Forecasting the carbon price sequence in the Hubei emissions exchange using a hybrid model based on ensemble empirical mode decomposition

Qunli Wu1,2 | Ziting Liu1

1Department of Economics and Management, North China Electric Power University, Baoding, China
2Beijing Key Laboratory of New Energy and Low-Carbon Development, North China Electric Power University, Beijing, China

Abstract
The prediction of carbon price is exceedingly essential for the regulators, investors, and participants of the carbon trading market. It is the basis for formulating market policies and improving risk management capabilities. China’s carbon price series are nonlinear and nonstationary, so it is difficult to predict accurately with traditional models. This paper proposes a multiscale ensemble prediction model based on ensemble empirical mode decomposition (EEMD-ADD) to improve the prediction accuracy of carbon price. Firstly, EEMD is used to decompose the carbon price sequence into several intrinsic mode functions (IMFs), and these IMFs are divided into high-frequency component, low-frequency component and the trend component. Then, LSSVM, PSO-LSSVM, and BA-LSSVM are used to predict the three components respectively after comparative analysis. Finally, the results are combined to obtain the final prediction value. In the empirical analysis of the Hubei Emissions Exchange, the proposed model outperforms other comparative models. The RMSE, MAE, and MAPE values of the EEMD-ADD model are 0.6180, 0.4726, and 1.6342, and the DS, CP, and CD values are 94.36, 92.16, and 96.48. In addition, the model performed best in other time periods. The results suggest that the proposed model is effective and could predict carbon prices more accurately.

KEYWORDS
a hybrid model, carbon price sequence, ensemble empirical mode decomposition, Hubei emission exchange

1 | INTRODUCTION
Global warming caused by carbon emissions is becoming increasingly prominent, which seriously threatens human survival and social development. China, as the largest carbon emitter, has played a major role in global climate change, and it has been confronted with increasing pressure to control carbon emissions. Therefore, our country vigorously promotes energy transformation and strives to develop into a low-carbon sustainable economy. In 2005-2017, the per Gross Domestic Product (GDP) fell by 45% and has already met its pledge at Copenhagen to cut it to 40%-45% by 2020. At the same time, China sets the Intended Nationally Determined Contribution goals in the Paris Agreement: in 2030, the per GDP carbon emission intensity will be reduced by 60%-65% compared with 2005, the nonfossil energy accounts
for 20% of primary energy consumption, carbon emissions will reach the peak around 2030 and strive for an early peak. Nevertheless, China's rapid economic growth has led to a sharp increase in energy demand and a gradual increase in carbon emissions. And the traditional methods are difficult to meet the needs of emission reduction. So, the Chinese government has proposed the carbon emission trading scheme (ETS) as an important measure to mitigate climate change. In other words, the market regulates carbon emissions by itself, allowing participants to realize their interests and needs. This method is considered to be the most effective way to reduce emissions and has become an important tool for the Chinese government to formulate policies. Since 2011, China has successively set up several carbon emission exchanges. With the development and improvement of carbon emission trading market, the frequency and liveness of carbon trading have gradually increased. The forecasting of carbon prices plays an increasingly significant role in understanding China's carbon market dynamics and in making decisions about carbon emissions reductions. Therefore, it is intensely crucial to select appropriate methods to achieve higher prediction accuracy.

With the maturity of carbon trading market at home and abroad, the research on carbon price has increased dramatically. These studies are mainly divided into two categories, one of which is based on econometric and statistical models for prediction. Among them, the generalized autoregressive conditional heteroskedasticity (GARCH) model is widely used. Byun and Cho applied GARCH model to predict the carbon price of EU ETS. Li and Lu used empirical mode decomposition (EMD) and generalized autoregressive conditional heteroskedasticity (GARCH) model to predict the carbon price of the first stage in EU ETS. Benz and Truck adopted MS-AR-GARCH model to conduct a recursive rolling prediction of carbon price gains in the first stage of EU ETS. However, these methods cannot meet the requirements of high accuracy because of the nonlinear and nonstationary of carbon price series. To overcome the shortcomings of traditional methods, intelligent algorithms with self-learning ability have been introduced by researchers. Zhu and Wei used LSSVM model to forecast the carbon price of EU ETS from April 2005 to February 2011. In 2013, Zhu and Wei proposed ARIMA-LSSVM model to forecast DEC 10 and DEC 12 carbon futures contracts of the European Climate Exchange (ECX). Tian and Hao established an improved carbon price analysis and prediction system composed of analysis module and prediction module, and the empirical results show that the HOA-based model has the best prediction performance compared with other traditional models. In order to predict carbon price more accurately, Hao and Tian proposed a kernel-based extremum learning machine optimization model with good generalization and stability and based on this, established a mixed prediction framework considering multiple factors. Lu et al used six different artificial intelligence models to predict the daily carbon price and trading volume of eight carbon markets in China, by comparing the prediction results, a more suitable model for specific carbon market was obtained. In order to make up for the defects of ARMA model and accurately predict the carbon price, Wang used Boosting algorithm for optimal subset search. Artificial intelligence algorithm effectively overcomes the limitations of traditional models, but its forecasting ability for nonlinear and nonstationary carbon price series is still insufficient. In recent years, the multiscale ensemble hybrid forecasting model represented by EMD has been widely used. Yao and Lv studied the formation mechanism and carbon price prediction of carbon market in Hubei province by virtue of EMD-SVM model. Zhu et al established a carbon price forecasting model based on EMD and neural networks. Zhu and Han used EMD-LSSVM method to predict carbon price. Gao used the EMD-PSO-SVM model to predict the carbon price. Huang and Sun proposed a novel hybrid carbon price forecasting model based on EMD-VMD-GA-BP. Although EMD can improve the model's fitting degree, it is easier to cause mode mixing, thus reducing the prediction accuracy. Wu and Huang proposed the EEMD which adds the Gaussian white noise so as to avoid the mode mixing problem. For this reason, mounting number of scholars use EEMD-types model to predict nonlinear and nonstationary sequences. It is also widely used in carbon price forecasting, for instance, Zhang and Hu used EEMD and BP to forecast the error sequence, and the results show that the method has higher precision accuracy on carbon price. Feng et al used variance ratio and EEMD to analyze the carbon price fluctuation factors in the EU ETS.

However, the research on carbon price prediction is in its initial stage, and there are still some shortcomings: (a) the existing prediction methods are not robust for accurately forecasting carbon prices that have the characteristics of the nonlinear and nonstationary. Krogh et al pointed out that when the single model constituting the combination model is accurate and diversified, the combination model is bound to achieve better results than the single model. Therefore, this paper analyses the characteristics of different frequency components decomposed by EEMD and chooses a suitable prediction model for each frequency component. Finally, all classification prediction results are integrated. (b) Most of the existing studies focus on the international carbon trading market and seldom take China's carbon price as a sample. Through Augmented-Dickey-Fuller (ADF) and Brock-Dechert-Scheinkman (BDS) tests, it is known that the regional carbon price in China ETS has the characteristics of nonlinear and nonstationary, which is similar to the fluctuation trend of the international carbon market. So, this paper
can use the method of studying the international carbon price to analyze and predict the China ETS.\textsuperscript{31}

In this paper, a multiscale ensemble forecasting model is constructed based on EEMD, which is called EEMD-ADD. The original price sequence of Hubei Emissions Exchange is taken as sample for empirical analysis. To verify the validity of the model, the EEMD-ADD model is compared with LSSVM, PSO, BA, PSO-LSSVM, and BA-LSSVM. The innovations and main contributions of this paper are as follows: (a) EEMD algorithm is applied to predict carbon price. The IMFs obtained from EEMD is adopted to study the carbon price series, which reduces the data irregularity and improves the prediction accuracy. (b) Combining EEMD, LSSVM, PSO-LSSVM, and BA-LSSVM, the hybrid model of carbon price forecasting is first proposed, which might provide a new idea for carbon price forecasting, and effectively make up for the shortcomings of the single model. (c) Data from the carbon markets in Hubei are used for empirical analysis, while five other relevant models and other time periods are used for performance comparison to prove the robustness and effectiveness of the models. In addition, different evaluation indicators are used to analysis the model performance, including RMSE, MAE, MAPE, and DS, CP, CD.

The rest of this article is arranged as follows: Section 2 introduces the algorithms mentioned in this article. Section 3 introduces the specific structure of the hybrid prediction model. Section 4 carries on the demonstration analysis based on the carbon price sequence of the Hubei Emission Exchange. Section 5 verifies the prediction performance of the EEMD-ADD. Finally, Section 6 provides some conclusions and future work.

2 | METHODOLOGY

This paper first introduces four basic algorithms EEMD, LSSVM, PSO, BA and then proposes two combined algorithms based on them. The details are as follows.

2.1 | Ensemble empirical mode decomposition

Empirical mode decomposition (EMD) is a method suitable for analyzing and processing of nonlinear and nonstationary signals.\textsuperscript{32-34} The method filters arbitrary signals into deterministic IMFs of different scales and a residual component. The core is to screen the signals layer by layer and decompose the signals of different frequencies at different levels. The IMFs of each order must meet the following two requirements: (a) in the whole data sequence, the number of extremum and Zero-crossings must be equal or at most one difference; (b) the local mean of the data sequence is zero in any time. However, EMD is relatively easy to cause mode mixing. In order to avoid this problem, Wu and Huang\textsuperscript{22} proposed EEMD that increases Gaussian white noise. Gauss white noise has the characteristics of an independent random sequence and a zero mean. And different scales of the signal can be compared and identified in an eigenmode function, then eliminate the white noise sequence by averaging the whole signal, so as to reduce the appearance of mode mixing problem.\textsuperscript{35} The basic steps of EEMD algorithm are as follows:

Step 1: Add a set of Gaussian white noise into the original sequence
Step 2: Decompose the integrated sequence using the EMD algorithm to obtain IMFs
Step 3: Repeat steps 1-2 with different scales of Gaussian white noise each time to obtain the corresponding IMFs

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

where \( x(t) \) is the original signal, each \( C_i(t) \) represents the \( i \)-th IMF, and \( R(t) \) is the residual component.

2.2 | Least squares support vector machine

On the basis of support vector machine, Suykens et al\textsuperscript{36} proposed the least square support vector machine (LSSVM), which is an improvement on the standard support vector machine algorithm, making the algorithm simpler, easy to implement and enhance the computing speed, thus reducing the problem of high complexity of support vector machine. In LSSVM model, the training sample set \( S = \{ (x_i, y_i) | i = 1, 2, \ldots, t \} \) \( x_i = R^n, y_i = R \). Then, the optimal function is constructed by a high dimensional feature space, it described as follows:

\[ f(x) = \omega^T \varphi(x) + b \]  

where \( f(x) \) stands for prediction value, \( \varphi(x) \) represents the nonlinear mapping function from the input space to the high-dimensional feature space, \( \omega \) is weight and \( b \) is bias.

Its structural risk minimization is shown as follows:

\[ R = \frac{1}{2} ||\omega^2|| + cR_{emp} \]  

where \( \omega^2 \) is the complexity; \( c \) stands for error penalty factor;\( R_{emp} \) denotes the error range function.

The objective function of LSSVM is as follows:

\[ \min Z(\omega, \xi) = \frac{1}{2} ||\omega^2|| + c \sum_{i=1}^{t} \xi_i^2 \]  

The flight speed can be adjusted at any time according to the optimization purpose. In order to find the optimal solution conveniently, we assume that bats forage in D-dimensional space, set at t-time, the position is \( X_i^t \), the speed is \( V_i^t \), and the current global optimal position is \( X^* \). The updated formulas of bat's position and velocity at \((t+1)\)-time are as follows:

\[
F_i = F_{\text{min}} + (F_{\text{max}} - F_{\text{min}}) \times \beta 
\]

\[
V_i^{t+1} = V_i^t + (X_{\text{best}} - X_i^t) \times F_i 
\]

\[
X_i^{t+1} = X_i^t + V_i^{t+1} 
\]

The random number \( \beta \in [0,1] \), and \( F_i \) is the search frequency emitted by bat \( i \), \( F_i \in \left[ F_{\text{min}}, F_{\text{max}} \right] \).

The formula for updating the intensity and frequency of sound wave (pulse) are:

\[
A_i^{t+1} = A_i^t \times \alpha 
\]

\[
r_i^{t+1} = r_0 \left[ 1 - \exp(-\gamma \times r) \right] 
\]
2.5 | **PSO-LSSVM**

For the LSSVM model with radial basis function, the kernel parameter $\alpha$ and the normalized parameter $\gamma$ have a great influence on the final prediction accuracy.\(^{39}\) Therefore, we adopt PSO algorithm to optimize it. The steps are as follows:

Step 1: Initialize the parameters of PSO and LSSVM

Step 2: Calculate the fitness of each particle. Compare the current fitness value of each particle with the optimum fitness value and select the optimal one

Step 3: Comparing the fitness value of each particle's optimal position with that of the population's optimal position, select the best and keep the current state of the particle

Step 4: Judgment of optimizing termination conditions. If the error requirements or the maximum number of evolutions are reached, the optimization will end. Otherwise, repeat Step 2 to continue the optimization

2.6 | **BA-LSSVM**

Because the forecasting ability of LSSVM depends on the parameters, the BA is used to obtain the optimal parameters. Based on the BA-LSSVM model; the basic steps of parameter optimization can be described as follows:

Step 1: Parameters setting. They are bat number $M$, maximum iteration number $N$, maximum pulse intensity $A$, and pulse frequency $\gamma$

Step 2: Initialization population. Initialize the position $X$ and speed $V$ of the bat population. Each bat position is composed of LSSVM ($\gamma, \sigma$). The formula is:

$$X = X_{\text{min}} + \text{rand}(1,d) \times (X_{\text{max}} - X_{\text{min}})$$  \tag{16}

Step 3: Update parameters. Calculate the population fitness, find the current optimal solution, and update the flight speed and position of bats according to formula (11)-(13)

Step 4: Update loudness and pulse frequency according to formula (14) and (15)

Step 5: Output global optimal parameters. Repeat Steps 2-4 until the maximum number of iterations is reached and output the global optimal parameters. Then, a carbon price prediction model is performed.

3 | **FRAMEWORK OF THE PROPOSED HYBRID MODEL**

In this paper, we used a multiscale ensemble prediction model EEMD-ADD to predict the carbon price sequence of Hubei ETS, as shown in Figure 1. The steps are as follows:

Step 1: The carbon price sequence is decomposed into several IMFs and a residual by EEMD

Step 2: According to the fine-to-coarse reconstruction algorithm, the IMFs are divided into low-frequency component and high-frequency component, the residual as the trend component

Step 3: Through comparative analysis LSSVM is used to predict the low-frequency component, high-frequency component is predicted by PSO-LSSVM, and the trend component is predicted by BA-LSSVM

Step 4: The final result is obtained by summarizing the predicted values of the three components and comparing them with the original carbon price sequence

4 | **EMPIRICAL ANALYSIS**

4.1 | **Data selection and test**

To push forward low-carbon procedure, the National Development and Reform Commission of China (NDRC) has established seven ETS pilots including Beijing, Tianjin, Shenzhen, Shanghai, Chongqing, Guangdong Province, and Hubei Province.\(^{40}\) Among them, Hubei has the largest trading volume and a comprehensive business structure. The total volume of trade is 52.879 million tons, accounting for 35.4% of China; and the total volume of trade is 111.445 million-yuan, accounting for 31.4% of the total volume of trade in China. This paper analyses the carbon price series of Hubei 2014 (HB2014) from 2 April 2014 to 5 December 2018. Excluding the daily data of zero trading volume, there are 1129 data (data from China Carbon Emission Trading Network). The carbon price sequence is shown in Figure 2. Statistical descriptions on carbon price series are shown in Table 1.

HB2014 carbon price sequence is clearly irregular. In the initial stage, the sequence was relatively stable, followed by rapid decline and sharp fluctuations, then goes up by big margins, and finally tended to be stable. ADF is used to test the stability of carbon price series and BDS is used to test its linear characteristics. Table 2 is the ADF stability test result. This statistic meets the critical values of the 1%, 5% and 10% level. And the $P$-value is .1916. Obviously, the
The price sequence is nonstationary. Table 3 is the result of BDS test. The BDS value increases with the increase of embedding dimension, and the corresponding Z-statistics value is larger than the critical value. And the relevant $P$-value is 0, less than .05. Consequently, the carbon price sequence is nonlinear. To sum up, the price sequence of Hubei carbon trading market is nonstationary and nonlinear, making the sequence fluctuation characteristics basically the same as that of EU ETS.

4.2 | Data processing and analysis

4.2.1 | The analysis of IMF statistics

In this paper, the price sequence of Hubei carbon market is decomposed into nine IMFs and one residual by EEMD. As shown in Figure 3, the IMFs are arranged in order of frequency from high to low. To further investigate the results of modal decomposition, the mean period of each IMF, the
The correlation between each IMF and the original sequence, and the variance and variance percentage of each IMF were calculated, see Table 4. The average period of IMF1-IMF5 is shorter than 100 days, which indicates that these fluctuations are quite intense. And the average cycle of the IMF6-IMF9 is over 200 days, with less fluctuation. The linear correlation between two sequences coefficient is described by the Pearson correlation, and whether the two sequences have a trend of simultaneous change is tested via the Kendall correlation coefficient. The correlation between the residual and the original sequence was the highest, reaching 0.3403 and 0.1985, but there was little fluctuation. The correlation coefficient of IMF6-IMF9 is relatively high, while that of IMF9 is negatively correlated with the original signal. The higher the variance ratio, the greater the influence of correlation IMF on the original sequence. The variance ratio of the residuals was 4.11%, which greatly influenced the original sequence. The trend of residuals is similar to the original sequence. Meanwhile, the sum of variance ratios of IMF6-IMF9 is 0.2%, it has little resemblance to the original sequence. It should be emphasized that due to the rounding error and the nonlinearity of the original sequence, the total variance of the IMF and residuals is not equal to the variance of the original sequence.

Through the fine-to-coarse reconstruction algorithm, the IMFs are decomposed into high-frequency and low-frequency components, and the residual as the trend component. Calculate the mean and the T test of the sum of superposition sequences $S_i (i = 1, 2, \cdots, 9)$ from $C_i (t)$ to $C_i (t)$:

$$S_i = \sum_{n=1}^{i} C_n (t)$$

**TABLE 1** Statistical descriptions of the carbon price of HB2014

| Sample | Mean   | Standard deviation | Median | Maximum | Minimum | Skewness | Kurtosis |
|--------|--------|--------------------|--------|---------|---------|----------|----------|
| 1129   | 20.29  | 4.89               | 21.5   | 32.71   | 10.07   | 0.07     | 1.76     |

**TABLE 2** The stability test of carbon price sequence of HB2014

| Variable   | ADF statistics | Critical value (significance level) |
|------------|----------------|------------------------------------|
| Carbon price | −2.2421        | −3.4359 (1% level) −2.8639 (5% level) −2.5681 (10% level) |

**TABLE 3** The linearity test of the carbon price sequence of HB2014

| Dimension | BDS statistic | Std. error | z-Statistic | Prob. |
|-----------|---------------|------------|-------------|-------|
| 2         | 0.1754        | 0.0012     | 145.2175    | 0.0000|
| 3         | 0.3035        | 0.0019     | 159.1741    | 0.0000|
| 4         | 0.3935        | 0.0023     | 174.6301    | 0.0000|
| 5         | 0.4553        | 0.0023     | 195.4166    | 0.0000|
| 6         | 0.4969        | 0.0022     | 222.9683    | 0.0000|
| 7         | 0.5248        | 0.0020     | 259.1476    | 0.0000|
| 8         | 0.5430        | 0.0017     | 305.9583    | 0.0000|
| 9         | 0.5544        | 0.0015     | 366.1991    | 0.0000|
| 10        | 0.5611        | 0.0013     | 443.7803    | 0.0000|
The results are as Table 5.

The $T$ test of IMF 5 (mean > 0.01) was significantly deviated from zero. Therefore, IMF1-IMF5 was taken as the high-frequency component, IMF6-IMF9 was taken as the low-frequency component, and residual as the trend component. Figure 4 shows the three components and the original sequence. Table 6 shows the statistical indicators of these sequences, including the Pearson and Kendall correlation between each component and the original sequence, the variance and variance percentage.

### 4.2.2 Low-frequency component—impacts of major events

The impact of major events on carbon price is mainly reflected in the low-frequency component. Such heterogeneous environment has marvelous impact on carbon prices, lasting for a long time, with a larger range, but with a lower frequency. For example, in April 2016, China set a goal of Intended Nationally Determined Contribution in the Paris Agreement, which caused significant fluctuations in carbon

![Figure 3](image)

**Figure 3** The decomposition of the HB2014 sequence based on EEMD

| Sequence      | Mean period (d) | Pearson correlation | Kendall correlation | Variance of original | Variance as % of $\Sigma $IMFs + residual |
|---------------|-----------------|---------------------|---------------------|----------------------|------------------------------------------|
| Original      | Null            | Null                | Null                | 23.883               | Null                                     |
| IMF1          | 3.25            | 0.0865              | 0.0637              | 0.289                | 1.21%                                    |
| IMF2          | 12.98           | 0.0623              | 0.0543              | 0.154                | 0.64%                                    |
| IMF3          | 16.85           | 0.1412*             | 0.0921              | 0.148                | 0.62%                                    |
| IMF4          | 43.42           | 0.0766              | 0.0369              | 0.288                | 1.21%                                    |
| IMF5          | 75.26           | 0.0535              | 0.0469              | 0.205                | 0.86%                                    |
| IMF6          | 225.8           | 0.3317*             | 0.2102*             | 1.935                | 8.10%                                    |
| IMF7          | 376.34          | 0.1441*             | 0.1053*             | 1.64                 | 6.87%                                    |
| IMF8          | 1129            | 0.6637*             | 0.5056*             | 20.219               | 8.47%                                    |
| IMF9          | Null            | 0.4956*             | 0.3300*             | 0.001                | 0.01%                                    |
| Residual      | Null            | 0.3403*             | 0.1985*             | 9.821                | 4.11%                                    |
| Sum           | Null            | Null                | Null                | 76.87%               | 100.00%                                  |

*Correlation is significant at the 0.01 level (2-tailed).
prices, reduced to less than 20 yuan and lasted until mid-July 2018. During this period, China held the 19th National People’s Congress of the Communist Party of China. The government attaches great importance to energy conservation and emission reduction, resulting in the continuous rise in carbon prices.

IMF6-IMF9 belongs to low-frequency component, with an average period of more than 200 days, which indicates that the impact of major events is difficult to eliminate and has a great impact on carbon price fluctuation. The Pearson and Kendall coefficients of low-frequency component are 0.7639 and 0.5443, shows that there is an obvious correlation between low-frequency component and the original sequence. And its variance contribution rate is higher than that of high-frequency component, so the fluctuations could be explained by major events. All in all, market participants should pay more attention to the occurrence of major events and judge the trend of carbon price according to its natures.

4.2.3 High-frequency component—short-term market fluctuations

Although the products in the carbon trading market are special, they still have general market attributes, that is, they are affected by the supply and demand mechanism of the market. The price rises when demand increases; the price decreases when demand decreases. Under the influence of market mechanism, the supply and demand of carbon emission rights are affected by energy prices, weather, political situation, and so on. These factors are short-lived and high in frequency, so their effect on carbon prices is mainly reflected in the high-frequency component. However, its impact tends to be relatively small.

The short average period and small variance contribution of high-frequency components indicate that they have little effect on the long-term trend of carbon prices, but have a greater short-term impact. Consequently, the establishment of short-term price rules is of great significance.

### TABLE 5 The mean of Si and T test value

| Si  | S1     | S2     | S3     | S4     | S5     | S6     | S7     | S8     | S9     |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Mean| −0.0112| −0.0051| −0.0004| −0.0113| −0.0107| 0.0243 | −0.0088| 0.436  | 0.0381 |
| T-value| 0.0002| −0.2636| −0.0394| −1.2219| −1.3220| 2.4128 | −0.7798| 2.1267 | 2.1205 |

### FIGURE 4 The three components and the original price sequence

![Graph showing three components and original price sequence](image)

### TABLE 6 Statistical indicators of the original and three components

| New sequence      | Pearson correlation | Kendall correlation | Variance | Variance as % of original | Variance as % of (ΣIMFs + residual) |
|-------------------|---------------------|---------------------|----------|---------------------------|-------------------------------------|
| Original          | Null                | Null                | 23.883   | Null                      | Null                                |
| High-frequency component | 0.1761*            | 0.1004*             | 0.0469   | 0.20%                     | 0.42%                               |
| Low-frequency component | 0.7639*            | 0.5443*             | 1.4128   | 5.92%                     | 12.52%                              |
| Trend component   | 0.3403*            | 0.1985*             | 9.821    | 4.11%                     | 87.06%                              |
| Sum               | Null                | Null                | Null     | 10.23%                    | 100.00%                             |

*Correlation is significant at the 0.01 level (2-tailed).
4.3.4 Trend component- potential operating rule

The trend component is the determiner of the long-term trend of carbon price, which could reflect the important influence of the potential operating rules on carbon price. The Pearson and Kendall coefficients of the trend component are 0.3403 and 0.1985, respectively, indicating that the trend components are highly correlated with the original carbon price series. Therefore, the trend component can better reflect the long-term trend of carbon price, fully explain its fluctuation, and provide a reliable reference for the long-term equilibrium price of the carbon market.

4.3 Forecasting the carbon price sequence

4.3.1 Model performance evaluation

In order to compare the predicted performance of EEMD-ADD with LSSVM, PSO, BA, PSO-LSSVM, and BA-LSSVM, the following error types and trend types are used for performance testing: The model prediction error was evaluated by the root mean square error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE). The performance of trend types was evaluated by directional symmetry (DS), correct up-trend (CP), and correct down-trend (CD). DS represents the accuracy of the predicted direction. CP and CD represent the accuracy of the predicted rising and falling trends. The calculation formula is as follows:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (18)
\]

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (19)
\]

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (20)
\]

\[
DS = \frac{100}{n} \sum_{i=1}^{n} a_i, a_i = \begin{cases} 1, & \text{if } (y_{ts} - y_{(i-1)ts}) (y_i - y_{i-1}) \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (21)
\]

\[
CP = \frac{100}{n_1} \sum_{i=1}^{n_1} a_i, a_i = \begin{cases} 1, & \text{if } (y_{ts} - y_{(i-1)ts}) > 0 \text{ and } (y_{ts} - y_{(i-1)ts}) (y_i - y_{i-1}) \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (22)
\]

\[
CD = \frac{100}{n_2} \sum_{i=1}^{n_2} a_i, a_i = \begin{cases} 1, & \text{if } (y_{ts} - y_{(i-1)ts}) < 0 \text{ and } (y_{ts} - y_{(i-1)ts}) (y_i - y_{i-1}) \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (23)
\]

4.3.2 Parameter setting and input selection

In this paper, the three prediction models of LSSVM, PSO-LSSVM, and BA-LSSVM are applied. And the parameters related to these algorithms are shown in Table 7.

From April 2, 2014, to September 18, 2018, there were 1079 observations in the training data set, as well as the test data from 19 September 2018 to 5 December 2018, a total of 50 observations. It is shown in Table 8.

| TABLE 7 | The parameter setting of the three algorithms |
|---------|---------------------------------------------|
| Algorithm | Parameter | Value |
| LSSVM | Type | f |
| | Kernel | RBF |
| | Gam | 10 |
| | Sig₂ | 0.1 |
| PSO-LSSVM | Initial population size | 30 |
| | Max-generation | 100 |
| | k | 0.5 |
| | Variance range | [−5, 5] |
| | Velocity range | [−5, 5] |
| BA-LSSVM | Population size | 10 |
| | Loudness | 0.25 |
| | Pulse rate | 0.5 |
| | Frequency | [0, 5] |
| | Stop tolerance | 10⁻⁵ |

| TABLE 8 | The data set of HB2014 |
|---------|------------------------|
| Label | Data | Number |
| Data sets | 2014/4/2-2018/12/5 | 1129 |
| Training set | 2014/4/2-2018/9/18 | 1079 |
| Testing set | 2018/9/19-2018/12/5 | 50 |

where \( y_i \) is the \( t \)-th actual value, \( y_{ts} \) is the \( t \)-th prediction value, \( n \) indicates the number of teased data points, \( n_1 \) is the number of data points up, and \( n_2 \) is the number of data points down. The smaller the values of RMSE, MAE and MAPE, the smaller the error. The higher the values of DS, CP and CD, the higher the accuracy of the prediction.

The data set of HB2014, the number of data sets is 1129, with 1079 observations in the training data set, as well as the test data from 19 September 2018 to 5 December 2018, a total of 50 observations. It is shown in Table 8.
4.3.3 | Results

As shown in Figure 5, the relative errors of the three components and the original are described via LSSVM, PSO, BA, PSO-LSSVM, BA-LSSVM, and EEMD-ADD. This analysis shows that: (a) The relative error of the low-frequency components prediction by LSSVM, BA-LSSVM and PSO-LSSVM are relatively stable, and LSSVM has the smallest error. Nevertheless, there are some fluctuations in the error results of PSO and BA. (b) In the prediction of high-frequency components, the error of PSO-LSSVM is relatively stable and minimum, and the error of BA-LSSVM, LSSVM,
PSO, BA is gradually increasing and becoming more and more unstable. (c) Removing the highest point, the relative error range of the trend component predicted based on BA-LSSVM is from −1.1 to 0.6, which is the lowest and smoothest. And the error of PSO-LSSVM, BA, LSSVM, and PSO increases gradually. (d) The error of BA, PSO, LSSVM, BA-LSSVM, and PSO-LSSVM in forecasting carbon price series fluctuates greatly. However, the error range of carbon price prediction based on EEMD-ADD is the lowest and the most stable. Therefore, LSSVM is used to predict the low-frequency component, high-frequency component is predicted by PSO-LSSVM, and the trend component is predicted by BA-LSSVM. And the fitting effect and predicted result of the multiscale ensemble forecasting model are found to be great.

Figure 6 shows the prediction results of Hubei carbon price sequence by six methods: EEMD-ADD, LSSVM, PSO, BA, PSO-LSSVM, and BA-LSSVM. The calculated error and trend type performance are listed in Table 9. The analysis shows that: (a) The EEMD-ADD model has the highest accuracy, and the accuracy of PSO-LSSVM, BA-LSSVM, LSSVM, PSO, and BA decreases in turn. As can be seen from Table 9, the RMSE, MAE, and MAPE values of the EEMD-ADD model are 0.6180, 0.4726, and 1.6342, which are smaller than the other five models, confirming that the multiscale set prediction model's accuracy is the highest. DS, CP and CD values are 94.36, 92.16, and 96.48, which are larger than the other models, proving

### Table 9 The statistical values of the carbon price predictions of the six models

| Model      | RMSE  | MAE   | MAPE  | DS   | CP   | CD   |
|------------|-------|-------|-------|------|------|------|
| EEMD-ADD   | 0.6180| 0.4726| 1.6342| 94.36| 92.16| 96.48|
| PSO-LSSVM  | 1.8818| 1.6172| 5.6066| 80.94| 75.64| 83.49|
| BA-LSSVM   | 2.0675| 1.7733| 6.1552| 75.49| 72.24| 78.19|
| LSSVM      | 2.5963| 2.2930| 7.9598| 71.59| 68.96| 73.47|
| PSO        | 2.8731| 2.4973| 8.4359| 70.43| 66.97| 71.39|
| BA         | 3.1069| 3.0729| 8.8847| 68.79| 66.13| 72.04|

### Table 10 Data sets for the other three-year carbon price sequence

| Product | Data sets                  | Total number | Training set size | Testing set size |
|---------|---------------------------|--------------|-------------------|------------------|
| HB2015  | 2015/1/5-2018/12/5         | 944          | 894               | 50               |
| HB2016  | 2016/1/4-2018/12/5         | 707          | 657               | 50               |
| HB2017  | 2017/1/3-2018/12/5         | 471          | 421               | 50               |

### Table 11 The statistics of the other three-year carbon price prediction

| Product | Model      | RMSE  | MAE   | MAPE  | DS   | CP   | CD   |
|---------|------------|-------|-------|-------|------|------|------|
| HB2015  | EEMD-ADD   | 0.7351| 0.5689| 1.7183| 92.16| 90.47| 93.28|
|          | PSO-LSSVM  | 2.0164| 1.7954| 6.7432| 80.16| 77.98| 80.43|
|          | BA-LSSVM   | 2.3654| 1.8129| 7.1564| 77.59| 70.16| 74.94|
|          | LSSVM      | 3.1006| 2.4675| 9.0160| 74.31| 66.43| 70.18|
|          | PSO        | 3.2894| 2.8437| 9.1463| 74.59| 66.49| 69.84|
|          | BA         | 3.4169| 3.0126| 9.3478| 73.41| 64.21| 67.53|
| HB2016  | EEMD-ADD   | 0.8751| 0.4297| 2.0146| 90.15| 84.26| 89.96|
|          | PSO-LSSVM  | 3.1272| 1.3649| 8.4479| 83.46| 79.51| 81.47|
|          | BA-LSSVM   | 2.6837| 2.4157| 8.5691| 73.26| 71.74| 75.23|
|          | LSSVM      | 3.4568| 2.6910| 9.2648| 69.49| 64.03| 68.77|
|          | PSO        | 3.6492| 3.0347| 9.1126| 67.18| 63.87| 64.28|
|          | BA         | 3.8715| 2.9843| 9.3764| 68.53| 62.19| 61.93|
| HB2017  | EEMD-ADD   | 1.0584| 1.2483| 4.1623| 87.16| 80.47| 83.16|
|          | PSO-LSSVM  | 2.8123| 2.9451| 7.0052| 79.37| 76.14| 77.84|
|          | BA-LSSVM   | 3.7416| 2.1426| 8.1564| 70.45| 69.48| 71.43|
|          | LSSVM      | 3.2018| 3.5461| 9.0160| 65.89| 57.91| 60.15|
|          | PSO        | 3.4482| 3.7196| 10.5702| 63.48| 54.13| 59.64|
|          | BA         | 3.8499| 3.9942| 9.9984| 61.75| 50.89| 56.83|
that the prediction trend performance of the multiscale ensemble prediction model is the highest. The above results show that the EEMD-ADD model is superior to signal models and some hybrid models at predicting the carbon price sequence.

5 | STABILITY TEST

In order to further verify the predictive performance of EEMD-ADD model, the other three-year carbon price sequence of Hubei Province and five other models are given for stability detection. The data are shown in Table 10.

As shown in Table 11, the analyses and their conclusions are similar to those in 4.3.3. For these sequences, EEMD-ADD has the highest prediction accuracy and the best trend prediction performance. Moreover, when the sample size gradually decreases, the prediction accuracy will gradually decline, which shows that the prediction accuracy of large samples is higher. The empirical analysis in this paper not only collected a large number of sample data, but also decomposed the carbon price series into low-frequency, high-frequency components and trend terms through EEMD, which greatly reduced the impact of price fluctuation, so as to achieve higher accuracy. Thus, it can be verified that the EEMD-ADD prediction model is more effective than signal models and some hybrid models.

6 | CONCLUSIONS AND FUTURE WORK

In this paper, a multiscale ensemble forecasting model is proposed for carbon price predicting. First, EEMD is used to decompose the original carbon price sequence into nine IMFs and one residual. Second, reconstruct these IMFs into three different frequency components by a fine-to-coarse reconstruction algorithm. Third, after comparative analysis, LSSVM, PSO-LSSVM, and BA-LSSVM are used to predict the three components, respectively. Fourth, combine the results to get the final predicted value. Using the carbon price of HB2014 for samples, the prediction performance of EEMD-ADD model has been compared with LSSVM, PSO, BA, PSO-LSSVM, and BA-LSSVM. And three additional carbon price series can also prove the performance of EEMD-ADD.

In general, we can draw some conclusions from the empirical analysis results: (a) EEMD can decompose the carbon price sequence into components with different frequencies, so it can capture the characteristics of each component more accurately. (b) By making full use of the advantages of different models, the complex characteristics of carbon spot price can be well described, so as to obtain more reliable and accurate prediction results. Therefore, the proposed model could provide significant improvements for carbon price prediction and is competitive at predicting nonlinear and nonstationary carbon prices.

In the future, (a) we should establish more advanced optimization algorithms and prediction models to further improve prediction performances. (b) We should study the changing characteristics of carbon price from regional and stage perspectives so as to predict its changing trend. (c) We need to analyze carbon prices in accordance with national policies and market rules in order to maximize profits.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support from the National Social Science Fund of China (Grant No. 17BGL252) and the support of the high-quality curriculum construction for postgraduate of North China Electric Power University

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Ziting Liu https://orcid.org/0000-0002-1057-5778

REFERENCES

1. Zhou K, Li Y. An empirical analysis of carbon emission price in China. Energy Proc. 2018;152:823-828.
2. Fan JH, Todorova N. Dynamics of China’s carbon prices in the pilot trading phase. Appl Energy. 2017;208:1452-1467.
3. Cheng K, Pan G, Smith P, et al. Carbon footprint of China’s crop production—an estimation using agro-statistics data over 1993–2007. Agr Ecosyst Environ. 2011;142:231-237.
4. Weng Q, Xu H. A review of China’s carbon trading market. Renew Sustain Energy Rev. 2018;91:613-619.
5. Liao Z, Zhu X, Shi J. Case study on initial allocation of Shanghai carbon emission trading based on Shapley value. J Clean Prod. 2015;103:338-344.
6. Han M, Ding LL, Zhao X, Kang WL. Forecasting carbon prices in the Shenzhen market, China: the role of mixed-frequency factors. Energy. 2019;171:69-76.
7. Byun SJ, Cho H. Forecasting carbon futures volatility using GARCH models with energy volatilities. Energy Econ. 2013;40:207-221.
8. Di W, Lu C. The research on setting a unified interval of carbon price benchmark in the national carbon trading market of China. Appl Energy. 2015;155:728-739.
9. Paolella MS, Taschini L. An econometric analysis of emission allowance prices. J Bank Finance. 2008;32(10):2022-2032.
10. Benz E, Truck S. Modeling the price dynamics of CO2 emission allowances. Energy Econ. 2009;31(1):4-15.
11. Zhu BZ, Wei YM. Carbon price prediction based on integration of GMDH, particle swarm optimization and least squares support vector machines. Syst Eng-Theory Pract. 2011;31(12):2264-2271.
12. Zhu BZ, Wei YM. Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. Omega-Int J Manage Sci. 2013;41(3):517-524.
13. Tian CS, Hao Y. Point and interval forecasting for carbon price based on an improved analysis-forecast system. Appl Math Model. 2020;79:126-144.
14. Hao Y, Tian CS. A hybrid framework for carbon trading price forecasting: the role of multiple influence factors. *J Clean Prod*. 2020. https://doi.org/10.1016/j.jclepro.2020.120378

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

16. Wang N. Forecasting of carbon price based on boosting-ARMA model. *Stat Inform Forum, Chin*. 2017;32(03):28-34.

17. Yao Y, Lv J, Zhang CG. Price formation mechanism and price forecast of Hubei carbon market. *Stat Dec*. 2017;19:166-169.

18. Zhu BZ, Han D, Wang P, Wu Z, Zhang T, Wei Y-M. Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. *Appl Energy*. 2017;191:521-530.

19. Gao Y, Li J. International carbon finance market price prediction based on EMD-PSO-SVM error correction model. *Chin Popul Resour Environ*. 2014;24(06):163-170.

20. Sun W, Huang CC. A carbon price prediction model based on secondary decomposition algorithm and optimized back propagation neural network. *J Clean Prod*. 2020;243:118671.

21. Wu Z, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv Adapt Data Anal*. 2009;1:1-41.

22. Jiang G. *Multi-scale Study of Exchange Rate Price Based on EEMD Model*. Guangdong, China: Jinan University; 2018.

23. Liu Y, Yang Y, Nie L, Song QY. The EEMD-ARIMA prediction of runoff at mountain pass of Manas River. *Res Soil Water Conserv*. 2017;24(06):273-280+285.

24. Yang X. *Analysis and Prediction of Multi-Scale Gold Price Fluctuation Based on EEMD Model*. Changsha, China: Changsha University of Science & Technology; 2014.

25. Lin H. *Multi-scale Analysis of Time Series of Shanghai Composite Index Based on EEMD*. Changchun, China: Changchun University of Technology; 2014.

26. Wang J, Wang X, Wang H, et al. Monthly runoff forecasting of reservoir based on EEMD and ANN hybrid model. *Yellow River*. 2019;41(05):43-46.

27. Zhang C, Hu B. Expectation of multiple factors BP international carbon market price based on error correction. *Prices Month*. 2017;01:11-18.

28. Feng Z, Liu C, Wei Y. How does carbon price change? Evidences from EU ETS. *Int J Glob Energy Iss*. 2010;35:132-144.

29. Krogh A, Vedelsby J. Neural network ensembles, crosses validation and active learning. *Neural Comput Appl*. 1995;25:231-238.

30. Zhou JG, Yu XC, Yuan XL. Predicting the carbon price sequence in the Shenzhen emissions exchange using a multiscale ensemble forecasting model based on ensemble empirical mode decomposition. *Energies*. 2018;11(7):1907.

31. Huang NE, Shen Z, Long SR, et al. The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. *Proc R Soc Lond: Ser A*. 1998;454:903-995.

32. Huang NE, Wu MC, Qu WD, Long SR, Shen SSP. Applications of Hilbert-Huang transform to non-stationary financial time series analysis. *Appl Stochastic Models Bus Industry*. 2003;19(3):245-268.

33. Jiang TJ, Zhang HQ, Wang XJ. EEMD and evolutionary KPCR based adaptive prediction modeling on complex time series. *Syst Eng-Theory Pract*. 2014;34(10):2722-2730.

34. Zhang N, Zhang YP. Theory and application of least square support vector machine. *J Daqing Univ*. 2014;34(6):30-32.

35. Yang D, Li S. Study on PSO for the optimization of SVR parameters selection. *Papers of the Annual Conference on Control and Decision-making in China*; 2007:447-452.

36. Yang XS. A new metaheuristic bat-inspired algorithm. *Nat Inspired Cooperative Strategies Optimization (NISCO)*. 2010;284:65-74.

37. Li YS, Jiang ZW, Gan YH. Soft measurement of oil-fired industrial boiler efficiency based on PSO-LSSVM. *Industry Control Appl*. 2017;36(10):41-45.

38. Zhu JM, Wu P, Chen HY, Liu JP, Zhou LG. Carbon price forecasting with variational mode decomposition and optimal combined model. *Physica A-Stat Mech Appl*. 2018;519:140-158.

39. Zhang X, Lai KK, Wang SY. A new approach for crude oil price analysis based on Empirical Mode Decomposition. *Energy Econ*. 2008;30(3):905-918.

**How to cite this article:** Wu Q, Liu Z. Forecasting the carbon price sequence in the Hubei emissions exchange using a hybrid model based on ensemble empirical mode decomposition. *Energy Sci Eng*. 2020;8:2708–2721. https://doi.org/10.1002/ese3.703