INTRODUCTION

Global climate change, which was triggered by the excessive emissions of greenhouse gases, has become an obstacle in sustainable development. Among all greenhouse gases, carbon dioxide is the most prominent and is artificially controllable. China ranks as the largest emitter of carbon dioxide worldwide and is a pioneer in economic growth. As the international community is advocating in reducing carbon emissions, Chinese carbon emissions and its corresponding
changes have become the focus of attention, which burdens the Chinese government with tremendous pressure. The EU Emissions Trading Scheme (ETS) is supposed to be an internationally effective means of reducing carbon emissions according to a previous study. Inspired by this, and in order to cope with growing economic development demands and alleviate the adverse effects of climate change, a total of eight carbon trading pilots have been established in China since 2013. This was followed by the diffusion of national emissions trading scheme (ETS), which was introduced on December 19, 2017, and nearly doubled the coverage of the carbon pricing system. The carbon price is well known to be at the heart of establishing a stable and effective ETS. Besides, accurately grasping the trend of price fluctuations in advance may both assist the government in constructing a stable and mature carbon price mechanism and help investors take timely steps in reducing investment risk. In this light, exploring an accurate, stable, and universal carbon price forecasting technique is of prime significance for Chinese ETS.

### 1.1 Overview of studies on carbon price forecasting and other relevant prediction techniques

An increasing number of academic studies exist on carbon price forecasting as well as techniques applied in embracing two types: time series prediction and multi-factor prediction. Multi-factor prediction refers to several related exogenous variables being used in the forecasting process. This method has achieved definite success, though in certain cases, adverse consequences have emerged, owing to two obvious issues. First, the final prediction results are based on the prediction values of other exogenous variables, which may cause accumulations in errors. Second, multicollinearity may be present among the selected variables, inevitably leading to over-fitting. Time series forecasting is independent of exogenous variables and can acquire future trends by using intrinsic characteristics of its historical data. Therefore, time series prediction is both feasible and highly efficient. Moreover, the wide application of time series prediction has demonstrated its validity in forecasting. Accordingly, in this study, multi-factor forecasting was abandoned while time series forecasting was adopted in order to anticipate the carbon price.

The time series forecasting technique is usually data-based and can be categorized into two branches involving statistical methods and machine learning algorithms. Statistical methods previously employed for ETS carbon price forecasting include multiple linear regression, generalized autoregressive conditional heteroskedasticity (GARCH)-type model, heterogeneous autoregressive model of realized volatility (HAR-RV) model, and gray model GM(1,1). Nevertheless, the nonlinear and high volatility of carbon prices may not be dealt with well enough via the statistical method. The advent of machine learning has spawned a considerable number of studies on future trends of the carbon price, in which nonlinear methods are involved, for example, the artificial intelligence neural network (ANN), least squares support vector machine (LSSVM), and other hybrid methods. Yi et al confirmed its effectiveness in analyzing future trends of EUA carbon price by executing a type of feed-forward neural network (FFNN) technique. Tsai and Kuo carried out the radial basis function neural network (RBFNN) with parameters optimized by ant colony optimization so as to forecast the carbon price. Additionally, an empirical study done on carbon price forecasting for four seasons demonstrated that RBFNN is superior to other contrast models. Zhu et al derived a model with a combination of empirical mode decomposition and particle swarm optimization, which improved LSSVM to forecast the EU carbon price. The corresponding results proved that the proposed model was more appropriate compared with conventional models in terms of accuracy and stability. In order to obtain a more accessible method in predicting the carbon price, Zhu proposed an ANN-based multiscale ensemble forecasting model which integrated empirical mode decomposition, genetic algorithms, and artificial neural networks for carbon price forecasting, which resulted in superior forecasting results. Among these methods, the traditional ANN and hybrid models extended from it are extensively used and have yielded quite adequate results in strong fitting capability and fault tolerance for nonlinear and complex time series. However, modeling of the nonlinear carbon price pattern is still underdeveloped as existing techniques are static in nature and unable to handle long-term dependencies.

The recurrent neural network (RNN) is a special neural network structure belonging to deep learning. Unlike traditional ANN, for example, FFNN and RBFNN, it incorporates feedback loops, enabling information flow that forms a closed loop between two adjacent hidden layers, which may subsequently learn features in the long-term and extract hidden information more comprehensively. However, its inability to solve vanishing or exploding gradient problems in traditional RNN could lead to inaccurate results; thus, an alternative novel method called LSTM emerged in order to effectively overcome this issue. As a promising approach, the LSTM model has attained state-of-the-art performance in challenging prediction issues. Zhang et al innovatively adopted the LSTM model to predict wind turbine power, and case studies have proved that its forecasting accuracy could be greatly improved. Moreover, the convergence speed of LSTM is faster compared to that of FFNN, RBFNN, and other approaches. Xia et al presented an LSTM-driven deep learning framework to forecast the remaining useful life of the machine, and incorporating multiple time windows to
further enhance its applicability and effectiveness. Their empirical study verified its superiority by comparing it with individual benchmark models.\textsuperscript{16} Ma et al\textsuperscript{17} combined the Grid concept and LSTM (G-LSTM) for the forecasting of fuel cell degradation. More than that, the latest researches applied LSTM to more hot areas of prediction, for example, electricity price forecasting,\textsuperscript{18} flood forecasting,\textsuperscript{19} wind speed forecasting,\textsuperscript{20} air pollution forecasting,\textsuperscript{21} voltages forecasting,\textsuperscript{22} demand forecasting,\textsuperscript{23} photovoltaic power forecasting,\textsuperscript{24} and so forth.\textsuperscript{25,26} The LSTM was wielded to alleviate the vanished gradient in a multi-layer network architecture. Above all, the prediction results indicated validation of this developed model. However, to our knowledge, few studies have applied this technique in carbon price forecasting. Consequently, this study attempted to extend the LSTM model into carbon price modeling and forecasting and provide new references for accurate carbon price prediction.

Considering the high volatility and intermittence of the researched carbon price series, a single model is insufficient in obtaining satisfactory forecasting results. Inspired by ensemble learning theory, where various models or multi-mode learning extract inherent characteristic information from multiple dimensions resulting in better forecasting quality compared with individual models or algorithms.\textsuperscript{27} Wang et al\textsuperscript{27} ensembled three kinds of machine learning approaches involving the random forest model, long short-term memory model, and Gaussian process regression model to forecast wind gusts. The sufficient accuracy and generalization performance of the established model was then verified by a series of empirical studies.\textsuperscript{28} Yu et al\textsuperscript{28} presented a novel hybrid ensemble model in forecasting monthly biofuel production. In this model, the original time series was decomposed and reconstructed into multi-mode by the empirical mode decomposition method and fine-to-coarse approach, after which the modes were separately predicted and the results of the individual predictions were accumulated to generate the final prediction results. The aforementioned study indicated that the proposed hybrid ensemble forecasting technique competitively predicted biofuel production.\textsuperscript{29} Ribeiro and dos Santos\textsuperscript{30} analyzed and studied the role of different types of ensemble learning methods (Bagging, Boosting, and Stacking) on predicting price series, and finally concluded that ensemble technique showed statistically significant gains, reducing prediction errors to a large extent. In addition, many recent studies\textsuperscript{20,31-33} have analyzed ensemble learning methods and have made great progress in this regard. Thus, ensemble learning based on mode transformation methods was introduced to forecast the carbon price in this study. Model transformation methods used in existing studies have mainly included empirical mode decomposition (EMD),\textsuperscript{3} variational mode decomposition (VMD),\textsuperscript{34} ensemble empirical mode decomposition (EEMD),\textsuperscript{35} and wavelet transform (WT).\textsuperscript{36} However, these methods are not sufficiently advanced. For example, EMD possesses problems in modal aliasing; EEMD overcomes this issue but contains residual noise. The complementary ensemble empirical mode decomposition (CEEMD) is a type of self-adaptive approach used to decompose time series into multi-mode, which are named intrinsic mode functions (IMFs).\textsuperscript{37} As the most up-to-date advancements pertaining to EMD,\textsuperscript{38} CEEMD alleviates the mode mixing issue of the traditional EMD by adding adaptive white noise and implementing averaging calculation while addressing the issue where added white noise in EEMD was not completely neutralized.\textsuperscript{39} Moreover, the decomposition level of CEEMD is only determined by the length of the time series and independent of the set basis function.\textsuperscript{40}

Thus, the CEEMD is preferable in integrating with LSTM in order to establish an ensemble learning framework (ED-LSTM). Additionally, to the best of our knowledge, CEEMD has never been used in carbon price time series forecasting. In light of these aspects, an innovative ensemble-driven LSTM model (ED-LSTM) based on CEEMD was proposed in this paper. The superiority of the ED-LSTM model mainly derives from the two aspects: LSTM has memory units and can acquire serialization features and long-term dependencies; the ensemble learning architecture enables the model to enhance feature representation, information extraction, and the application of long short-term relationships in order for the data to be deeply explored and for their features to be utilized to the greatest potential. The modeling process entailed the following three steps: (a) the CEEMD was introduced to multi-mode feature extraction, transforming a complex single mode into multi-mode; (b) LSTM was implemented on each mode to learn features; (c) the inverse CEEMD calculation was utilized in integrating the results of multi-mode learning.

### 1.2 Overview of empirical research on carbon price of China carbon trading pilot

Gokhan et al\textsuperscript{41} pointed out that under the premise of ensuring forecasting accuracy, generality and stability are extremely important for one model. The generality of a model refers to the breadth of its application. Conducting research over a long period range can measure whether the forecasting model has enough generality to a certain degree. Chinese carbon trading pilot emerged relatively late, hence, studies are limited. Table 1 lists related papers according to the perspectives on the amount of China carbon trading pilots covered as well as the corresponding time horizon. Moreover, the table demonstrates that most research attempts to predict the carbon price of one or several carbon markets and deduce the good applicability of this method to other carbon price series. However, it is not reasonable...
enough as the inherent characteristics of carbon price data vary with changes in different Chinese carbon markets. In past studies, only one paper\textsuperscript{54} predicted all carbon prices of the eight carbon markets, which shows that thus far, very few related studies exist on all Chinese carbon price forecasting. The researched period range on the study,\textsuperscript{54} however, was not long enough, and the sample capacity was limited, which lacked a certain comprehensiveness for

| Study | Number of researched sites | Site covered | Research period range | Total size |
|-------|---------------------------|--------------|-----------------------|------------|
| 42    | 5                         | Shenzhen     | 2013/6/19-2014/9/30  | 233        |
|       |                           | Shanghai     | 2013/12/19-2014/9/30 | 128        |
|       |                           | Beijing      | 2013/11/28-2014/9/30 | 155        |
|       |                           | Guangdong    | 2013/12/19-2014/9/30 | 92         |
|       |                           | Tianjin      | 2013/12/26-2014/9/30 | 157        |
| 43    | 4                         | Shenzhen     | 2013/11/1-2017/6/20  | 851        |
|       |                           | Guangdong    | 2013/12/19-2017/12/29| 1184       |
|       |                           | Hubei        | 2014/4/2-2017/12/29  | 1077       |
|       |                           | Shanghai     | 2013/12/19-2017/12/29| 1184       |
| 44    | 4                         | Hubei        | 2014/4/02-2018/8/14  | 1228       |
|       |                           | Beijing      | 2013/11/28-2018/8/14 | 1356       |
|       |                           | Guangdong    | 2013/12/19-2018/8/14 | 1338       |
|       |                           | Shanghai     | 2013/12/19-2018/8/14 | 1336       |
| 45    | 3                         | Hubei        | 2014/4/2-2019/1/11   | 1078       |
|       |                           | Beijing      | 2013/11/28-2019/1/11 | 1204       |
|       |                           | Shanghai     | 2013/12/19-2019/1/11 | 1186       |
| 46    | 3                         | Shenzhen     | 2016/9/1-2018/9/11   | 496        |
|       |                           | Guangdong    | 2016/9/1-2018/9/11   | 496        |
|       |                           | Hubei        | 2016/9/1-2018/9/11   | 496        |
| 47    | 3                         | Beijing      | 2013/11/28-2017/12/29| 1173       |
|       |                           | Shenzhen     | 2013/11/1-2017/12/29 | 1097       |
|       |                           | Hubei        | 2014/4/2-2017/6/20   | 1071       |
| 36    | 2                         | Hubei        | 2014/4/2-2018/5/31   | 1176       |
|       |                           | Shanghai     | 2013/10/19-2018/5/31 | 1282       |
| 48    | 2                         | Shenzhen     | 2015/1/1-2018/6/30   | 864        |
|       |                           | Hubei        | 2015/1/1-2018/6/30   | 860        |
| 35    | 1                         | Shenzhen     | 2014/1/14-2016/8/24  | 800        |
| 49    | 1                         | Shenzhen     | 2013/11/1-2017/6/30  | 882        |
| 50    | 1                         | Hubei        | 2016/1/4-2018/6/22   | 602        |
| 51    | 1                         | Shenzhen     | 2013/6/19-2014/11/3  | 344        |
| 52    | 1                         | Hubei        | 2014/4/2-2018/12/5   | 1129       |
| 53    | 1                         | Hubei        | 2014/4/3-2018/11/28  | 1131       |
| 54    | 8                         | Beijing      | 2016/8/23-2019/9/20  | 450        |
|       |                           | Shanghai     | 2016/11/18-2019/9/20 | 440        |
|       |                           | Guangdong    | 2014/3/11-2019/9/26  | 1128       |
|       |                           | Shenzhen     | 2018/9/6-2019/9/25   | 225        |
|       |                           | Hubei        | 2016/9/1-2018/10/11  | 504        |
|       |                           | Fujian       | 2017/1/9-2019/9/18   | 450        |
|       |                           | Chongqing    | 2018/6/29-2019/9/27  | 134        |
|       |                           | Tianjin      | 2013/12/27-2016/6/2  | 434        |

The study 8 Seen in Table 2
carbon trading markets in China. In addition, the study proposed two kinds of models to predict carbon prices, one of which could not fully adapt to all data characteristics and obtain good prediction results. Therefore, the proposed model was not sufficiently universal.

In order to strengthen the existing literature, this paper aims to explore a general and stable forecasting model that could precisely anticipate future carbon prices for eight carbon trading pilots in China given the time series data. An ensemble-driven LSTM model (ED-LSTM) based on complementary ensemble empirical mode decomposition (CEEMD) that builds on multiple LSTM models in different modes for ensemble learning was explored to forecast carbon price. Methodologically, CEEMD was initially employed for multi-mode feature extraction, and PACF was executed in each mode so as to select time-lagged factors as input. Subsequently, LSTM was implemented for multi-mode feature learning, and inverse CEEMD computation was ultimately utilized to integrate the results of the forecasted multi-mode. Empirically, all eight-pilot carbon trading markets in China were simultaneously researched, and the proposed model was applied to forecast carbon prices in all markets during a long forecasting period. The range of each carbon price series data essentially covered the daily carbon price from its inception of the transaction to the most recent period. The contributions and findings of this study are concisely summarized as follows:

- This paper focuses on the accuracy of the model as well as its stability and universality.
- This study newly proposes a general carbon price forecasting technique and applies it to accurately and simultaneously forecast all Chinese carbon prices, spanning data over a considerably long period.
- A novel ED-LSTM model is established in order to forecast the carbon price. LSTM is introduced to learn

![FIGURE 1](https://example.com/figure1.png)

**FIGURE 1** A. The simple architecture of FFNN and RBFNN. B. The simple architecture of RNN

![FIGURE 2](https://example.com/figure2.png)

**FIGURE 2** The structure of the LSTM network
long-term characteristics. Moreover, inspired by ensemble learning, an ED-LSTM model based on CEEMD is formed, capable of deeply extracting inherent features of data from various aspects, thus greatly enhancing its predictive performance.

- The explored model is shown to be accurate, stable, and universal, which may serve as a promising tool for Chinese carbon price forecasting.

The rest of the paper is organized as follows. Section 2 detailed describes the methodologies used throughout this paper. In Section 3, we introduce the proposed model and its building process. Section 4 presents eight case studies of China carbon market using the proposed method as well as related models executed for comparison. In this section, the forecasting results of utilized models are also analyzed and compared. In Section 5, we draw some key conclusions.
2 RELATED METHODOLOGIES

2.1 Complementary Ensemble Empirical Mode Decomposition (CEEMD)

The CEEMD, a novel adaptive signal decomposition approach, is developed in 2010.\textsuperscript{55} As illustrated in Equation (1), it is utilized to decompose the original complex time series signal into a limited amount of intrinsic mode function (IMFs) and a residue component.

\[ x(t) = \sum_{i=1}^{n} C_i(t) + R(t) \]  

where \( C_i(t) \) denotes the \( i \)th IMF, along with \( R(t) \) is the residue component that represents the main trend of the original signal.

Complementary ensemble empirical mode decomposition is a promising substitute for EMD and EEMD giving that it not only can effectively solve the problems of decomposition instability, pattern aliasing and endpoint effects in EMD methods but also is an impactful method to suppress the error caused by the artificially added white noise in EEMD, meanwhile, it is computational-efficient. In the process of executing the CEEMD algorithm, two vital parameters are involved. The amplitude \( k \) of white noise is usually set at 0.05-0.5 times. Driven by data characteristics and empirical research, this paper sets \( k \) as 0.4 and the iteration time \( P \) as 100. This setting can not only ensure the effect of decomposition but also ensure the decomposition speed. The succinct calculation procedures of CEEMD are as follows.

1. Two kinds of new signals \( x_p^+(t) \) and \( x_p^-(t) \) are obtained by adding a pair of standard white noise of the same magnitude and 180° phase angle difference to the original signal \( x(t) \). Where \( x_p^+(t) \) represents the summation of the original data and positive noise, while \( x_p^-(t) \) means summation of the original data and the negative one.

\[ \begin{align*}
    x_p^+(t) &= x(t) + \lambda_p(t) \quad p = 1, 2, \ldots, P \\
    x_p^-(t) &= x(t) - \lambda_p(t)
\end{align*} \]  

(2) \( C_{ip}^+ \) and \( C_{ip}^- \) are generated by executing EMD on \( x_p^+(t) \) and \( x_p^-(t) \), respectively.

3. Repeat steps (1) and (2) for \( P \) times until a smooth decomposition signal is acquired. The ensemble pattern of all corresponding IMFs is computed by the following Equation (3).

\[ C_i = \frac{1}{2P} \sum_{p=1}^{P} (C_{ip}^+ + C_{ip}^-) \]  

where \( C_i \) represents the \( i \)th final IMF component derived from CEEMD. Where \( P \) is the iteration time set in this paper.

4. Finally, obtain the post-decomposition recombination sequence as explained in Equation (1).

\[ \text{FIGURE 3} \quad \text{Architecture of the ED-LSTM model for carbon price forecasting} \]
2.2 The partial autocorrelation function (PACF)

Partial autocorrelation function is capable to deal with correlations between the given time series and its lagged values, which have made it widely used in describing structural characteristics of the stochastic process.

For time series \( x_t \), the so-called bias autocorrelation coefficient of lag \( k \) is the correlation metric that \( x_{t-k} \) affects \( x_t \) under the condition without regarding the intermediate \( k - 1 \) variables.

Overall, the \( k \)-order autoregressive model is described as follows:

\[
x_t = \varphi_{k1} x_{t-1} + \varphi_{k2} x_{t-2} + \cdots + \varphi_{kk} x_{t-k} + u_t
\]

where \( \varphi_{kj} \) denotes the \( j \)-th regression coefficient of the \( k \)-th order autoregressive equation, and \( \varphi_{kk} \) is the last coefficient among them.

2.3 Long short-term memory network (LSTM)

Long short-term memory network is a kind of recurrent neural network (RNN). Traditional neural networks (NN), for example, FFNN and RBFNN as shown in Figure 1(A), deems the information as independent ignoring relevancy between the data and considers the only input is the input data at the current time. Significantly distinct from traditional NN, RNN, shown in Figure 1(B), incorporates feedback loops

| Carbon price | Abbreviation | Time Period | All Samples | Train Samples | Test Samples | Min | Max | Mean | Std | SE  |
|--------------|--------------|-------------|-------------|---------------|--------------|-----|-----|------|-----|-----|
| Shenzhen     | SZ           | 2013/6/19-2019/1/7 | 1370 | 1110 | 260 | 19.84 | 122.97 | 45.60 | 15.615 | 0.199 |
| Beijing      | BJ           | 2013/11/28-2019/1/7 | 1439 | 1179 | 260 | 30.00 | 77.00 | 51.61 | 7.672 | 0.631 |
| Guangdong    | GD           | 2014/3/6-2019/1/7 | 1350 | 1090 | 260 | 7.57 | 77.00 | 22.96 | 16.139 | 0.799 |
| Shanghai     | SH           | 2013/12/19-2019/1/7 | 1432 | 1152 | 280 | 4.20 | 48.00 | 28.26 | 11.809 | 0.293 |
| Tianjin      | TJ           | 2013/12/26-2019/1/7 | 1419 | 1159 | 260 | 7.00 | 50.10 | 19.31 | 8.110 | 0.067 |
| Hubei        | HB           | 2014/4/2-2019/3/12 | 1363 | 1103 | 260 | 10.07 | 35.13 | 21.17 | 4.987 | 0.735 |
| Chongqing    | CQ           | 2014/6/19-2019/1/7 | 1247 | 1027 | 220 | 1.00 | 47.52 | 17.83 | 11.487 | 0.044 |
| Fujian       | FJ           | 2017/11-2019/1/7 | 486 | 406 | 80 | 12.93 | 42.28 | 25.34 | 7.133 |
that allow information flow forming a closed loop between two adjacent hidden layers, constructing a self-joining hidden layer. Theoretically, the structure can make hidden units share parameters across time indexes and long sequences memory is thus built to recognize and predict. In practice, however, the issue of vanishing gradient prohibits its extensive application in complex long-term tasks.

Therefore, LSTM proposed by Hochreiter and Schmidhuber is gradually developed into an available method. It can gain serialized features and long-term dependencies. Meanwhile, it can eliminate the gradient vanishing issue exiting in the traditional RNN algorithm by introducing the memory cell and self-connected gate mechanism in the hidden units. Among these core

### TABLE 3

| City | IMF1 | IMF2 | IMF3 | IMF4 |
|------|------|------|------|------|
| SZ   | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}$ |
| BJ   | $x_{t-1}, x_{t-2}, x_{t-3}$ | $x_{t-1}, x_{t-2}, x_{t-3}$ | $x_{t-1}, x_{t-2}, x_{t-3}$ | $x_{t-1}, x_{t-2}, x_{t-3}$ |
| GD   | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}$ |
| SH   | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ |
| TJ   | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ |
| HB   | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ |
| CQ   | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ | $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$ |

**TABLE 4** Parameter setting of the ED-LSTM technique

| Building process | Parameter setting |
|------------------|-------------------|
| CEEMD decomposition | Amplitude of white noise $(k) = 0.4$; Iteration time $(P) = 100$ |
| LSTM prediction | Input node is time-lagged factors which determined by PACF technique (seen in Table 3), and output node are corresponding decomposed values of original series; 2 layers LSTM architecture (first layer node depends on the amount of time-lagged factors and second layer node is 10); Batch-size = 20; Time-step = 20; Learning rate = 0.0006; Training epoch = 500 |
| Inverse CEEMD application | Input node $= (\hat{M}(t), \hat{R}(t))$; Output node $= \hat{Y}(t)$ |
improvements, memory cells are exploited for historical information recording. Gate structure does not provide information, in contrast, it is essentially a multi-level feature selection method to remove or add messages to the cell state. Figure 2 depicts the structure of LSTM memory units. Four main operational phases are described in conjunction with Figure 2.

I. Forgetting phase. Forget the input from the previous node selectively. That is, the forgetting gate controls the retention or forgetting of information from the last cell state. The performance of the forget gate is shown in Equation 5.

\[ \Gamma_f = \sigma(W_f[a^{(t-1)}, x^{(t)}]+b_f) \] (5)

II. Selecting and memorizing phase. Determine what new information will be reserved in the current cell state. The sigmoid layer is the input gate layer determining what value will be updated; the tanh( ) layer creates a new candidate value vector \( \tilde{C}^{(t)} \), and adds it to the current state.

\[ \tilde{C}^{(t)} = \tanh(W_c[a^{(t-1)}, x^{(t)}]+b_c) \]

\[ \Gamma_i = \sigma(W_i[a^{(t-1)}, x^{(t)}]+b_i) \] (6)

III. Updating phase. Update the cell state from \( C^{(t-1)} \) to \( C^{(t)} \) by the following Equation 7. The cell state \( C^{(t-1)} \) multiply by \( \Gamma_f \) to forget the information needed to forget and plus current input cell state \( \tilde{C}^{(t)} \) multiplied by the

![Table 5: Parameter setting of related benchmark models](image)

**TABLE 5** Parameter setting of related benchmark models

| Model  | Parameter setting                                                                 |
|--------|----------------------------------------------------------------------------------|
| LSTM   | Exactly the same as LSTM prediction process shown in Table 4                     |
| FFNN   | Iteration times = 100; Learning rate = 0.1; epoch hidden layer node = 5; Goal = 0.00004 |
| RBFNN  | Gotten by network self-training                                                  |

**Figure 6** Forecasting results between different models for Shenzhen carbon price
input gate by element can acquire the new cell state $C^{(i)}$.

\[ C^{(i)} = \Gamma_i \ast \hat{C}^{(i)} + \Gamma_f \ast C^{(i-1)} \]  

(7)

IV. Outputting phase. The output is obtained according to the cell state. First, the sigmoid layer decides which state of the cell will be the output. Then, the cell state is coped by the $\tanh(\cdot)$ layer and multiplied by the above result to finally determine the output information.

\[ \Gamma_o = \sigma(W_o[a^{(i-1)}, x^{(i)}] + b_o) \]
\[ a^{(i)} = \Gamma_o \ast \tanh(C^{(i)}) \]  

(8)

where $\Gamma_f, \Gamma_i, \Gamma_o$ denotes the forget gate, the input gate, and the output gate, respectively, along with $a^{(i)}$ is the implicit vector of all useful information stored at time $t$ and before, $C^{(i)}$ represents cell state vector, $W_f, W_i, W_o,$ and $W_o$ denote the weight matrices of the forget gate, the input gate, the current input cell state, and the output gate, respectively, as well $b_f, b_i, b_o, \text{ and } b_o$ are corresponding offset vectors. $\sigma$ is the sigmoid function and $\tanh(\cdot)$ is the hyperbolic tangent function.

3 | THE PROPOSED MODEL AND ITS BUILDING PROCESS

3.1 | Ensemble-driven LSTM technique based on mode decomposition CEEMD (ED-LSTM)

The core concept of the ensemble learning framework is integrating one or more constituent algorithms or multi-mode learning to obtain improved performance. Given the non-linear complexity and high volatility nature of carbon price series in single mode, along with its vulnerability to external factors, the ensemble-driven LSTM model based on mode decomposition CEEMD was employed to improve the performance of LSTM.

According to the aforementioned description of CEEMD process, its proposed procedure of application in this paper was as follows:

**FIGURE 7** Forecasting results between different models for Beijing carbon price
1. Standard white noise was added into the original carbon price time series, resulting in multi-mode frequencies being separated more readily, facilitating the addressing of mode mixing related issues.

2. Different IMFs were identified that originated from the raw carbon price series.

3. The ensemble calculation in the corresponding decomposed IMFs and $R$ were realized as the ultimate result:

$$Y = \sum_{i=1}^{m} M_i + R$$  \hspace{1cm} (9)

where $Y$ represents the original carbon price series, which can subsequently be decomposed into $i$ IMFs (marked as $M$) and a residue mode $R$. After CEEMD, PACF was applied to extract latent input information for each IMF and $R$. Then, LSTM was performed in each mode. Specifically, $(m + 1)$ LSTM models were applied to learn features of different modes, respectively. The modeling architecture may be presented as:

$$\hat{M}_i(t) = LSTM_i(M(t-k)) \quad i \in (1, m)$$

$$\hat{R}(t) = LSTM_{m+1}(R(t-k))$$  \hspace{1cm} (10)

where $M(t-k)$ denotes the selected input factors of each IMF mode, exerting influence on $M$ under the given confidence interval. $k$ refers to the time-lagged feature of the $k$-order. Analogously, $R(t-k)$ are the input variables of R mode, where $\hat{M}_i(t)$ refers to the prediction results of each IMF and $\hat{R}(t)$ is the prediction results of the R mode.

Finally, the inverse CEEMD was used to restructure the forecasted carbon price value $\hat{Y}(t)$ by means of Equation 11.

$$\hat{Y}(t) = \sum_{i=1}^{m} \hat{M}_i(t) + \hat{R}(t)$$  \hspace{1cm} (11)

### 3.2 Building process of the proposed model

Essentially, by adhering to the technical route “mode decomposition, separate modeling, and ensemble learning”, a novel ED-LSTM model was explored to forecast the carbon price in Chinese carbon markets. The building process is presented in Figure 3, and the detailed summarization is given below.

**Step 1:** The CEEMD technique was initially implemented for mode transformation to decompose the original

![Figure 8](image-url)
complicated mode into a set of simple modes, which was conducive in acquiring ideal prediction results. Through this approach, m IMFs as well as a residue mode $R$ were generated.

Step 2: PACF was performed on each decomposed mode to mine suitable input features for the follow-up forecasting procedure.

Step 3: $(m + 1)$ LSTM models were constructed according to Equation (10) in order to acquire their forecasting values.

Step 4: The inverse CEEMD was applied to gain the predicted carbon price values according to Equation (11). Ultimate predicting results can be obtained by aggregating forecasting values of all IMFs and $R$.

4 | EMPIRICAL ANALYSIS

4.1 | China carbon markets and datasets description

Since the establishment of the first carbon trading market pilot —Shenzhen carbon market on June 19, 2013, the Chinese government has successively established seven regional carbon trading markets in Beijing, Guangdong, Shanghai, Tianjin, Hubei, Chongqing, and Fujian. Figure 4 displays the carbon trading volume (unit: 10 000 tons) from their inception to June 1, 2019. Among them, the larger steam drum represents the more trading volume. The total volume is up to 17 122 which is nearly a quarter of the total national carbon emissions in 2016, showing the prosperity of the carbon market. Besides, the carbon price is the hotpot in the course of carbon trading. Therefore, it is necessary to comprehensively predict the Chinese carbon prices of all carbon markets due to the differences in their economic development, geographical location, and policy regulations.

The target of this research is to apply an innovative approach to forecasting carbon price among all eight Chinese carbon markets. The original daily carbon price series are obtained from the China carbon trading website (http://www.tanjiayi.com/). The concrete values are shown in Appendix S1. These acquired time series are drawn in Figure 4, where three parallel lines from top to bottom respectively represent the maximum, average, and minimum values of each carbon price series, meanwhile, high volatility, nonlinear, uncertainty, and complexity are clearly
seen. More detailed statistical information is illustrated in Table 2, including the minimum value (Min), the maximum value (Max), the mean value (Mean), standard deviation (Std) and sample entropy (SE). Among them, Std is the standard for measuring the degree of dispersion of data distribution and SE denotes the complexity of a time series. The smaller SE implies the less complex of the given series and vice versa. Each series was branched into a training set and a test set in case studies where approximately 80% of the data were used to train the model and the rest were employed as the observation. Figure 4 shows the curve of each carbon price series containing the daily carbon price for each carbon market for a long period of time since its inception, along with the left and right sides of the vertical red line represent the training subset and test subset. Details are presented in Table 2.

4.2 Multi-mode feature extraction and input selection

The proposed mode transformation technique was employed to decompose the original carbon price series into 8 IMFs and one residual R. Shenzhen carbon price series decomposed by CEEMD was taken as an example to illustrate. These modes in Figure 5(A) are shown in order according to the frequency from high to low. The decomposition results of other carbon price series used in this paper are similar.

Subsequently, PACF was utilized to extract input features that have high relevance to each decomposed mode, regarding the short time interval and the intrinsic relationship of selected data. Set $x_i$ as output and once the PACF result at $k$ lag is out of the 90% confidence interval, $x_{i-k}$ is mined as an input feature. Similarly, the Figure 5(B) takes Shenzhen carbon price as an example to display input variables of each mode. Table 3 lists the input variables for each mode in carbon price series of eight different carbon markets.

And thus, nine LSTM models were built and trained to learn the features of each mode respectively.

4.3 Parameter setting

In this paper, the ED-LSTM model was explored for carbon price forecasting in eight Chinese carbon markets. Its
constructing and modeling procedures were based on the parameter setting in Table 4. The parameter setting of the LSTM technique is detailed illustrated here. Comprehensively consider performance and computational cost, the structure of LSTM with two layers was chosen in this paper. As for the number of the nodes in the second layer, if the amount of hidden layer node is too small, the network cannot have the necessary learning ability and information processing ability. On the contrary, if too many, it will not only greatly increase the complexity of the network structure rendering the network is more likely to fall into a local minimum during the learning process, but also make the learning speed of the network become very slow. Therefore, choosing a moderate number of hidden layer nodes is necessary. We set it as 10 in this paper. Since the total sample size of each empirical study in this paper is relatively large, in order to improve the computational efficiency, the Batch-size and Time-step are set relatively larger, and both set as 20. For the parameters of Learning rate and Training epoch, they were set as 0.0006 and 500, respectively based on the previous trial and error research. As for the inverse CEEMD arithmetic, the input node \( \hat{M}(t) \) and \( \hat{R}(t) \) denote the prediction results of each IMF and the \( R \) mode, respectively, meanwhile, the output node \( \hat{Y}(t) \) is the finally prediction results.

For the purpose of examining the performance of the proposed ED-LSTM, the traditional NN including FFNN and RBFNN, as well LSTM were respectively executed as benchmark models for carbon price forecasting. The detailed parameter setting of each contrast model is listed in Table 5. The hidden layer node of the FFNN model was set as 5, which is due to the hidden layer node is generally set to 75% of the number of input layer nodes. In this paper, the number of input nodes was mostly between 1 and 10, so the number of hidden layer nodes was set to 5. Iteration time was set as 100 for the reason that in case studies of this paper, once the number of iteration exceeds 100, the results change very little, so it is no need to increase the time of iteration. The Learning rate and Goal were determined by trial and error methods. When adjusting these two parameters to different values separately, we chose the value that gives the best results, that is, the Learning rate was 0.1 and Goal was 0.00004. The parameters of the RBFNN model were gotten by network self-training. As for modeling data, the same training data and the completely same input data were used.
4.4 Performance evaluation criteria

Firstly, traditional statistics criteria were utilized in this paper to measure predicting performance. Among them, mean absolute error (MAE) is the average of absolute error between the original and prediction data, mean absolute percentage error (MAPE) measures the average forecasting accuracy, root mean square error (RMSE) tests model stability, and the coefficient of determination ($R^2$) denotes fitting degree. Smaller MAE, MAPE, and RMSE, along with a bigger value of $R^2$ indicate a better prediction quality, vice versa. The definition of these criteria is as follows:

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^{N} |y_t - y_t^*|$$  \hspace{1cm} (12)  

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{y_t - y_t^*}{y_t} \right| \times 100\%$$  \hspace{1cm} (13)  

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - y_t^*)^2}$$  \hspace{1cm} (14)  

Figure 12 shows the forecasting results between different models for Chongqing carbon price.

Moreover, $A_{\text{MAE}}, A_{\text{MAPE}}, A_{\text{RMSE}}$, and $A_{R^2}$ were introduced to evaluate the amelioration of the proposed model. If all indices are greater than 0, which means the proposed model is better. Additionally, the values of $A_{\text{MAE}}, A_{\text{MAPE}}, A_{\text{RMSE}}$, and $A_{R^2}$ are closer to 0, the smaller amelioration the proposed model renders. The definition of $A_{\text{MAE}}, A_{\text{MAPE}}, A_{\text{RMSE}}$, and $A_{R^2}$ is described as Equations (16-19).

$$A_{\text{MAE}} = \left( \frac{\text{MAE}_2 - \text{MAE}_1}{\text{MAE}_1} \right) \times 100\%$$  \hspace{1cm} (16)  

$$A_{\text{MAPE}} = \left( \frac{\text{MAPE}_2 - \text{MAPE}_1}{\text{MAPE}_1} \right) \times 100\%$$  \hspace{1cm} (17)  

$$A_{\text{MRSE}} = \left( \frac{\text{MRSE}_2 - \text{MRSE}_1}{\text{MRSE}_1} \right) \times 100\%$$  \hspace{1cm} (18)  

$$R^2 = \frac{\left( N \sum_{t=1}^{N} y_t^* y_t - \sum_{t=1}^{N} y_t^* \sum_{t=1}^{N} y_t \right)^2}{\left( N \sum_{t=1}^{N} y_t^2 - \left( \sum_{t=1}^{N} y_t^* \right)^2 \right) \left( N \sum_{t=1}^{N} y_t^2 - \left( \sum_{t=1}^{N} y_t \right)^2 \right)}$$  \hspace{1cm} (15)
where the subscript 1 and 2 denote the statistic index of the proposed model and other benchmark models.

Besides, the stability test was also carried out in this study to evaluate the capabilities of models. The concrete principle depends on the formula (20).

\[
S_{\text{Var}}(t) = \text{var} \left( \frac{y_t - \hat{y}_t}{y_t} \right) \tag{20}
\]

The smaller $S_{\text{Var}}$ value represents higher stability, and vice versa.

### 4.5 Results and discussion

According to the aforementioned procedure, the proposed ED-LSTM model and other contrast models including LSTM, FFNN, and RBFNN were established. The final forecasting results of the related models are shown in Figures 6-13, which represent the carbon prices of Shenzhen, Beijing,
Guangdong, Shanghai, Tianjin, Hubei, Chongqing, and Fujian, respectively. The Shenzhen carbon price, shown in Figure 6, served as an example to discuss the forecasting results. The fitting curves of the forecasting values from different models, frequency histogram of RE (relative error), and scatter correlation diagrams are presented in this figure. The fitting curve of the ED-LSTM model was nearest to the actual curve, implying that it is the best fitted. Moreover, in the frequency histogram, RE of the proposed model is observed to be distributed more on the left side compared to other benchmark models. In other words, the proposed model had small errors in most instances. In regard to the scatter correlation diagram, the regression line denoted that the real value was the same as the prediction value; thus, the closer the scatters were to this line, the better the performance was. The scatters of the ED-LSTM model were observed to be nearest and more concentrated to the regression line.

Analogously, the forecasting results in the other seven case studies of Chinese carbon price series were also seen in the corresponding Figure (Figures 7-13), where it is clear that the forecasting results of the proposed model were superior due to the following three aspects: The fitting curve of ED-LSTM model was closest to the actual curve; RE of the ED-LSTM model was relatively small; scatters of the ED-LSTM model were nearest and more concentrated to the regression line.

The values of traditional statistical criteria are listed in Table 6, which may be seen more clearly in Figure 14. The optimal values are marked in bold in Table 6. In view of MAE, MAPE, and RMSE and R2, though the evaluation values of the ED-LSTM model vary with different carbon price series, these values served as the smallest MAE, MAPE, and RMSE and the highest R2, unlike other benchmark models. Besides, the statistic values of the LSTM model were also better than the FFNN model as well as the RBFNN model. Specifically, the values of MAE, MAPE, and RMSE of the LSTM model were smaller than the corresponding values of the FFNN and RBFNN models. Additionally, the values of R2 were found to be larger than the corresponding values of the FFNN and RBFNN models. Key findings can be derived based on the above results of statistical indicators. First, the prediction performance of the ED-LSTM model was superior to that of the single LSTM model as the ensemble technique was based on mode decomposition, which may effectively predict each additional stable mode so as to significantly enhance the prediction accuracy. It also indicated the merit of the “mode decomposition, separate modeling, and ensemble learning” principle. Second, the single LSTM model was shown to be prone to generating prediction results closer to the true values compared with the single FFNN and RBFNN models, which mainly contributed in its features of excellent serialization and long-term dependencies.

From the statistical index radar chart (A), (B), and (C) in Figure 14, the pink line representing the ED-LSTM model may be observed to be at the innermost side, which is followed by the yellow line representing the LSTM model. According to the radar chart (D) in Figure 14, the pink line is at the outermost side, followed by the yellow line, which all have values of reduced MAE, MAPE, and RMSE, but have a higher R2 in the proposed model regarding eight carbon price series forecasting. Combining Figure 15 and Table 6, the models that covered memory architecture (ED-LSTM and LSTM), especially ED-LSTM, greatly enhanced the forecasting accuracy compared to FFNN and RBFNN.

Another four evaluation indices covering AMAE, AMAPE, ARMSE, and AR2 were executed. As evident in Figure 15, all indexes were positive, demonstrating that the proposed model was superior to the benchmarks. Furthermore, the MAE of the proposed model decreased by 75.6%, 57.4%, 69.2%, 56.8%, 87.1%, 82.2%, 60.6%, and 78.4%, which was respectively in contrast to the worst performing RBFNN method. Additionally, it was reduced by 38.9%, 34.6%, 48.0%, 49.4%, 17.2%, 47.1%, 38.0%, and 42.9%, which was in contrast to the
The best performing LSTM model, respectively. Analogously, the AMAPE, ARMSE, and AR2 acquired accordant conclusions in that the proposed model greatly improved the forecasting results. The specific improvement values are as follows. The MAPE of the proposed model decreased by 75.0%, 59.1%, 70.2%, 56.7%, 85.2%, 78.7%, 63.1%, and 77.1% in contrast to the worst performing RBFNN method, while it was reduced by 40.8%, 40.3%, 48.7%, 50.0%, 14.6%, 42.0%, 47.2%, and 42.1% in contrast to the best performing LSTM model. The RMSE of the proposed model decreased by 51.9%, 20.7%, 61.9%, 58.4%, 85.4%, 82.9%, 60.7%, and 75.8% in contrast to the worst performing RBFNN method, while it was reduced by 37.4%, 9.0%, 48.7%, 54.2%, 34.1%, 54.7%, 48.0%, and 57.4%, and 42.9% in contrast to the best performing LSTM model. The R2 of the proposed model increased by 10.6%, 4.5%, 27.0%, 16.1%, 2.1%, 5.0%, 2.4%, and 2.6% in contrast to the worst performing RBFNN method and increased by 6.8%, 4.0%, 24.9%, 13.2%, 0.45%, 2.5%, 1.2%, and 2.5% in contrast to the best performing LSTM model.

The results of the stability test are listed in Table 7. The smallest $S_{\text{Var}}$ is highlighted in bold. Obviously, the $S_{\text{Var}}$ of the proposed model was smallest. Concrete $S_{\text{Var}}$ values of the ED-LSTM model were 0.00111, 0.00223, 0.00061, 0.00019, 0.00009, 0.00017, 0.00325, and 0.00008 in each case study, respectively. Thus, compared with the benchmark models, the developed model also possessed strong prediction stability.

Moreover, to verify the positive effect of PACF filtering input features on the prediction results, the contrast model ED-LSTM* that did not include PACF was also executed. In this model, the input features were set manually. In all experiments, the results of the evaluation of model accuracy and stability are listed in Table 6 and Table 7. Accordingly, it can be found that adding PACF can improve the performance of the ED-LSTM* model; hence, PACF is essential in constructing the forecasting model.

Overall, the following remarkable conclusions may infer from these results:

| TABLE 7 Results of $S_{\text{Var}}$ |
|-------------------------------------|
| Models | ED-LSTM | ED-LSTM* | LSTM | FFNN | RBFNN |
|----------------|---------|---------|------|------|-------|
| SZ | 0.00111 | 0.00739 | 0.00297 | 0.00312 | 0.00255 |
| BJ | 0.00223 | 0.00490 | 0.00061 | 0.00036 | 0.00348 |
| GD | 0.00061 | 0.00009 | 0.00013 | 0.00133 | 0.00123 |
| SH | 0.00019 | 0.00017 | 0.00013 | 0.00013 | 0.00123 |
| TJ | 0.00009 | 0.00123 | 0.00030 | 0.00038 | 0.00179 |
| HB | 0.00017 | 0.00024 | 0.00087 | 0.00218 | 0.00428 |
| CQ | 0.00325 | 0.01236 | 0.02306 | 0.02663 | 0.02791 |
| FJ | 0.00008 | 0.00049 | 0.00062 | 0.00075 | 0.00013 |
| Ave | 0.00097 | 0.00343 | 0.00451 | 0.00514 | 0.00552 |
a. By combining Table 6 and Figures 6-13, the forecasting results of Beijing and Guangdong carbon prices were relatively poor, while that of Tianjin and Hubei were good, which corresponds to the volatility and complexity of the original data shown in Table 2. Those represent a relatively larger volatility and complexity in the Beijing carbon price, the most unstable and relatively complex Guangdong carbon price, a relatively smaller volatility and complexity in the Tianjin carbon price, and the smallest fluctuation of the Hubei carbon price.
b. In contrast to traditional ANN (FFNN and RBFNN), LSTM and the proposed ED-LSTM model performed well, which may have occurred as traditional ANN models are static in nature due to being historically data-driven only, while models possessing memory units can gain serialization features and long-term dependencies.
c. It can be drawn from the overall assessments of performance that the ensemble-driven hybrid model based on mode decomposition is better than the individual model once the data largely fluctuates. This phenomenon may be because the model under ensemble learning architecture is capable of enhancing feature representation, information extraction, and the application of long short-term relationships in order for the data to be deeply explored and for their features to be utilized to the greatest potential.
d. The proposed model was applied to all eight different China carbon price series forecasts given the large sample data. The forecasting results demonstrated that the model was precise, stable, and sufficiently general, showing its capabilities in Chinese carbon price forecasting.

5 | CONCLUSIONS

The precise prediction of carbon price is essential nowadays as carbon trading markets are being widely established to promote environmental protection. This paper explored a general and stable forecasting model in order to efficiently anticipate future carbon prices given time series data, which was applied to all eight-pilot carbon trading markets in China. Moreover, an ED-LSTM model that built multiple LSTM models in different modes for ensemble learning was explored for carbon price forecasting. The established model was implemented, encompassing three steps: (a) CEEMD was initially employed to multi-mode feature extraction, and PACF was executed on each mode to select time-lagged factors as input; (b) LSTM was subsequently employed for multi-mode feature learning; (c) inverse CEEMD computation was ultimately utilized to integrate the forecasted results of multi-mode. Practically, the performance of the developed ED-LSTM was compared with single LSTM as well as two other kinds of conventional ANN by considering evaluation indicators such as MAE, MAPE, RMSE, R2, AMAE, AMAPE, ARMSE, and AR2. The proposed model improved by 38.9%, 34.6%, 48.0%, 49.4%, 17.2%, 47.1%, 38.0%, and 42.9% compared with LSTM in the eight carbon price series. The ED-LSTM model was more enhanced compared to the two conventional static ANN models. Additionally, the stability test demonstrated that the developed model was superior in prediction accuracy as well as prediction stability. Taking the daily carbon price of all eight China carbon trading pilots for an extended period of time as the research object, the obtained forecasting results indicated the satisfactory performance of the proposed model, which possessed sufficient accuracy, stability, and strong universality. Overall, the proposed method may serve as a promising tool in Chinese carbon price forecasting.

CONFLICT OF INTEREST
The authors declare no conflicts of interest.

ORCID
Zhaoqi Li https://orcid.org/0000-0002-3748-8914

REFERENCES
1. Chang C. A multivariate causality test of carbon dioxide emissions, energy consumption and economic growth in China. Appl Energy. 2010;87(11):3533-3537.
2. Zhang YJ, Wei YM. An overview of current research on EU ETS: evidence from its operating mechanism and economic effect. Appl Energy. 2010;87(6):1804-1814.
3. Zhu B, Han D, Wang P, Wu Z, Zhang T, Wei YM. Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. Appl Energy. 2017;191:521-530.
4. Sun W, Sun J. Daily PM2.5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm. J Environ Manage. 2017;188:144-152.
5. Zhu B, Ye S, Wang P, He K, Zhang T, Wei YM. A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting. Energy Econ. 2018;70:143-157.
6. Zhang J, Li D, Hao Y, Tan Z. A hybrid model using signal processing technology, econometric models and neural network for carbon spot price forecasting. J Clean Prod. 2018;204:958-964.
7. Guobrandsdottir HN, Haraldsson HO. Predicting the price of EU ETS carbon credits. Syst Eng Proc. 2011;1:481-489.
8. Byun SJ, Cho H. Forecasting carbon futures volatility using GARCH models with energy volatilities. Energy Econ. 2013;40:207-221.
9. Chevallier J, Sévi B. On the realized volatility of the ECX CO2 emissions 2008 futures contract: distribution, dynamics and forecasting. Ann Finance. 2009;7(1):1-29.
10. Guan XT. Research on Carbon Market Transaction Price Forecast based on Grey Theory. Chengdu, China: Southwest Jiaotong University; 2016.
11. Lan Y, Zhaopeng L, Li Y, Jie L. The scenario simulation analysis of the EU ETS carbon price trend and the enlightenment to China. J Environ Econ. 2017;3:4.
12. Tsai MT, Kuo YT. A forecasting system of carbon price in the carbon trading markets using artificial neural network. Int J Environ Sci Dev. 2013;4(2):163-167.
13. Zhu B. A novel multiscale ensemble carbon price prediction model integrating empirical mode decomposition, genetic algorithm and artificial neural network. *Energies*. 2012;5(2):355-370.
14. Bai Y, Zeng B, Li C, Zhang J. An ensemble long short-term memory neural network for hourly PM$_{2.5}$ concentration forecasting. *Chemosphere*. 2019;222:286-294.
15. Zhang J, Yan J, Infield D, Liu Y, Lien FS. Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model. *Appl Energy*. 2019;241:229-244.
16. Xia T, Song Y, Zheng Y, Pan E, Xi L. An ensemble framework based on convolutional bi-directional LSTM with multiple time windows for remaining useful life estimation. *Comput Ind*. 2020;115:103182.
17. Ma R, Yang T, Breaa E, Li Z, Broois P, Gao F. Data-driven proton exchange membrane fuel cell degradation prediction through deep learning method. *Appl Energy*. 2018;231:102-115.
18. Chang Z, Zhang Y, Chen W. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. *Energy*. 2019;187:115804.
19. Ding Y, Zhu Y, Feng J, Zhang P, Cheng Z. Interpretable spatio-temporal attention LSTM model for flood forecasting. *Neurocomputing*. 2020;403:348-359.
20. Moreno SR, da Silva RG, Mariani VC, dos Santos CL. Multi-step wind speed forecasting based on hybrid multi-stage decomposition model and long short-term memory neural network. *Energy Conv Manage*. 2020;213:112869.
21. Chang YS, Chiao HT, Abimannan S, Huang YP, Tsai YT, Lin KM. An LSTM-based aggregated model for air pollution forecasting. *Atmos Pollut Res*. 2020;11(8):1451-1463.
22. Chen Y. Voltages prediction algorithm based on LSTM recurrent neural network. *Optik*. 2020;164869:220
23. Abbasimehr H, Shabani M, Yousefi M. An optimized model using LSTM network for demand forecasting. *Comput Ind Eng*. 2020;143:106435.
24. Wang F, Xuan Z, Zhen Z, Li K, Wang T, Shi M. A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. *Energy Conv Manage*. 2020;212:112766.
25. Luo H, Huang M, Zhou Z. A dual-tree complex wavelet enhanced convolutional LSTM neural network for structural health monitoring of automotive suspension. *Measurement*. 2019;137:14-27.
26. Li Y, He Y, Zhang M. Prediction of Chinese energy structure based on Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM). *Energy Sci Eng*. 2020;8(8). https://doi.org/10.1002/esce.698
27. Opitz D, Maclin R. Popular ensemble methods: an empirical study. *J Artificial Intelligence Res*. 1999;11:169-198.
28. Wang H, Zhang YM, Mao JX, Wan HP. A probabilistic approach for short-term prediction of wind gust speed using ensemble learning. *J Wind Eng Ind Aerodyn*. 2020;202:104198.
29. Yu L, Liang S, Chen R, Lai KK. Predicting monthly biofuel production using a hybrid ensemble forecasting methodology. *Int J Forecast*. 2019. https://doi.org/10.1016/j.ijforecast.2019.08.014
30. Ribeiro MHDM, dos Santos CL. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Appl Soft Comput*. 2020;86:105837.
31. Ribeiro GT, Mariani VC, dos Santos CL. Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting. *Eng Appl Artif Intell*. 2019;82:272-281.
32. Jiang M, Liu J, Zhang L, Liu C. An improved stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms. *Phys A*. 2020;541:122272.
33. Xiao J, Xiao Z, Wang D, Bai J, Hayarimana V, Zeng F. Short-term traffic volume prediction by ensemble learning in concept drifting environments. *Knowledge-Based Syst*. 2019;164:213-225.
34. Sun G, Chen T, Wei Z, Sun Y, Zang H, Chen S. A carbon price forecasting model based on variational mode decomposition and spiking neural networks. *Energies*. 2016;9(1):54.
35. Zhou J, Yu X, Yuan X. Predicting the carbon price sequence in the Shenzhen emissions exchange using a multiscale ensemble forecasting model based on ensemble empirical mode decomposition. *Energies*. 2018;11(7):1907.
36. Wei S, Chongchong Z, Cuiping S. Carbon pricing prediction based on wavelet transform and K-ELM optimized by bat optimization algorithm in China ETS: the case of Shanghai and Hubei carbon markets. *Carbon Manage*. 2018;9(6):605-617.
37. Lu Y, Xie R, Liang SY. CEEMD-assisted bearing degradation assessment using tight clustering. *Int J Adv Manuf Technol*. 2019;104(1-4):1259-1267.
38. Zhao L, Yu W, Yan R. Gearbox fault diagnosis using complementary ensemble empirical mode decomposition and permutation entropy. *Shock Vib*. 2016;2016:1-8.
39. Wang Z, Jia L, Qin Y. Bearing fault diagnosis using multiclass self-adaptive support vector classifiers based on CEEMD–SVD. *Wirel Pers Commun*. 2018;102(2):1669-1682.
40. Zheng H, Dang C, Gu S, Peng D, Chen K. A quantified self-adaptive filtering method: effective IFMs selection based on CEEMD. *Meas Sci Technol*. 2018;29(8):085701.
41. Nagi GM, Yang D, Srinivasan D. Automatic hourly solar forecasting using machine learning models. *Renew Sust Energ Rev*. 2019;105:487-498.
42. Li W, Lu C. The research on setting a unified interval of carbon price benchmark in the national carbon trading market of China. *Appl Energy*. 2015;155:728-739.
43. Sun W, Zhang C. Analysis and forecasting of the carbon price using multi-resolution singular value decomposition and extreme learning machine optimized by adaptive whale optimization algorithm. *Appl Energy*. 2018;231:1354-1371.
44. Zhou J, Huo X, Xu X, Li Y. Forecasting the carbon price using extreme-point symmetric mode decomposition and extreme learning machine optimized by the grey wolf optimizer algorithm. *Energies*. 2019;12(5):950.
45. Sun W, Huang C. A carbon price prediction model based on secondary decomposition algorithm and optimized back propagation neural network. *J Clean Prod*. 2020;243:118671.
46. Xiong S, Wang C, Fang Z, Ma D. Multi-step-ahead carbon price forecasting based on variational mode decomposition and fast multi-output relevance vector regression optimized by the multi-objective whale optimization algorithm. *Energies*. 2019;12(1):147.
47. Sun W, Duan M. Analysis and forecasting of the carbon price in China's regional carbon markets based on fast ensemble empirical mode decomposition, phase space reconstruction, and an improved extreme learning machine. *Energies*. 2019;12(2):277.
48. Zhu J, Wu P, Chen H, Lui J, Zhou L. Carbon price forecasting with variational mode decomposition and optimal combined model. *Phys A*. 2019;519:140-158.
SUN AND LI

49. Hao Y, Tian C, Wu C. Modelling of carbon price in two real carbon trading markets. J Clean Prod. 2020;244:118556.

50. Zhang X, Zhang C, Wei Z. Carbon price forecasting based on multi-resolution singular value decomposition and extreme learning machine optimized by the moth-flame optimization algorithm considering energy and economic factors. Energies. 2019;12(22):4283.

51. Xianzi ZCY. Forecasting of China’s regional carbon market price based on multi-frequency combined model. Syst Eng-Theory Pract. 2016;36:3017-3025.

52. Wu Q, Liu Z. Forecasting the carbon price sequence in the Hubei emissions exchange using a hybrid model based on ensemble empirical mode decomposition. Energy Sci Eng. 2020. https://doi.org/10.1002/ese3.703

53. Hao Y, Tian C. A hybrid framework for carbon trading price forecasting: the role of multiple influence factor. J Clean Prod. 2020;262:120378.

54. Lu H, Ma X, Huang K, Azimi M. Carbon trading volume and price forecasting in China using multiple machine learning models. J. Clean Prod. 2020;249:119386.

55. Yeh JR, Shieh JS, Huang NE. Complementary ensemble empirical mode decomposition: a novel noise enhanced data analysis method. Adv Adaptive Data Anal. 2010;2(02):135-156.

56. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735-1780.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Sun W, Li Z. An ensemble-driven long short-term memory model based on mode decomposition for carbon price forecasting of all eight carbon trading pilots in China. Energy Sci Eng. 2020;8:4094–4115. https://doi.org/10.1002/ese3.799