Science of the Total Environment 762 (2021) 143099

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting

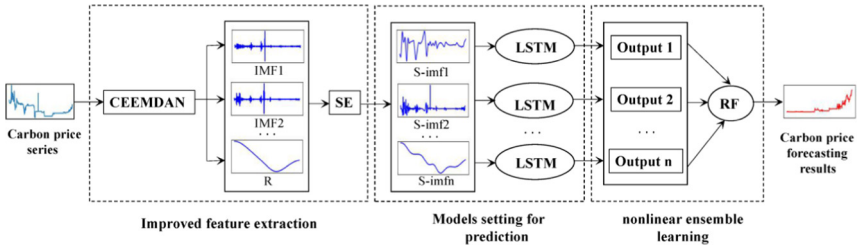
Jujie Wang [a,b,*], Xin Sun [a], Qian Cheng [a], Quan Cui [a]

[a] School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China
[b] Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing 210044, China

HIGHLIGHTS

- Propose a new model based on nonstationary and nonlinear data: CEEMDAN-SE-LATM-RF
- The efficacy was tested in different carbon trading markets in China.
- Proposed model outperformed the other 4 benchmark methods.
- Improved feature extraction heightens the forecast accuracy.
- Nonlinear ensemble model represents better generalization ability for robustness.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 19 July 2020
Received in revised form 4 October 2020
Accepted 11 October 2020
Available online 16 October 2020

Editor: Kuishuang Feng

Keywords:
Carbon price prediction
Hybrid model
Improved feature extraction
Long short-term memory network
Nonlinear ensemble algorithm
Random forest

ABSTRACT

Carbon price is the basis of developing a low carbon economy. The accurate carbon price forecast can not only stimulate the actions of enterprises and families, but also encourage the study and development of low carbon technology. However, as the original carbon price series is non-stationary and nonlinear, traditional methods are less robust to predict it. In this study, an innovative nonlinear ensemble paradigm of improved feature extraction and deep learning algorithm is proposed for carbon price forecasting, which includes complete ensemble empirical mode decomposition (CEEMDAN), sample entropy (SE), long short-term memory (LSTM) and random forest (RF). As the core of the proposed model, LSTM enhanced from the recurrent neural network is utilized to establish appropriate prediction models by extracting memory features of the long and short term. Improved feature extraction, as assistant data preprocessing, represents its unique advantage for improving calculating efficiency and accuracy. Removing irrelevant features from original time series through CEEMDAN lets learning easier and it's even better for using SE to recombine similar-complexity modes. Furthermore, compared with simple linear ensemble learning, RF increases the generalization ability for robustness to achieve the final nonlinear output results. Two markets' real data of carbon trading in china are as the experiment cases to test the effectiveness of the above model. The final simulation results indicate that the proposed model performs better than

Abbreviations: ADF, Augmented Dickey-Fuller; AE, Approximate entropy; ANN, Artificial neural network; ARIMA, Autoregressive integrated moving average; BDS, Brock-Decher-Scheikman; BPNN, Back propagation neural network; CEEMD, Complete ensemble empirical mode decomposition; CEEMDAN, Complete ensemble empirical mode decomposition with adaptive noise; EMD, Empirical mode decomposition; EEMD, Ensemble empirical mode decomposition; EWT, Empirical wavelet transform; GARCH, Generalized autoregressive conditional heteroscedasticity; GPR, Gaussian process regression; GRU, Gated recurrent unit; IMF, Intrinsic mode function; KNN, K-nearest neighbor; LSSVM, Least squares support vector machine; LSTM, Long short-term memory; MAE, Mean absolute error; MAPE, Mean absolute percentage error; RF, Random forest; RMSE, Root mean square error; RNN, Recurrent neural network; SE, Sample entropy; SVM, Support vector machine; VAR, Vector auto regression; VARIMA, Vector autoregressive integrated with moving average; WT, Wavelet transform; XGboost, Extreme gradient boosting.

* Corresponding author at: School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China.
E-mail address: jujiewang@126.com (J. Wang).

https://doi.org/10.1016/j.scitotenv.2020.143099

the other four benchmark methods reflected by the smaller statistical errors. Overall, the developed approach provides an effective method for predicting carbon price.

1. Introduction

From a global perspective, climate changes and natural disasters caused by greenhouse gas emissions are more and more serious in recent years (Huang and He, 2020). This phenomenon has much to do with the continuous economic activities of humans, but it is inevitable. The best approach at present is the market-based policy formulation of pricing for carbon emissions, which can calculate the cost of emissions by directly considering the economic efficiency (Sun and Wang, 2020; Song et al., 2019). Therefore, recent years have seen the trading system of carbon emissions gaining global popularity (Wu and Liu, 2020). Looking at the international market, it admits of no doubt that the European Union Emissions Trading System (EU-ETS) is the largest, most liquid and most influential greenhouse gas emission reduction mechanism in the world, which has provided an effective operating paradigm for carbon emissions trading around the world and accumulated a wealth of data and experience (Zhang and Wei, 2010). Afterwards, as one of the major contributors in reducing global carbon emissions, China has successively established eight carbon markets in the pilot cities including Beijing (2013), Shanghai (2013), Chongqing (2014), Guangdong (2013), Tianjin (2013), Shenzhen (2013), Hubei (2014) and Fujian (Song et al., 2018). Moreover, in 2017, the national development and reform commission officially announced that China will launch carbon trading market pilots and put this project on a vital position in the 13th five-year plan, which indicates our firm confidence in developing a green economy. In light of this, enhancing carbon price forecasting precision can not only assist in policy implementation for manufacturing processes and investment decisions, but also create an effective and stable carbon pricing mechanism (Zhu et al., 2019). However, as a burgeoning policy-setting market created by artificial regulation, the price series of carbon have the characteristics of nonstationary and nonlinear, which are caused by internal market mechanisms as well as external environmental uncertainties (Yang and Liang, 2017; Yang et al., 2020). Therefore, it is a great challenge to predict carbon prices and more researchers get involved in this research issue.

Numerous studies show that the carbon price is, to some extent, predictable only if use some certain methods, which are essentially the time-series prediction technology. This technology is broadly divided into two classes: statistical and economics models and artificial intelligence models. Among them, statistical and economics models are relatively traditional, including autoregressive moving integrated moving average (ARIMA), vector autoregression (VAR), generalized autoregressive conditional heteroscedasticity (GARCH) and the like. Byun and Cho (2013) predicted the volatility of carbon price through GARCH-type models and K-nearest neighbor model (KNN). It can be found that an improved GARCH model obtained the most accurate result in all of the selected models. García-Martos et al. (2013) verified that the approach of vector autoregressive integrated with moving average (VARIMA) gave better results for very short forecasting horizons (one-day-ahead forecasts) in the case of nonstationary and nonlinear price series. Benz and Truck (2009) predicted the carbon spot price and volatility in the different phases of returns with the Markov-switching method and a standard AR-GARCH model. Although these statistical and economics methods are able to achieve effective prediction results and possess good theoretical description based on statistical logic, their implementations are rooted on the linear hypothesis and mainly for very short forecasting span, which don't have adequate ability to catch the nonlinear characteristics hidden in the time-series. For the sake of solving this problem, artificial intelligence models sprung up and were applied to predict time-series, which are the data-driving method (Sun et al., 2020). Artificial intelligence models mainly cover artificial neural network (ANN) and support vector machine (SVM). Fan et al. (2015) applied ANN with a multi-layer perceptron model for carbon price forecasting, which proved prediction accuracy and fitting capacity are higher than many single and variant models. Zhu et al. (2016) developed an ensemble paradigm of forecasting energy price with the feature of nonstationary and nonlinear through the least squares SVM (LSSVM) model and the final research result also showed the superiority over the models such as ARIMA and ANN. Patel et al. (2015) combined ANN with SVM as the hybrid model, which resulted in the best overall prediction performance. Also, the hybrid model incorporating the artificial intelligence model and statistical model can represent relatively better performance. Zhu and Wei (2013) then made a combination between LSSVM and ARIMA, which proved better carbon price prediction results compared by simple statistical models. Furthermore, for the reason that time series prediction is relative not only to the current data but also to the data at a passed time, whose information will be lost if only applying the latest data, the recurrent neural network (RNN) connecting with hidden units was established for enabling the network to keep the memory of recent information, which is unlike traditional ANNs (Lin et al., 1998). It has been widely used in many fields, especially for nonstationary and nonlinear series. Rather et al. (2015) found that according to the prediction error and correlation between the original data and output data, RNN generates much better prediction results. Chen et al. (2013) developed an enhanced RNN for multi-step-ahead flood forecasting with the feature of nonstationary and nonlinear, representing that RNN could deal with the association between the past and future data by the memory features. Meanwhile, in the time series forecast field, the long short-term memory (LSTM) as an enhanced special RNN has a wide use. The “gate” mechanism is the core of the algorithm, which can selectively filter the input information and more important historical data information will be extracted for the following prediction (Li, 2020). Huang et al. (2019a) established the LSTM prediction model for carbon emissions in China and the emulation result showed that the prediction accuracy is higher compared with back propagation neural network (BPNN) and Gaussian process regression (GPR). Krishan et al. (2019) used LSTM on the field of air quality, which referred to the carbon emission and results showed that LSTM did have a good capacity to capture complicated features in the nonstationary and nonlinear series. Liu and Shen (2019) also established an improved LSTM model for obtaining higher carbon price forecasting accuracy to some extent. These extinguished results encouraged us to explore the possibilities of predicting carbon prices in China via LSTM. Nonetheless, given a large fluctuation towards the original carbon price series, the effect through a single AI technique is not very outstanding.

And except for single AI technique, data preprocessing plays an important role in the time-series prediction, among which feature extraction is exactly vital. On the one hand, feature extraction can reduce the feature numbers and dimensions to prevent the emergence of an overfitting problem. On the other hand, it can remove irrelevant features and let learning easier. Generally, the methods of feature extraction are broadly divided into wrapper methods and filter methods. Wrapper methods usually cost highly and have a slow learning rate because of many times' learner training, whereas filter methods are relied on original data and have a faster calculating rate than wrapper methods (Wang and Li, 2018). Hence, considering the convenience and speed case, researchers may prefer the later. Zhu et al. (2013) manifested that the exploitation of various internal logic and necessary features implied at different frequencies can be made possible through decomposing methods. Wavelet transform (WT) is one of the decomposition methods widely utilized in the area of prediction. Chevallier

(2011) applied the wavelet packet transforms into carbon price forecasting, which represented a good performance. Sun et al. (2018) also used WT as the basis decomposing model in the research of China Emissions-Trading Scheme. But these representations of effectivity mainly depend on the wavelet basis function selected by researchers' subjectivity without a specific theory foundation. Empirical mode decomposition (EMD) is an adaptive method overcoming the drawback of reliance on the subjective experience of setting a basis function previously. Zhu et al. (2017) proved that using EMD can well capture several components with different features. Gao and Jian (2014) proposed a hybrid model comprising particle swarm optimization (PSO), SVM and EMD. In the EMD part, several stationary intrinsic mode functions (IMFs) and a residual series will be put into a neural network for training. For the sake of innovation, Gilles (2013) built a new self-adaptive signal decomposition method named empirical wavelet transform (EWT) by combing EMD with WT, whose final result showed a better performance. However, the process of decomposition through EMD is easy to emerge modal mixing problem and its physical meaning is lacking (Tian and Hao, 2020). To tackle the problem, Wu and Huang (2009) carried out a study and improved EMD, which was named as ensemble empirical mode decomposition (EEMD). Qin et al. (2015) utilized EEMD as a data preprocessing method for improving the prediction effect of the carbon price. Wu et al. (2019) also combined EEMD with LSTM to predict the spot price of west texas intermediate crude oil. It can be found that EEMD is an enhanced EMD, which can improve the phenomenon of modal mixing effectively by offsetting and restraining the effects of noises in many times' experiences. Despite robustness and effectiveness of forecasting based on EEMD, there is still a drawback towards it. Increasing the times of integration can reduce the error of reconstruction, whereas it expands the scale of calculation and remains residual noises to a certain amplitude. Besides, Wu and Huang (2009) said that the problem of modal splitting may occur. To overcome this defect, complete ensemble empirical mode decomposition (CEEMD) as an improved method of EEMD is applied. Zhang et al. (2018) proved that complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) as a signal processing technology can not only solve the modal aliasing problem, but also lessen the white noise interference and save the computing time. What's more, Cao et al. (2019) combined CEEMDAN with LSTM, indicating that CEEMDAN can exploit more hidden information than EMD and the hybrid model surpasses the single one. Therefore, CEEMDAN can be seen as a relatively progressive decomposition method at present and utilized by this paper for the reason that its error of reconstruction is nearly zero by adding adaptive white noise into each phase.

Extant studies have shown that the AI prediction models with the feature extraction part can not only achieve the effects of data preprocessing and improve the calculating efficiency, but also establish an appropriate prediction model for the time series. But several major drawbacks still remain. First of all, after decomposing the carbon price series, each sub-sequence has been put into a prediction model for the output results, which didn't consider the similar complexity and correlation among them so as to lower efficiency and accuracy. Secondly, the prediction model for each sub-sequence is the same without the realization that each mode is different for its unique feature and frequency, so the respective establishment of models with more proper parameters is of vital importance (Che, 2015). Thirdly, after achieving the prediction results of each sub-sequence, existing final ensemble models mainly limit to the linear form such as obtaining the final forecast result through combining the prediction values of all the decomposed modes (Zhu et al., 2018). For the reason that it is not usually applicable for all the cases, a linear ensemble approach may affect the accuracy of predicting (Liao and Tsao, 2006). There are two main types of nonlinear integration methods. One of them is serialization methods with strong dependencies among individual learners and the other is parallelization methods generated simultaneously without strong dependencies among individual learners. Representative of the former is boosting and the latter is bagging, which develop the extreme gradient boosting (XGboost) and the random forest (RF) respectively. By comparing these two methods, we can find that the XGboost is more sensitive to overfitting if the data is noisy and it is often takes longer for being built in sequence (Fan et al., 2020). What's more, RF is more adjustable.

In order to solve these existing problems towards carbon price forecast, a novel hybrid model incorporating CEEMDAN, Sample entropy (SE), LSTM and Random forest (RF) is put forward. From the perspective of methodology, it develops an innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning algorithm for higher accuracy in the case of nonstationary and nonlinear carbon price forecast. Firstly, the original carbon price series is decomposed into several simple stationary modes with the application of CEEMDAN algorithm. Then, the obtained simple modes with similar complexity are recombined according to the SE algorithm, so as to boost calculating efficiency and accuracy. Considering that different modes have their own frequency and characteristic, LSTM can then be applied to set an appropriate prediction model for each reconstructed component because of its strong long and short term memory. At last, after forecast results of reconstructed components have been achieved through the deep learning algorithm, RF as a nonlinear ensemble bagging learning model is utilized to aggregate the final carbon price forecast result for the further improved prediction accuracy.

From the above, the main innovations and contributions of this research compared to the findings in the literature are shown in the following four points:

a. Considering the neglect of similar complexity and correlation among decomposed modes, an improved feature extraction incorporating CEEMDAN and SE is adopted for screening different features effectively from the original carbon price series so as to the higher efficiency and accuracy.
b. With the realization that respective establishment of models is of vital importance and in order to capture more complicated features, LSTM replaces RNN as the crucial prediction model.
c. For the reason that nonlinear ensemble learning can get smaller errors and more stability than a linear approach, this research applies RF as integrated algorithm to improve the forecast accuracy.
d. The novel hybrid model for carbon price forecast setting as an adaptive nonlinear ensemble learning paradigm is firstly proposed, which excels single model and represents its unique robustness.

The structure of the rest of this paper is as follows: the methodologies and brief proposed model structure are outlined in Section 2. The case study with data collecting, preprocessing and relative measurement indices are elaborated in Section 3. Section 4 describes the forecast results as well as discussions in more detail. At last, Section 5 draw a conclusion.

2. Methodology

2.1. Complete ensemble empirical mode decomposition (CEEMDAN)

EMD proposed by Huang et al. (1998) has been widely utilized in many fields, which is an adaptive signal decomposition method without any assumptions about data. However, the problem of modal aliasing causes the decomposed intrinsic functions affecting each other, which deprives the physical meaning of the IMF. To solve this problem, Wu and Huang (2009) proposed EEMD, which can offset the effects of noise during the procession of decomposition by making several times' experiments. Unfortunately, there is residual noise in the components, which lowers efficiency. Overall,

CEEMDAN is an enhanced method overcoming the above two drawbacks. Through adding adaptive white noise, CEEMDAN can not only eliminate the reconstruction error, but also decrease the iterations and calculation cost of the algorithms effectively (Cao et al., 2019). Being the same times as EEMD, CEEMDAN makes experiments on the signal $s(n) + \varepsilon_0 v^i(n)$ and decomposes the original sequence into $\widetilde{IMF}_1$, which is the first IMF. Besides, the values of trial and epsilon need to be set in advance.

Define that $E_k()$ is the k_{th} model function decomposed by EMD and $\widetilde{IMF}_k$ is the k_{th} intrinsic model function decomposed by CEEMDAN. The algorithm implementation of CEEMDAN is as following:

Step1: When k is equal to 1, evaluate the first only residual signal $e_1(n)$:

$$\widetilde{IMF}_1 = \frac{1}{I}\sum_{i=1}^{I} IMF_1^i \tag{1}$$

$$e_1(n) = s(n) - \widetilde{IMF}_1 \tag{2}$$

Step2: Continue experimenting for I times ($i = 1, 2, \cdots\cdots, I$) and decompose signal $S'_{(t)}$ in each experiment until to achieve the first model function of EMD. On the basis of it, evaluate the $\widetilde{IMF}_2$, which is the second IMF:

$$\widetilde{IMF}_2 = \frac{1}{I}\sum_{i=1}^{I} E_1\left(S'_{(t)}\right) \tag{3}$$

$$S'_{(t)} = e_1(n) + \varepsilon_1 E_1\left(v^i(n)\right) \tag{4}$$

Step3: Repeat the step 1 and step 2. Until the extremum points of residual signal aren't over two, the algorithm will be terminated and the final residual signal R(n):

$$R(n) = s(n) - \sum_{k=1}^{K} \widetilde{IMF}_k \tag{5}$$

Step4: The original signal sequence s(n) decomposed by CEEMDAN can be finally represented as following:

$$s(n) = \sum_{k=1}^{K} \widetilde{IMF}_k + R(n) \tag{6}$$

where the original signal and noise sequences are marked as s(n) and ε_0 respectively.

2.2. Sample entropy algorithm (SE)

SE is originated from approximate entropy (AE), which is proposed by Richman and Randall (2000). They are both the measures of complexity for the nonstationary time series, whose core idea is detecting the probability of the creation of a new subsequence in a time series. SE can be applied to recombine the features according to the values and generally speaking, the larger of the SE value, the lower sequence autocorrelation is. In the process of feature extraction, SE's appliance improves the prediction accuracy and balances the calculation efficiency.

Calculating SE needs three compositions: the time series with N uniform data points, the m data points of sub-fragments to be cut and the tolerance of similitude r among sub-fragments.

Define $\{X_{(n)}\} = x(1), x(2), \cdots, x(N)$ as a time series. The algorithm implementation of SE is as follows:

Step1: Mark off (N-m + 1) sub-fragments named $X_m(i)$ from $X_{(n)}$ according to the m data points.

Step2: Calculate the distances between $X_m(i)$ and other $(N - m + 1)$ sub-fragments and select the largest distance value, written as d[X(i),X(j)].

$$d[X(i), X(j)] = \max_{k=0,\cdots,m-1}(|x(i+k) - x(j+k)|) \tag{7}$$

Step3: For the given $X_m(i)$, count the number of $j(1 \le j \le N - m, j \ne i)$ when $d[X(i),X(j)]$ is less than or equal to r and this number is written as B_i:

$$B_m^i = \frac{\text{number of } X(j) \text{ such that } d[X(i), X(j)] \le r}{N - m}, i \ne j \tag{8}$$

Step4: Calculate the mean value of B_i^m and record it as $B^m(r)$

$$B^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} B_i^m \tag{9}$$

Step5: For the label $k = m + 1$, calculate $A^k(r)$ by repeating step 2 to 4.

Step6: According to the above calculation, the final sample entropy can be expressed as $SampEn(m,r,N)$:

$$SampEn(m, r, n) = -\ln\left[\frac{A^k(r)}{B^m(r)}\right] \tag{10}$$

In general, $m = 1$ or $m = 2$ and $r = 0.1$~$0.25SD$, where SD represents the standard deviation of the original series.

2.3. Long short-term memory (LSTM)

RNN applies to the issues of machine learning for time series, unlike common neural network, it has the memory function. However, the existent problems of gradient disappearance and explosion have led the network not to retain information over a long period. To deal with this drawback, Hochreiter and Schmidhuber (1997) came up with the LSTM algorithm.

The structure of LSTM is similar to RNN, whereas it's more complicated for filtering and controlling the information flow through the parts of gating. The core of gating mechanisms contains one cell and three gates, where the amount of information that will be forgotten or remembered is decided by selecting old information and adding new information respectively. Fig. 1 briefly shows the basic structure of RNN and improved memory unit structure of LSTM with a memory cell and three gates, which determines the cell state.

Suppose that x_t and s_{t-1} are actual values as input and output from the last moment respectively, at time t, i_t, f_t and o_t represent the information of the three gates and $\widetilde{c}_t$ represents a new candidate value that may be added to cell state. The calculations for each part are mainly performed by following formulas:

$$i_t = \sigma(A_i x_t + U_i s_{t-1} + B_i) \tag{11}$$

$$f_t = \sigma(A_f x_t + U_f s_{t-1} + B_f) \tag{12}$$

$$o_t = \sigma(A_o x_t + U_o s_{t-1} + B_o) \tag{13}$$

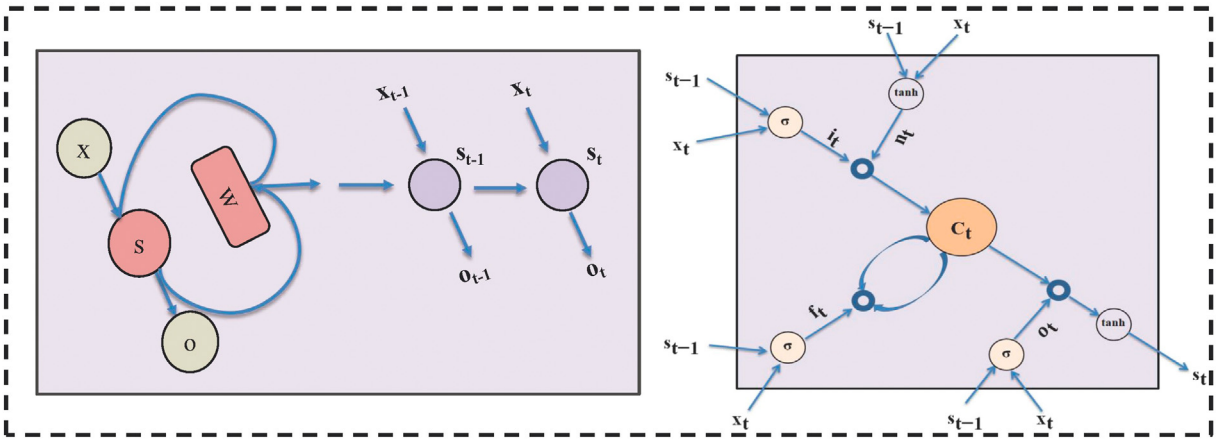


Fig. 1. Basic structures of RNN and LSTM.

$$\tilde{c}_t = tanh\left(A_{\tilde{c}} x_t + U_{\tilde{c}} s_{t-1} + B_{\tilde{c}}\right) \quad (14)$$

$$c_t = c_{t-1} * f_t + \tilde{c}_t * i_t \quad (15)$$

$$h_t = \tanh(c_t) * o_t \quad (16)$$

where the indices of weight are represented by A and U, bias vector is represented by B, '*' represents the scalar product. Tanh is an activation used for generating the new candidate and restricting the scope. σ in Fig. means the gating sigmoid activation, where the output value is adjusted into the range of [0,1]. Besides, as the output value increases,

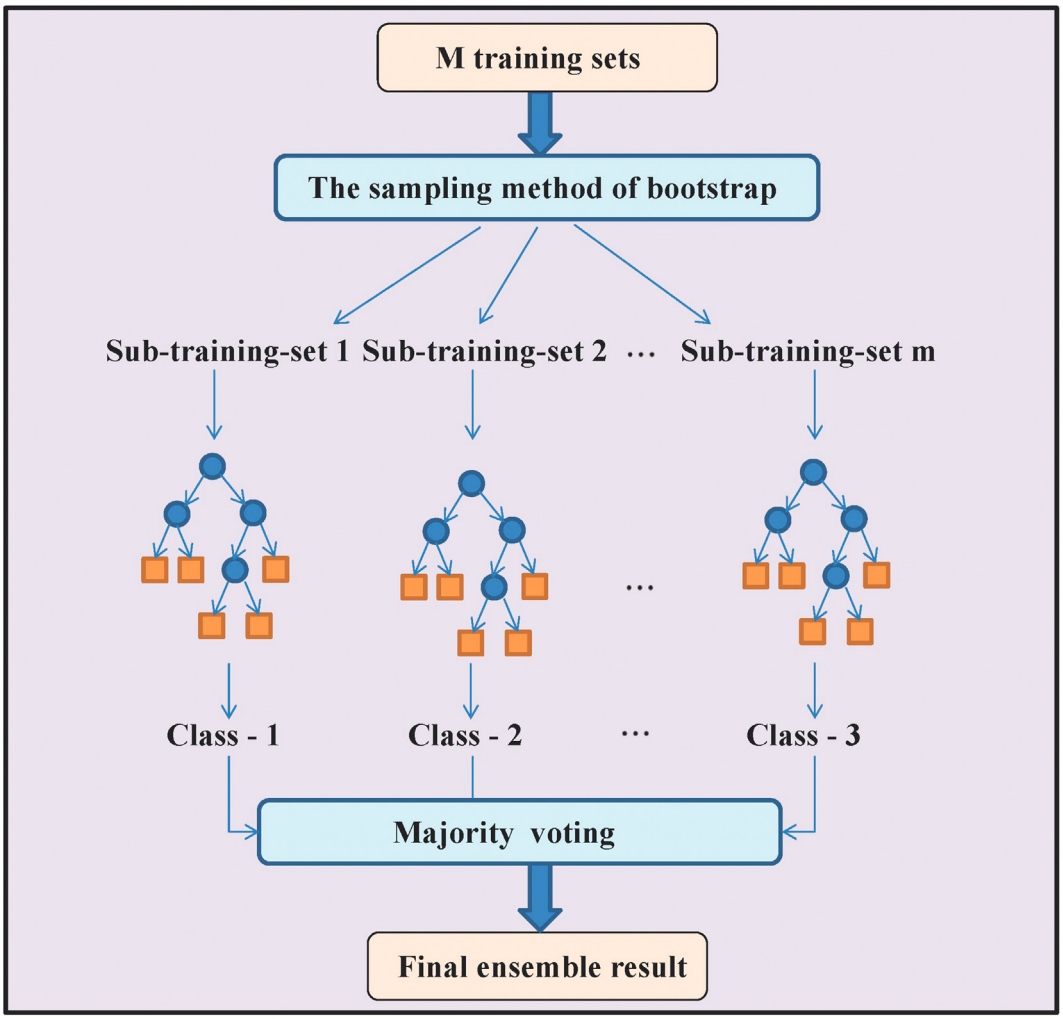


Fig. 2. Model structure of RF.

more information is allowed. For this function of reflecting selective information, LSTM has a superior capacity to learn longer temporal patterns.

2.4. Random forest (RF)

Ensemble learning can be mainly classified into three categories: boosting, bagging and stacking.

Random forest, obtaining the final output with combining many decision trees to form forests, is improved and developed from a bagging algorithm by Breiman (2001). Firstly, the CART tree is selected as its weak learner. Secondly, compared to the common decision tree, RF selects an optimal feature as the classification standard of right and left subtrees by choosing a part of features, not all, on a mode randomly, which improved the generalization ability of the model for robustness and strength. Furthermore, the Gini coefficient, as the standard of selecting the optimal feature, is defined as (Huang et al., 2019b):

$$\mathrm{Gini(e)} = 2e - 2e^2 \tag{17}$$

where Gini represents the Gini coefficient.

Fig. 2 displays the basic structure of RF and suppose that input is the sample set $D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}$, the iterations are T times ($t = 1, 2, \ldots, T$) and the output is the final strong classifier f(x). The algorithm implementation is as follows:

Step1: Process the t_{th} time random sampling for a sample set to achieve D_t comprising m samples.

Step2: Use D_t to train the t_{th} decision tree model. When training the node of a decision tree, select the part of characters from all for the optimal one to classify the right and left subtrees.

Step3: Final output will be obtained by the arithmetic mean of the regression results from weak learners.

2.5. Proposed model

Three stages are included in this novel hybrid model. Improved feature extraction is the first stage, where CEEMDAN is utilized to decompose the original carbon price series into several IMFs and a residual and then decomposed modes with similar complexity are recombined according to the SE algorithm. In the second stage, forecasting models are established through LSTM for each sub-sequences prediction after feature extraction. Nonlinear ensemble learning as the final stage, one of the most effective bagging algorithms named RF is utilized to aggregate the fitted subsequences through LSTM for achieving the end prediction values. The flowchart of the proposed model is shown in Fig. 3.

3. Case analysis

3.1. Data sources

This paper selects two carbon trading markets in China including Tianjin and Guangdong for case studying. The carbon price data sets are downloaded from China carbon trading data network (http://www.ktanjiaoyi.com/). The time span of Tianjin ranges from December 26, 2013 to May 26, 2020, besides that, Guangdong ranges from December 20, 2013 to April 27, 2020. There are 1578 and 1566 data respectively. Both two carbon markets were established early and their development is relatively mature. The price trends of two carbon trading markets and their respective geography positions are shown in Fig. 4, which are in the form of line charts, boxplots and histograms. Meanwhile, the descriptive statistics like mean, maximum, quantile, minimum, and standard deviation of sample data are shown in Table 1. Additionally, both datasets are divided into training sets and test sets, which account for 75% and 25% respectively.

Moreover, in the neural network, the dimensionless data can not only improve the performance of the model to a certain extent, but also increase the calculation efficiency to let the model run faster. In this study, the normalization method will be adopted to preprocess the data and the specific formula is shown as following:

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{18}$$

where x^* represents the normalized data, x represents the original data, and x_{max}, x_{min}represent the maximum and minimum value respectively.

3.2. Nonstationary and nonlinear test for data sets

In order to verify the characters of nonstationary and nonlinear towards original carbon price series, Augmented Dickey-Fuller (ADF) and Brock-Decher-Scheikman (BDS) are applied to make tests, which are the most common methods in this field. The consequences of these two tests for both datasets in Tianjin and Guangdong are respectively shown in Table 2 and Table 3. As are vividly shown, two original carbon price series are both non-stationary within the 1% level. Meanwhile, as the values of probability are smaller than 1%, the data from Tianjin and Guangdong are nonlinear as well. Above all, the sample data can be classified as non-stationary and nonlinear data.

3.3. Improved feature extraction

3.3.1. Decomposition of original carbon price series via CEEMDAN

Before carrying out CEEMDAN on the carbon price series, two parameters need to be set in advance. One is trials, meaning the number of trials and the other is epsilon, which represents self-adapting noise (ε) by multiplying std. σ. And after many times' attempts, the values of the trials and epsilon are set to 100 and 0.05 respectively. The decomposition results and correlation coefficients of each sub-sequence are intuitively shown in Figs. 5 and 6 where each decomposed modes are named as IMFi ($i = 1,2, \ldots,7$) and Residual respectively. It can be found that the original carbon price series of Tianjin, as well as Guangdong, are decomposed into eight sub-sequences including seven IMFs and one residual series. Beyond that, it can also be found that the amplitudes become higher as the procession of decomposition is continuing. The specific descriptive statistics of each sub-sequence including mean and variance are similarly shown in Table 4.

3.3.2. Reconstruction of decomposed sub-sequences via SE

As previously mentioned, the sample entropy can measure the complexity of time series. In order to improve the accuracy of prediction and efficiency of calculating, the sub-sequences with close sample entropy value are recombined according to the theory of SE.

The line charts in Figs. 7 and 8 show the sample entropy results of each sub-sequence. It can be concluded from the graphs that for Tianjin decomposed sub-sequences, (IMF1, IMF2, IMF3, IMF4), (IMF5, IMF6) and (IMF7, Residual) can be reconstructed to S-imf1, S-imf2 and S-imf3, which is shown in right tendency chart. Meanwhile, for Guangdong, (IMF1, IMF2, IMF3), (IMF4, IMF5) and (IMF6, IMF7, Residual) can be reconstructed into three sub-sequences including S-imf1, S-imf2 and S-imf3.

3.4. Accuracy assessment

In this paper, the mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) are applied to evaluate the numerical accuracy of prediction performance. The smaller the indicator, the better the forecast will be. The formulae of these indicators are listed as follows:

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i| \tag{19}$$

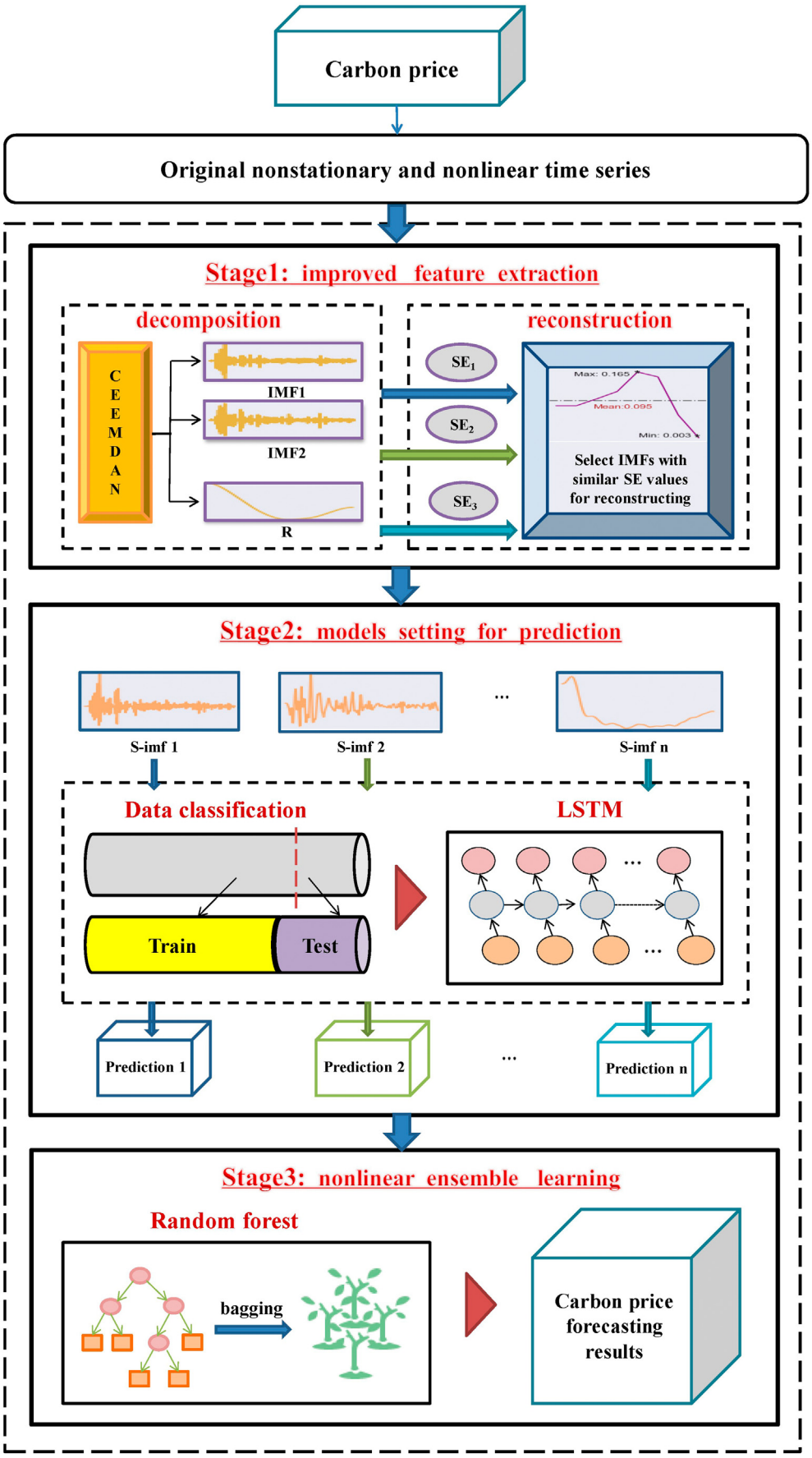


Fig. 3. Flowchart of the proposed model.

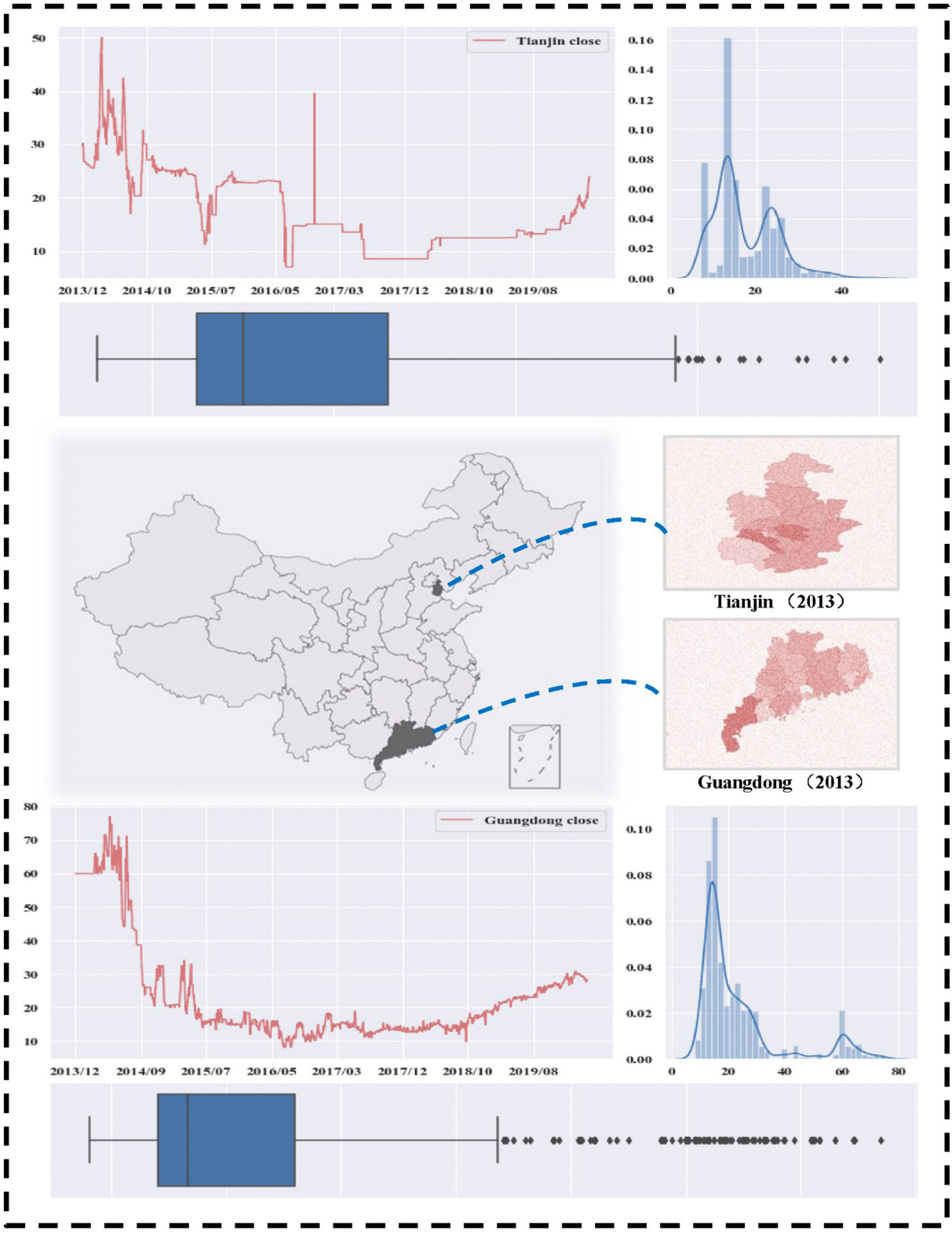


Fig. 4. Daily carbon price tendencies and geographic positions of Tianjin and Guangdong.

Table 1
The basic descriptive statistics of sample data sets.

	Mean	Standard deviation	Minimum	Lower quantile	Median	Upper quantile	Maximum	Skewness	Kurtosis
Tianjin	17.533	4.087	7.000	12.500	15.050	23.060	50.100	0.898	0.689
Guangdong	23.181	14.885	8.100	14.103	16.720	26.000	77.000	1.889	2.587

Table 2
Test results of ADF.

Carbon prices	t-Stat	Prob.	Stationary
Tianjin	−2.575046	0.0984	✘
Guangdong	−2.546157	0.1048	✘

Note: ✘ means that carbon price is non-stationary at the 1% level.

Table 3
Test results of BDS.

Carbon prices	m-dimensional space							
	2		3		⋯	6		linear
	Z-Stat	Prob.	Z-Stat	Prob.	⋯	Z-Stat	Prob.	
Tianjin	118.83	0.00	127.57	0.00	⋯	174.95	0.00	✘
Guangdong	62.12	0.00	66.31	0.00	⋯	88.53	0.00	✘

Note: ✘ shows that carbon price is nonlinear at the 1% level.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{m}} \tag{20}$$

$$\mathrm{MAPE} = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\% \tag{21}$$

where y_i represents the true value, $\hat{y}_i$ represents the predicted value obtained by the model and m is the number of samples.

4. Results and discussions

4.1. Forecast results of the carbon price in Tianjin and Guangdong

4.1.1. Models setting for forecasting through LSTM

Different time sequences have their own appropriate prediction models, which are distinguished by different values of parameters. In the LSTM algorithm, there are several parameters including epoch, batch size, neurons, dropout, dense, activation and so on. In order to achieve the most accurate forecasting output, choosing the best parameters is of vital importance for neural network training. After dividing the data sets into training and testing parts, where training sets account for 75% and the remaining 25% belong to testing sets, many times' trails have been made for finding out the optimal parameters. At last, look back equals to 3, batch size equals to 30, neurons equal to 100, dropout equals to 0.25, dense equals to 1 and activation chooses 'relu' for more accurate forecasting. Moreover, the upper limit of the epoch for each experiment is 1000. Hence, the proper prediction models for each reconstructed sub-sequence have been build and the results of model accuracy are shown in Table 5. It's not difficult to see that LSTM has fairly good performance on the reconstructed sub-sequences forecasting.

4.1.2. The final forecast results through the nonlinear ensemble learning

After achieving the prediction results of each recombined sub-series via LSTM, all output prediction series of one city have been put together for the final integration. For instance, the Tianjin data set builds three prediction models including LSTM1, LSTM2 and LSTM3, whose output results will be put together as inputs for random forest ensemble and achieve the final one result. One main parameter named after n_estimators should be set previously, which means the number of trees in the random forest. To some extent, the higher n_estimators is, the performance will be better. However, the model ultimately pursues higher generalization ability, that is, better performance in both the training set and the test set. Blindly increasing n_estimators will make the model more complex and prone to overfitting problems. So we should find out the relatively proper parameter. After many trials, 200 as the final parameter value has been chosen for training. The line charts in Fig. 9 show the true test values and corresponding prediction values of Tianjin and Guangdong respectively. Although some values didn't achieve optimal stimulation effects, the whole tendency has been represented fairly fitting, which also can be seen in the histogram in the right according to three evaluation indicators.

4.1.3. Wilcoxon rank-sum test for the final forecast results

Wilcoxon rank-sum test is used to test whether there are significant differences between original data and forecasted data. Functionally, the Wilcoxon test rank-sum test is completely consistent with the paired sample *t*-test. The difference between them is whether the data series is normally distributed. If the data is normally distributed, paired sample t-test is used; otherwise, Wilcoxon rank-sum test is used.

First of all, we tested the normality of the data in Guangdong and Tianjin, and the results showed that they were all non-normally distributed. Therefore, we adopted the Wilcoxon test rank-sum test as the statistical significance test for the final forecasting results. As is vividly shown in the Table 6, at the 1% level, the probability values are both

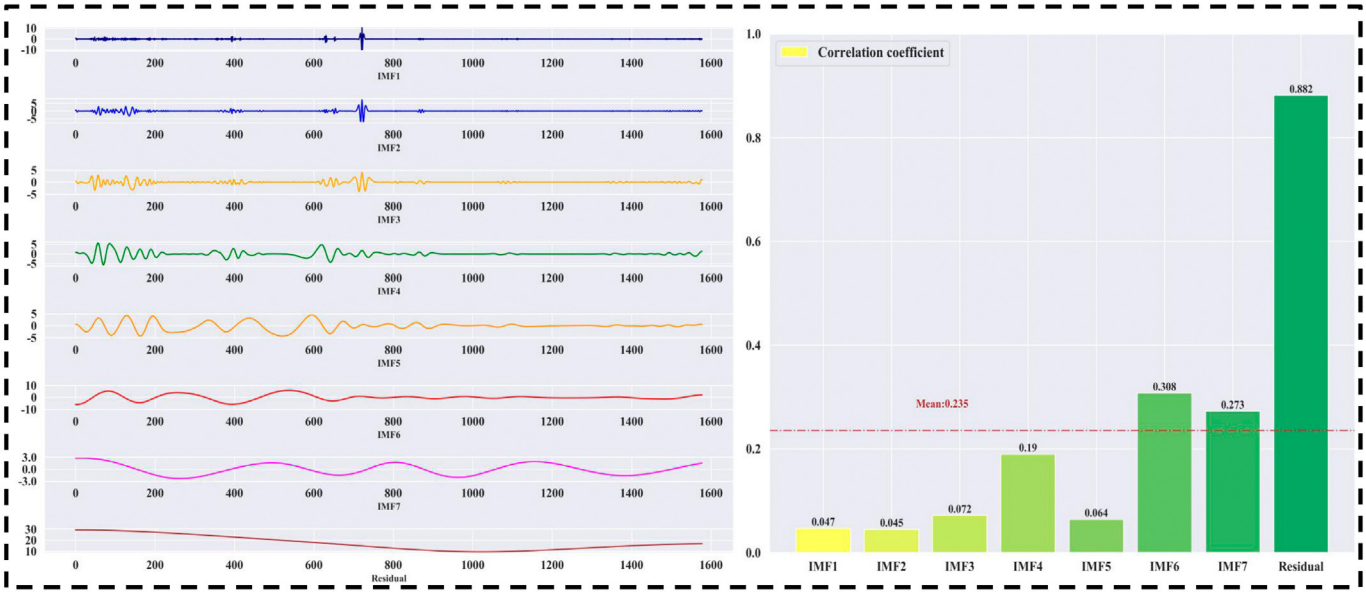


Fig. 5. The decomposition consequence of Tianjin and the correlation coefficients of each sub-sequence.

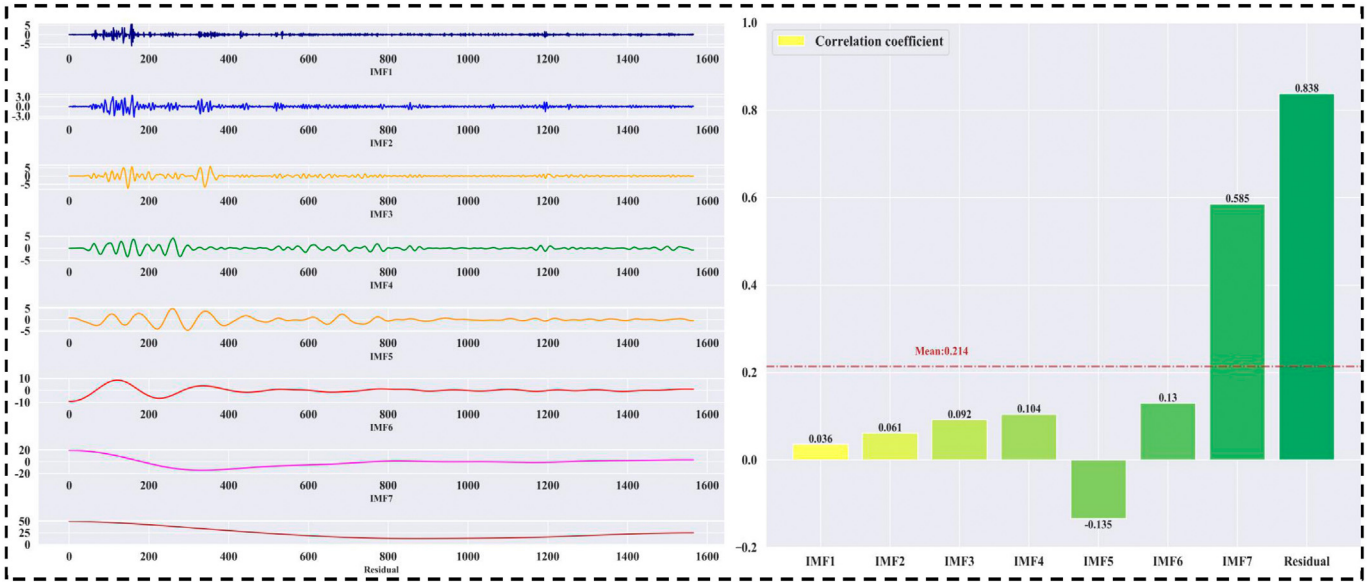


Fig. 6. The decomposition consequence of Guangdong and the correlation coefficients of each sub-sequence.

Table 4
The descriptive statistic of each sub-sequence.

		IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	Residual
Tianjin	Mean	0.009	−0.001	−0.007	−0.001	−0.126	0.073	0.016	17.569
	Variance	0.395	0.422	0.443	1.669	2.772	5.812	1.925	41.589
Guangdong	Mean	−0.003	0.002	0.006	0.013	−0.100	0.025	−0.796	24.034
	Variance	0.351	0.276	1.214	0.881	2.002	6.548	57.374	139.149

beyond it, rejecting the null hypothesis. Therefore, the final forecasting results of Tianjin and Guangdong are both no significant differences between their original data.

4.2. The results for the further three evaluation periods

In order to compare the hybrid model performance over time, we have added the latest three consecutive data as a new validation interval on the basis of the original series, for the reason that prediction model adopts the three steps. And then we achieve the corresponding forecasted result through using appropriate LSTM model established before for three times. Table 7 has shown the three original data, forecasted results and the RMSE of each city. It is vividly shown that the prediction results are roughly in line with the data trend and the RMSE of each city falls within the range of previous indicator. Therefore, the performance of the proposed model is persuasive.

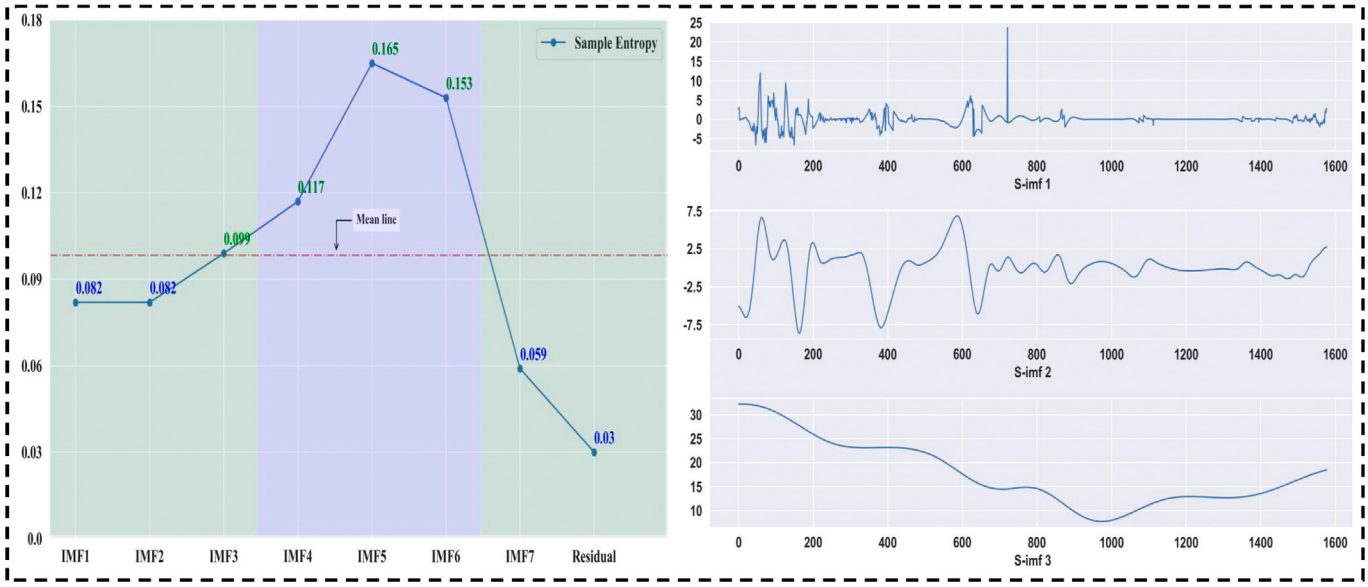


Fig. 7. The sample entropy result of each sub-sequence of Tianjin and reconstruction consequence.

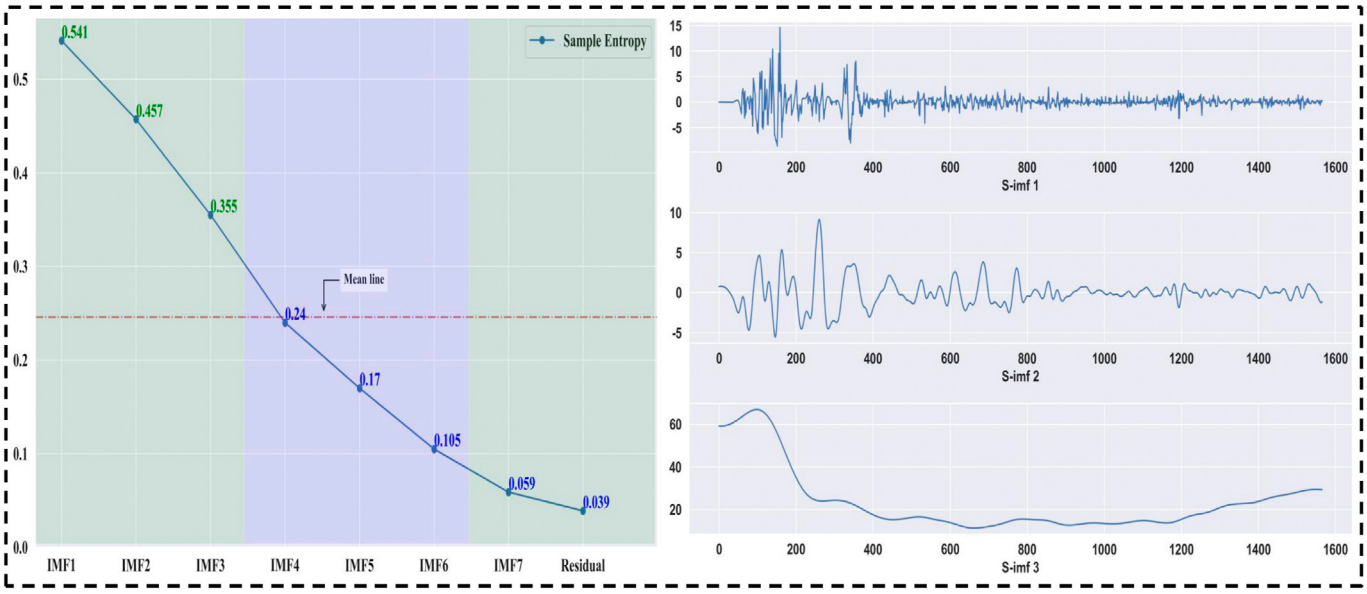


Fig. 8. The sample entropy result of each sub-sequence of Guangdong and reconstruction consequence.

Table 5
The results of model accuracy for each reconstructed sub-sequences.

		LSTM1	LSTM2	LSTM3
Tianjin	MAE	3.044	1.457	0.644
	RMSE	3.235	2.027	0.996
	MAPE	0.229	0.097	0.043
Guangdong	MAE	3.767	3.760	0.717
	RMSE	4.574	4.527	0.949
	MAPE	0.184	0.182	0.239

Table 6
The analysis results of Wilcoxon rank-sum test.

Area	Pairing (median)		Difference	Z-Stat	Prob.	Significant difference
	true	predict				
Tianjin	13.200	13.100	0.100	1.998	0.046	✘
Guangdong	23.020	23.005	0.015	2.946	0.003	✘

Note: ✘ shows that there is no significant difference between the original data and the final forecasting results at the 1% level.

4.3. Carbon price study through the other four benchmark methods

To prove that the hybrid novel model is preponderant, this study proposed four other models as benchmark methods, which include EMD-SE-LSTM-RF, CEEMDAN-LSTM-RF, CEEMDAN-SE-RNN-RF and CEEMDAN-SE-LSTM-XGboost. As the standard of comparison, the more details about the evaluation indexes of MAE, RMSE and MAPE are shown in Table 8.

In the part of feature extraction, two benchmark methods including EMD-SE-LSTM-RF and CEEMDAN-LSTM-RF are proposed to represent

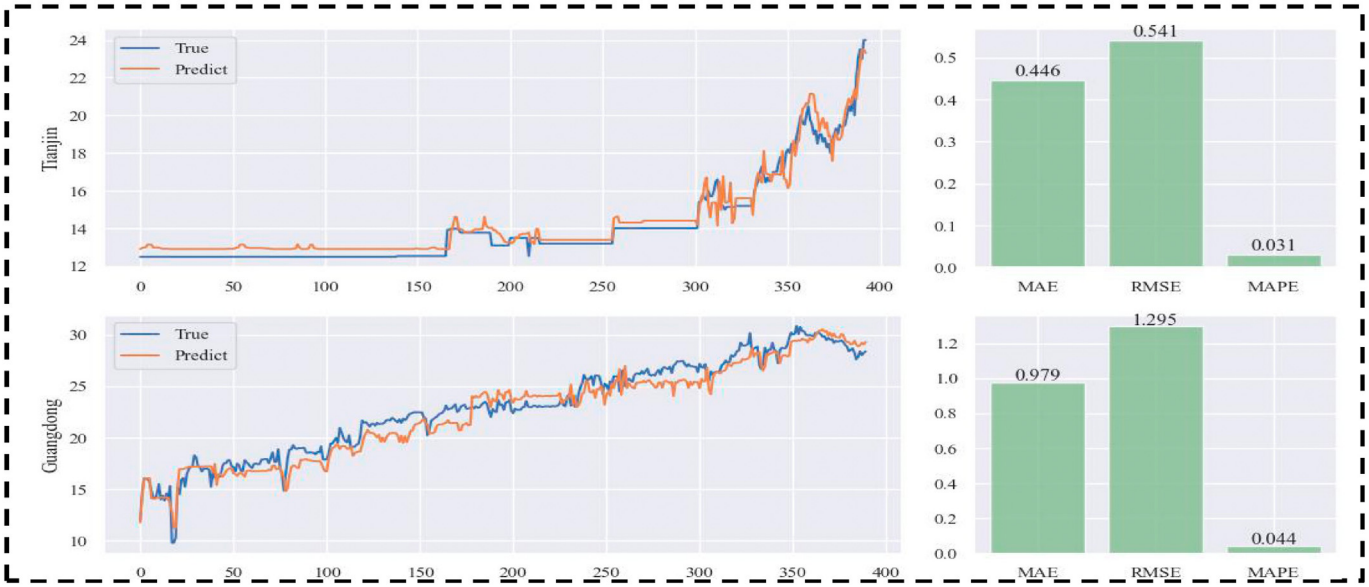


Fig. 9. The final ensemble results and evaluation indicators for carbon prediction in Tianjin and Guangdong.

Table 7
The result of the further three evaluation periods.

Tianjin		May 27, 2020	May 28, 2020	June 1, 2020	RMSE
	True	25.000	25.000	23.000	0.517
	Prediction	24.604	24.423	23.560	
		April 28, 2020	April 29, 2020	April 30, 2020	RMSE
Guangdong	True	27.990	28.140	27.900	0.788
	Prediction	28.826	29.105	28.379	

Table 8
The performance of the comparison models and proposed model.

Area	Model	MAE	RMSE	MAPE (%)
Tianjin	EMD-SE-LSTM-RF	1.192	1.357	0.089
	CEEMDAN-LSTM-RF	1.552	2.161	0.108
	CEEMDAN-SE-RNN-RF	0.659	0.805	0.047
	CEEMDAN-SE-LSTM-XGboost	0.882	1.047	0.063
	CEEMDAN-SE-LSTM-RF	0.446	0.541	0.031
Guangdong	EMD-SE-LSTM-RF	1.779	2.158	0.087
	CEEMDAN-LSTM-RF	1.282	1.858	0.060
	CEEMDAN-SE-RNN-RF	2.964	3.229	0.137
	CEEMDAN-SE-LSTM-XGboost	1.404	1.730	0.069
	CEEMDAN-SE-LSTM-RF	0.979	1.295	0.044

the advantages with a combination of CEEMDAN and SE. On one hand, for the decomposing method, CEENDAN is superior to EMD according to the indexes of EMD-SE-LSTM-RF are nearly twice as much as the hybrid model. On the other hand, improved feature extraction represents its unique strengths. The method without reconstruction needs to establish much more prediction models, which will decrease the accuracy according to the evaluation indexes between CEEMDAN-LSTM-RF and the hybrid model.

In the part of the forecasting method, the comparison between RNN and LSTM is put forward. Taking Tianjin as an example, MAE, RMSE and MAPE achieve 0.659, 0.805 and 0.047 respectively. Even though it represented a relatively good result, the model of CEEMDAN-SE-RNN-RF was still difficult to go beyond the hybrid novel model for the descend range of 32.32%, 32.80% and 34.04%.

In the end, towards the nonlinear ensemble method, this paper compared the boosting ensemble algorithm of XGboost with the bagging ensemble algorithm of RF. It also can be found that the performance of CEEMDAN-SE-LSTM-RF excels CEEMDAN-SE-LSTM-XGboost. Taking MAPE as an example, the hybrid model has improved by 49.43% and 30.27% respectively.

Overall, as is vividly shown in Fig. 10, the remarkable prediction results of proposed model uncover that it is applicable for carbon price prediction.

4.4. Discussions

As the nonstationary and nonlinear time series, improving carbon price prediction accuracy is vital for developing a low carbon economy. This study mainly adopts the data-driving method and sets an innovative random forecast-based nonlinear ensemble paradigm of feature extraction and deep learning algorithm.

It is found that the two parts of the feature extraction stage optimize the prediction model well, especially the recombination through the SE algorithm let the single modes with similar complexity merge for better calculation efficiency. After many times' attempts, the best parameters chosen for the LSTM structure contribute to good intermediate forecast results. Besides, the final nonlinear ensemble learning provided an innovative method for aggregating each sub-sequence's results.

This novel hybrid model is first proposed in the field of carbon price forecast. With the purpose of exploring the internal periodic law with nonstationary and nonlinear series, it has achieved a certain effect. Therefore, it offers a new forecasting tool for enterprises and governments when referring to the topic of carbon trading markets.

5. Conclusions

In this paper, a novel hybrid model is established for nonstationary and nonlinear carbon price series prediction. CEEMDAN-SE-LSTM-RF integrates three parts, which include improved feature extraction, forecasting and ensemble forecast. In the first part, this paper uses CEEMDAN to decompose the carbon price series into several IMFs and one residual with smooth fluctuations, simple structure and strong regularity, which are their constituent components. Supported by SE theory, constituent components with similar complicity are reconstructed as several modes for better prediction accuracy. In the crucial prediction

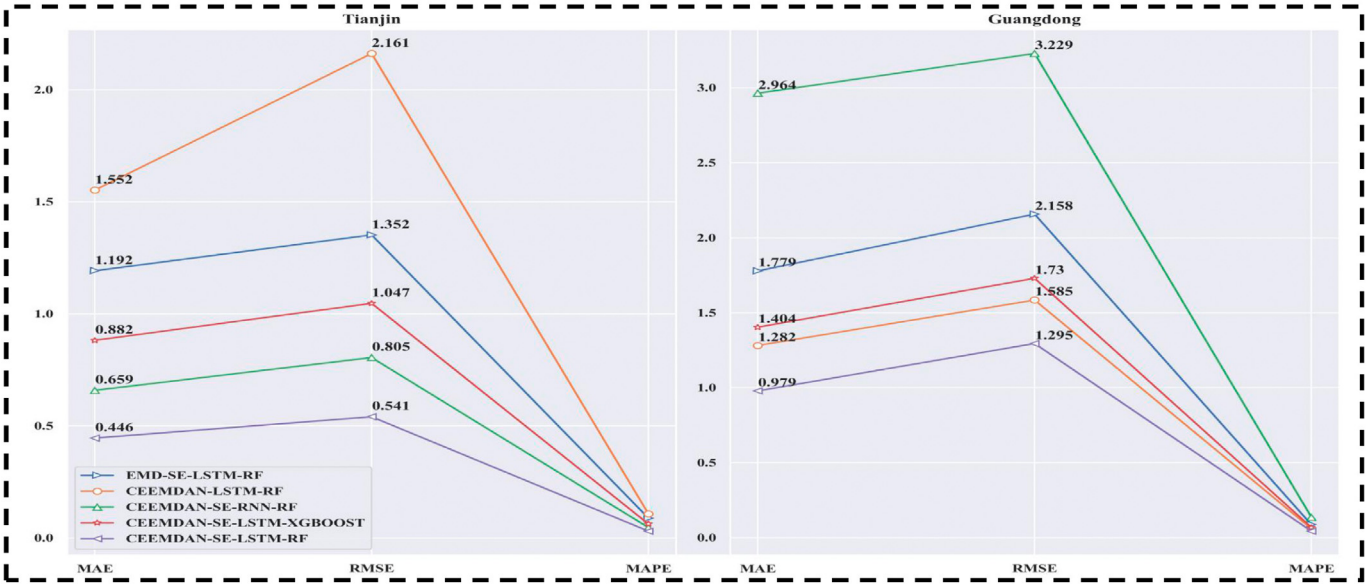


Fig. 10. A visual results presentation between the benchmark models and proposed model.

part, the LSTM algorithm is utilized to capture the long and short term memory, so as to establish an appropriate model for each reconstructed mode. Finally, in the third part, this paper uses RF nonlinear ensemble to aggregate the forecast values and produce the ultima carbon price prediction results. What can be found from the results is that the proposed hybrid model could obtain higher accuracy and robustness predictions compared with those common popular forecast methods. Therefore, it is relatively competitive for predicting the high nonstationary and nonlinear carbon price series.

However, this study only applies data-driving models and doesn't consider the related influencing factors such as crude oil, natural gas prices, technique indexes and forex market and so on. Additionally, suitable individual forecasting models can be selected by an algorithm rather than simple attempts. How to establish optimal proper forecast models can be the topic of further study. Effectively fulfilling these two tasks is deemed to further make the accuracy of nonstationary and nonlinear time series forecast improved. In the future, with the development of technology and the maturity of methods, new methods of signal analysis and machine learning will have been proposed, the proposed model will be updated and tested via various market data.

CRediT authorship contribution statement

Jujie Wang, Xin Sun, Qian Cheng, and Quan Cui: conceived of the presented idea, developed the theory and performed the computations, discussed the results, wrote the paper and approved the final manuscript.

Declaration of competing 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.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant No. 71971122 and 71501101), National Social Science Foundation of China (Grant No. 18ZDA052), and Postgraduate research and innovation plan project of Jiangsu Province (Grant No. SJCX20_0283).

References

Benz, E., Truck, S., 2009. Modeling the price dynamics of CO2 emission allowances. Energ. Econ. 31 (1), 4–15.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Byun, S.J., Cho, H., 2013. Forecasting carbon futures volatility using GARCH models with energy volatilities. Energ. Econ. 40, 207–221.
Cao, J., Li, Z., Li, J., 2019. Financial time series forecasting model based on CEEMDAN and LSTM. Physica A 519, 127–139.
Che, J.X., 2015. Optimal sub-models selection algorithm for combination forecasting model. Neurocomputing 151, 364–375.
Chen, P.A., Chang, L.C., Chang, F.J., 2013. Reinforced recurrent neural networks for multi-step-ahead flood forecasts. J. Hydrol. 497, 71–79.
Chevallier, J., 2011. Wavelet packet transforms analysis applied to carbon prices. Econ. Bull. 31 (2), 1731–1747.
Fan, X.H., Li, S.S., Tian, L.X., 2015. Chaotic characteristic identification for carbon price and an multi-layer perceptron network prediction model. Expert Syst. Appl. 42, 3945–3952.
Fan, J.L., Wu, L.F., Ma, X., Zhou, H.M., Zhang, F.C., 2020. Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions. Renew. Energ. 145, 2034–2045.
Gao, Y., Jian, L.I., 2014. International carbon finance market price prediction based on EMD-PSO-SVM error correction model. China Popul. Resour. Env. 24, 163–170.
García-Martos, C., Rodríguez, J., Sánchez, M.J., 2013. Modelling and forecasting fossil fuels, CO2 and electricity prices and their volatilities. Appl. Energ. 101, 363–375.
Gilles, J., 2013. Empirical wavelet transform. IEEE T. Signal Proces. 61 (16), 3999–4010.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.
Huang, Y.C., He, Z., 2020. Carbon price forecasting with optimization prediction method based on unstructured combination. Sci. Total Environ. 725.
Huang, N.E., Shen, Z., Long, S.R., Wu, C.M., Shi, H.H., Zheng, Q.A., Yen, N., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. P. Math. Phys. Eng. S. 454, 903–995.
Huang, Y.S., Shen, L., Liu, H., 2019a. Grey relational analysis, principal component analysis and forecasting of carbon emissions based on long short-term memory in China. J. Clean. Prod. 209, 415–423.
Huang, G.M., Wu, L.F., Ma, X., Zhang, W.Q., Fan, J.L., Yu, X., Zeng, W.Z., Zhou, H.M., 2019b. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. J. Hydrol. 574, 1029–1041.
Krishan, M., Jha, S., Das, J., Singh, A., Goyal, M.K., Sekar, C., 2019. Air quality modelling using long short-term memory (LSTM) over NCT-Delhi, India. Air qual. Atmos. Hlth. 12 (8), 899–908.
Li, Y., 2020. Forecasting Chinese carbon emissions based on a novel time series prediction method. Energy Sci. Eng. 8, 2274–2285.
Liao, G.C., Tsao, T.P., 2006. Application of a fuzzy neural network combined with a chaos genetic algorithm and simulated annealing to short-term load forecasting. IEEE T. Evolut. Comput. 10 (1), 330–340.
Lin, T., Horne, B.G., Giles, C.L., 1998. How embedded memory in recurrent neural network architectures helps learning long-term temporal dependencies. Neural Netw. 11 (5), 861–868.
Liu, H., Shen, L., 2019. Forecasting carbon price using empirical wavelet transform and gated recurrent unit neural network. Carbon manag 11 (1), 25–37.
Patel, J., Shah, S., Thakkar, P., Kotecha, K., 2015. Predicting stock market index using fusion of machine learning techniques. Expert Syst. Appl. 42 (4), 2162–2172.
Qin, Q.D., He, H.D., He, L.Y., 2015. A novel decomposition-ensemble based carbon price forecasting model integrated with local poly nominal prediction. Comput. Econ. 55, 1249–1273.
Rather, A.M., Agarwal, A., Sastry, V.N., 2015. Recurrent neural network and a hybrid model for prediction of stock returns. Expert Syst. Appl. 42 (6), 3234–3241.
Richman, J.S., Randall, M.J., 2000. Physiological time-series analysis using approximate entropy and sample entropy. Am. J. Physiol. Heart Circ. Physiol. 278, 2039–2049.
Song, Y.Z., Liang, D.P., Liu, T.S., Song, X.Q., 2018. How China's current carbon trading policy affects carbon price? An investigation of the Shanghai Emission Trading Scheme pilot. J. Clean. Prod. 181, 374–384.
Song, Y.Z., Liu, T.S., Liang, D.P., Li, Y., Song, X.Q., 2019. A Fuzzy Stochastic Model for Carbon Price Prediction under the Effect of Demand-Related Policy in China's Carbon Market.
Sun, W., Wang, Y.W., 2020. Factor analysis and carbon price prediction based on empirical mode decomposition and least squares support vector machine optimized by improved particle swarm optimization. Carbon Manag 11 (3), 315–329.
Sun, W., Zhang, C.C., Sun, C.P., 2018. Carbon pricing prediction based on wavelet transform an K-ELM optimized by bat optimization algorithm in China ETS: the case of Shanghai and Hubei carbon markets. Carbon Manag 9 (6), 605–617.
Sun, W., Sun, C.P., Li, Z.Q., 2020. A hybrid carbon price forecasting model with external and internal influencing factors considered comprehensively: a case study from China. Pol. J. Environ. Stud. 29, 3305–3316.
Tian, C.S., Hao, Y., 2020. Point and interval forecasting for carbon price based on an improved analysis-forecast system. Appl. Math. Model. 79, 126–144.
Wang, J.J., Li, Y.N., 2018. Multi-step ahead wind speed prediction based on optimal feature extraction, long short term memory neural network and error correction strategy. Appl. Energ 230, 429–443.
Wu, Z., Huang, N.E., 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv. Adap. Data Anal 1, 1–41.
Wu, Q.L., Liu, Z.T., 2020. Forecasting the carbon price sequence in the Hubei emissions exchange using a hybrid model based on ensemble empirical mode decomposition. Energy Sci. Eng. 00, 1–14.
Wu, Y.X., Wu, Q.B., Zhu, J.Q., 2019. Improved EEMD-based crude oil price forecasting using LSTM networks. Physica A 516, 114–124.
Yang, X., Liang, J., 2017. Analysis and test of fractal and chaotic behavior characteristics of the international carbon emissions market. Syst. Eng. Princ. Pract. 37, 1420–1431.
Yang, S.M., Chen, D.J., Li, S.L., Wang, W.J., 2020. Carbon price forecasting based on modified ensemble empirical mode decomposition and long short-term memory optimized by improved whale optimization algorithm. Sci. Total Environ. 716.
Zhang, Y.J., Wei, Y.M., 2010. An overview of current research on EU ETS: evidence from its operating mechanism and economic effect. Appl. Energ. 87 (6), 1804–1814.
Zhang, J.L., Li, D., Hao, Y., Tan, Z.F., 2018. A hybrid model using signal processing technology, econometric models and neural network for carbon spot price forecasting. J. Clean. Prod. 204, 958–964.
Zhu, B.Z., Wei, Y.M., 2013. Carbon price prediction with a hybrid ARIMA and least squares support vector machines methodology. Omega 41, 517–524.
Zhu, B.Z., Wang, P., Chevallier, J., Wei, Y., 2013. Carbon price analysis using empirical mode decomposition. Comput. Econ. 45 (2), 195–206.
Zhu, B.Z., Shi, X.T., Chevallier, J., Wang, P., Wei, Y.M., 2016. An adaptive multiscale ensemble learning paradigm for nonstationary and nonlinear energy price time series forecasting. J. Forecast. 35 (7), 633–651.
Zhu, B.Z., Han, D., Wang, P., 2017. Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. Appl. Energ 191, 521–530.
Zhu, B.Z., Ye, S.X., Wang, P., He, K.J., Zhang, T., Wei, Y.M., 2018. A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting. Energ. Econ. 70, 143–157.
Zhu, J.M., Wu, P., Chen, H.Y., Liu, J.P., Zhou, L.G., 2019. Carbon price forecasting with variational mode decomposition and optimal combined model. Physica A 519, 140–158.