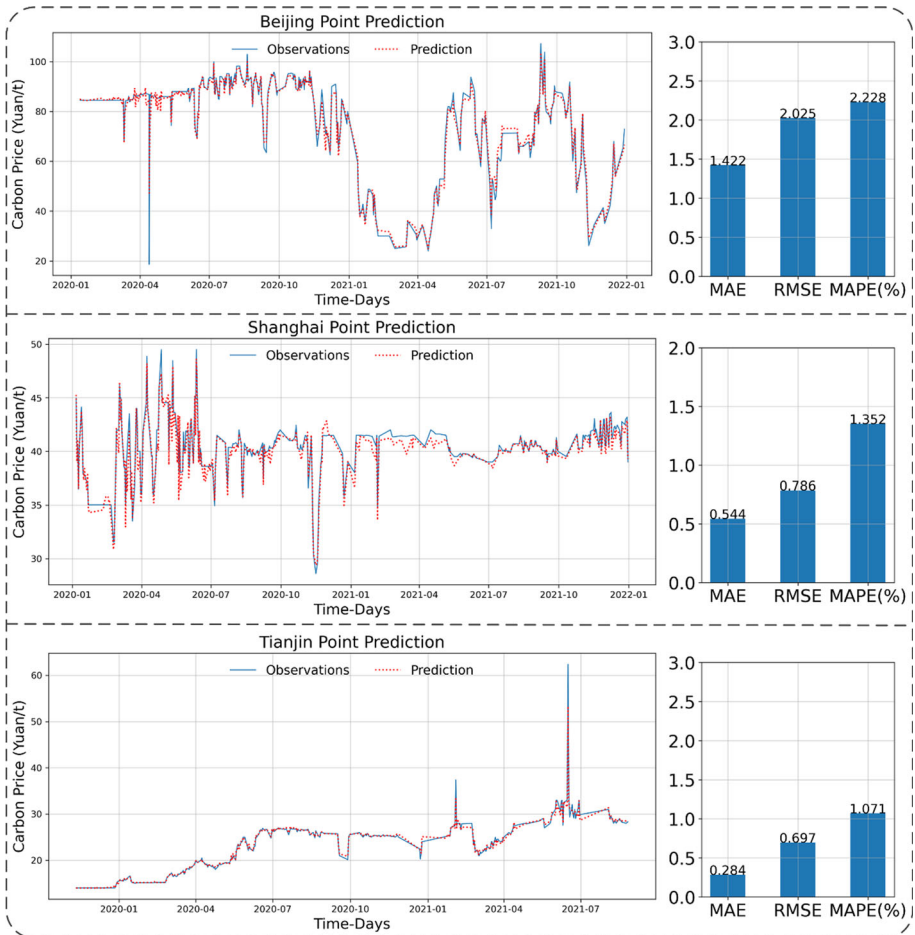


Fig. 8 Point prediction results for Beijing, Shanghai, and Tianjin

compared algorithms are shown in Table 9. DEA and PSO have fewer parameters and fast convergence, but have the disadvantage of easily falling into local optimal points. Therefore, DEA and PSO are difficult to find the optimal hyperparameters, and the prediction performance of their optimized BiLSTM is lower than that of other optimization algorithms. GA and ACO have a strong global optimization-seeking ability and can effectively find suitable hyperparameters, but the convergence speed is slow and prone to stagnation. Therefore, the training time for optimizing hyperparameters using GA and ACO is longer, while the training time of CS-LSTM is reduced by 28.22% and 30.80% compared with GA and ACO, respectively. The CS algorithm has local search and global search capability and converges faster. Meanwhile, CS is a global search using Lévy flight, which has infinite mean and variance, which can effectively ensure that the CS algorithm can discover the global optimal hyperparameters more effectively. The experimental results prove that CS-BiLSTM outperforms the other four compared algorithms in both the training time and prediction performance of the optimized model.

Table 8 Point prediction results under different observation windows

| Dataset | Observation window | MAE | RMSE | MAPE (%) |
|-----------|--------------------|---------------|---------------|---------------|
| Beijing | 3 | 2.2222 | 2.5423 | 3.6561 |
| | 4 | 1.4228 | 2.0256 | 2.2289 |
| | 5 | 1.7795 | 2.2124 | 2.4325 |
| | 6 | 2.0319 | 2.4451 | 3.4715 |
| Tianjin | 3 | 0.5411 | 0.9213 | 1.5431 |
| | 4 | 0.2844 | 0.6974 | 1.0705 |
| | 5 | 0.3596 | 0.7559 | 1.2556 |
| Shanghai | 3 | 0.5741 | 0.8456 | 1.4555 |
| | 4 | 0.5446 | 0.7865 | 1.3528 |
| | 5 | 0.6212 | 0.9321 | 1.5787 |
| Shenzhen | 3 | 2.1044 | 2.3121 | 5.7123 |
| | 4 | 1.4873 | 1.6678 | 4.6977 |
| | 5 | 1.4930 | 1.6824 | 4.7323 |
| Guangdong | 3 | 1.6511 | 1.8535 | 5.0431 |
| | 4 | 2.2341 | 2.8892 | 5.4205 |
| | 4 | 1.2804 | 1.7352 | 3.7756 |
| Hubei | 3 | 1.3251 | 1.7894 | 3.8125 |
| | 4 | 0.8448 | 1.2489 | 2.6084 |
| | 5 | 2.0451 | 2.4529 | 4.6521 |
| Chongqing | 3 | 1.0341 | 1.5672 | 2.8011 |
| | 4 | 1.1170 | 1.4643 | 4.0875 |
| | 4 | 0.9927 | 1.3201 | 3.8505 |
| | 5 | 1.4427 | 1.7854 | 4.3764 |
| | 6 | 1.8745 | 2.2316 | 5.4215 |

The bold symbol represents the observation window with the best prediction performance in different data sets

5.5.2 Comparison of benchmark models

To further validate the superiority and stability of the hybrid framework proposed in this paper, eight benchmark models were developed in this study. Four of the benchmark models were derived from the hybrid framework constructed in this study, and the other four benchmark models were derived from excellent research papers in the same research area. The specific evaluation results of the compared models are shown in Tables 10 and 11.

As shown in Table 10, the prediction performance of the hybrid framework proposed in this paper outperformed the other four benchmark models in all metrics. By comparing the prediction performance of these four benchmark models, it can be found that the decomposition technique and the nonlinear integration algorithm can effectively improve the prediction

Table 9 Comparison of hyperparametric optimization methods

| Optimization method | Population number | Number of iterations | Average training time per BiLSTM (minutes) | MAE | RMSE | MAPE |
|---------------------|-------------------|----------------------|--|-------|-------|-------|
| DEA-BiLSTM | 15 | 50 | 3.97 | 10.13 | 10.94 | 13.84 |
| GA-BiLSTM | 15 | 50 | 4.57 | 8.02 | 9.14 | 11.67 |
| PSO-BiLSTM | 15 | 50 | 3.89 | 10.73 | 11.45 | 14.72 |
| ACO-BiLSTM | 15 | 50 | 4.74 | 8.42 | 9.65 | 12.46 |
| CS-BiLSTM | 15 | 50 | 3.28 | 7.24 | 8.53 | 10.42 |

stability and accuracy of the models on different data sets. Taking the Beijing dataset as an example, the MAPE of the VMD-SAE-BiLSTM- RF model is reduced by about 55.34% and 10.99% compared with the BP model and the VMD-SAE-BiLSTM model, respectively. In addition, the VMD-SAE-BiLSTM-BiLSTM model has 17.45%, 9.62%, and 8.52% reduction in MAE, RMSE, and MAPE, respectively, compared to the VMD-SAE-BiLSTM-RF model, which indicates the higher integration performance of BiLSTM for nonlinear integration.

As shown in Table 11, among these four benchmark models, although the random forest model can effectively predict carbon prices in datasets such as Shanghai and Guangdong, the stability of its prediction is not good. The prediction accuracy of the random forest model in datasets such as Shenzhen and Hubei is poor, and there is still room for improvement. Although EMD-PSO-LSSVR outperformed random forest on all datasets, its predictive stability was also poor. CEEDMAN-Sample entropy (SEn)-LSTM-RF has a more advanced decomposition technique, LSTM prediction model, and nonlinear integrated model, which makes its comprehensive performance better than EMD-PSO-LSSVR in different datasets. In the EMD-VMD-PACF-GA-BP, the model introduces a double decomposition technique and an intelligent optimization algorithm, which shows good stability on different datasets. However, the hybrid model proposed in this paper outperforms the previous models on all datasets. The above comparative experiments can prove that the hybrid framework proposed in this paper is an effective tool for carbon price prediction, with good prediction accuracy and stability.

5.6 Interval prediction

In the interval prediction stage, this paper uses a hybrid kernel function with a combination of SE kernel function and RQ kernel function, and carbon price interval prediction is performed using a GPR model based on the hybrid kernel function. Figure 9 shows the predicted results and evaluation indicators of carbon price intervals for Beijing, Shanghai, and Tianjin at the significance level of 0.05. The interval prediction performance of the seven carbon trading pilot data at 0.05, 0.1, and 0.2 significance levels are shown in Table 12. At the significance level of 0.05, the PIAW is 3.42, 2.2, and 2.12 for Beijing, Shanghai, and Tianjin, respectively, and the PICP is 98.22%, 100%, and 100%, respectively. As can be seen from Fig. 9, almost all of the true values of carbon prices fall in the prediction interval, and only a few fall outside the prediction interval. As shown in Table 12, the Gaussian process regression based on the hybrid kernel maintains good interval prediction performance on different data sets.

Table 10 Four comparison models based on hybrid models

| Dataset | Model | MAE | RMSE | MAPE (%) |
|-----------|-----------------------|---------|---------|----------|
| Beijing | BP | 8.9726 | 10.6100 | 13.0877 |
| | VMD-SAE-BiLSTM | 4.5648 | 5.5950 | 6.5664 |
| | VMD-SAE-BiLSTM-RF | 3.9561 | 4.6891 | 5.8449 |
| | VMD-SAE-BiLSTM-BiLSTM | 3.2357 | 4.2366 | 5.3472 |
| | Proposed model | 1.4228 | 2.0256 | 2.2289 |
| Tianjin | BP | 3.8565 | 5.0626 | 15.0265 |
| | VMD-SAE-BiLSTM | 1.0242 | 1.7564 | 3.9483 |
| | VMD-SAE-BiLSTM-RF | 0.8561 | 1.4246 | 3.2031 |
| | VMD-SAE-BiLSTM-BiLSTM | 0.6781 | 1.1127 | 2.9459 |
| | Proposed model | 0.2844 | 0.6974 | 1.0705 |
| Shanghai | BP | 6.3129 | 6.6509 | 15.3784 |
| | VMD-SAE-BiLSTM | 1.2451 | 1.9867 | 3.5642 |
| | VMD-SAE-BiLSTM-RF | 1.0420 | 1.4563 | 2.9604 |
| | VMD-SAE-BiLSTM-BiLSTM | 0.7899 | 1.1205 | 2.2412 |
| | Proposed model | 0.5446 | 0.7865 | 1.3528 |
| Shenzhen | BP | 11.3482 | 13.1831 | 37.9152 |
| | VMD-SAE-BiLSTM | 7.4439 | 8.5933 | 13.4212 |
| | VMD-SAE-BiLSTM-RF | 4.7639 | 5.4431 | 9.0192 |
| | VMD-SAE-BiLSTM-BiLSTM | 3.5201 | 3.9987 | 7.5481 |
| | Proposed model | 1.4873 | 1.6678 | 4.6977 |
| Guangdong | BP | 6.6119 | 7.5429 | 17.9026 |
| | VMD-SAE-BiLSTM | 3.7609 | 4.3011 | 10.4567 |
| | VMD-SAE-BiLSTM-RF | 2.9005 | 3.8791 | 7.5671 |
| | VMD-SAE-BiLSTM-BiLSTM | 2.1477 | 3.0801 | 6.1044 |
| | Proposed model | 1.2804 | 1.7352 | 3.7756 |
| Hubei | BP | 4.6684 | 5.5948 | 13.4900 |
| | VMD-SAE-BiLSTM | 3.0711 | 3.9551 | 7.5541 |
| | VMD-SAE-BiLSTM-RF | 2.3478 | 3.1125 | 5.2414 |
| | VMD-SAE-BiLSTM-BiLSTM | 1.7831 | 2.5477 | 4.8199 |
| | Proposed model | 0.8448 | 1.2489 | 2.6084 |
| Chongqing | BP | 3.8807 | 4.7501 | 12.4191 |
| | VMD-SAE-BiLSTM | 3.1235 | 3.9022 | 7.8910 |
| | VMD-SAE-BiLSTM-RF | 2.0558 | 2.9052 | 6.5041 |
| | VMD-SAE-BiLSTM-BiLSTM | 1.4599 | 2.4351 | 5.7871 |
| | Proposed model | 0.9927 | 1.3201 | 3.8505 |

Table 11 Four comparative models based on relevant literature

| Dataset | Model | MAE | RMSE | MAPE (%) |
|-----------|---|--------|---------|----------|
| Beijing | Random forest (Yahşi et al., 2019) | 5.7266 | 8.5909 | 9.2349 |
| | CEEDMAN-SEn-LSTM-RF (Wang et al., 2021) | 1.6787 | 2.2081 | 2.5451 |
| | EMD-PSO-LSSVR (Zhu et al., 2017) | 2.0342 | 2.6341 | 3.2315 |
| | EMD-VMD-PACF-GA-BP (Sun & Huang, 2020) | 1.8948 | 2.4712 | 2.6591 |
| | Proposed model | 1.4228 | 2.0256 | 2.2289 |
| Tianjin | Random forest | 1.6077 | 2.5835 | 7.0618 |
| | CEEDMAN-SEn-LSTM-RF | 0.6789 | 1.1552 | 1.5633 |
| | EMD-PSO-LSSVR | 0.4233 | 0.8974 | 1.2457 |
| | EMD-VMD-PACF-GA-BP | 0.5239 | 0.9865 | 1.3469 |
| | Proposed model | 0.2844 | 0.6974 | 1.0705 |
| Shanghai | Random forest | 8.8958 | 10.6867 | 4.9639 |
| | CEEDMAN-SEn-LSTM-RF | 0.9103 | 1.1095 | 1.7707 |
| | EMD-PSO-LSSVR | 1.5231 | 1.9976 | 2.5647 |
| | EMD-VMD-PACF-GA-BP | 1.0119 | 1.2239 | 1.8702 |
| | Proposed model | 0.5446 | 0.7865 | 1.3528 |
| Shenzhen | Random forest | 8.8958 | 10.6867 | 54.2009 |
| | CEEDMAN-SEn-LSTM-RF | 2.0951 | 2.9856 | 9.4266 |
| | EMD-PSO-LSSVR | 5.7605 | 6.7731 | 27.5530 |
| | EMD-VMD-PACF-GA-BP | 1.7892 | 2.2358 | 6.4759 |
| | Proposed model | 1.4873 | 1.6678 | 4.6977 |
| Guangdong | Random forest | 2.3733 | 3.5875 | 6.9866 |
| | CEEDMAN-SEn-LSTM-RF | 1.6540 | 2.5672 | 5.0461 |
| | EMD-PSO-LSSVR | 2.1246 | 3.3211 | 6.5954 |
| | EMD-VMD-PACF-GA-BP | 1.312 | 2.3587 | 4.7853 |
| | Proposed model | 1.2804 | 1.7352 | 3.7756 |
| Hubei | Random forest | 3.4285 | 3.8541 | 10.4070 |
| | CEEDMAN-SEn-LSTM-RF | 1.2245 | 1.9874 | 3.5641 |
| | EMD-PSO-LSSVR | 2.0941 | 2.4639 | 5.4321 |
| | EMD-VMD-PACF-GA-BP | 1.1208 | 1.9040 | 3.4567 |
| | Proposed model | 0.8448 | 1.2489 | 2.6084 |
| Chongqing | Random forest | 2.1455 | 2.9197 | 7.9991 |
| | CEEDMAN-SEn-LSTM-RF | 1.2852 | 1.9872 | 4.6327 |
| | EMD-PSO-LSSVR | 1.8977 | 2.2001 | 5.6601 |
| | EMD-VMD-PACF-GA-BP | 1.1042 | 1.7692 | 4.3321 |
| | Proposed model | 0.9927 | 1.3201 | 3.8505 |

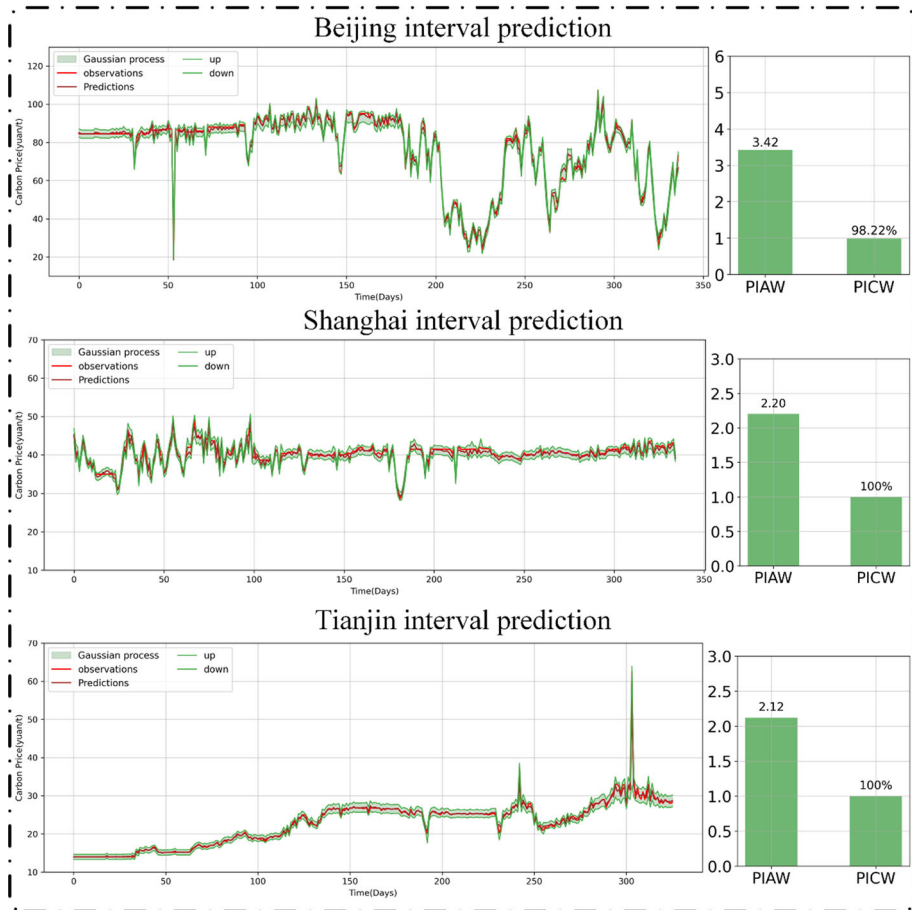


Fig. 9 Interval prediction results for Beijing, Shanghai, and Tianjin

In order to further verify the superiority and prediction stability of the hybrid kernel function-based GPR, the effects of six different kernel functions on the interval prediction performance are compared separately. The interval forecasts of carbon prices for the seven carbon trading pilots were performed using SE, RQ, and Matern as single kernel functions and three hybrid kernel functions, respectively. The predictive performance of GPR based on different kernel functions at the 0.05 level of significance for the seven pilots is shown in Table 13. From Table 13, it can be found that although the GPR based on a single kernel function can effectively perform interval prediction, it is difficult to maintain excellent prediction results on different data sets. Hybrid kernel functions can enhance the generalization ability of GPR compared to single kernel functions, making it more effective in dealing with nonlinear nonsmooth carbon price series. Furthermore, among these three hybrid kernel functions, the hybrid kernel function based on SE and RQ used in this paper has stronger prediction performance than the other kernel functions on different data sets and has stronger robustness for carbon price interval prediction.

Table 12 Interval forecast performance of seven carbon trading pilots

| Dataset | Evaluation metrics | Significance level | | |
|-----------|--------------------|--------------------|--------|--------|
| | | 0.05 | 0.1 | 0.2 |
| Beijing | PIAW | 3.42 | 2.87 | 2.24 |
| | PICW | 98.22% | 92.58% | 83.81% |
| Tianjin | PIAW | 2.12 | 1.78 | 1.39 |
| | PICW | 100.00% | 96.01% | 91.33% |
| Shanghai | PIAW | 2.20 | 1.85 | 1.45 |
| | PICW | 100.00% | 90.59% | 82.23% |
| Shenzhen | PIAW | 6.17 | 5.18 | 4.05 |
| | PICW | 100.00% | 98.05% | 95.27% |
| Guangdong | PIAW | 4.54 | 3.82 | 2.99 |
| | PICW | 98.91% | 91.10% | 80.47% |
| Hubei | PIAW | 2.73 | 2.30 | 1.80 |
| | PICW | 99.44% | 90.35% | 81.48% |
| Chongqing | PIAW | 2.93 | 2.46 | 1.92 |
| | PICW | 99.40% | 89.58% | 80.10% |

Table 13 Prediction performance of interval prediction models with different kernel functions

| Dataset | Evaluation Metrics | Single kernel functions | | | Combined kernel functions | | |
|-----------|--------------------|-------------------------|--------|---------|---------------------------|-------------|-------------|
| | | SE | RQ | Matern | SE + RQ | Matern + RQ | Matern + SE |
| Beijing | PIAW | 5.61 | 5.14 | 6.99 | 3.42 | 3.78 | 3.67 |
| | PICW | 75.66% | 78.79% | 82.78% | 98.22% | 94.33% | 83.29% |
| Tianjin | PIAW | 2.12 | 2.04 | 2.61 | 2.12 | 2.24 | 2.10 |
| | PICW | 97.85% | 96.55 | 98.46% | 100.00% | 100.00% | 97.85% |
| Shanghai | PIAW | 2.21 | 2.37 | 2.57 | 2.20 | 2.22 | 2.48 |
| | PICW | 85.97% | 85.97% | 88.35% | 100.00% | 89.97% | 98.50% |
| Shenzhen | PIAW | 6.18 | 2.44 | 8.87 | 6.17 | 6.18 | 6.18 |
| | PICW | 99.44% | 45.83% | 100.00% | 100.00% | 99.82% | 100.00% |
| Guangdong | PIAW | 4.56 | 4.56 | 5.83 | 4.54 | 4.56 | 4.91 |
| | PICW | 80.11% | 80.10% | 87.47% | 98.91% | 89.11% | 98.63% |
| Hubei | PIAW | 2.75 | 8.95 | 3.28 | 2.73 | 3.13 | 2.92 |
| | PICW | 65.84% | 98.34% | 73.82% | 99.44% | 96.96% | 98.62% |
| Chongqing | PIAW | 2.934 | 5.30 | 3.59 | 2.93 | 3.12 | 3.04 |
| | PICW | 71.04% | 89.85% | 79.10% | 99.40% | 96.72% | 98.50% |

6 Conclusion

With the continuous in-depth exploration of China's carbon emission trading pilot, valuable experience has been accumulated for the construction of China's carbon market. However, key issues such as carbon market trading platform setup, rules for quota allocation and use, and the development of carbon financial systems derived from the carbon trading market are still controversial and need further research. For investors, an accurate carbon price forecast can provide some guidance for their investment decisions. For policymakers, accurate prediction of carbon price can enable them to better analyze the changing trends and problems of the carbon trading market, make ex-ante policy impact assessments and formulate more reasonable policies for this purpose.

In this paper, an intelligent optimized nonlinear integrated carbon price forecasting framework based on multi-factor and two-stage dimension feature extraction is proposed, containing VMD algorithm, stacked autoencoder, random forest, bidirectional long and short term memory artificial neural network, cuckoo search algorithm, and Gaussian process regression. To verify the effectiveness of the proposed hybrid framework, four hyperparametric optimization algorithms are tested, and eight benchmark models are developed to evaluate the proposed hybrid framework systematically and comprehensively. The prediction results show that the proposed framework outperforms all benchmark models, and the following conclusions are drawn: (1) The introduction of potential influencing factors for forecasting on the basis of historical carbon price data can improve the forecasting performance and robustness of the model, and effectively enhance the forecasting of carbon price fluctuation trends. (2) The two-stage feature dimension reduction method can extract the effective information in the data and can well solve the problems of overfitting and error accumulation easily caused by the introduction of exogenous variables. (3) The hyperparameter optimization method based on the cuckoo search algorithm can effectively improve the prediction performance of the model and make it more robust. (4) Interval forecasts provide more information on the fluctuations of carbon prices than point forecasts. (5) The hybrid framework has good practical significance and practical application value.

Although the hybrid forecasting framework proposed in this study has good carbon price forecasting performance and forecasting accuracy, there are many directions that can be improved. In the future, novel and effective individual forecasting models can be more reasonably selected for forecasting. Unstructured data based on text and images may contain factors that are difficult to quantify, such as investor sentiment, and their role in carbon price volatility could be the subject of further research. Further, intelligent forecasting systems and carbon financial trading decision systems can be developed for the carbon market, providing a novel forecasting tool for governments to specify sound policies and investors with appropriate trading strategies to achieve returns on their investments.

Acknowledgements This research was supported by the National Natural Science Foundation of China (Grant No. 71971122 and 71501101) and NUIST Students' Platform for Innovation and Entrepreneurship Training Program.

Declarations

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

References

- Abedin, M. Z., Guotai, C., & Colombage, S. (2018). Credit default prediction using a support vector machine and a probabilistic neural network. *Journal of Credit Risk*, *14*, 1–27. <https://doi.org/10.21314/JCR.2017.233>
- Abedin, M. Z., Moon, M. H., Hassan, M. K., & Hajek, P. (2021). Deep learning-based exchange rate prediction during the COVID-19 pandemic. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-021-04420-6>
- Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial Markets and Portfolio Management*, *25*, 239. <https://doi.org/10.1007/s11408-011-0165-y>
- Bremnes, J. B. (2004). Probabilistic wind power forecasts using local quantile regression. *Wind Energy*, *7*, 47–54. <https://doi.org/10.1002/we.107>
- Byun, S. J., & Cho, H. (2013). Forecasting carbon futures volatility using GARCH models with energy volatilities. *Energy Economics*, *40*, 207–221. <https://doi.org/10.1016/j.eneco.2013.06.017>
- Çanaköglu, E., Adıyke, E., & Ağralı, S. (2018). Modeling of carbon credit prices using regime switching approach. *Journal of Renewable and Sustainable Energy*, *10*, 035901. <https://doi.org/10.1063/1.4996653>
- Chai, S., Zhang, Z., & Zhang, Z. (2021). Carbon price prediction for China's ETS pilots using variational mode decomposition and optimized extreme learning machine. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-021-04392-7>
- Chen, P., Chang, L., & Chang, F. (2013). Reinforced recurrent neural networks for multi-step-ahead flood forecasts. *Journal of Hydrology*, *497*, 71–79. <https://doi.org/10.1016/j.jhydrol.2013.05.038>
- Chi, G., Uddin, M. S., Abedin, M. Z., & Yuan, K. (2019). Hybrid model for credit risk prediction: An application of neural network approaches. *International Journal on Artificial Intelligence Tools*, *28*, 1950017. <https://doi.org/10.1142/S0218213019500179>
- Dragomiretskiy, K., & Zosso, D. (2014). Variational mode decomposition. *IEEE Transactions on Signal Processing*, *62*, 531–544. <https://doi.org/10.1109/TSP.2013.2288675>
- Dutta, A., Bouri, E., & Noor, M. H. (2018). Return and volatility linkages between CO2 emission and clean energy stock prices. *Energy*, *164*, 803–810. <https://doi.org/10.1016/j.energy.2018.09.055>
- Fan, X., Li, S., & Tian, L. (2015). Chaotic characteristic identification for carbon price and an multi-layer perceptron network prediction model. *Expert Systems with Applications*, *42*, 3945–3952. <https://doi.org/10.1016/j.eswa.2014.12.047>
- Farouq, I. S., Umar Sambo, N., Ahmad, A. U., Jakada, A. H., & Danmaraya, I. A. (2021). Does financial globalization uncertainty affect CO2 emissions? Empirical evidence from some selected SSA countries. *Quantitative Finance and Economics*, *5*, 247–263. <https://doi.org/10.3934/QFE.2021011>
- García-Martos, C., Rodríguez, J., & Sánchez, M. J. (2013). Modelling and forecasting fossil fuels, CO2 and electricity prices and their volatilities. *Sustainable Development of Energy, Water and Environment Systems*, *101*, 363–375. <https://doi.org/10.1016/j.apenergy.2012.03.046>
- Guo, W., Liu, Q., Luo, Z., & Tse, Y. (2022). Forecasts for international financial series with VMD algorithms. *Journal of Asian Economics*, *80*, 101458. <https://doi.org/10.1016/j.asieco.2022.101458>
- Hao, Y., & Tian, C. (2020). A hybrid framework for carbon trading price forecasting: The role of multiple influence factor. *Journal of Cleaner Production*, *262*, 120378. <https://doi.org/10.1016/j.jclepro.2020.120378>
- Hintermann, B., Peterson, S., & Rickels, W. (2016). Price and market behavior in phase II of the EU ETS: A review of the literature. *Review of Environmental Economics and Policy*, *10*, 108–128. <https://doi.org/10.1093/reep/rev015>
- Huang, Y., Shen, L., & Liu, H. (2019). Grey relational analysis, principal component analysis and forecasting of carbon emissions based on long short-term memory in China. *Journal of Cleaner Production*, *209*, 415–423. <https://doi.org/10.1016/j.jclepro.2018.10.128>
- Jiang, M., Jia, L., Chen, Z., & Chen, W. (2022). The two-stage machine learning ensemble models for stock price prediction by combining mode decomposition, extreme learning machine and improved harmony search algorithm. *Annals of Operations Research*, *309*, 553–585. <https://doi.org/10.1007/s10479-020-03690-w>
- Jin, H., Shi, L., Chen, X., Qian, B., Yang, B., & Jin, H. (2021). Probabilistic wind power forecasting using selective ensemble of finite mixture Gaussian process regression models. *Renewable Energy*, *174*, 1–18. <https://doi.org/10.1016/j.renene.2021.04.028>
- Khodayar, M., Saffari, M., Williams, M., & Jalali, S. M. J. (2022). Interval deep learning architecture with rough pattern recognition and fuzzy inference for short-term wind speed forecasting. *Energy*. <https://doi.org/10.1016/j.energy.2022.124143>

- Li, H., Jiang, Z., Shi, Z., Han, Y., Yu, C., & Mi, X. (2022a). Wind-speed prediction model based on variational mode decomposition, temporal convolutional network, and sequential triplet loss. *Sustainable Energy Technologies and Assessments*, 52, 101980. <https://doi.org/10.1016/j.seta.2022.101980>
- Li, J., Hao, J., Feng, Q., Sun, X., & Liu, M. (2021a). Optimal selection of heterogeneous ensemble strategies of time series forecasting with multi-objective programming. *Expert Systems with Applications*, 166, 114091. <https://doi.org/10.1016/j.eswa.2020.114091>
- Li, J., Hao, J., Sun, X., & Feng, Q. (2021b). Forecasting China's sovereign CDS with a decomposition reconstruction strategy. *Applied Soft Computing*, 105, 107291. <https://doi.org/10.1016/j.asoc.2021.107291>
- Li, Y., Chen, J., Dan, H., & Wang, H. (2022b). Probability prediction of pavement surface low temperature in winter based on bayesian structural time series and neural network. *Cold Regions Science and Technology*, 194, 103434. <https://doi.org/10.1016/j.coldregions.2021.103434>
- Liu, J., Wang, P., Chen, H., & Zhu, J. (2022). A combination forecasting model based on hybrid interval multi-scale decomposition: Application to interval-valued carbon price forecasting. *Expert Systems with Applications*, 191, 116267. <https://doi.org/10.1016/j.eswa.2021.116267>
- Maia, A. L. S., & de Carvalho, F. D. A. T. (2011). Holt's exponential smoothing and neural network models for forecasting interval-valued time series. *Special Section 1 Forecasting with Artificial Neural Networks and Computational Intelligence*, 27, 740–759. <https://doi.org/10.1016/j.ijforecast.2010.02.012>
- Meng, B., Zhou, L., Qu, L., & Abedin, M. Z. (2019). Measurement of urban green economy development — an empirical analysis from 31 provinces in China. *Ekoloji*, 28, 2069–2082.
- Momenh, S., & Nourani, V. (2022). Application of a novel technique of the multi-discrete wavelet transforms in hybrid with artificial neural network to forecast the daily and monthly streamflow. *Modeling Earth Systems and Environment*. <https://doi.org/10.1007/s40808-022-01387-6>
- Mori, H., Jiang, W., (2008). An ANN-based risk assessment method for carbon pricing. In *2008 5th international conference on the European electricity market. Presented at the 2008 5th international conference on the european electricity market* (pp. 1–6). <https://doi.org/10.1109/EEM.2008.4579094>
- Peng, C., Tao, Y., Chen, Z., Zhang, Y., & Sun, X. (2022). Multi-source transfer learning guided ensemble LSTM for building multi-load forecasting. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2022.117194>
- Peng, H., & Bai, X. (2019). Gaussian Processes for improving orbit prediction accuracy. *Acta Astronautica*, 161, 44–56. <https://doi.org/10.1016/j.actaastro.2019.05.014>
- Petelin, D., Kocijan, J., (2014). Evolving Gaussian process models for predicting chaotic time-series. In *2014 IEEE conference on evolving and adaptive intelligent systems (EAIS). Presented at the 2014 IEEE conference on evolving and adaptive intelligent systems (EAIS)* (pp. 1–8). <https://doi.org/10.1109/EAIS.2014.6867476>
- Quan, H., Srinivasan, D., & Khosravi, A. (2014). Short-term load and wind power forecasting using neural network-based prediction intervals. *IEEE Transactions on Neural Networks and Learning Systems*, 25, 303–315. <https://doi.org/10.1109/TNNLS.2013.2276053>
- Şaylı, M., & Yılmaz, E. (2017). Anti-periodic solutions for state-dependent impulsive recurrent neural networks with time-varying and continuously distributed delays. *Annals of Operations Research*, 258, 159–185. <https://doi.org/10.1007/s10479-016-2192-6>
- Segnon, M., Lux, T., & Gupta, R. (2017). Modeling and forecasting the volatility of carbon dioxide emission allowance prices: A review and comparison of modern volatility models. *Renewable and Sustainable Energy Reviews*, 69, 692–704. <https://doi.org/10.1016/j.rser.2016.11.060>
- Shajalal, M., Hajek, P., & Abedin, M. Z. (2021). Product backorder prediction using deep neural network on imbalanced data. *International Journal of Production Research*. <https://doi.org/10.1080/00207543.2021.1901153>
- Souza, L. C., Souza, R. M. C. R., Amaral, G. J. A., & Silva Filho, T. M. (2017). A parametrized approach for linear regression of interval data. *Knowledge-Based Systems*, 131, 149–159. <https://doi.org/10.1016/j.knosys.2017.06.012>
- Sun, S., Jin, F., Li, H., & Li, Y. (2021). A new hybrid optimization ensemble learning approach for carbon price forecasting. *Applied Mathematical Modelling*, 97, 182–205. <https://doi.org/10.1016/j.apm.2021.03.020>
- Sun, W., & Huang, C. (2020). A carbon price prediction model based on secondary decomposition algorithm and optimized back propagation neural network. *Journal of Cleaner Production*, 243, 118671. <https://doi.org/10.1016/j.jclepro.2019.118671>
- Sun, W., Zhang, C., & Sun, C. (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 Management*, 9, 605–617. <https://doi.org/10.1080/17583004.2018.1522095>

- Sun, X., Hao, J., & Li, J. (2022). Multi-objective optimization of crude oil-supply portfolio based on interval prediction data. *Annals of Operations Research*, 309, 611–639. <https://doi.org/10.1007/s10479-020-03701-w>
- Wang, J., Sun, X., Cheng, Q., & Cui, Q. (2021). An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting. *Science of the Total Environment*, 762, 143099. <https://doi.org/10.1016/j.scitotenv.2020.143099>
- Wang, J., Xu, W., Zhang, Y., & Dong, J. (2022a). A novel air quality prediction and early warning system based on combined model of optimal feature extraction and intelligent optimization. *Chaos Soliton Fract*, 158, 112098. <https://doi.org/10.1016/j.chaos.2022.112098>
- Wang, K., Wang, J., Zeng, B., & Lu, H. (2022b). An integrated power load point-interval forecasting system based on information entropy and multi-objective optimization. *Applied Energy*, 314, 118938. <https://doi.org/10.1016/j.apenergy.2022.118938>
- Wei, S., Chongchong, Z., & 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 Management*, 9, 605–617. <https://doi.org/10.1080/17583004.2018.1522095>
- Wen, L., Yuan, X. (2020). Forecasting CO₂ emissions in Chinas commercial department, through BP neural network based on random forest and PSO. *Science of The Total Environment*, 718, 137–194. <https://doi.org/10.1016/j.scitotenv.2020.137194>
- Xiong, T., Bao, Y., Hu, Z., & Chiong, R. (2015). Forecasting interval time series using a fully complex-valued RBF neural network with DPSO and PSO algorithms. *Information Sciences*, 305, 77–92. <https://doi.org/10.1016/j.ins.2015.01.029>
- Xu, X., & Ren, W. (2022). A hybrid model of stacked autoencoder and modified particle swarm optimization for multivariate chaotic time series forecasting. *Applied Soft Computing*, 116, 108321. <https://doi.org/10.1016/j.asoc.2021.108321>
- Yahşi, M., Çanakoğlu, E., & Ağralı, S. (2019). Carbon price forecasting models based on big data analytics. *Carbon Management*, 10, 175–187. <https://doi.org/10.1080/17583004.2019.1568138>
- Yang, S., Chen, D., Li, S., & Wang, W. (2020). Carbon price forecasting based on modified ensemble empirical mode decomposition and long short-term memory optimized by improved whale optimization algorithm. *Science of the Total Environment*, 716, 137117. <https://doi.org/10.1016/j.scitotenv.2020.137117>
- Yang, X., & Deb, S. (2014). Cuckoo search: Recent advances and applications. *Neural Computing and Applications*, 24, 169–174. <https://doi.org/10.1007/s00521-013-1367-1>
- Ye, T., Zhao, N., Yang, X., Ouyang, Z., Liu, X., Chen, Q., Hu, K., Yue, W., Qi, J., Li, Z., & Jia, P. (2019). Improved population mapping for China using remotely sensed and points-of-interest data within a random forests model. *Science of the Total Environment*, 658, 936–946. <https://doi.org/10.1016/j.scitotenv.2018.12.276>
- Yildiz, A. R. (2013). Cuckoo search algorithm for the selection of optimal machining parameters in milling operations. *The International Journal of Advanced Manufacturing Technology*, 64, 55–61. <https://doi.org/10.1007/s00170-012-4013-7>
- Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*, 30, 2623–2635. <https://doi.org/10.1016/j.eneco.2008.05.003>
- Zhang, C., Wei, H., Zhao, X., Liu, T., & Zhang, K. (2016). A Gaussian process regression based hybrid approach for short-term wind speed prediction. *Energy Conversion and Management*, 126, 1084–1092. <https://doi.org/10.1016/j.enconman.2016.08.086>
- Zhang, F., & Xia, Y. (2022). Carbon price prediction models based on online news information analytics. *Finance Research Letters*, 46, 102809. <https://doi.org/10.1016/j.frl.2022.102809>
- Zhang, L., Lu, S., Ding, Y., Duan, D., Wang, Y., Wang, P., Yang, L., Fan, H., & Cheng, Y. (2022a). Probability prediction of short-term user-level load based on random forest and kernel density estimation. *ICPE 2021-the International Conference on Power Engineering*, 8, 1130–1138. <https://doi.org/10.1016/j.egy.2022a.02.256>
- Zhang, S., Wang, C., Liao, P., Xiao, L., & Fu, T. (2022b). Wind speed forecasting based on model selection, fuzzy cluster, and multi-objective algorithm and wind energy simulation by Betz's theory. *Expert Systems with Applications*, 193, 116509. <https://doi.org/10.1016/j.eswa.2022.116509>
- Zhang, T., Tang, Z., Wu, J., Du, X., & Chen, K. (2022c). Short term electricity price forecasting using a new hybrid model based on two-layer decomposition technique and ensemble learning. *Electric Power Systems Research*, 205, 107762. <https://doi.org/10.1016/j.eprs.2021.107762>
- Zhao, X., Han, M., Ding, L., & Kang, W. (2018). Usefulness of economic and energy data at different frequencies for carbon price forecasting in the EU ETS. *Applied Energy*, 216, 132–141. <https://doi.org/10.1016/j.apenergy.2018.02.003>

- Zhu, B., Han, D., Wang, P., Wu, Z., Zhang, T., & Wei, Y.-M. (2017). Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. *Applied Energy*, *191*, 521–530. <https://doi.org/10.1016/j.apenergy.2017.01.076>
- Zhu, B., Wang, P., Chevallier, J., & Wei, Y. (2015). Carbon price analysis using empirical mode decomposition. *Computational Economics*, *45*, 195–206. <https://doi.org/10.1007/s10614-013-9417-4>
- Zhu, B., & Wei, Y. (2013). Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. *Omega*, *41*, 517–524. <https://doi.org/10.1016/j.omega.2012.06.005>
- Zhu, B., Ye, S., He, K., Chevallier, J., & Xie, R. (2019a). Measuring the risk of European carbon market: An empirical mode decomposition-based value at risk approach. *Annals of Operations Research*, *281*, 373–395. <https://doi.org/10.1007/s10479-018-2982-0>
- Zhu, J., Wu, P., Chen, H., Liu, J., & Zhou, L. (2019b). Carbon price forecasting with variational mode decomposition and optimal combined model. *Physica a: Statistical Mechanics and Its Applications*, *519*, 140–158. <https://doi.org/10.1016/j.physa.2018.12.017>

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