Heliyon 8 (2022) e12562

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Carbon price prediction based on multi-factor MEEMD-LSTM model

Yang Min *, Zhu Shuzhen, Li Wuwei

Glorious Sun School of Business and Management, Donghua University, Shanghai 200051, China

ARTICLE INFO

Keywords:
Carbon price prediction
Multi-factor
MEEMD
LSTM
Production rule

ABSTRACT

China's national carbon market has already become the largest carbon market in the world. The prediction of carbon price is extremely important for policymakers and market participants. Therefore, the prediction of carbon price in China is of great significance. To achieve a better prediction effect, a multi-factor hybrid model combined with modified ensemble empirical mode decomposition (MEEMD) and long short-term memory (LSTM) neural network optimized by machine reasoning system on the basis of production rules is proposed in this paper. In addition to historical carbon price, other factors, such as energy, macroeconomy, environmental condition, temperature, exchange rate which affect carbon price fluctuation, are formed as multi-factor. The change characteristics of carbon price time series data and other associated factors are extracted in the carbon price prediction. The MEEMD is used to decompose data which is taken as potential input variables into LSTM neural network for prediction and the machine reasoning system based on production rules can automatically search and optimize the parameters of LSTM to further improve the prediction results. The experimental results demonstrate that the proposed method has better prediction effect, robustness and adaptability than the LSTM model without MEEMD decomposition, the single factor MEEMD-LSTM method and other benchmark models. Overall it seems that the proposed method is an advanced approach for predicting the non-stationary and non-linear carbon price time series.

1. Introduction

The International Energy Agency pointed out that due to the factors, such as the sharp rise in natural gas price, the Covid-19 pandemic and the Ukraine crisis, the global carbon dioxide emissions in the energy sector have reached 36.3 billion tons and grown 6 percent year on year to an all-time high in 2021. The carbon emission trading market (hereinafter referred to as the carbon market), which is the product of the Kyoto Protocol and considerable a powerful mean to global warming, encourages all participants to save energy. 24 carbon markets in four continents have been launched by the end of 2019 and more and more governments of countries and regions have seen carbon market as the most important tool to deal with globe climate change. As the largest carbon emitter, China has successively launched pilot carbon markets in Beijing, Tianjin, Shanghai, Chongqing and Shenzhen these five cities and three carbon provinces, Guangdong, Hubei and Fujian. With nearly eight years of development, the cumulative trading volume and turnover exceeded 179 million tons and 7.661 billion ¥Yuan respectively by the end of June 2019. The carbon price varies greatly among carbon markets and fluctuates frequently and violently because of different natural conditions and local policies. In September 2020, President Xi Jinping announced at the general debate of the 75th session of the United Nations General Assembly that China aims to peak carbon dioxide emissions before 2030 and achieves carbon neutrality before 2060. To meet the target of carbon peak and carbon neutrality, China' national carbon market started trading on July 16, 2021 and it has already replaced the European Union's carbon trading market (EU ETS) to the world's largest because of covering over 4 billion metric tons of carbon dioxide a year now. Prediction carbon price accurately cannot only provide a practical guidance for policy-makers to develop an efficient stabilization mechanism for carbon price but also give participants and investors at home and abroad to have a better understanding of the national carbon market [1]. Therefore, how to improve the prediction accuracy has become an urgent issue for Chinese researchers.

However, carbon marketers in China are policy-based artificial markets which cause carbon prices are influenced by internal mechanism, such as the historical carbon price, trading volume and external environmental heterogeneity, for example, macro-economy, energy prices [2, 3] by chevalier and Tan who used different models to prove that there is a strong connection between macro-economy, energy price and carbon price, temperature is regarded as one of the influencing factors using HP Filter by Guo [4], Air Quality index (AQI) proved by Li who believed AQI

* Corresponding author.
E-mail address: 2488187592@qq.com (Y. Min).

https://doi.org/10.1016/j.heliyon.2022.e12562
Received 11 January 2022; Received in revised form 23 April 2022; Accepted 14 December 2022

and carbon price affected each other with a negative direction [5], exchange rate factors confirmed by LV et al [6] and prices in other carbon markets. All of these cause the nonsationary and nonlinear characteristics of carbon price which make it a big challenge for researchers to predict carbon price accurately.

The primary focus of this paper is to put forward a new model to better the prediction performance by taking the carbon price data in Hubei, Shanghai, Beijing, Guangdong and Shenzhen. The reminder of this paper is organized as follows. Section 2 contains a brief literature review and drawbacks. Section 3 introduce MEEMD algorithm, LSTM model and explains the framework of the proposed hybrid prediction approach. Section 4 preprocesses the selected data and the effectiveness and robustness of the proposed model is tested by empirical analysis. Section 5 is about the conclusions and future work.

2. Literature review

According to the existing literatures, there are various methods adopted for carbon price prediction. These methods are mainly divided into two categories: statistical methods and artificial network methods.

Before the rapid development of computers, the most common prediction method for carbon pricing was statistical models, Byun, Maria and Ji used ARMA and GARCH models to predict carbon price on the basis of the assumptions that the carbon price time series is normal and linear [7, 8, 9]. In reality, however, the carbon price time series has the non-stationary and non-linear characteristics, the traditional statistical models usually cannot deal with this kind of situation effectively pointed out by Feng et al. [10].

To get better prediction effect and with the development of machine learning, more and more intelligent algorithms, such as Support Vector Machine (SVM), Markov and Back propagation (BP) neural networks have been introduced to predict complex time series data. They do not need to satisfy statistical hypothesizes and can sensitively capture the hidden nonlinear features in time series. Their strong parallel processing information ability of self-adaptive and self-learning greatly improves the accuracy of prediction results that make up for the shortcomings of the traditional statistical model. Tang et al found the prediction of mixed models with different methods usually has been better than that of a single model [10]. After that, some scholars used combination models to predict carbon price and got a better prediction result. For example, Zhang (2016), Yao (2017), Jiang (2018) and Hu, (2018) proposed combination models Grey-Markov, EMD-SVM and PSO-BP respectively. Their studies all discovered that the prediction accuracy of proposed combination model significantly improved by compared with single model GARCH, Markov, SVM and BP [11, 12, 13, 14].

However, the intelligent algorithms above belong to static network and are difficult to fully reflect the dynamic financial markets in reality which leads room for improvement. Deep learning is considered to be much closer to artificial intelligence because it simulates the mechanism of the human brain, thus making the prediction more accurate. Long Short-term memory (LSTM) is an efficient and recursive iterated neural network of Deep learning [15]. It developed from the Recurrent Neural Network (RNN) meet the requirements of dynamic system because of its memory which means the input of the previous step can be fed back to the input of the next step. Therefore, in recent years, LSTM has been applied to some financial markets other than Language System, Image Recognition and Machine Translation, such as Exchange Rate of Euro against US dollar prediction [16], Stock Price prediction [17] and Commodity Price prediction [18]. These literatures all proved that LSTM had better predictive ability than the existing methods, that is, LSTM is feasible and effective in financial market prediction field.

This study will establish a carbon price prediction model based on a LSTM neural network. However, the application of LSTM has two difficulties. One problem is how to select the appropriate input, that is, single variable or multi-variable. As mentioned above, there are many factors that affect the fluctuation of carbon price, however, almost all existing researches only use the historical carbon price data as the input without considering the possibilities of using multi-variables which will cause errors of prediction results.

The other problem is that LSTM is a typical black box algorithm with many parameters, how can we ensure the optimal selection of them. To improve the parameter settings of the LSTM, researchers have suggested combining. symbol learning, interpretable algorithms with deep learning to form production rules that can provide prior knowledge and solve the learning problems of small samples in a dynamic environment, leading to better expression ability, adaptability and interpretability of the LSTM [19, 20, 21]. Srivastava et al [22] and He et al [23] proposed that the working mechanism and performance of neural network prediction has changed by different topology, control gate and parameter settings of the LSTM. Chen et al [24], Zhang and Zhao [25] and Pei and Zhu [26] argued that the settings and adjustment of parameters, like network layers, neurons and time window, would affect the prediction effect of the LSTM.

Last but not least, empirical mode decomposition (EMD) has been proven to be an effective preprocessing approach to improve the prediction accuracy by decompose the nonlinear original time series data into subseries with high stability and regularity [27]. EMD is self-adaptive. It does not need artificially set any basis function, but decomposes the signal on the basis of its own time-scale characteristics which can effectively solve the problem about strong randomicity of carbon price, reflecting all information as much as possible. It was first successfully applied to natural science and engineering fields such as bioengineering, image processing and atmospheric science, then introduced into the economic field which indicates EMD has a good performance in data analysis and interpretation. In 2012, EMD was first used in the analysis of carbon price by Zhu [28]. Since then, EMD was utilized more often in data preprocess and got good results. But there is still room for improvement, because mode mixing is quite easily appear in EMD algorithm. To solve this problem, Wu et al. proposed EEMD (ensemble empirical mode decompostition) [29] and Yeh et al. proposed CEEMD (complementary ensemble empirical mode decompostition). These two methods can suppress mode aliasing to a certain extent, but increasing computing time and pseudo-components [30]. On the basis of these above EMD family algorithm, Zheng et al (2013) built a modified ensemble EMD (MEEMD)which uses the singalrandomness based on permutation entropy (PE) presented by Bandt to detech high-frequency or intermittent signals [31].. To sum up, MEEMD not only suppresses the mode mixing caused by EMD, but also effectively avoids the unnecessary integrated average in EEMD and CEEMD, reducing the amount of calculation and reconstruction error.

In this paper, we propose a multi-factor hybrid prediction approach which is incorporated MEEMD and LSTM including the internal factors (e.g. opening price, highest price, lowest price and trading volume) and the external factors (e.g. CSI300 index, coal price, crude oil price, carbon price in different carbon markets (e.g. EU carbon price, Shanghai carbon price, Guangdong carbon price, Shenzhen carbon price), AQI, daily temperature and the rate of Euro against renminbi [RMB]). Taking the Hubei carbon price as the research object, the Root Mean Square Error (RMSE), mean absolute error (MAE) and the coefficient of determination (R^2) are introduced to evaluate the prediction effect and the machine reasoning system is used to automatically search and optimize the parameters of LSTM. To prove the robust of the proposed model, the price of Shanghai and Guangdong are taken as the complementary cases. To confirm the validity of the proposed model, it is compared with LSTM and single-factor MEEMD-LSTM. To further improve the prediction effect of the proposed model, a reasoning machine system on automatically searching production rules is used.

The main four innovations and contributions of this paper are listed below:

(a) The application of MEEMD algorithm not only restrain the mode mixing of EMD algorithm effectively, but also decrease the reconstruction error and the computation amount of EEMD and CEEMD, leading to better prediction effect.

(b) Taking into account the historical carbon price data and other influencing factors as the input of LSTM neural network to get a multi-variable combination model to predict carbon price might provide a new idea.

(c) A reasoning machine system on automatically searching production rules is utilized to realize the automatic adjustment and optimization of parameters of the LSTM model which helps further improve the prediction accuracy.

(d) Data from three carbon markets, Hubei, Shanghai and Guangdong are selected for empirical cases by using relevant models to prove the effectiveness and robustness of the proposed model with two error evaluation indicators RMSE, MAE and R^2.

3. Research method

3.1. Construction of the carbon price prediction model

Our model is composed of a main control module and three process modules: the MEEMD multi-scale decomposition, the LSTM multivariable prediction and the machine reasoning system. The multi-scale decomposition by MEEMD can maintain the inherent characteristics of the original data and reduce the noise effectively which helps to ensure the effectiveness of the input data.

We entered data for the carbon price and its influencing factors which are collected from the China Carbon Trading Website (http://www.tanjiaoyi.com) into the LSTM as variables to form a multi-factor prediction model. We calculated the prediction index errors that represent the correlation degree between different variables and the prediction of carbon price and gave priority to the variables with a higher correlation degree. We then sorted the prediction errors of various multivariable combinations, based on the evaluation indexes, to obtain the optimal result. With the support of the machine reasoning system based on production rules, we identified optimized parameters to further improve accuracy of LSTM. Figure 1 shows the overall flow chart of the multi-factor MEEMD-LSTM carbon price prediction model.

3.2. The MEEMD method

Two additional sets of white noise of equal amplitude and standard deviant and parallel to the original signal with opposite directions are added into the original data time series. Then the new signal is composed by EEMD and EMD in order of priority. MEEMD not only makes the Intrinsic Mode Functions (IMFs) have clearer, more complete and independent physical meanings, but also lowers the amount of calculation and reduces the reconstruction error. For a non-stationary signal S(t), the MEEMD algorithm includes six steps:

1. A pair of white noise $n_i(t)$ and $-n_i(t)$ is added to the original signals S(t), as shown in formula (1)

$$S_i^+(t) = S(t) + a_i n_i(t)$$
$$S_i^-(t) = S(t) - a_i n_i(t) \tag{1}$$

where $n_i(t)$ is the added white noise; a_i is the amplitude of added noise,$i = 1, 2, \ldots N_e$ andN_e is the number of pairs of added white noise.

2. $S_i(t)$ and $-S_i(t)$ are decomposed by CEEMD to obtain p number of IMFs Components $\{I_{ip}^+(t)\}$ and $\{I_{ip}^-(t)\}(i = 1, 2, \ldots N_e)$
3. Average $\{I_{ip}^+(t)\}$ and $\{I_{ip}^-(t)\}$ obtain $I_1(t) = \frac{1}{2N}\sum_{i=1}^{N_e}[I_{ip}^+(t) + I_{ip}^-(t)]$ and check whether $I_p(t)$ is an abnormal component.
4. If $I_p(t)$ is an abnormal component, return to step1 until $I_p(t)$ is no longer an abnormal component
5. Separate the first composed P-1 number of components from the original data, as shown in formula (2)

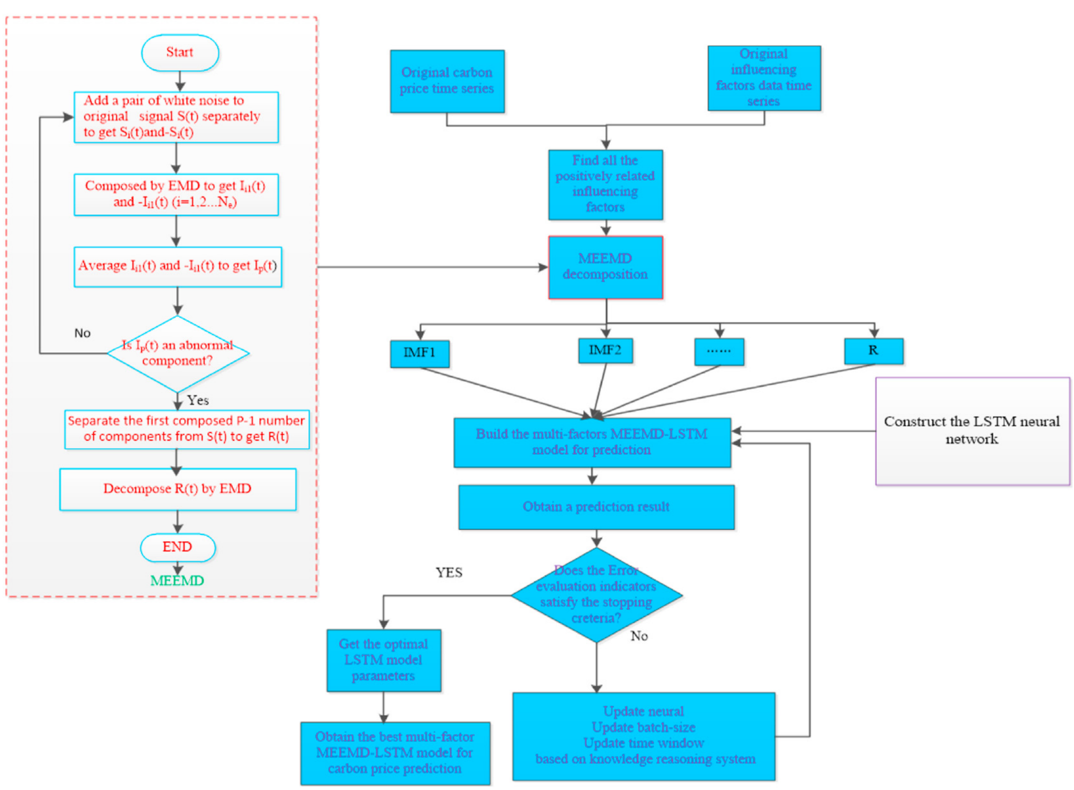

Figure 1. Flow chart of the multi-factor MEEMD-LSTM carbon price prediction model.

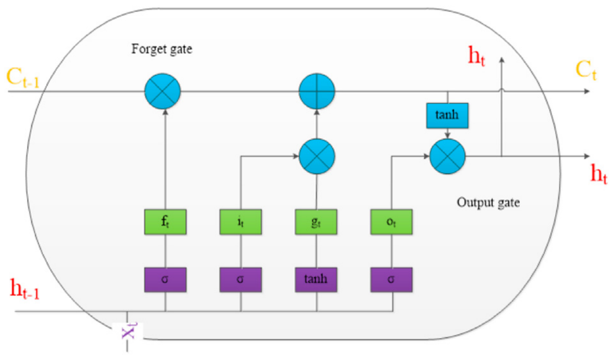

Figure 2. Unit structure of the LSTM.

$$r(t) = S(t) - \sum_{j=1}^{p-1} I_j(t) \tag{2}$$

6. Use EMD to decompose $r(t)$ and arrange the IMFs components from high to low frequency.

3.3. LSTM model

LSTM model, first proposed by Hochreiter in 1997, is an improvement of recurrent neural network (RNN) by effectively solving gradient vanishing and gradient exploding. LSTM model introduce a mechanism called "gate" which can selectively add new information and forget the previous information, thus reducing the sequence length and the number of grid layers [32]. Figure 2 shows the unit structure of the LSTM.

The flow of the LSTM are firstly according to the previous external state h_{t-1} and the current input x_t, calculate the gates of forget, input and output, secondly activate forget gate f_t, input gate i_t and g_t to update the current internal state of the memory unit c_t and lastly activate output gate o_t and output the external state h_t and internal state c_t to the next unit at the same time. The specific calculation formulas for each "gate" are shown in formula (3).

$$\begin{aligned} f_t &= \sigma(U_f x_t + V_f h_{t-1} + b_f) \\ i_t &= \sigma(U_i x_t + V_i h_{t-1} + b_i) \\ g_t &= \varphi(U_g x_t + V_g h_{t-1} + b_g) \\ C_t &= C_{t-1} \odot f_t + i_t \odot g_t \\ h_t &= o_t \odot \varphi(C_t) \end{aligned} \tag{3}$$

In formula (3), U_f, V_f, U_i, V_i, U_g and V_g are the state-state weight matrix and state-output weight matrix of forget gate, input gate and output gate; b_f, b_i and b_g are the bias matrices of the forgetting gate, input gate and output gate respectively; $\odot$ is the Hadamard product and $\sigma(\cdot)$ and $\varphi(\cdot)$ are activation functions of sigmoid and tanh respectively.

3.4. Machine reasoning method based on production rules

The production rule is one of the most widely used knowledge representation methods in artificial intelligence [33]. The main advantages of machine reasoning system based on production rules are the simplicity of expressing uncertain knowledge, natural expression, clear control structure, strong modularity and good consistency. The steps are as follows:

1. Initialize the integrated database and input the known facts of problems to be resolved in the database;
2. Check the rule base to see if there are any unused rules whose premises can match the known facts in the integrated database. If so, then select; otherwise, go to step 5;
3. Execute the selected rule and mark it. If we obtain conclusions, then store them in the integrated database as new facts. If we obtain operations, then execute them;
4. Check whether the solution of the problem is included in the integrated database. If included, then the problem has been solved and the solving process is over. Otherwise, go to step 2;
5. When there are still unused rules in the rule base and none of them can match the existing facts in the integrated database, then more known facts are required. If more known facts can be provided, go to step 2. Otherwise, the problem has no solution and the solving process is terminated;
6. If there are no more unused rules in the knowledge base, the problem has no solution, and the solving process is also terminated.

3.5. Prediction error indexes

In this paper, we use RMSE, MAE and R^2 to evaluate the prediction effect. The calculation formulas are shown in (4), 5 and (6).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \tag{4}$$

$$MAE = \frac{1}{N} = \sum_{i=1}^{N} |y_t - \hat{y}_t| \tag{5}$$

where $\hat{y}_i$ is the predicted value and y_i is the actual value

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{6}$$

where $SS_{res} = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$ and$SS_{tot} = \sum_{i=1}^{n} (y_i - \overline{y}_i)^2$

RMSE reflects the deviation between the predicted and actual value, the smaller the RMSE, the higher the prediction accuracy. We used R^2 to show the correlation between the input and output variables. In general, when R^2 is less than 0.85, there is little correlation between the input and output variables. On the contrary, when R^2 is greater than or equal to 0.85, there is correlation between the input and output variables. Furthermore when R^2 is greater than 0.95, there is strong correlation between input and output variables. Therefore, when the prediction results had smaller RMSE values and larger R^2 values, we considered that the prediction model was effective.

4. Data description and preprocessing

4.1. Data description

The dataset consist of carbon prices and other influencing factors from energy price, macroecnomic, temperature, environmental condition and exchange rate. The carbon prices in China and EU ETS are all available from the Tanpaifang website. Oil price (Brent Crude

Table 1. The abbreviations of the dataset.

Item	Abbreviation	Item	Abbreviation	Item	Abbreviation
Hubei carbon price	HB	Shanghai carbon price	SH	Beijing carbon price	BJ
Guangdong carbon price	GD	Shenzhen carbon price	SZ	EU ETS carbon price	EU
Internal factor	IF	CSI 300 Index	HZ300	Crude oil price	oil
Thermal coal price	coal	Exchange rate of EURO against RMB	HL	Temperature	TEMP
Air quality index	AQI				

Table 2. Proportion of three carbon markets.

Item	Hubei	Shanghai	Guangdong	Total
Trading volume	32.4%	7.2%	32.1%	71.7%
Turnover	28.8%	8.8%	27.1%	64.7%

price) and coal price (daily continuous thermal coal futures settlement price) are utilized to represent energy price. This paper selects Shanghai-Shenzhen 300 index to stand for the macroeconomic. Because of the correlation between the carbon price in EU ETS and the carbon price in China, this paper chooses exchange rate of EURO against RMB. All the above are obtained in Wind database. As for the temperature and environmental condition, this paper selects the daily average temperature and air quality index in Hubei, Shanghai, Beijing, Guangdong and Shenzhen which are available in Tianqi2345 website.

Table 1 shows the abbreviations of the dataset.

The economic structure and industrialization degree of Hubei Province approximate the current average level of China and the trading volume, turnover and other indicators of Hubei carbon market are in a leading position. We took 1561 daily closing price data from April 2, 2014 to November 13, 2020 as a case for empirical analysis. Every sample set is divided into a training set and a test set. The first 1356 data before 2020 were used to train the model and the rest data during 2020 were used to test the prediction performance of the model. To prove the robustness of the model, carbon price of Shanghai and Guangdong market are selected as supplementary cases. The reason for choosing these above two markets is that the trading volume, turnover of Guangdong carbon market are second only to Hubei carbon market and the trading platform and settlement system of national carbon market has been built in Shanghai. As shown in Table 2, the trading volume and turnover of the three carbon markets have reached to 71.7% of the total trading volume and 64.7% of the total turnover of the carbon markets. So to a great extent, these three carbon markets can represent the overall situation of China's carbon market. Figure 3 shows the price fluctuation trend and partial autocorrelation analysis results of the three carbon markets. The graphs indicate the carbon price is nonlinearity, complexity and non-stationarity.

5.2. Data preprocessing

Because the dimensions of the variables were inconsistent, we normalized all samples first which made the step size of each parameter decrease with the gradient corresponding to its order of magnitude, this was done in order to improve the convergence speed and the prediction accuracy of the model. Formula (7) shows the computational process of normalization.

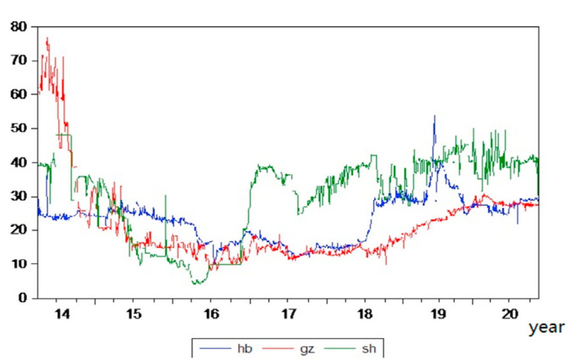

Figure 3. Price fluctuation trend and partial autocorrelation analysis results of the three carbon markets.

Table 3a. Results of the LSTM.and EMD-LSTM.

Different combination of inputs	LSTM			EMD-LSTM		
	RMSE	MAE	R^2	RMSE	MAE	R^2
HB	0.1262	0.0862	0.5554	0.1200	0.0886	0.5935
HBIF	0.0324	0.0127	0.9414	0.0302	0.0109	0.9472
HBAQI	0.1022	0.0683	0.6128	0.0827	0.0600	0.7322
HBHZ300	0.0973	0.0731	0.6278	0.0772	0.0431	0.7528
HBTemp	0.0910	0.0609	0.6541	0.0714	0.0424	0.7549
HBHL	0.0709	0.0597	0.8186	0.0624	0.0405	0.8055
HBEU	0.1088	0.0798	0.6018	0.0869	0.0643	0.7327
HBSH	0.0977	0.0708	0.6237	0.0760	0.0442	0.7953
HBGD	0.0867	0.0641	0.6609	0.0677	0.0495	0.8099
HBSZ	0.1021	0.0773	0.6191	0.0825	0.0598	0.7491
HBBJ	0.0980	0.0705	0.6229	0.0975	0.0674	0.6516
HBoil	0.1065	0.0806	0.6111	0.0842	0.0621	0.7483
HBcoal	0.1124	0.0651	0.5932	0.0986	0.0611	0.6281

$$x = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{7}$$

where x_i is thei^{th} sample value, $x_{\min}$ is the minimum value of the sample, and$x_{\max}$ is the maximum value of the sample. According to formula (7), the data of all prediction indices are from 0 to 1. After normalization, the time series of the carbon closing price and the internal and external factors were considered to be candidate input variables which were dealt with by LSTM to determine the correlation degree, (i.e., to what extent they influenced the target variable).

5. Empirical study

5.1. Case one-Hubei carbon price prediction

We set the learning rate, the time step, number of neurons, batch-size and the time window to 0.001, 5, 100, 100 and 2 respectively. Tables 3a and 3b shows the results of the LSTM with and without MEEMD decomposition.

1) For the same group of input variables, the RMSE value of the LSTM after MEEMD was smaller than that of the LSTM only, that is, the prediction of the MEEMD-LSTM was better than that of the LSTM. Because the carbon price time series was nonlinear and decomposed into several stationary components which led to higher accuracy.

Table 3b. Results of the EEMD-LSTM and MEEMD-LSTM.

Different combination of inputs	EEMD-LSTM			MEEMD-LSTM		
	RMSE	MAE	R^2	RMSE	MAE	R^2
HB	0.0974	0.0662	0.6416	0.0817	0.0590	0.6957
HBIF	0.0258	0.0130	0.9488	0.0277	0.0114	0.9687
HBAQI	0.0817	0.0590	0.8570	0.0743	0.0424	0.7925
HBHZ300	0.0768	0.0548	0.7548	0.0761	0.0570	0.7889
HBTemp	0.0786	0.0674	0.7294	0.0822	0.0710	0.6660
HBHL	0.0617	0.0487	0.8626	0.0611	0.0459	0.8735
HBEU	0.0754	0.0527	0.7698	0.0378	0.0258	0.9422
HBSH	0.0702	0.0501	0.8021	0.0684	0.0474	0.8483
HBGD	0.0695	0.0496	0.8301	0.0692	0.0558	0.8371
HBSZ	0.0792	0.0692	0.7411	0.0642	0.0494	0.8518
HBBJ	0.0956	0.0612	0.6282	0.0940	0.0741	0.6453
HBoil	0.0682	0.0472	0.8544	0.0578	0.0444	0.8826
HBcoal	0.0791	0.0562	0.7412	0.0619	0.0492	0.8697

Three conclusions can be drawn from these results.

Table 4. Results of the multivariable MEEMD-LSTM.

Variables	Prediction evaluation index	
HB_{total}	RMSE	R^2
	0.0208	0.9847

2) The input of two variables, Hubei daily temperature and the Beijing closing carbon price, made the value of RMSE higher than that of single variable HB, and their R^2 values were 0.666 and 0.6453 respectively which were far less than 0.85, indicating no correlation between the two input variables and the target variable.
3) Other than the two variables in Eq. (2), the input of the other variables decreased the value of RMSE and increased the value of R^2 close to or more than 0.85 which indicated higher correlation. Ordering the R^2 from largest to smallest, the internal factors (e.g. opening price, maximum price, minimum price and trading volume) had the highest correlation with the carbon closing price, followed by the EU carbon price, crude oil price, the exchange rate, coal price, the Shenzhen carbon price, the Shanghai carbon price, the Guangdong carbon price, AQI, CSI300 index.

Based on this analysis, all factors listed in Eq. (3) were taken as multivariable input, named HB_{total}, into the MEEMD-LSTM model to predict the carbon price. Table 4 presents the results of the multivariable MEEMD-LSTM model.

As shown in Tables 3a and 3b, with the increase of the effective input variables, the value of RMSE was significantly smaller while the value of R^2 increased. Therefore, the prediction effect of the multivariable MEEMD-LSTM model was significantly improved compared with that of the single-variable LSTM model.

The parameter selection of LSTM also affected the final prediction with the help of a reasoning machine system on automatically searching production rules, it changed the number of neurons, batch size and the time window and investigated the change of the prediction error value, to realize the automatic adjustment and optimization of parameters of the LSTM model.

Figure 4 respectively shows the changes of the RMSE and the R^2 value when one variable is changed and the other two variables remain unchanged. In conclusion, when the batch size, neurons number and the time window were 60, 40 and 2, the RMSE value was the smallest and the R^2 value was the largest.

5.2. Case two-Shanghai carbon price prediction

The actual data of Shanghai carbon market and the external and internal factors affecting its fluctuation are taken as the candidate inputs of the model. Tables 5a and 5b shows the performance comparison results of MEEMD-LSTM and LSTM model based on different combination of inputs.

In Table 6, all relevant effective variables are named as SH_{total} as input into MEEMD-LSTM which has minimum RMSE and maximum R^2.

Figure 5 respectively shows the changes of the RMSE and the R^2 value when one variable is changed and the other two variables remain unchanged.

Table 5a. Results of the LSTM.and EMD-LSTM.

Different combination of inputs	LSTM			EMD-LSTM		
	RMSE	MAE	R^2	RMSE	MAE	R^2
SH	0.1331	0.1050	0.7861	0.1201	0.0987	0.8092
SHIF	0.0948	0.0731	0.8526	0.0752	0.0682	0.8901
SHAQI	0.1146	0.0839	0.8184	0.1006	0.0724	0.8226
SHHZ300	0.0851	0.0646	0.8689	0.0820	0.0700	0.8721
SHTemp	0.0683	0.0119	0.8946	0.065	0.0631	0.9002
SHHL	0.1050	0.0694	0.8195	0.0928	0.0701	0.8320
SHEU	0.1627	0.1280	0.7946	0.1539	0.1155	0.7982
SHcoal	0.1531	0.1012	0.8014	0.1429	0.1007	0.7998
SHGD	0.1554	0.1104	0.8071	0.1414	0.0994	0.8020
SHSZ	0.1196	0.0988	0.8161	0.0948	0.0731	0.8284
SHBJ	0.1378	0.0903	0.7824	0.1196	0.0988	0.8261
SHoil	0.0926	0.0605	0.8899	0.0842	0.0642	0.8396
SHHB	0.1539	0.1155	0.8082	0.1102	0.0911	0.8148

Table 5b. Results of the EEMD-LSTM and MEEMD-LSTM.

Different combination of inputs	EEMD-LSTM			MEEMD-LSTM		
	RMSE	MAE	R^2	RMSE	MAE	R^2
SH	0.0988	0.0774	0.8892	0.0951	0.0143	0.8514
SHIF	0.0638	0.0132	0.9021	0.0564	0.0686	0.9303
SHAQI	0.0928	0.0728	0.8921	0.0819	0.0621	0.8756
SHHZ300	0.0708	0.0167	0.8992	0.0658	0.0466	0.9029
SHTemp	0.0689	0.0149	0.9001	0.0641	0.0270	0.9093
SHHL	0.0823	0.0407	0.8998	0.0866	0.0681	0.8669
SHEU	0.0750	0.0348	0.8548	0.0900	0.0719	0.8616
SHcoal	0.0745	0.0434	0.8821	0.0836	0.0621	0.8789
SHGD	0.0928	0.0754	0.8932	0.1091	0.0838	0.8119
SHSZ	0.0821	0.0402	0.8998	0.0914	0.0638	0.8554
SHBJ	0.0872	0.0601	0.8953	0.0890	0.0605	0.8632
SHoil	0.0913	0.0712	0.8948	0.0846	0.0646	0.8638
HBcoal	0.0729	0.0331	0.9029	0.0619	0.0492	0.8697

Table 6. Results of the multivariable MEEMD-LSTM.

Variables	Prediction evaluation index	
SH_{total}	RMSE	R^2
	0.0470	0.9489

From Tables 5 and 6 and Figure 5, the same analysis can be obtained as Case 1.

5.3. Case three-Guangdong carbon price prediction

The actual data of Guangdong carbon market and the external and internal factors affecting its fluctuation are taken as the candidate inputs of the model. Tables 7a and 7b shows the performance comparison

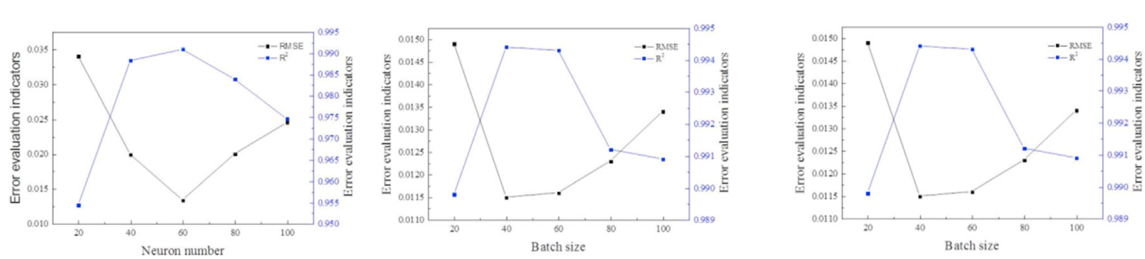

Figure 4. Impact of the batch size, neurons number and time window on prediction of Hubei carbon price.

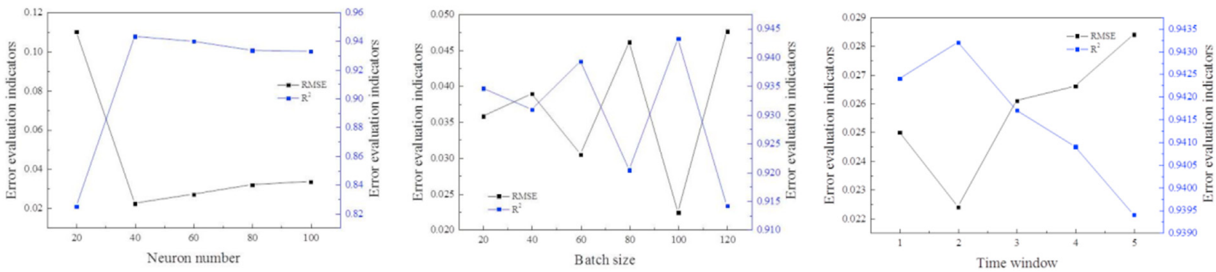

Figure 5. Impact of the batch size, neurons number and time window on prediction of Shanghai carbon price.

Table7a. Results of the LSTM.and EMD-LSTM

Different combination of inputs	LSTM			EMD-LSTM		
	RMSE	MAE	R^2	RMSE	MAE	R^2
GD	0.1090	0.0387	0.8504	0.1024	0.0361	0.8401
GDIF	0.0502	0.0211	0.9683	0.0498	0.0302	0.8998
GDAQI	0.1045	0.0379	0.8565	0.0982	0.0359	0.8545
GDHZ300	0.0743	0.0287	0.9087	0.0741	0.0342	0.8610
GDTemp	0.0625	0.0265	0.8884	0.0611	0.0333	0.8702
GDHL	0.0742	0.0277	0.9088	0.0724	0.0348	0.8667
GDEU	0.1007	0.0317	0.8653	0.0892	0.0352	0.8577
GDcoal	0.1140	0.0556	0.8471	0.0912	0.0350	0.8566
GDSH	0.0938	0.0385	0.8728	0.0900	0.0349	0.8567
GDSZ	0.0910	0.0360	0.8787	0.0871	0.0352	0.8580
GDBJ	0.0975	0.0392	0.8661	0.0912	0.0361	0.8600
GDoil	0.0876	0.0343	0.8801	0.0800	0.0320	0.8623
GDHB	0.1120	0.0501	0.8481	0.1101	0.0372	0.8589

Table7b. Results of the EEMD-LSTM and MEEMD-LSTM

Different combination of inputs	EEMD-LSTM			MEEMD-LSTM		
	RMSE	MAE	R^2	RMSE	MAE	R^2
GD	0.0964	0.0342	0.8589	0.0924	0.0330	0.8685
GDIF	0.0496	0.0300	0.9002	0.0492	0.0299	0.9631
GDAQI	0.0888	0.0352	0.8551	0.0823	0.0407	0.8894
GDHZ300	0.0738	0.0338	0.8612	0.0734	0.0487	0.9101
GDTemp	0.0589	0.0325	0.8734	0.0542	0.0079	0.9347
GDHL	0.0719	0.0354	0.8676	0.0706	0.0483	0.9138
GDEU	0.0790	0.0376	0.8669	0.0670	0.0321	0.9296
GDcoal	0.0898	0.0350	0.8544	0.0861	0.0462	0.8858
GDSH	0.0855	0.0341	0.8601	0.0873	0.0661	0.8809
GDSZ	0.0700	0.0331	0.8631	0.0743	0.0453	0.9086
GDBJ	0.0900	0.0357	0.8592	0.0911	0.0424	0.8787
GDoil	0.0789	0.0377	08669	0.0755	0.0414	0.8994
GDHB	0.1020	0.0387	0.8422	0.1050	0.0502	0.8564

results of MEEMD-LSTM and LSTM model based on different combination of inputs.

In Table 8, all relevant effective variables are named as GD_{total} as input into MEEMD-LSTM which has minimum RMSE and maximum R^2.

Table 8. Results of the multivariable MEEMD-LSTM.

Variables	Prediction evaluation index	
GD_{total}	RMSE	R^2
	0.0384	0.9739

Figure 6 respectively shows the changes of the RMSE and the R^2 value when one variable is changed and the other two variables remain unchanged.

From Tables 7 and 8 and Figure 6, the same analysis can be obtained as Case 1 and Case 2.

Figure 7 shows the optimum prediction curve and the actual value curve using the improved multivariable data-driven MEEMD-LSTM model on Hubei carbon market, Shanghai carbon market and Guangdong carbon market respectively. These results demonstrated that the model proposed in this paper has high prediction accuracy.

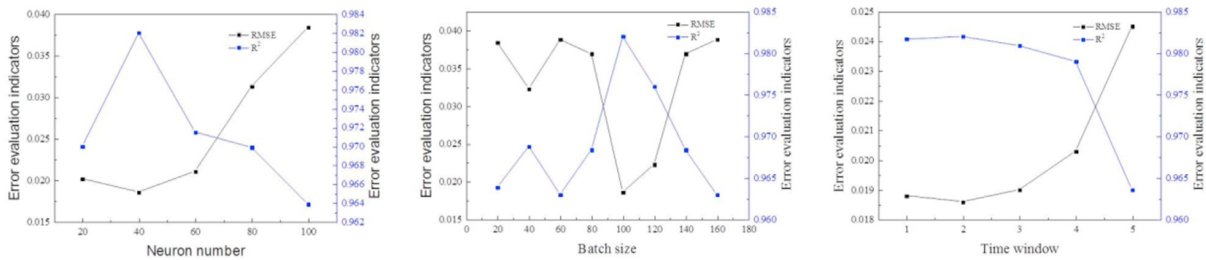

Figure 6. Impact of the batch size, neurons number and time window on prediction of Guangdong Carbon price.

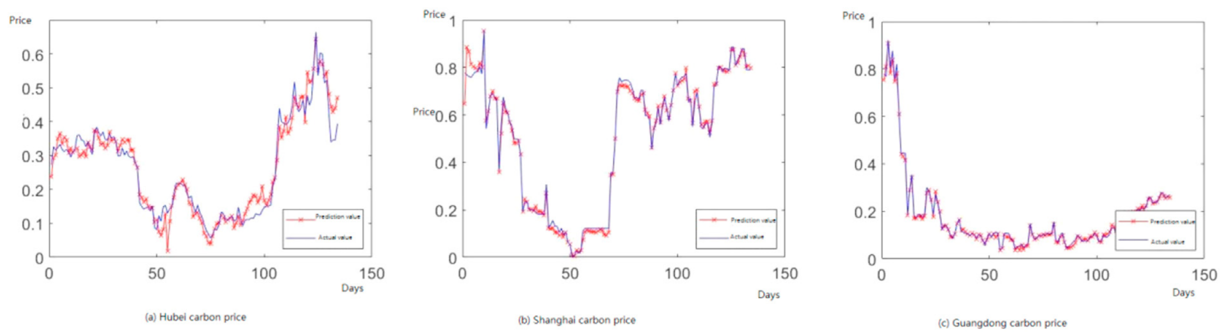

Figure 7. Optimum prediction results of the improved multivariable data-driven MEEMD-LSTM model.

6. Conclusions and suggestions for future work

In order to effectively improve the prediction accuracy of carbon price in China, we proposed a multi-factor MEEMD-LSTM carbon price prediction method. We selected data for Hubei, Shanghai and Guangdong carbon prices to verify the effectiveness of the presented model compared with the prediction performance of LSTM model without MEEMD and single-factor MEEMD-LSTM model.

The empirical results showed that 1) the MEEMD decomposed the time series data into some more regular components, to reduce noise, better describe the physical characteristics which improve the prediction accuracy. 2) compared with single-factor input, multi-factor input can improve the prediction accuracy, but it is notable that including more variables did not affect the prediction accuracy, only when the input information was effective and highly correlated to the target variable, could the prediction effect be improved. 3) the machine reasoning system based on production rules could automatically select and optimize the LSTM parameters, leading to better accuracy and adaptability. 4) in all three cases, we found that our proposed model performed much better than other models and in general it avoided the blindness of parameter selection, effectively improved the prediction accuracy and had better robustness and adaptability.

In the future, we can enhance the proposed model by selecting more relevant influencing factors, like carbon policy, to build a factor system for carbon price prediction. Another direction is to develop an intelligent prediction support and decision system for carbon policy makers and carbon market participants on the basis of prediction results.

Declarations

Author contribution statement

Min Yang, Ph. D: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Shuzhen Zhu: Conceived and designed the experiments.

Wuwei Li: Contributed reagents, materials, analysis tools or data.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data associated with this study has been deposited at www.tanpaifang.com.

Declaration of interest's statement

The authors declare no competing interests.

Additional information

No additional information is available for this paper.

Acknowledgements

We thank LetPub (www.letpub.com) for linguistic assistance and pre-submission expert review.

References

[1] Y.J. Zhang, Y.M. Wei, An overview of current research on EU ETS: evidence from its opeating mechanism and economic effect, Appl. Energy 87 (2010) 1804–1814.
[2] J. Chevallier, A model of carbon price interactions with macroeconomic and energy dynamics, Energy Econ. 33 (6) (2011) 1295–1312.
[3] X.P. Tan, X.Y. Wang, Dependence changes between the carbon price and its fundamentals: a Quantile Regression Approach, Appl. Energy 190 (2017) 306–325.
[4] B.Y. Guo, R.Y. Zhou, Analysis on price cycle and volatility characteristics of China's carbon market, Stat. Decis. 21 (2016) 154–157.
[5] Y. Li, An empirical study on the external influencing factors of carbon emission trading pricing, Price: Theor. Pract. 6 (2020) 146–149.
[6] J.Y. Lv, H. Yang, Z. Guo, Research on the impact of RMB exchange rate on China's regional carbon price-Based on VAR model, Price Mon. 6 (2019) 77–83.
[7] S.J. Byun, H.J. Cho, Forecasting carbon futures volatility using GARCH models with energy volatilities, Energy Econ. 40 (2013) 207–221.
[8] E.S. Maria, V. Francesco, M.B. Maria, Understanding volatility dynamics in the EU-ETS market, Energy Pol. 82 (2015) 321–331.

[9] Q.H. Ji, Y.Z. Sun, H. Yu, X.F. Guo, Y.P. Sun, Q. Liu, Study on forecast model of carbon emission quota price based on multiple linear regression analysis, Mod. Chem. Ind. 38 (4) (2018) 220–224.
[10] Z.H. Feng, L.L. Zou, Y.M. Wei, Carbon price volatility: evidence from EU ETS, Appl. Energy 88 (2011) 590–598.
[11] C. Zhang, X.Z. Yang, Study on prediction model of regional carbon price based on Improved Grey-Markov, Stat. Decis. 453 (9) (2016) 92–95.
[12] Y. Yao, J. Lv, C.G. Zhang, Formation mechanism and prediction of carbon price in Hubei Province, Stat. Decis. 487 (19) (2017) 166–169.
[13] F. Jiang, Z.J. Peng, Forecasting of carbon price based on BP neural network optimized by chaotic PSO algorithm, Stat. Inf. Forum 33 (5) (2018) 92–95.
[14] G.H. Hu, F.,M. Zhu, Modeling and predicting the volatility of carbon price: based on infinite activity Levy processes, J. Appl. Stat. Manag. 37 (5) (2018) 892–903.
[15] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.
[16] J.L. Yan, L.G. Xiang, H.Y. Wu, S.Y. Sun, Urban road traffic speed prediction based on LSTM, Geomatics World 26 (5) (2019) 79–85.
[17] Y.X. Feng, Y.M. Li, A research on the CSI300 index prediction model based on LSTM neural network, Math. Pract. Theory 49 (7) (2019) 308–315.
[18] J.Y. Liu, Research on the improved LSTM method for agricultural futures prices forecasting, Jiangsu Sci. Technol. Inf. 36 (27) (2019) 48–52.
[19] S.P. Domingao, The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World, Basic books, a member of the Perseus Books Group, New York, USA, 2015.
[20] M. Wu, M.C. Hughes, S. Parbhoo, M. Zazzi, V. Roth, F. Doshi-Velez, Beyond Sparsity: Tree Regularization of Deep Models for Interpretability, 2017, pp. 1–26. Eprint Arxiv.
[21] T. Salimans, J. Ho, X. Chen, S. Sidor, I. Sutskever, Evolution strategies as a scalable alternatives to reinforcement learing, Comput. Sci. (2017). Eprint Arxiv.
[22] R.K. Srivastava, K. Greff, J. Schmidhuber, Training very deep networks, Adv. Neural Inf. Process. Syst. (2015) 2377–2385.
[23] K. He, X. Zhang, S. Ren, et al., Depresidual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[24] K. Chen, Y. Zhou, F.A. Dai, LSTM-Based Method for Stock Returns Forecasting: A Case Study of China Stock Market, IEEE International Conference on Big Data, 2015, pp. 2823–2824.
[25] S.L. Zhang, X.F. Zhao, The application of LSTM model in a stock market of China, China Circ. Econ. (35) (2018) 94–95.
[26] D.W. Pei, M. Zhu, Stock price prediction based on multiple-factor and multi-variable long short term memory, Comput. Syst. Appl. 28 (8) (2019) 30–38.
[27] B.Z. Zhu, S.X. Ye, P. Wang, K.J. He, T. Zhang, Y.M. Wei, A novel multiscale nonlinear ensemble learning paradigm for carbon price forecasting, Energy Econ. 70 (2018) 143–157.
[28] B.Z. Zhu, P. Wang, Y.M. Wei, Multi-scale analysis of influencing factors of carbon price based on EMD, Econ. Perspect. (6) (2012) 92–97.
[29] Z. Wu, N.E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, Adv. Adapt. Data Anal. I (1) (2009) 1–41.
[30] J.R. Yeh, J.S. Shierh, N.E. Huang, Complementary ensemble empirical mode decomposition: a novel noise enhanced data analysis method, Adv. Adapt. Data Anal. 2 (2) (2010) 135–156.
[31] J.D. Zheng, J.S. Cheng, Y. Yang, Modified EEMD algorithm and its application, J. Vib. Shock 32 (21) (2013) 21–26.
[32] N.E. Huang, et al., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Process R. Soc. London A454 (1998) 903–995.
[33] J. Zhang, P. Wang, R. Yan, et al., Long short-term memory for machine remaining life prediction, J. Manuf. Syst. 48 (2018) 78–86.