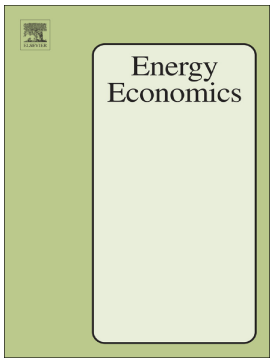
A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting

Zhu, Bangzhu; Ye, Shunxin; Wang, Ping; He, Kaijian; Zhang, Tao; Wei, Yi-Ming

DOI:
10.1016/j.eneco.2017.12.030

Document Version
Peer reviewed version

Citation for published version (Harvard):
Zhu, B, Ye, S, Wang, P, He, K, Zhang, T & Wei, Y-M 2018, 'A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting', *Energy Economics*, vol. 70, pp. 143-157.
https://doi.org/10.1016/j.eneco.2017.12.030



A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting

Bangzhu Zhu, Shunxin Ye, Ping Wang, Kaijian He, Tao Zhang, Yi-Ming Wei

PII: S0140-9883(18)30003-3
DOI: https://doi.org/10.1016/j.eneco.2017.12.030
Reference: ENEECO 3869

To appear in:

Received date: 17 May 2017
Revised date: 23 December 2017
Accepted date: 28 December 2017

Please cite this article as: Bangzhu Zhu, Shunxin Ye, Ping Wang, Kaijian He, Tao Zhang, Yi-Ming Wei , A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Eneeco(2018), https://doi.org/10.1016/j.eneco.2017.12.030

A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting

Bangzhu Zhu[a*], Shunxin Ye, Ping Wang[a*], Kaijian He[b*], Tao Zhang [c], Yi-Ming Wei[d]

[a] School of Management, Jinan University, Guangzhou, Guangdong 510632, China

[b] School of Business, Hunan University of Science and Technology, Xiangtan, Hunan, 411201, China Birmingham

[c] Business School, University of Birmingham, Edgbaston, Birmingham, UK, B15 2TT

[d] Center for Energy and Environmental Policy Research, Beijing Institute of Technology, Beijing 100081, China

Abstract: In this study, a novel multiscale nonlinear ensemble leaning paradigm incorporating empirical mode decomposition (EMD) and least square support vector machine (LSSVM) with kernel function prototype is proposed for carbon price forecasting. The EMD algorithm is used to decompose the carbon price into simple intrinsic mode functions (IMFs) and one residue, which are identified as the components of high frequency, low frequency and trend by using the Lempel-Ziv complexity algorithm. The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is used to forecast the high frequency IMFs with ARCH effects. The LSSVM model with kernel function prototype is employed to forecast the high frequency IMFs without ARCH effects, the low frequency and trend components. The forecasting values of all the components are aggregated into the ones of original carbon price by the LSSVM with kernel function prototype-based nonlinear ensemble approach. Furthermore, particle swarm optimization is used for model selections of the LSSVM with kernel function prototype. Taking the popular prediction methods as benchmarks, the empirical analysis demonstrates that the proposed model can achieve higher level and directional predictions and higher robustness. The findings show that the proposed model seems an advanced approach for predicting the high nonstationary, nonlinear and irregular carbon price.

Keywords: carbon price forecasting, least squares support vector machine, empirical mode decomposition, particle swarm optimization, kernel function prototype

1.Introduction

With its increasingly severe consequences, global climate change has become a serious threat to human sustainable development. Carbon market, as an effective mechanism dealing with the global climate change, attracts significant attention of governments and organizations worldwide. In the past few years, global carbon market, represented by the European Union Emissions Trading System (EU ETS), have witnessed a rapid development. However, carbon price violent variations remarkably impact emissions reduction performances and market values. Predicting carbon price accurately can on the one hand enable us to understand the changeable patterns of carbon price and develop an efficient stabilization mechanism for carbon price, on the other hand help investors avoid carbon market risks and increase the value of carbon assets (Zhu et al., 2016). However, as emerging policy-based artificial markets, carbon markets are impacted by both internal market mechanisms and external environmental heterogeneity (Zhang and Wei,2010), which causes the nonstationary and nonlinear characteristics of carbon prices. Therefore carbon price prediction presents a great challenge to researchers and becomes one of the priority topics in the fields of energy and climate economics.

In the literature there are various methods adopted for carbon price forecasting. Early studies mainly used the qualitative analysis to predict carbon price (Reilly and Paltsev, 2005; Kanen, 2006). Recent studies used more complex methods for carbon price forecasting, which can roughly be divided into three categories: statistical and econometric models, artificial intelligence models, and ensemble (hybrid) models. The traditional statistical and econometric models

* Corresponding author: Bangzhu Zhu, Tel: +86 15915761388; Email:wpzbz@126.com.

were widely used for carbon price forecasting. Chevallier (2011) used a nonparametric method to predict carbon prices, and found that the method could reduce the prediction error by almost 15% compared with linear autoregression models. Byun and Cho (2013) used the Generalized Autoregressive Conditional Heteroskedasticity (GARCH)-family models to predict carbon prices, and found that GJR-GARCH was more effective than TGARCH and GARCH in their particular case. Koop and Tole (2013) used the dynamic model averaging (DMA) method to forecast carbon prices, and obtained a high prediction accuracy. María et al. (2015) applied the ARMAX-GARCH model with a time-varying jump probability to predict carbon prices, and obtained higher prediction accuracy than the standard ARMAX-GARCH model. Although these models based on stationary data and linear hypotheses can obtain high prediction accuracy, they cannot effectively deal with the nonlinearity in carbon price changes. To cope with this major limitation of statistical and econometric models, artificial intelligence models, including artificial neural networks (ANN), support vector machines (SVM) and least square SVM (LSSVM) were applied for carbon price prediction. For example, Zhu and Wei (2011) used the LSSVM method to predict carbon prices, and obtained better results than ARIMA and ANN. Fan et al.(2016) used the multi layered perceptron (MLP)-ANN model to predict carbon prices, and gained a good performance. These artificial intelligence models can effectively capture the nonlinear patterns hidden in carbon price variations, thus obtaining better prediction results than statistical and econometric models. In the meantime, in order to overcome the drawbacks of single models and further improve the accuracy of carbon price prediction, ensemble (hybrid) models were adopted for carbon price forecasting. Zhu (2012) used an empirical mode decomposition (EMD)-based ANN to predict carbon prices, and obtained a higher prediction accuracy than single ANN model. Zhu and Wei (2013) used a hybrid forecasting model, incorporating ARIMA and LSSVM, for carbon price forecasting, and found that the hybrid model was more effective than single ANN and ARIMA models. Wei and Can (2015) used the EMD–GARCH method to forecast the carbon prices of five pilots carbon markets in China, and gained the alternative interval of a lower bound of 30 yuan/ton and an upper bound of 50 yuan/ton in the national carbon trading market. Zhu et al.(2016) used an ensemble EMD (EEMD) -based LSSVM to predict the carbon price, and found that the proposed method could obtain high level and directional predictions. Atsalakis (2016) proposed a computational intelligence model with a novel hybrid neuro-fuzzy controller that forms a closed-loop feedback mechanism(PATSOS) to forecasting the daily carbon price, and get a higher accurate. Sun et al. (2016) combined variational mode decomposition (VMD) and spiking neural networks (SNNs) to improve forecasting accuracy and reliability. Zhu et al. (2017) used EMD and evolutionary LSSVM to improve the robustness of carbon price prediction, and obtained more robust performance than the other popular forecasting methods.

Extant studies have shown that multiscale ensemble forecasting can decompose the complex carbon price into several simple modes, so as to significantly improve the prediction accuracy of carbon price. However, there are several major drawbacks in existing studies. Firstly, in existent multiscale ensemble prediction, the same model is used to predict all the simple modes. As each mode has its own data characteristics, models with more appropriate mode-specific assumptions model is normally required for predicting it (Zhang et al.,2008; Zhu et al, 2015; Zhu et al., 2016). Secondly, although LSSVM has a high nonlinear modeling ability, its predictive power is subject to model selection. Moreover, existing studies have predetermined the radical basis function (RBF) as the kernel function so as to search the model parameters of LSSVM (Yu et al. 2009; Zhu et al., 2013; Silva et al, 2015; Zhang et al., 2015). Few works has investigated the suitability of the kernel function for a specific problem. The inappropriate kernel function would negatively affect the accuracy of carbon price prediction. Thirdly, existing multiscale ensemble prediction models are limited to linear ensemble form, i.e. aggregating the forecasted values of all the modes into the forecasting ones of the original carbon price. A linear ensemble approach is not always appropriate for all the circumstances (Liao

and Tsao, 2006; Alonso et al., 2007), thus affecting the accuracy of carbon price forecasting.

To address the existing drawbacks of carbon price prediction, this research develops a novel multiscale nonlinear ensemble leaning paradigm incorporating EMD, LSSVM with kernel function prototype, and particle swarm optimization (PSO) to improve the forecasting accuracy of nonstationary and nonlinear carbon price. It is a breakthrough study, making both methodological and empirical innovations. *Methodologically*, it develops a novel adaptive multiscale nonlinear ensemble learning paradigm for carbon price forecasting. Firstly, the EMD algorithm is used to decompose the carbon price into simple modes. Secondly, the obtained simple modes are identified as the components of high frequency, low frequency and trend by using the Lempel-Ziv complexity algorithm. GARCH can then be applied to predict the high frequency components with ARCH effects because of its strong short-term memory; and LSSVM with a universal kernel function prototype, is applied to forecast the high frequency components without ARCH effects, the low frequency and trend components. The proposed model can adaptively select the optimal kernel function type and model parameters according to the specific data using PSO, which can make good use of various kernel functions types and overcome the drawbacks of single kernel function. Finally, the LSSVM-based nonlinear ensemble approach is used to aggregate the prediction values of all the components acquired by different models into the forecasting values of the original carbon price, so as to further improve the prediction accuracy. *Empirically*, the proposed multiscale nonlinear ensemble learning paradigm has been tested with the data of the daily European Union Allowance futures prices from January 2, 2013 to April 14, 2015, obtained from the Intercontinental Exchange (ICE). Compared with popular prediction methods, the empirical analysis shows that the proposed model is the optimal, and can effectively deal with the nonlinearity and nonstationarity of carbon prices.

The rest of the paper is organized as follows: Sections 2, 3, 4 and 5 construct the novel adaptive multiscale nonlinear ensemble learning paradigm. Section 6 conducts empirical analysis. Section 7 concludes the study.

2. Methodology

2.1Empirical mode decomposition

EMD was proposed as a multiscale decomposition technique that takes advantage of the local characteristics scales of the underlying data components and extracts these components known as Intrinsic Modes from the data. It was originally proposed by Huang et al. (1998) as an effective empirical method for the nonlinear and non-stationary data. These intrinsic modes are defined as the intrinsic mode function (IMF), satisfying the following conditions: (i) The difference between the extrema and zero-crossings does not exceed 1. (ii) The functions are zero mean locally and symmetric. IMFs satisfying these conditions are zero mean and nearly periodic. Thus they are harmonic with changing amplitudes and frequencies at different timescales.

Compared with the traditional Fourier and wavelet decompositions, EMD technique has several distinct advantages. Firstly, it is relatively easy to understand and implement EMD. Secondly, since the decomposition is based on the local characteristic timescales of the data and only extrema are used in the sifting process, EMD is local, self-adaptive, and very implicative. It is highly efficient for nonlinear and nonstationary time series decomposition, therefore can adaptively and robustly decompose carbon price time series into several IMFs and one residue that display linear and nonlinear behaviours depend only on the nature of carbon prices. Thirdly, the IMFs derived from EMD have a clear instantaneous frequency as the derivative of the phase function. Thus the Hilbert transformation can be applied to the IMFs, allowing us to explore the data in a time-frequency-energy space. Last but not the least, in wavelet decomposition, a filter base function must be determined beforehand. However, it is difficult for some unknown series to determine the filter base function. Unlike wavelet decomposition, EMD does not have to determine a filter base function before decomposition. The four merits discussed above makes EMD an effective decomposition tool.

2.2 LSSVM with kernel function prototype

A kernel function is the critical parameter for the LSSVM predictor. When applied in practice, the kernel function used in LSSVM predictor is selected based on the prior data characteristics and the problem domain. However, the search for the optimal kernel function is a difficult research problem. Researchers usually rely on the empirical analysis and prior experience to select the optimal kernel function. Although not optimal, they often resort back to the common kernel functions such as RBF as a compromise.

The commonly used kernel functions include: linear kernel $K_{lin}(x_i, x_j) = x_i \cdot x_j$, polynomial kernel $K_{poly}(x_i, x_j) = (x_i \cdot x_j + t)^d$, radial basis function (RBF) kernel $K_{rbf}(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2})$ and sigmoid kernel $K_{sig}(x_i, x_j) = \tanh[s \cdot (x_i \cdot x_j) + h]$. According to the Mercer theory, a new kernel function can be constructed with linear combination of various kernel functions. Data contains the complete information about a specific problem. A kernel function, selected by using the machine learning methods under the data-driven circumstance without any prior information, is proved to be the optimal one for the specific problem. Therefore, we introduce a new kernel function prototype in this study, which is defined as (Zhu et al.,2016):

$$K(x, x') = \lambda_1 K_{sig}(x, x') + \lambda_2 K_{rbf}(x, x') + \lambda_3 K_{poly}(x, x') + \lambda_4 K_{lin}(x, x')$$

The proposed kernel function prototype is universal. With different parameters fitted to the specific data, it can produce either commonly used single kernel functions, or new kernel function. Thus, it takes advantage of different kernel functions and avoids the drawbacks of using a single kernel function.

There are two model parameters for LSSVM with the kernel function prototype: kernel function type parameter λ , and kernel function parameter u . Different kernel functions would be produced with different λ and u . Using PSO algorithm, these two parameters can be coded into the particles to search for their optimal values adaptively. Afterwards, the optimal kernel function for the problem is selected adaptively. A single kernel function (for example, $\lambda_1 = 1, \lambda_i = 0, i = 2,3,4$, or a mixture kernel function (for example, $\lambda_i \neq 0, i = 1,2,3,4$) can be obtained.

In this paper, we also use phase space reconstruction (PSR) for the carbon price prediction. Embedding dimension (m) and delay (τ) are two critical parameters for PSR. If *m* is too small, the fine structure of chaotic system will not reveal for carbon prices. If *m* is too large, it will lead to very complicated computation and noise. Likewise, if τ is too small, information redundancy would result as the adjacent delay coordination elements differ only slightly in the phase space. If τ is too large, information loss and folding signal trajectories would result as the association between adjacent delay coordination elements would disappear.

Therefore, PSO algorithm is used to select adaptively the model parameters of LSSVM. These parameters include: kernel function type parameter λ , kernel function parameter u , PSR m and τ , and penalty factor γ . Using PSO algorithm, these parameters are coded into the particles. The optimal solutions survive the evolutionary optimization process in PSO algorithm.

2.3 The proposed adaptive LSSVM learning through particle swarm optimization

Inspired by the foraging behavior of bird flocks, PSO is a new type of swarm intelligent algorithm (Pehlivanoglu, 2013). In the PSO algorithm, candidate solutions are defined as the particles. Firstly, the particles and their speeds are randomly initialized. s denoted the population size, i.e. the number of particles, the location of the *i*th particle in *D*–dimensional space is denoted by $x_i = (x_{i1}, x_{i2}, \cdots x_{iD})$, $i = 1,2 \cdots, s$. The speeds of particles are denoted by $v_i = (v_{i1}, v_{i2}, \cdots v_{iD})$, $i = 1,2 \cdots, s$. *D* is the number of parameters. The fitness function is predefined to calculate and

assign the fitness value to each particle. Afterwards, according to the fitness value of each particle, the local optimum $P_{best} = (P_1, P_2, \cdots P_D)$ and global optimum $G_{best} = (G_1, G_2, \cdots G_D)$ of each particle are updated. Finally, the local and global optimums of each particle are calculated dynamically using formulas (1) and (2), so that the speed and location is updated simultaneously:

$$v_{ij}(t+1) = w(t) \cdot v_i(t) + c_1 \cdot r_1 \cdot [P_j(t) - x_{ij}(t)] + c_2 \cdot r_2 \cdot [G_j(t) - x_{ij}(t)] \quad (1)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (2)$$

where t is the current iteration number. r_1, r_2 are the random numbers distributed uniformly in interval of (0,1) respectively. c_1, c_2 are the acceleration factors. The inertia weight $w(t)$ is calculated using formula (3):

$$w(t) = w_{\max} - \frac{w_{\max} - w_{\min}}{t_{\max}} \cdot t \quad (3)$$

where $w_{\max}$ and $w_{\min}$ are respectively the maximal and minimal inertial weights.

In the updating process, the speed of each particle is limited into a preset interval: If $v_{ij}(t+1) > v_{\max}$, then $v_{ij}(t+1) = v_{\max}$; if $v_{ij}(t+1) < -v_{\max}$, then $v_{ij}(t+1) = -v_{\max}$. The termination condition for the iteration of the PSO algorithm is: reaching the preset maximum iteration number.

In this paper, we proposed an adaptive model selection algorithm, called PSO-LSSVM, as illustrated in Fig.1.

Step 1: Coding. The integer coding is used for m, τ, d ,and the real coding is adopted for other parameters here. All the model parameters are coded into a particle, $x_i = (x_{i1}, x_{i2}, \cdots x_{iD})$, D=12.

Step 2: Initialization. Randomly generate the particles and their speeds, and set the maximum speed, maximum iterations, and ranges of all other parameters.

Step 3: Define the fitness function of each particle. The fitness function is set to the minimization of the Root Mean Square Error (RMSE) and is defined as: $\min RMSE(\lambda, \mu, m, \tau, \gamma)$ where $RSME = \sqrt{\frac{1}{n}\sum_{t=1}^{n}[y_t - \varphi(x_i, \lambda, \mu, m, \tau, \gamma)]}$, $\varphi(\cdot)$ is a nonlinear mapping function determined by the LSSVM predictor with given parameters and training samples. x_i is the samples of training set. y_i is the target value corresponding to the training sample x_i , and n is the size of training set.

Step 4: Calculate the fitness value of each particle. The fitness function is defined to calculate the fitness value of each particle, $F(P_{ipresent})$. The current location is set using the local optimum of each particle P_{ibest}. Then , the initial global optimum G_{best} is set to the particle with the optimal fitness value.

Step 5: Update the speed, location and fitness value of each particle by formulas (l), (2), (3) and (4).

Step 6: Compare $F(P_{ipresent})$ and $F(P_{ibest})$. If $F(P_{ipresent}) < F(P_{ibest})$, then $P_{ibest} = P_{ipresent}$.

Step 7: Compare the updated $F(P_{ipresent})$ and $F(G_{best})$. If $F(P_{ipresent}) < F(G_{best})$, then $G_{best} = P_{ipresent}$.

Step 8: Evaluate whether the termination condition is satisfied. If it is satisfied, the search process is stopped and the optimized LSSVM parameters are recorded. Otherwise, let t=t+1, and return to step 4.

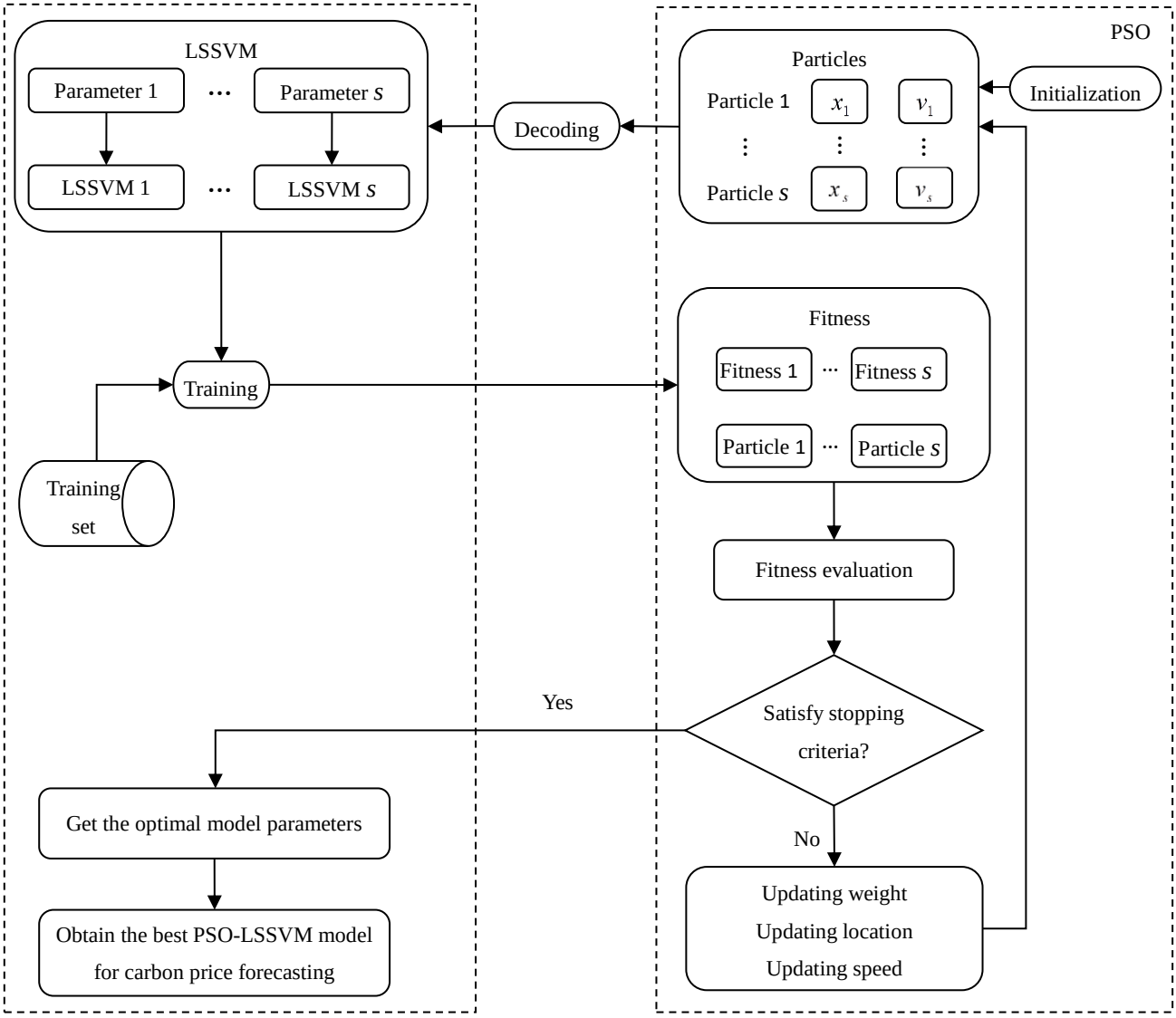


Fig.1. The process of proposed PSO-LSSVM model.

2.4 The proposed multiscale nonlinear ensemble learning paradigm for carbon price forecasting

The proposed multiscale nonlinear ensemble learning paradigm for carbon price forecasting, namely, EMD-HLT-LSSVM, integrating EMD, PSO, kernel function prototype, LSSVM and GARCH models organically, can be divided into four stages, as shown in Fig.2.

Stage 1: Decomposition of EMD

We decomposed the carbon price into m IMFs with different amplitudes and frequencies and one residue. IMFs are ordered from high frequency to low frequency. These IMFs are independent, characterized by simple structure, stable fluctuation, and strong regularity.

$$X(t) = \sum_{j=1}^{m} c_j(t) + Res(t)$$

Classified by the frequencies of IMFs, we divided IMFs into three categories: (i) High frequency components (HFs), with high frequencies and low amplitudes, can characterize the random changes of the carbon price induced by the supply and demand imbalance of short-term market in normal operation. Although HFs fluctuate frequently in short-term, they cannot generate a long-term effect. Additionally, the fast changing carbon price is accompanied by fast recovery speed. The rise (or decline) of price always precedes the decline (or rise) of price, i.e., the gathering effects of

fluctuations. (ii) Low frequency components (LFs) with low frequencies and high amplitudes, can characterize the periodic fluctuations data characteristics influenced by outside environments. The rapid rise or fall of LFs is caused by the shocks from the major unexpected events. Though the fluctuation frequencies of LFs are low, they may affect the carbon price greatly and even change the pricing mechanism of the carbon market. (iii) Trend component (T) can describe the stable change of the carbon price in the long-term. Each component has its unique characteristics. Therefore, the accuracy of carbon price prediction is improved if the appropriate forecasting model for each component is selected according to its data characteristics.

Stage 2: Identification of HFs, LFs and T

The different data characteristics exhibited by HFs, LFs, and T respectively show the different intrinsic properties of initial carbon price. Once the decomposition of EMD is finished, the Lempel-Ziv complexity algorithm (Lempel, 1976) is employed to identify the HFs, LFs, and T, so as to select the optimal models for forecasting each component. The identification process is given as follows:

Step 1: Calculating the complexity C_i of IMF_i, $i = 1,2,\cdots,m$.

Step 2: Setting the critical value $\lambda_0(\geq 0.8)$ so as to find out the min value of k which satisfies the condition of $\eta = \sum_{i=1}^{k} C_i / \sum_{i=1}^{m} C_i \geq \lambda_0, k \in m$.

Step 3: The IMFs from IMF_1 to IMF_k are identified as HFs, and the IMFs from IMF_{k+1} to IMF_m are identified as LFs. Meanwhile, the residue is identified as T.

Stage 3: Prediction of HFs, LFs, and T

The HFs, LFs, and T are forecasted by the following strategy: firstly, the appropriate model is selected based on different data characteristics to predict each component individually, and then the individual prediction results are aggregated to produce the carbon price forecasts.

The forecasting model for carbon price time series X_t $(t = 1,2,\cdots,n)$ can be defined as:

$$\hat{X}_{t+h} = f(X_t, X_{t-\tau}, X_{t-2\tau}, \cdots, X_{t-(m-1)\tau}) + \varepsilon_t$$

where $\hat{X}_t$ is the predicted value of initial time series, h is the prediction horizon, m, τ are the embedding dimension and delay of PSR respectively, ε_t is the perdition errors. When h=1, it is one-step-ahead forecasting. When h≥2, it is h-step-ahead forecasting.

(1) Prediction of HFs. HFs demonstrate the short-term fluctuations of the carbon price, characterized by high randomness, time-varying and aggregation. Make an ARCH test at first, if the HFs are with arch effect, the GARCH model is employed to forecast them, otherwise the PSO-LSSVM model is used to forecast them. GARCH(1,1), the econometric model with strong short-term memory on a random process with time-varying and aggregation, is applicable to forecasting HFs. The details of building the GARCH(1,1) model and the specific modeling processes are described in Tim Bollerslev (1986).

(2) Predication of LFs. Since LFs demonstrate strong periodic fluctuations, the PSO-LSSVM model is used to predict LFs due to its nonlinear modeling capability. We use the kernel function prototype to derive the kernel function for the LSSVM model. The proposed PSO-LSSVM method can adaptively select the optimal kernel function type and model parameters in a data driven approach, as described in Section 2.3.

(3) Prediction of T. The trend component exhibits clear changing patterns. Therefore, LSSVM model is used to model and forecast it. The model selection of LSSVM for T is the same as that in the prediction of LFs.

Stage 4: Nonlinear integration of the prediction values of HFs, LFs, and T

The carbon price is constructed by aggregating from the prediction values of HFs, LFs, and T:

$$\hat{X} = f(\hat{HF}\text{s}, \hat{LF}\text{s}, \hat{T}) = f(\hat{c}_1, \hat{c}_2, \cdots, \hat{c}_m, \hat{r})$$

where $\hat{X}$ is the predicted value of initial carbon price, and $\hat{HF}\text{s}$, $\hat{LF}\text{s}$, $\hat{T}$ are the predicted values of HFs, LFs and T respectively. $f(\cdot)$ is a nonlinear mapping function, determined by the proposed LSSVM model with the kernel function prototype, which can minimize the sum of square errors of the training set, defined by

$$min\left\{ [X - f(\hat{c}_1, \hat{c}_2, \cdots, \hat{c}_m, \hat{r})][X - f(\hat{c}_1, \hat{c}_2, \cdots, \hat{c}_m, \hat{r})]^T \right\}$$

It is worth pointing out that, to avoid that the proposed model is trained on future information, we forecast just one point every time and repeat the forecasting progress one by one, recursively. Namely, we use the training set to forecast next point, then add its real value of this point into the training set as a new training set, and repeat the four stages to forecast the next point.

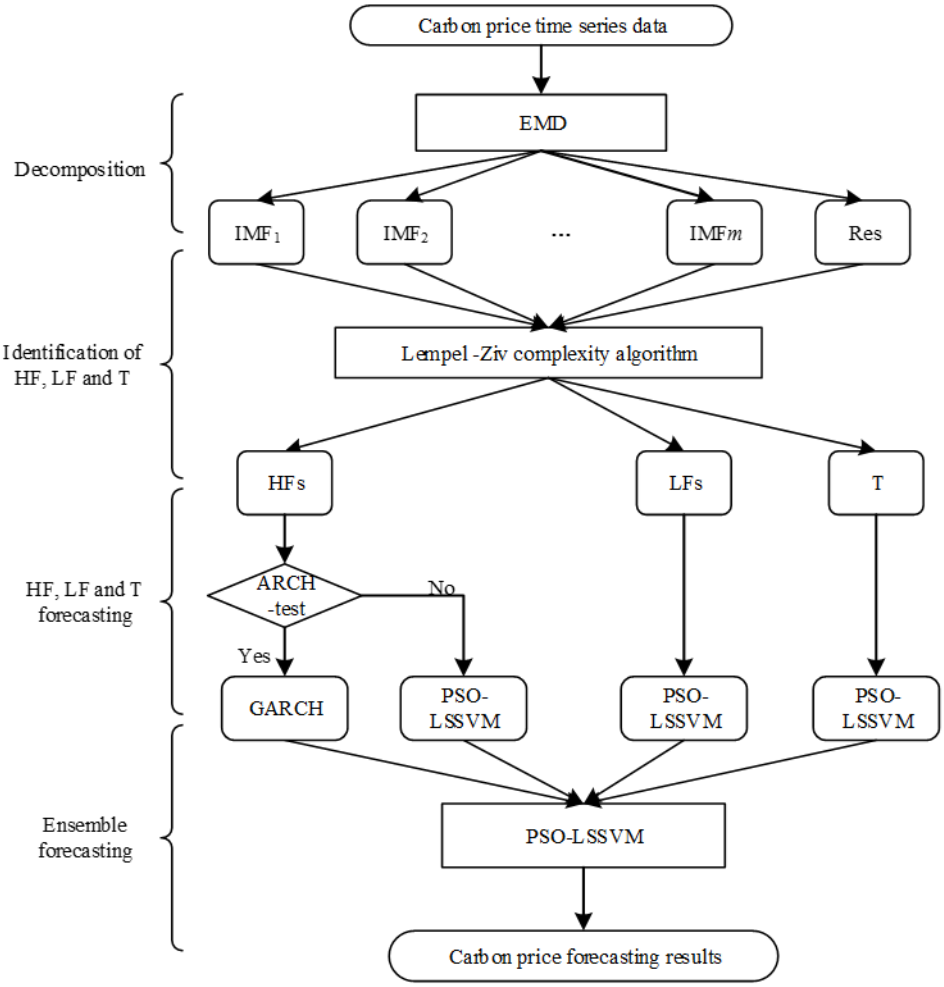


Fig.2. The forecasting process for carbon price with EMD-HLT-LSSVM model

3. Empirical analysis

3.1 Data

The validity of the proposed model is tested using the real carbon market data. We use the daily European Union Allowance futures prices with maturity date in December 2015 and 2016. They are known as DEC15 and DEC16 respectively. The original data source is the Intercontinental Exchange (ICE). It has the maximum trading volume under

the EU ETS. The data set is divided into two subsets: a training set and a testing set. The training set is used to estimate the model parameters. The testing set is used to evaluate the established models. Details of sample carbon price data set are reported in Table 1.

Table 1. Samples of carbon prices

Carbon prices		Size	Date
DEC15	Sample set	583	January 02, 2013-April 14, 2015
	Training set	460	January 02, 2013- October 16, 2014
	Testing set	123	October 17, 2014-April 14, 2015
DEC16	Sample set	532	January 02, 2013-January 29, 2015
	Training set	420	January 02, 2013- August 21, 2014
	Testing set	112	August 22, 2014-January 29, 2015

3.2Forecasting evaluation criteria

We use the *RMSE* and directional perdition statistics (D_{stat}) (Zhu and Wei, 2013) to evaluate the prediction performance of the established models. When evaluating the predictive power of models, one practical issue is what tests should be employed to decide one model is better than other models. To evaluate models more objectively and robustly, the superior predictive ability (SPA) test (Hansen and Lunde, 2005) and the model confidence set (MCS) test (Hansen et al., 2011) are applied in evaluating the accuracy of level forecasting, while rate test (RT) and Pesaran-Timmermann test (PT) (Pesaran and Timmermann, 1992) are applied in evaluating the accuracy of directional forecasting.

3.3Nonstationary and nonlinear tests of carbon prices

Among different statistical tests, the Augmented Dicky-Fuller (ADF) test is a popular stationarity test and Brock-Decher-Scheikman (BDS) test is a popular nonlinarity test. When the BDS test is used, the embedding dimension is set to 2-5, and the dimensional distance is set to 0.7 times of the variance of data. The Eviews 8.0 software package is employed to test the nonstationarity and nonlinearity of carbon prices, and the test results are shown in Tables 2 and 3. The ADF test demonstrates that carbon prices are nonstationary at the significance level of 5%. BDS test demonstrates that carbon prices are nonlinear at the significance level of 1% .

Table 2. ADF test results

Carbon prices	t-Stat	Prob.	Stationarty
DEC15	-2.138391	0.2298	×
DEC16	-2.569762	0.1004	×

Note:×shows that carbon price is nonstationary at the significance level of 5%.

Table 3. BDS test results

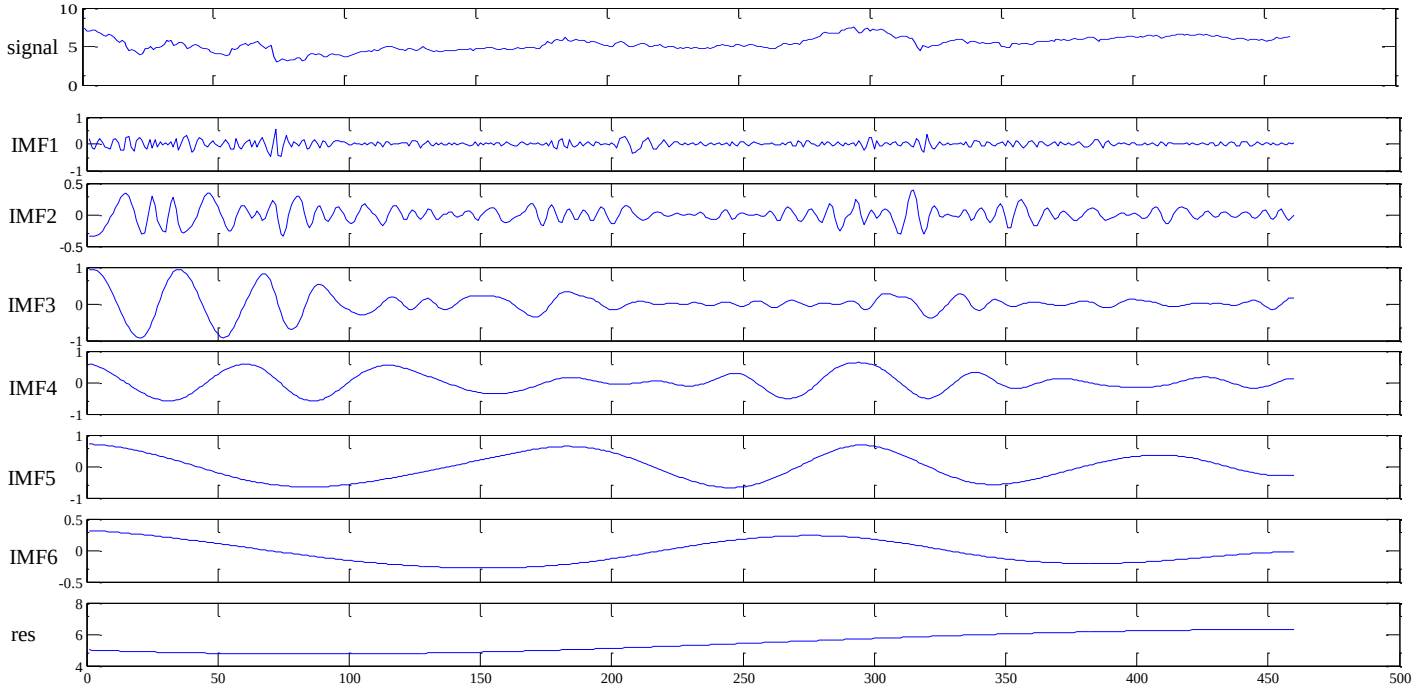
Carbon prices	*m*-dimensional space								Linearity
	2		3		4		5		
	Stat.	Prob.	Stat.	Prob.	Stat.	Prob.	Stat.	Prob.	
DEC15	0.1808	0.000	0.3048	0.000	0.3891	0.000	0.4452	0.000	×
DEC16	0.1753	0.000	0.2956	0.000	0.3768	0.000	0.3768	0.000	×

Note:×shows that carbon price is nonlinear at the significance level of 1%.

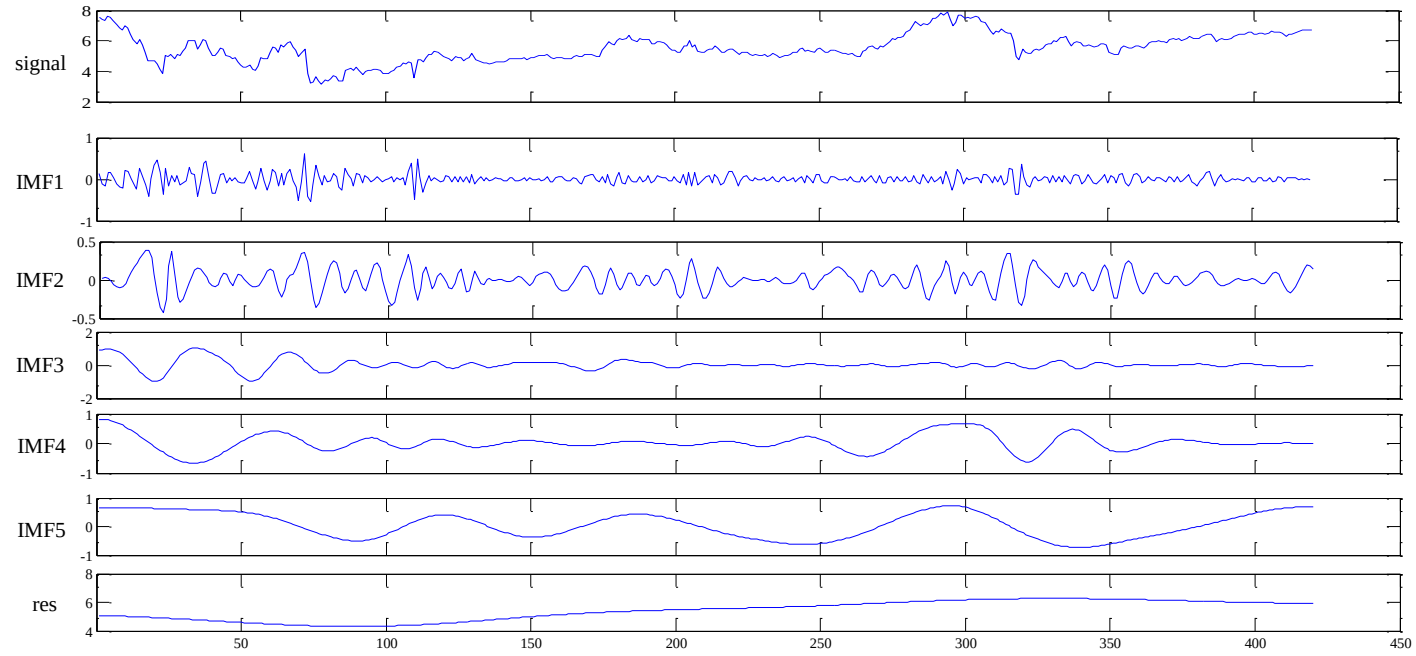
3.4 Decomposition of EMD

The EMD introduced in section 2.1 is used to decompose the carbon prices into IMFs and residue in this study. The model is one-step-ahead forecasting, so every decomposition of the carbon price time series is implemented once just for the forecasting of only one day's price. Thus we have to decompose DEC16 for 112 times and DEC15 for 123 times. For the sake of space limitation, only the first decomposition results of DEC15 and DEC16 is reported in Fig.3.

We arranged all the IMFs by their frequency, from highest one to the lowest one. The last one is the residue. We found that HFs are characterized by randomness, LFs are characterized by strong periodic fluctuations, and the residue is characterized by trend factor. Compared with the original carbon prices, these decomposed IMFs demonstrate simpler structure, more stationary fluctuation, and greater regularity. All of these characteristics lead to higher fitting and forecasting accuracy.



(a) The first decomposition result of DEC15 during January 02, 2013- October 16, 2014



(b) The first decomposition result of DEC16 during January 02, 2013- August 21, 2014

Fig.3. Results of the first EMD decomposition of carbon prices

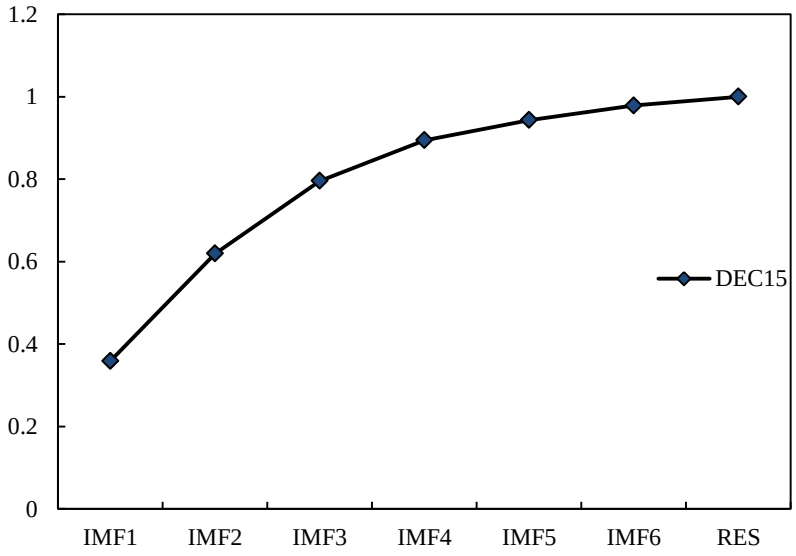
For comparisons, the EEMD and extrema symmetry expansion EMD (Ren et al. 2012) are implemented as well in this study. EEMD can restrain the mode mixing, and extrema symmetry expansion EMD can restrain the end effect. However, whether those can improve the accuracy effectively in our paradigm for carbon price forecasting is still an

open problem, this will be discussed in the following sections.

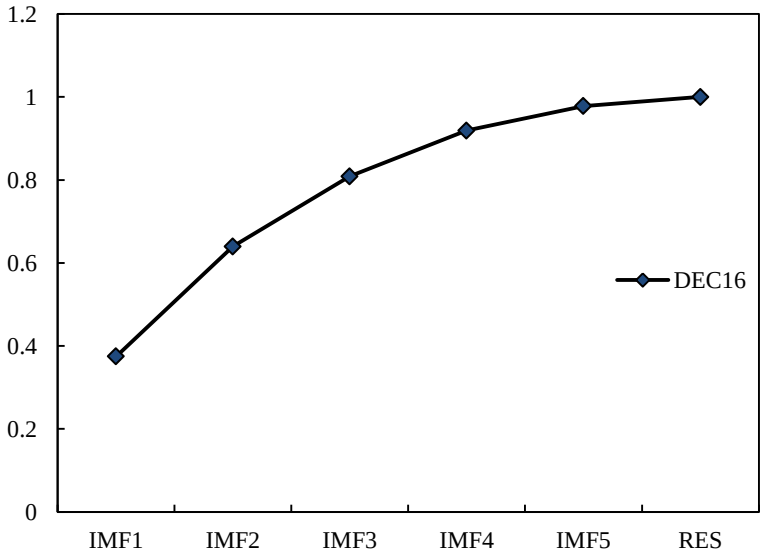
3.5 Identification of HFs, LFs, and T

The Lempel-Ziv complexity algorithm is applied to identify the HFs, LFs of DEC15 and DEC16. For the sake that there is a different identification result for each recursive decomposition. The identification result of the first group of IMFs and residue is reported in Fig. 4. In this study, the critical value λ_0 is set to 80%. It is obtained that k is 4 in the first group of IMFs and residue of DEC15 and η is 89.4%, thus IMF_1-IMF_4 belong to HFs, while IMF_5-IMF_6 are the LFs. Similarly k is 3 in the first group of IMFs and residue of DEC16 and η is 80.9%, so IMF_1-IMF_3 are HFs and IMF_4-IMF_5 are LFs. Besides that, the residues of two carbon prices are the Ts.

In addition, through the ARCH test，it is found that all of the HFs are with the ARCH effects at the significance level of 5%, however, all of the LFs are no ARCH effects at the significance level of 5% in the study. Thus, GARCH(1,1) is utilized for the prediction of HFs. Furthermore, for comparisons, the fine-to-coarse reconstruction (FTC) algorithm (Zhang et al., 2008; Zhu et al., 2015) is also used to identify HFs, LFs of DEC15 and DEC16. Which of the two methods is more appropriate for our paradigm for carbon price forecasting will be demonstrated by comparisons of accuracy in the following sections.



(a) The complexities of the first group of IMFs for DEC15 during January 02, 2013- October 16, 2014



(b) The complexities of the first group of IMFs for DEC16 during January 02, 2013- August 21, 2014

Fig.4. The complexities of the first group of IMFs for carbon prices

3.6 Forecasting results and discussions

In this study, the proposed EMD-HLT-LSSVM model is used for carbon price one-step-ahead forecasting. For comparisons, the single GARCH and LSSVM models, multiscale linear ensemble prediction models including EMD-GARCH-Σ (add the values of all IMFs and residue predicted by GARCH), EMD-LSSVM-Σ (add the values of all IMFs and residue predicted by LSSVM) and EMD-HLT-Σ (add the values of HFs predicted through GARCH and the LFs and T predicted by LSSVM), EMD-HLT(FTC)- Σ(use the fine-to-coarse reconstruction for identification of HFs and LFs, and add the values of HFs predicted through GARCH and the LFs and T predicted by LSSVM), as well as multiscale nonlinear ensemble prediction models including EMD-GARCH-LSSVM (LSSVM nonlinearly aggregate the values of all IMFs and residue predicted by GARCH) and EMD-LSSVM-LSSVM (LSSVM nonlinearly aggregate the values of all IMFs and residue predicted by LSSVM) are applied for carbon price forecasting. Meanwhile, EEMD-GARCH-Σ (add the values of all IMFs and residue predicted by GARCH), EEMD-LSSVM-Σ (add the values of all IMFs and residue predicted by LSSVM), EEMD-HLT-Σ, expansion EMD-GARCH-Σ (add the values of all IMFs and residue predicted by GARCH), expansion EMD-LSSVM-Σ (add the values of all IMFs and residue predicted by LSSVM) and expansion EMD-HLT-Σ are implemented to compare the decomposition methods. All the GARCH models are established in the Eviews 6.0 software package. In addition, the kernel function prototype is selected for all LSSVM models, and these models are implemented by the Matlab R2012b platform. In the PSO algorithm, the integer coding is used for m, τ, d ,and the real coding is adopted for other parameters. The numbers of particles and maximum iteration are set to 40 and 10 respectively. Meanwhile, $m \in [0,20]$, $\tau \in [0,5]$, $s \in [0.001,10000]$, $d \in [0,10]$, $h \in [-100,100]$, $\lambda_i \in [-50,50], i = 1,2,3,4$, $\gamma \in [0.0001,10000]$, $\sigma^2 = [0.0001,10000]$, $c_1 = c_2 = 2$, $w \in [0.1,0.9]$, and $v_{\max} = \frac{6}{\text{numbers of particles}} \cdot (x_{\max} - x_{\min})$. Table 4 illustrates the optimal model parameters of forecasting the first point in the test sets of DEC15 and DEC16 as an example. The comparisons of RMSE and D_{stat} between prediction models are shown in Table 5. The results of SPA test and MCS test are respectively shown in Tables 6 and 7. The results of PT test and RT test can be found in Tables 8 and 9. Fig.5 demonstrates the out-of-sample prediction results of DEC15 and DEC16 using the proposed EMD-HLT-LSSVM model.

In terms of level forecasting demonstrated by RMSE, it can be found that, firstly, all the multiscale ensemble prediction models obviously outperform single prediction models. This is mainly attributed to the fact that after EMD decomposition, the structure and fluctuations of IMFs (residue) carbon prices become simpler, more stable and more regular, which significantly improves the prediction accuracy. Secondly, for single prediction models, the prediction accuracy of GARCH model is superior over that of LSSVM model, because the former has a strong predicting ability for complicated randomness, time-varying and aggregation while original time series present such features. While for multiscale ensemble prediction models, LSSVM present advantage over GARCH, because IMFs' data characteristics of simpleness and regularity make the advantage of nonlinear modeling and self-learning visible. Thirdly, compared with linear ensemble (Σ), nonlinear ensemble (LSSVM) has a great advantage because it makes full use of different structures of all the IMFs, which overcomes the major limitation of the linear ensemble. The last but not least, the proposed EEMD-HLT-LSSVM model achieves the highest level accuracy than any other models, which shows that it has a significant advantage by integrating multiscale decomposition, hybrid modeling and nonlinear ensemble. From the perspective of directional prediction, similar conclusions can be drawn according to the results of D_{stat}. The proposed

EMD-HLT-LSSVM model achieves the optimal direction accuracy both in DEC15 and DEC16.

From the results of SPA test, the proposed EMD-HLT-LSSVM model obtains the best prediction accuracy on all 8 loss functions. It indicates that this model has a significant advantage in level prediction which can better depict and forecast the volatility characteristics of carbon prices. Similar results can be found in MCS test: compared with other models, the proposed EEMD-HLT-LSSVM model achieves the best prediction, which strongly confirms that robustness of the proposed model. From the results of PT test, all of the multiscale ensemble forecasting models can reject the null hypothesis at the confidence level of 99% for both the carbon prices, which shows that the forecasting directions are closely related to the real ones. RT test further reveals that multiscale ensemble prediction models of nonlinear ensemble are notably more robust than single prediction models. However, there are no significant differences between the multiscale ensemble forecasting models, although the D_{stat} of proposed EMD-HLT-LSSVM model is the highest.

It is noted that, the impact of mode mixing on carbon price forecasting is demonstrated by the comparison between EMD and EEMD which is designed to restrain the mode mixing. The results show that with the same conditions, the performance of EMD is better than that of EEMD in terms of RMSE and D_{stat}, while the difference between them is not significant at the significant level of 5%. The reason may lie in that: through EMD decomposition, carbon prices are decomposed into simple modes, which significantly improves the prediction accuracy. Otherwise, the mode mixing have no huge impacts on our predicted models and their prediction accuracies, which can also be drawn from the results of SPA test, MCS test, PT test and RT test.

As for the end effect of EMD, for comparison, this study utilizes the extrema symmetry expansion method to deal with this problem. It is found that the accuracy of expansion EMD is slightly higher than that of EMD, but the difference between them is not significant at the significant level of 5% as well. Namely, in our paradigm, expansion EMD cannot improve the prediction significantly. The main reason may lie in that the end effect of EMD decomposition is not so significant as not to lower the performance of the proposed models, which can also be drawn from the results of SPA test, MCS test, PT test and RT test.

Compared with the fine-to-coarse reconstruction algorithm, the Lempel-Ziv complexity algorithm has a superiority in the proposed models. However, the difference between them is not significant at the significant level of 5% as well, which is drawn from the results of SPA test, MCS test, PT test and RT test.

We can draw the following conclusions based on our results from the empirical analysis: firstly, the proposed EMD-HLT-LSSVM model produce forecasts at higher level of accuracy than those of other models, in terms of the level and directional predictions. Secondly, multiscale ensemble prediction models demonstrate the superior performance than single prediction models, when it is used to predict carbon prices. This results suggest that in the case of carbon price prediction, decomposition-ensemble strategy can lead to significantly improved prediction accuracy. Thirdly, because of its superior capability to handle the high nonstationarity and nonlinearity of carbon prices, the proposed EMD-HLT-LSSVM model, integrating multiscale decomposition, hybrid modeling and nonlinear ensemble, is more applicable to forecast carbon prices. Fourthly, the EEMD, extrema symmetry expansion EMD and fine-to-coarse reconstruction algorithm have no significant effect on our proposed models and carbon price forecasting. Finally, overall the proposed EMD-HLT-LSSVM model obtain the highest level of carbon price prediction accuracy. It can offer a competitive carbon price prediction approach.

Table 4 The optimal parameters of proposed model for the first prediction

Carbon prices	Parameters	Original series	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	RES	Ensemble (LSSVM)
DEC15	*d*	3.0	GARCH	GARCH	GARCH	GARCH	9	8	1	0

(October 17, 2014)	t	-0.3201	(1,1) [*1]	(1,1)	(1,1)	(1,1)	17.6943	9.0981	7.3660	7.672
	σ^2	0.01					2696.60	8207.8	202.668	43.52
	s	4465.6					8355.42	8887.94	6705.00	3650.67
	h	-51.18					-13.4478	-33.7888	3.35906	-8.750
	λ_1	-0.7856					-4.4547	1.0647	-1.2799	-0.1335
	λ_2	1.064					4.000	4.4843	-6.1840	-3.0601
	λ_3	2.350					-3.2264	5.5932	3.9282	4.4781
	λ_4	-3.306					0.2478	-1.0372	-5.6424	1.0427
	γ	3850.29					955.85	1731.22	5386.54	4430.70
	m	8.0					5	6	3	
	τ	1.0					1	1	1	
	d	7				10	10	*[2]	2	0
	t	-13.6104				-5.8449	-15.3125		2.093	1.8399
	σ^2	6610.22				5399.51	709.2759		7551.26	5586.5
	s	5309.58				0.0001	1786.51		9273. 60	4627.8
	h	-77.5338				-24.9955	-18.5010		45.2077	16.3845
DEC16 (August 21, 2014)	λ_1	4.781117	GARCH (1,1)	GARCH (1,1)	GARCH (1,1)	-8.63625	-0.79254		2.0912	0.005
	λ_2	3.401495				-2.2915	6.5770		3.3021	1.5827
	λ_3	2.971208				-0.4763	0.09029		-0.9549	8.4280
	λ_4	-2.02444				-0.5852	-3.05356		5.07029	0.6017
	γ	5540.994				7496.12	643.2569		7417.05	2231.5
	m	5				6	5		4	
	τ	1				1	1		1	

Note:*1GARCH(p, q) , in which p, q are respectively the orders of autoregressive and moving average.

*2 The training set of forecasting first point in DEC16's testing set is decomposed into only five IMFs.

Table 5 Comparison of the RMSEs and the D_{stat}s

MODEL	DEC15		DEC16	
	RSME	D_{stat1}	RSME	D_{stat1}
GARCH	0.110	0.626	0.111	0.598
LSSVM	0.123	0.593	0.117	0.580
EMD-GARCH-Σ	0.104	0.667	0.105	0.634
EMD-LSSVM-Σ	0.103	0.683	0.103	0.670
EMD-HLT-Σ	0.099	0.715	0.101	0.705
EMD-HLT(FTC)-Σ	0.102	0.699	0.103	0.696
EMD-GARCH-LSSVM	0.096	0.715	0.096	0.714
EMD-LSSVM-LSSVM	0.094	0.724	0.095	0.723
EMD-HLT-LSSVM	0.089	0.748	0.090	0.741
expansion EMD-GARCH-Σ	0.104	0.675	0.105	0.661
expansion EMD-LSSVM-Σ	0.102	0.683	0.102	0.670
expansion EMD-HLT-Σ	0.099	0.715	0.099	0.688
EEMD-GARCH-Σ	0.105	0.659	0.108	0.625
EEMD-LSSVM-Σ	0.104	0.675	0.107	0.634
EMD-HLT-Σ	0.103	0.691	0.105	0.688

Table 6 Results of SPA tests

	BENCHMARK	LOSS FUNCTION							
		MSE1	MAE1	MSE2	MAE2	HMSE	HMAE	QLIKE	RLOG
DEC15	GARCH	0.0614	0.0801	0.0603	0.0761	0.0572	0.0307	0.0675	0.0434
	LSSVM	0.0166	0.0095	0.0156	0.0096	0.0155	0.0096	0.0178	0.0168
	EEMD-GARCH-Σ	0.1168	0.0856	0.063	0.1235	0.0822	0.0961	0.0858	0.0762
	EEMD-LSSVM-Σ	0.1697	0.2249	0.1686	0.2179	0.17	0.2367	0.1697	0.1698
	EEMD-HLT-Σ	0.166	0.1063	0.1665	0.1045	0.1642	0.1095	0.1672	0.1655
	EMD-GARCH-Σ	0.1312	0.1898	0.1468	0.1981	0.1214	0.1871	0.124	0.1222
	EMD-LSSVM-Σ	0.1332	0.0965	0.1281	0.1319	0.1313	0.1197	0.1387	0.1757
	EMD-HLT-Σ	0.2071	0.2293	0.2188	0.2358	0.1858	0.2165	0.2117	0.2008
	EMD-HLT(FTC)-Σ	0.1794	0.2596	0.1435	0.2498	0.155	0.2736	0.1618	0.1558
	expansion EMD-GARCH-Σ	0.1481	0.1591	0.1048	0.157	0.139	0.1201	0.1272	0.1492
	expansion EMD-LSSVM-Σ	0.1907	0.2701	0.1871	0.2652	0.1809	0.2682	0.2053	0.1945
	expansion EMD-HLT-Σ	0.3147	0.3938	0.3256	0.4065	0.3023	0.3849	0.3067	0.305
	EMD-GARCH-LSSVM	0.5785	0.4364	0.562	0.4323	0.5607	0.4372	0.5759	0.5688
	EMD-LSSVM-LSSVM	0.6844	0.7095	0.6652	0.7152	0.6971	0.7229	0.7118	0.6979
	EMD-HLT-LSSVM	0.9698	0.967	0.974	0.9698	0.9695	0.9711	0.9674	0.9683
	BENCHMARK	LOSS FUNCTION							
		MSE1	MAE1	MSE2	MAE2	HMSE	HMAE	QLIKE	RLOG
DEC16	GARCH	0.0482	0.0375	0.0285	0.0264	0.00257	0.0519	0.01004	0.0198
	LSSVM	0.0134	0.0196	0.0086	0.0083	0.0129	0.0081	0.0037	0.0084
	EEMD-GARCH-Σ	0.0748	0.0754	0.0561	0.07003	0.0529	0.106	0.045	0.0534
	EEMD-LSSVM-Σ	0.0928	0.0106	0.0586	0.117	0.0642	0.1263	0.0501	0.0689
	EEMD-HLT-Σ	0.1655	0.2097	0.1767	0.15411	0.1545	0.18072	0.1604	0.1583
	EMD-GARCH-Σ	0.0859	0.1305	0.0804	0.1296	0.0966	0.1312	0.0991	0.0987
	EMD-LSSVM-Σ	0.145	0.1453	0.143	0.1446	0.1527	0.2477	0.1484	0.1498
	EMD-HLT-Σ	0.2361	0.2575	0.2187	0.26519	0.2474	0.4006	0.2708	0.265
	EMD-HLT(FTC)-Σ	0.2262	0.2399	0.2054	0.2375	0.2544	0.3432	0.2551	0.2549
	expansion EMD-GARCH-Σ	0.13	0.1334	0.1149	0.1328	0.152	0.1464	0.1936	0.1512
	expansion EMD-LSSVM-Σ	0.2983	0.1726	0.3048	0.176	0.2977	0.2706	0.2929	0.3892
	expansion EMD-HLT-Σ	0.3546	0.278	0.3141	0.285	0.3752	0.3845	0.3924	0.672
	EMD-GARCH-LSSVM	0.6017	0.3198	0.5489	0.3207	0.6647	0.3522	0.6953	0.672
	EMD-LSSVM-LSSVM	0.6333	0.3433	0.6205	0.3346	0.6471	0.3594	0.6412	0.6432
	EMD-HLT-LSSVM	0.9844	0.9795	0.9814	0.9782	0.9865	0.9821	0.9867	0.9864

Table 7 Results of MCS tests

	BENCHMARK	LOSS FUNCTION							
		MSE1	MAE1	MSE2	MAE2	HMSE	HMAE	QLIKE	RLOG
DEC15	GARCH	0.4468	0.2710	0.4518	0.2677	0.4316	0.2728	0.4564	0.4475
	LSSVM	0.2318	0.2042	0.2246	0.2037	0.2252	0.2041	0.2557	0.2437
	EEMD-GARCH-Σ	0.5302	0.5292	0.5150	0.5172	0.5133	0.5180	0.5667	0.5451
	EEMD-LSSVM-Σ	0.5392	0.6363	0.5150	0.6336	0.5916	0.6617	0.5726	0.5451

	EEMD-HLT-Σ	0.5934	0.5292	0.5328	0.5172	0.5916	0.5180	0.5726	0.5952
	EMD-GARCH-Σ	0.5934	0.5292	0.5241	0.6336	0.5133	0.6152	0.6001	0.5952
	EMD-LSSVM-Σ	0.5934	0.5292	0.5242	0.5172	0.5916	0.5180	0.6001	0.5952
	EMD-HLT-Σ	0.6587	0.6857	0.6008	0.6531	0.6248	0.6617	0.6192	0.6083
	EMD-HLT(FTC)-Σ	6.4130	0.6363	0.5974	0.6336	0.6248	0.6617	0.6192	0.6083
	EXPANSION EMD-GARCH-Σ	0.5604	0.5292	0.6008	0.5172	0.5133	0.5180	0.5667	0.6233
	EXPANSION EMD-LSSVM-Σ	0.5934	0.6363	0.6008	0.6336	0.5916	0.6617	0.6359	0.6235
	EXPANSION EMD-HLT-Σ	0.6672	0.6363	0.6351	0.6336	0.6783	0.6617	0.6359	0.6235
	EMD-GARCH-LSSVM	0.6672	0.6857	0.6504	0.7061	0.6783	0.6952	6.613	0.6524
	EMD-LSSVM-LSSVM	0.6955	0.6857	0.6504	0.7082	0.6783	0.6952	0.6613	0.6524
	EMD-HLT-LSSVM	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	BENCHMARK	LOSS FUNCTION							
		MSE1	MAE1	MSE2	MAE2	HMSE	HMAE	QLIKE	RLOG
	GARCH	0.3053	0.0571	0.3108	0.0521	0.2643	0.0667	0.0859	0.1022
	LSSVM	0.0556	0.0024	0.0784	0.0026	0.0371	0.0015	0.0426	0.0415
	EEMD-GARCH-Σ	0.4523	0.1237	0.4051	0.1159	0.5210	0.0953	0.5540	0.2822
	EEMD-LSSVM-Σ	0.3053	0.1237	0.3415	0.0816	0.2643	0.1167	0.5540	0.4192
	EEMD-HLT-Σ	0.5761	0.4532	0.5334	0.1457	0.6317	0.5077	0.6272	0.6269
	EMD-GARCH-Σ	0.4523	0.1237	0.4051	0.1159	0.5210	0.1363	0.5486	0.5162
DEC16	EMD-LSSVM-Σ	0.5761	0.1600	0.4156	0.1457	0.5210	0.1363	0.6259	0.6269
	EMD-HLT-Σ	0.5761	0.4532	0.5334	0.4024	0.6317	0.5077	0.6272	0.6269
	EMD-HLT(FTC)-Σ	0.5761	0.3421	0.5334	0.1159	0.6317	0.1788	0.6259	0.6269
	Expansion EMD-GARCH-Σ	0.4523	0.1600	0.4051	0.1127	0.5210	0.1538	0.5652	0.5162
	Expansion EMD-LSSVM-Σ	0.5761	0.2047	0.5334	0.1922	0.6317	0.2200	0.6272	0.6269
	Expansion EMD-HLT-Σ	0.5761	0.4532	0.5334	0.4024	0.6317	0.5077	0.6272	0.6269
	EMD-GARCH-LSSVM	0.5946	0.2816	0.6052	0.4024	0.7072	0.5077	0.7097	0.7088
	EMD-LSSVM-LSSVM	0.5946	0.4532	0.6052	0.4024	0.7072	0.5077	0.7097	0.7088
	EMD-HLT-LSSVM	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 8 Comparisons of PT tests

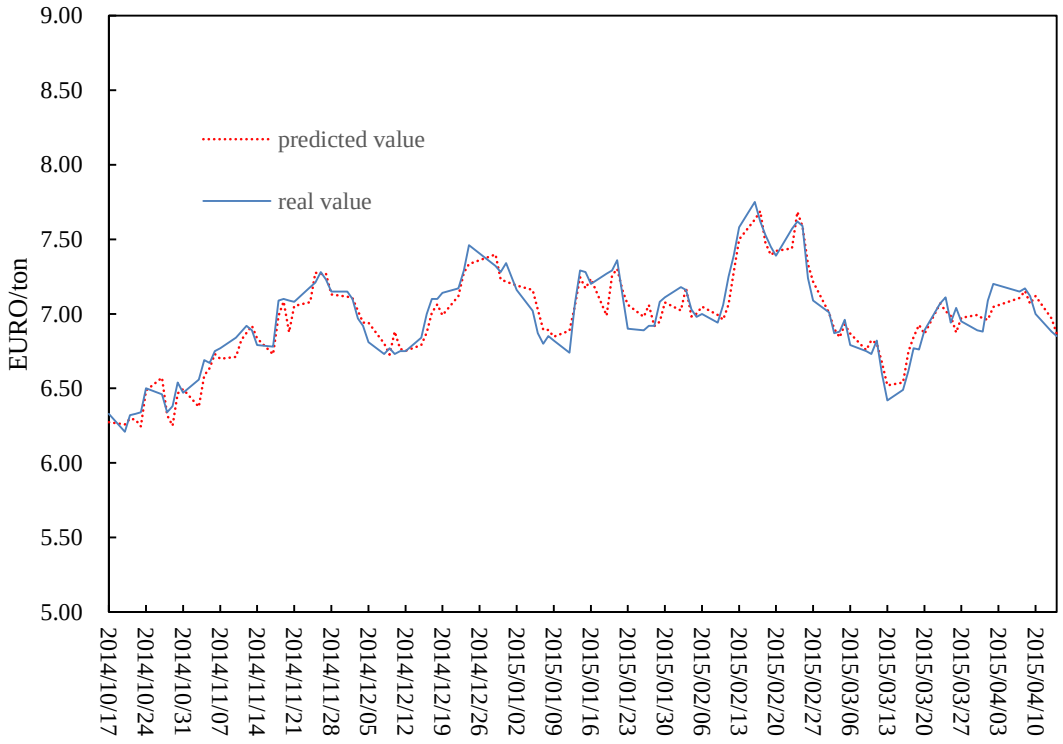
Model	Carbon price			
	DEC15		DEC16	
	t-value	P-value	t-value	P-value
GARCH	2.573	0.011	1.998	0.048
LSSVM	2.017	0.046	1.633	0.105
EEMD-GARCH-Σ	3.401	0.001	3.321	0.001
EEMD-LSSVM-Σ	3.752	0.000	3.416	0.001
EEMD-HLT-Σ	4.082	0.000	3.928	0.000
EMD-GARCH-Σ	3.58	0.001	2.794	0.006
EMD-LSSVM-Σ	3.924	0.000	3.443	0.001
EMD-HLT-Σ	4.529	0.000	4.199	0.000
EMD-HLT(FTC)-Σ	4.206	0.000	4.117	0.000
Expansion EMD-GARCH-Σ	3.71	0.000	2.524	0.013
Expansion EMD-LSSVM-Σ	3.896	0.000	2.687	0.008
Expansion EMD-HLT-Σ	4.62	0.000	3.807	0.000

EMD-GARCH-LSSVM	4.592	0.000	4.468	0.000
EMD-LSSVM-LSSVM	4.771	0.000	4.792	0.000
EMD-HLT-LSSVM	5.28	0.000	5.049	0.000

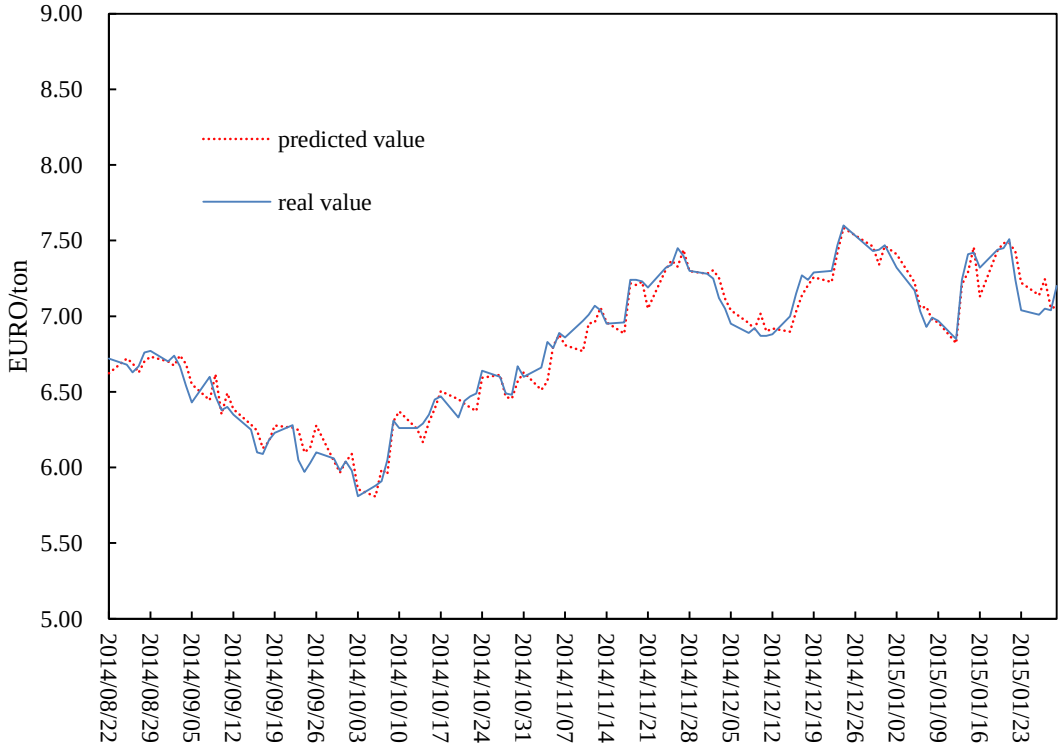
Table 9 Comparisons of RT tests

Carbon price	Test model	Base model													
		GARCH	LSSVM	EEMD-GARCH-Σ	EEMD-LSSVM-Σ	EEMD-HLT-Σ	EMD-GARCH-Σ	EMD-LSSVM-Σ	EMD-HLT-Σ	EMD-HLT(FTC)-Σ	Expansion EMD-GARCH-Σ	Expansion EMD-LSSVM-Σ	Expansion EMD-HLT-Σ	EMD-GARCH-LSSVM	EMD-LSSVM-LSSVM
DEC15	LSSVM	-0.523 (0.601)													
	EEMD-GARCH-Σ	0.532 (0.595)	1.057 (0.292)												
	EEMD-LSSVM-Σ	0.803 (0.423)	1.328 (0.185)	0.271 (0.787)											
	EEMD-HLT-Σ	1.078 (0.282)	1.604 (0.11)	0.545 (0.586)	0.274 (0.784)										
	EMD-GARCH-Σ	0.667 (0.505)	1.192 (0.235)	0.135 (0.893)	-0.136 (0.892)	-0.41 (0.682)									
	EMD-LSSVM-Σ	0.94 (0.348)	1.466 (0.144)	0.407 (0.684)	0.137 (0.892)	-0.137 (0.891)	0.272 (0.786)								
	EMD-HLT-Σ	1.499 (0.135)	2.028 (0.044)	0.964 (0.336)	0.693 (0.489)	0.419 (0.676)	0.829 (0.408)	0.556 (0.578)							
	EMD-HLT(FTC)-Σ	1.217 (0.225)	1.744 (0.082)	0.683 (0.495)	0.413 (0.68)	0.139 (0.89)	0.548 (0.584)	0.276 (0.783)	-0.28 (0.78)						
	Expansion EMD-GARCH-Σ	0.803 (0.423)	1.328 (0.185)	0.271 (0.787)	0.000 (1)	-0.274 (0.784)	0.136 (0.892)	-0.137 (0.892)	-0.693 (0.489)	-0.413 (0.68)					
	Expansion EMD-LSSVM-Σ	0.94 (0.348)	1.466 (0.144)	0.407 (0.684)	0.137 (0.892)	-0.137 (0.891)	0.272 (0.786)	0.000 (1)	-0.556 (0.578)	-0.276 (0.783)	0.137 (0.892)				
	Expansion EMD-HLT-Σ	1.499 (0.135)	2.028 (0.044)	0.964 (0.336)	0.693 (0.489)	0.419 (0.676)	0.829 (0.408)	0.556 (0.578)	0.000 (1)	0.28 (0.78)	0.693 (0.489)	0.556 (0.578)			
	EMD-GARCH-LSSVM	1.499 (0.135)	2.028 (0.044)	0.964 (0.336)	0.693 (0.489)	0.419 (0.676)	0.829 (0.408)	0.556 (0.578)	0.000 (1)	0.28 (0.78)	0.693 (0.489)	0.556 (0.578)	0.000 (1)		
	EMD-LSSVM-LSSVM	1.642 (0.102)	2.172 (0.031)	1.107 (0.27)	0.835 (0.404)	0.561 (0.575)	0.971 (0.332)	0.699 (0.486)	0.142 (0.887)	0.422 (0.673)	0.835 (0.404)	0.699 (0.486)	0.142 (0.887)	0.142 (0.887)	
	EMD-HLT-LSSVM	2.08 (0.039)	2.613 (0.01)	1.543 (0.124)	1.271 (0.205)	0.995 (0.321)	1.407 (0.161)	1.133 (0.258)	0.576 (0.565)	0.857 (0.393)	1.271 (0.205)	1.133 (0.258)	0.576 (0.565)	0.576 (0.565)	0.252 (0.801)
DEC16	LSSVM	-0.272 (0.786)													
	EEMD-GARCH-Σ	0.97 (0.333)	1.243 (0.215)												
	EEMD-LSSVM-Σ	1.113 (0.267)	1.386 (0.167)	0.142 (0.888)											
	EEMD-HLT-Σ	1.401 (0.163)	1.675 (0.095)	0.428 (0.669)	0.286 (0.775)										
	EMD-GARCH-Σ	0.55 (0.583)	0.822 (0.412)	-0.42 (0.675)	-0.561 (0.575)	-0.848 (0.397)									
	EMD-LSSVM-Σ	1.113 (0.267)	1.386 (0.167)	0.142 (0.888)	0 (1)	-0.286 (0.775)	0.561 (0.575)								
	EMD-HLT-Σ	1.694 (0.092)	1.969 (0.05)	0.719 (0.473)	0.577 (0.565)	0.291 (0.772)	1.14 (0.256)	0.577 (0.565)							
	EMD-HLT(FTC)-Σ	1.546 (0.123)	1.821 (0.07)	0.573 (0.567)	0.431 (0.667)	0.145 (0.885)	0.993 (0.322)	0.431 (0.667)	-0.146 (0.884)						
	Expansion EMD-GARCH-Σ	0.411 (0.681)	0.683 (0.495)	-0.558 (0.577)	-0.7 (0.485)	-0.987 (0.325)	-0.138 (0.89)	-0.7 (0.485)	-1.279 (0.202)	-1.132 (0.259)					
	Expansion EMD-LSSVM-Σ	0.55 (0.583)	0.822 (0.412)	-0.42 (0.675)	-0.561 (0.575)	-0.848 (0.397)	0.000 (1)	-0.561 (0.575)	-1.14 (0.256)	-0.993 (0.322)	0.138 (0.89)				
	Expansion EMD-HLT-Σ	1.401 (0.163)	1.675 (0.095)	0.428 (0.669)	0.286 (0.775)	0.000 (1)	0.848 (0.397)	0.286 (0.775)	-0.291 (0.772)	-0.145 (0.885)	0.987 (0.325)	0.848 (0.397)			
	EMD-GARCH-LSSVM	1.843 (0.067)	2.118 (0.035)	0.866 (0.387)	0.724 (0.47)	0.438 (0.662)	1.288 (0.199)	0.724 (0.47)	0.147 (0.883)	0.293 (0.77)	1.427 (0.155)	1.288 (0.199)	0.438 (0.662)		
	EMD-LSSVM-LSSVM	1.993 (0.047)	2.27 (0.024)	1.015 (0.311)	0.873 (0.383)	0.587 (0.558)	1.437 (0.152)	0.873 (0.383)	0.296 (0.768)	0.442 (0.659)	1.577 (0.116)	1.437 (0.152)	0.587 (0.558)	0.149 (0.882)	
	EMD-HLT-LSSVM	2.3 (0.022)	2.578 (0.011)	1.318 (0.189)	1.176 (0.241)	0.889 (0.375)	1.741 (0.083)	1.176 (0.241)	0.598 (0.551)	0.744 (0.458)	1.881 (0.061)	1.741 (0.083)	0.889 (0.375)	0.45 (0.653)	0.176 (0.861)

Note: **z_{RT}(*p-value*).



(a) Out-of-sample prediction results of DEC15



(b) Out-of-sample prediction results of DEC16

Fig.5. Predicted results of EEMD-HLT-LSSVM model

4. Conclusion

In this paper, we propose a novel multiscale nonlinear ensemble learning paradigm for carbon price prediction. It integrates EMD, LSSVM with kernel function prototype and PSO models. Firstly, we use the EMD algorithm to

decompose the carbon prices into their constituent components. These include several IMFs and one residue with strong regularity, simple structure, and smooth fluctuations. Secondly, the Lempel-Ziv complexity algorithm is used to identify HFs, LFs, and T. Thirdly, we use the GARCH model to predict the HFs with ARCH effects, and used the LSSVM model with kernel function prototype to predict the other IMFs (include LFs and the HFs without ARCH effects) and T according to their own data characteristics. Fourthly, we use LSSVM-based nonlinear ensemble to aggregate the forecasts of HFs, LFs, and T and produced the final prediction values of the original carbon price. We found that the proposed model could obtain higher level and directional predictions and higher robustness compared with the common popular prediction methods. Thus our results suggest that the proposed model is competitive for predicting the high nonstationary, nonlinear and irregular carbon price.

This study only performs the carbon price one-step-ahead forecasting. How to run the carbon price multi-step-ahead forecasting is one further challenge. Besides, this study only uses carbon price time series data for modeling and forecasting. How to introduce more factors to build the better prediction model is a second further challenge. Effectively addressing the two challenges can further enhance the prediction accuracy of the high nonstationary, nonlinear and irregular time series data. In the future, the proposed method will be tested using various real market data.

Acknowledgements

We thank Xuetao Shi for collecting and dealing with the data. Our heartfelt thanks should also be given to the National Natural Science Foundation of China (71771105, 71473180, and 71303174), National Philosophy and Social Science Foundation of China (15ZDA054 and 16ZZD049), Guangdong Young Zhujiang Scholar (Yue Jiaoshi [2016]95), Natural Science Foundation for Distinguished Young Talents of Guangdong (2014A030306031), Guangdong Key Base of Humanities and Social Science—Enterprise Development Research Institute, Institute of Resource, Environment and Sustainable Development Research, and Guangzhou key Base of Humanities and Social Science—Centre for Low Carbon Economic Research for their funding supports.

References

Alonso J.M, Alvarruiz F, Desantes J.M, Hernandez L, Hernandez V, Molto G. Combining neural networks and genetic algorithms to predict and reduce diesel engine emission. IEEE Transactions on Evolutionary Computation, 2007, 11 ,46-55.

Atsalakis G S. Using computational intelligence to forecast carbon prices. Applied Soft Computing, 2016, 43:107-116.

Bollerslev T. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 1986, 31(3): 307-327.

Byun SJ, Cho H. Forecasting carbon futures volatility using GARCH models with energy volatilities. Energy Economics, 2013, 40: 207-221.

Chevallier J. Nonparametric modeling of carbon prices. Energy Economics, 2011, 33(6): 1267-1282.

Fan XH, Li SS, Tian LX. Chaotic characteristic identification for carbon price and an multi-layer perceptron network prediction model. Expert Systems with Applications, 2015, 42: 3945-3952.

Hansen P R, Lunde A， James M N． The model confidence set．Econometrica， 2011，79(2) : 453 - 497.

Hansen P R, Lunde A. A forecast comparison of volatility models: does anything beat a GARCH(1, 1) . Journal of Applied Econometrics, 2005 , 20:873-889.

Hansen P R，Lunde A．A test for superior predictive ability．Journal of Business and Economic Statistics，2005，

23(4) : 365 - 380 .

Huang NE, Shen Z, Long SR. A new view of nonlinear water waves: the Hilbert spectrum. Annual Review of Fluid Mechanics, 1999, 31: 417-457.

Huang NE, Shen Z, Long SR. The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. Proceedings of the Royal Society of London, 1998, 454: 903-995.

Kanen JLM. Carbon Trading and Pricing. Environmental Finance Publications, London, 2006.

Koop G, Tole L, Forecasting the European carbon market. Journal of the Royal Statistical Society Series A, 2013, 176(3): 723-741.

Lempel A, Ziv J. On the complexity of finite sequenced. IEEE Transactions on Information Theory , 1976,22(1): 75-81.

Liao GC , Tsao TP. Application of a fuzzy neural network combined with a chaos genetic algorithm and simulated annealing to short-term load forecasting. 2006, 10(3):330-340.

María ES, Francesco V, María MB. Understanding volatility dynamics in the EU-ETS market. Energy Policy, 2015, (82): 321-331.

Pehlivanoglu YV. A New Particle Swarm Optimization Method Enhanced With a Periodic Mutation Strategy and Neural Networks. 2013, 17(3):436-452.

Pesaran MH, Timmermann P. A simple nonparametric test of predictive performance. Journal of Business & Economic Statistics, 1992, 10:461-465.

Reilly JM, Paltsev S. An analysis of the European emission trading scheme. Report No.127, MIT Joint Program on the Science and Policy of Global Change, 2005.

Ren DQ, Yang SX,Wu ZT, Yang B. Research on end effect of LMD based time-frequency analysis in rotating machinery fault diagnosis. China Mechanical Engineering, 2012, 23(8): 951-956.

Silva DA, Silva JP, Neto ARR. Novel approaches using evolutionary computation for sparse least square support vector machines . Neurocomputing, 2015, 168:908-916.

Sun G, Chen T, Wei Z, et al. A Carbon Price Forecasting Model Based on Variational Mode Decomposition and Spiking Neural Networks. Energies, 2016, 9(1):54-62.

Wei L, Can L. The research on setting a unified interval of carbon price benchmark in the national carbon trading market of China. Applied Energy, 2015, 155:728-739.

Xiong T, Bao YK, Hu ZY. Does restraining end effect matter in EMD-based modeling frame work for time series prediction? Some experimental evidences. Neurocomputing, 2014, 123: 174-184.

Yu L, Chen HH, Wang SY, Lai KK. Evolving Least Squares Support Vector Machines for Stock Market Trend Mining. IEEE Transactions on Evolutionary Computation, 2009, 13(1):87-102.

Zhang JL ,Zhang YJ, Zhang L. A novel hybrid method for crude oil price forecasting. Energy Economics, 2015, 49: 649-659.

Zhang X, Lai KK, Wang SY. A new approach for crude oil price analysis based on Empirical Mode Decomposition. Energy Economics, 2008, 30: 905-918.

Zhang YJ, Wei YM. An overview of current research on EU ETS: Evidence from its operating mechanism and economic effect. Applied Energy, 2010, 87(6): 1804-1814.

Zhu B, Han D, Wang P, et al. Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. Applied Energy, 2017, 191:521-530.

Zhu BZ, Shi XT, Chevallier J, Wang P, Wei YM. An adaptive multiscale ensemble learning paradigm for nonstationary and nonlinear energy price time series forecasting. Journal of Forecasting, 2016, 35,633-651.

Zhu BZ, Wang P, Chevallier J, Wei YM. Carbon Price Analysis Using Empirical Mode Decomposition. 2015, 45(2):195-206.

Zhu BZ, Wei YM. Carbon price prediction based on integration of GMDH, particle swarm optimization and least squares support vector machines. Systems Engineering-Theory & Practice, 2011, 31(12):2264-2271.

Zhu BZ, Wei YM. Carbon price prediction with a hybrid ARIMA and least squares support vector machines methodology . Omega, 2013, (41): 517-524.

Zhu BZ. A Novel Multiscale Ensemble Carbon Price Prediction Model Integrating Empirical Mode Decomposition, Genetic Algorithm and Artificial Neural Network. Energies, 2012, (5): 355-370.

APPENDIX

1.EMD

In practice, IMFs are usually extracted following the sifting process as follows:

(1) Identify all the maxima and minima of carbon price time series $x(t)$;

(2) Generate their upper and lower envelopes, $e_{\max}(t)$ and $e_{\min}(t)$, with cubic spline interpolation.

(3) Calculate the point-by-point mean, $(m(t))$, from the upper and lower envelopes: $m(t)=\frac{e_{\max}(t)+e_{\min}(t)}{2}$;

(4) Extract the mean from carbon price time series and define the difference between $x(t)$ and $m(t)$ as $d(t)$: $d(t)=x(t)-m(t)$;

(5) Check the properties of $d(t)$:

① If it is an IMF, denote $d(t)$ as the *i*th IMF and replace $x(t)$ with the residue $r(t)=x(t)-d(t)$. The *i*th IMF is often denoted as $c_i(t)$ and the i is called its index;

② If it is not an IMF, replace $x(t)$ with $d(t)$;

(6) Repeat steps (1) – (5) until the residue satisfies the stopping criteria. At the end of this sifting procedure, carbon price $x(t)$ can be expressed as $x(t)=\sum_{i=1}^{m}c_i(t)+r_m(t)$, where m is the number of IMFs, $r_m(t)$ is the final residue, and $c_i(t)(i=1,2,\cdots m)$ are the IMFs, which are nearly orthogonal to each other, and all have nearly zero means.

2.LSSVM

For a dataset $\{x_i,y_i\}$, $i=1,2,\cdots,l$ $S=\left\{(x_i,y_i),x_i\in R^n,y_i\in R\right\}_{i=1}^{l}$, in which x_i is an input vector, y_i is the output vector corresponding to x_i, and l is the sample size, LSSVM is defined as the following optimization problem :

$$\min Q(\omega,b,e)=\frac{1}{2}\|\omega\|^2+\frac{\gamma}{2}\sum_{i=1}^{l}e_i^2$$

s.t.

$$y_i=\omega^T\varphi(x_i)+b+e_i,i=1,2,\cdots l$$

where $\varphi(\cdot)$ is a nonlinear mapping function, $\omega\in R^n$ is a weight vector, $e_i\in R$ is an error vector, and $\gamma>0$ is a penalty

factor. By introducing the Lagrange multiplier a_i, we can obtain:

$$L(\omega,b,e,a)=Q(\omega,b,e)-\sum_{i=1}^{l}a_i[\omega^T\varphi(x_i)+b+e_i-y_i]$$

According to the Karush-Kuhn-Tucker (KKT) condition, we can get:

$$\begin{cases}\dfrac{\partial L}{\partial\omega}=0\Rightarrow\omega-\sum_{i=1}^{l}a_i\varphi(x_i)=0\\ \dfrac{\partial L}{\partial b}=0\Rightarrow\sum_{i=1}^{l}a_i=0\\ \dfrac{\partial L}{\partial e_i}=0\Rightarrow\gamma e_i-a_i=0\\ \dfrac{\partial L}{\partial a_i}=0\Rightarrow\omega^T\varphi(x_i)+b+e_i-y_i=0\end{cases}$$

By eliminating variables ω and e_i, the linear equations are obtained:

$$\begin{bmatrix}0 & 1_v^T\\ 1_v & \Omega+\gamma^{-1}I\end{bmatrix}\begin{bmatrix}b\\ a\end{bmatrix}=\begin{bmatrix}0\\ y\end{bmatrix}.$$

where $y=[y_1,y_2,\cdots y_l]^T$, $1_v=[1,1,\cdots 1]^T$, $a=[a_1,a_2,\cdots a_l]^T$, I is a first-order unit matrix, Ω is a non-negative positive definite matrix of $l\times l$ which can meet the Mercer condition that $\Omega_{ij}=K(x_i,x_j)=\varphi(x_i)^T\varphi(x_i), i,j=1,2,\cdots l$, $K(\cdot)$ is a kernel function. For any a kernel function, an Ω can be determined and a,b are calculated using the above linear equations. Therefore, an LSSVM predictor can be built:

$$f(x)=\sum_{i=1}^{l}a_iK(x,x_j)+b\,.$$

3.RMSE and directional perdition statistics (D_{stat}) .They are defined respectively as:

$$RMSE=\sqrt{\frac{1}{n}\sum_{t=1}^{n}[\hat{x}(t)-x(t)]^2}\quad\text{and}\quad D_{state}=\frac{1}{n}\sum_{t=1}^{n}a_t\times 100\%$$

where $x(t)$ and $\hat{x}(t)$ are respectively the actual value and predicted value, n is the size of testing set, and $a=\begin{cases}1, & [x(t+1)-x(t)][(\hat{x}(t+1)-x(t)]\ge 0\\ 0, & others\end{cases}$. Obviously, the smaller the *RMSE*, the greater the D_{stat} is, and the higher the prediction accuracy is.

4. SPA test. 8 different loss functions are selected to test the prediction accuracy of the models, i.e. $L_i(i=1,2,\cdots 8)$. L_1 and L_2 are mean squared error (MSE) and mean absolute error (MAE) respectively, which is commonly used in such judgments. Based on them, L_3 and L_4 take quadratic terms into consideration. L_5 and L_6 introduce the heteroscedasticity. L_7 is estimated by the Gauss likelihood function and L_8 is similar to the value of Mincer-Zarnowitz regression. The specific definition of each loss function is defined as:

$$L_1: MSE_1=n^{-1}\sum_{t=1}^{n}[x(t)-\hat{x}(t)]^2 \qquad L_2: MAE_1=n^{-1}\sum_{t=1}^{n}\left|x(t)-\hat{x}(t)\right|$$

$$L_3: MSE_2=n^{-1}\sum_{t=1}^{n}[x(t)^2-\hat{x}(t)^2]^2 \qquad L_4: MAE_2=n^{-1}\sum_{t=1}^{n}\left|x(t)^2-\hat{x}(t)^2\right|$$

$$L_5 : HMSE = n^{-1}\sum_{t=1}^{n}[1-\frac{\hat{x}(t)}{x(t)}]^2 \qquad L_6 : HMAE = n^{-1}\sum_{t=1}^{n}\left|1-\frac{\hat{x}(t)}{x(t)}\right|$$

$$L_7 : QLIKE = n^{-1}\sum_{t=1}^{n}[\ln(\hat{x}(t))^2+(\frac{x_t}{\hat{x}(t)})^2] \qquad L_8 : R^2LOG = n^{-1}\sum_{t=1}^{n}[\ln\frac{x(t)}{\hat{x}(t)}]^2$$

It can be calculated from every prediction model $M_k, (k = 0,1,\cdots, J)$ to get the 8 loss functions, denoted by $L_t^{i,k}$. Define M_0 as the base model for SPA test, for other *J* models ($k=1,2,\cdots j$), relative loss function is obtained as: $Y_t^{i,k} = L_t^{i,0} - L_t^{i,k}$. The null hypothesis H_0 is: compared with one other model M_k, M_0 has the best performance in prediction. The following statistic is defined as:

$$T^{SPA} = \max_{i,k}\frac{\overline{Y}_t^{i,k}}{\overline{\omega}_{kk}}$$

$$\overline{Y}_t^{i,k} = n^{-1}\sum_{t=1}^{n}Y_t^{i,k}$$

where $\overline{\omega}_{kk}$ is the consistent estimation of standard error for $\overline{Y}_t^{i,k}$. The empirical distribution of T^{SPA} can be obtained from the process of Bootstrap. Following Hansen and Lunde (2005), the parameters of Bootstrap are set as: 10000 and 0.5 for the resamples and the dependence respectively. A higher the p-value (closer to 1) of SPA test, the less likelihood that the null hypothesis can be rejected, which means a higher prediction accuracy of the base model.

5. **MCS test**. The process of choosing the base model in SPA test may cause the problem of multiple comparisons. To handle this drawback, Hansen et al. (2011) developed MCS test. Similar to SPA test, MCS test assumes that all prediction models are in a collection of M_T and the relative loss functions of any two models (*u* and *v*) is defined as: $Y_t^{i,uv} = L_t^{i,u} - L_t^{i,v}$. MCS test carries out a series of significance tests in the collection to eliminate the models with a poor predictive ability. The null hypothesis H_0 is: The two models have the same prediction ability. The following range statistic is defined:

$$T_R = \max_{u,v\in M}\frac{\left|\overline{Y}_t^{i,uv}\right|}{\sqrt{\mathrm{var}(Y_t^{i,uv})}}$$

$$\overline{Y}_t^{i,uv} = n^{-1}\sum_{t=1}^{n}Y_t^{i,uv}$$

In the MCS test, the empirical distribution of T_R can be obtained from the process of Bootstrap as well. Following Hansen et al.(2011), the parameters of Bootstrap are set as: 10000 and 2 for the resamples and the block length respectively. For a given model $k(k\in M_T)$, the higher the p-value is, the higher its prediction accuracy is.

6. RT test. It is expressed as

$$z_{RT} = \frac{p_A - p_B}{\sqrt{\frac{p_A(1-p_A)}{n}+\frac{p_B(1-p_B)}{n}}} \sim N(0,1),\ n\to\infty$$

where p_A and p_B are respectively the accuracies of directional prediction of models A and B. The null hypothesis of RT test is that the accuracies of directional prediction of models A and B are the same. As to the two-sided test, when the absolute value of z_{RT} exceeds 1.96, the null hypothesis is rejected at the significance level of 5%.

7. PT test. It is defined as

$$z_{PT} = \frac{\hat{p} - p^*}{\sqrt{p^*(1-p^*)/n}} \sim N(0,1),\ n \to \infty$$

where $\hat{p} = \frac{1}{n}\sum_{t=1}^{n} H_t[(x_{t+1} - x_t)(\hat{x}_{t+1} - x_t)]$, i.e., the accuracy of directional prediction of the model, measured by D_{stat}. Meanwhile, $p^* = p_1\hat{p}_1 + (1 - p_1)(1 - \hat{p}_1)$, $p_1 = \frac{1}{n}\sum_{t=1}^{n} H_t(x_{t+1} - x_t)$, $\hat{p}_1 = \frac{1}{n}\sum_{t=1}^{n} H_t(\hat{x}_{t+1} - x_t)$, and $H(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$. The null assumption of PT test is that the predicted direction and the actual direction are independent of each other. When the absolute value of z_{PT} is greater than 1.96 in the two-sided test, the null hypothesis is rejected at the significance level of 5%.

A novel multiscale nonlinear ensemble learning paradigm for carbon price forecasting

Bangzhu Zhu[a*], Ping Wang[a], Tao Zhang [b], Yi-Ming Wei[c]

[a] School of Management, Jinan University, Guangzhou, Guangdong 510632, China

[b] Birmingham Business School, University of Birmingham, Edgbaston, Birmingham, UK, B15 2TT

[c] Center for Energy and Environmental Policy Research, Beijing Institute of Technology, Beijing 100081, China

Highlights

>A novel multiscale nonlinear ensemble learning paradigm is proposed for carbon price forecasting.

>Carbon price is decomposed into regular modes via ensemble empirical mode decomposition .

>Particle swarm optimization is used for model section of least square support vector machine with kernel function prototype.

>The high frequency, low frequency and trend modes are identified by the Lempel-Ziv complexity.

>Empirical results show that the presented approach can exceed the popular forecasting methods.

* Corresponding author: Bangzhu Zhu, Tel: +86 15915761388; Email:wpzbz@126.com.

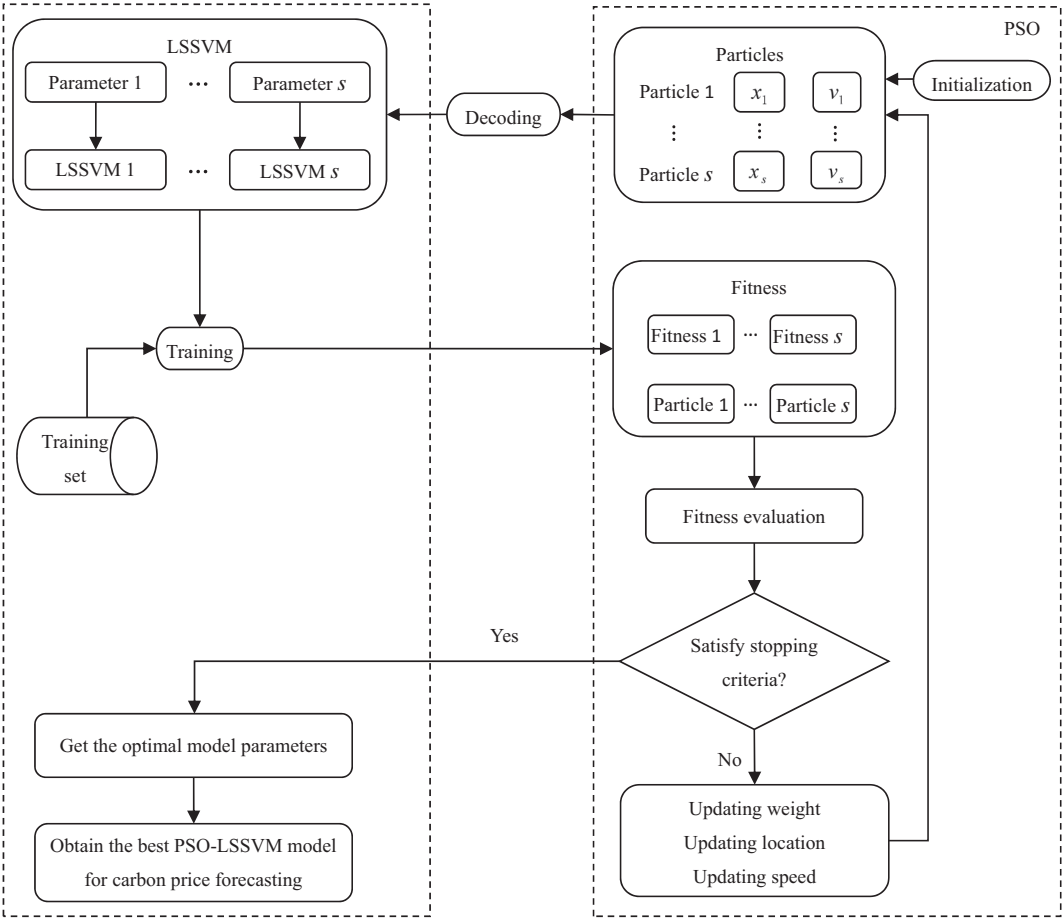

LSSVM
Parameter 1
Parameter s
LSSVM 1
LSSVM s
Decoding
PSO
Particles
Particle 1
Particle s
Initialization
Training
Training set
Fitness
Fitness 1
Fitness s
Particle 1
Particle s
Fitness evaluation
Satisfy stopping criteria?
Yes
No
Get the optimal model parameters
Obtain the best PSO-LSSVM model for carbon price forecasting
Updating weight
Updating location
Updating speed

Figure 1

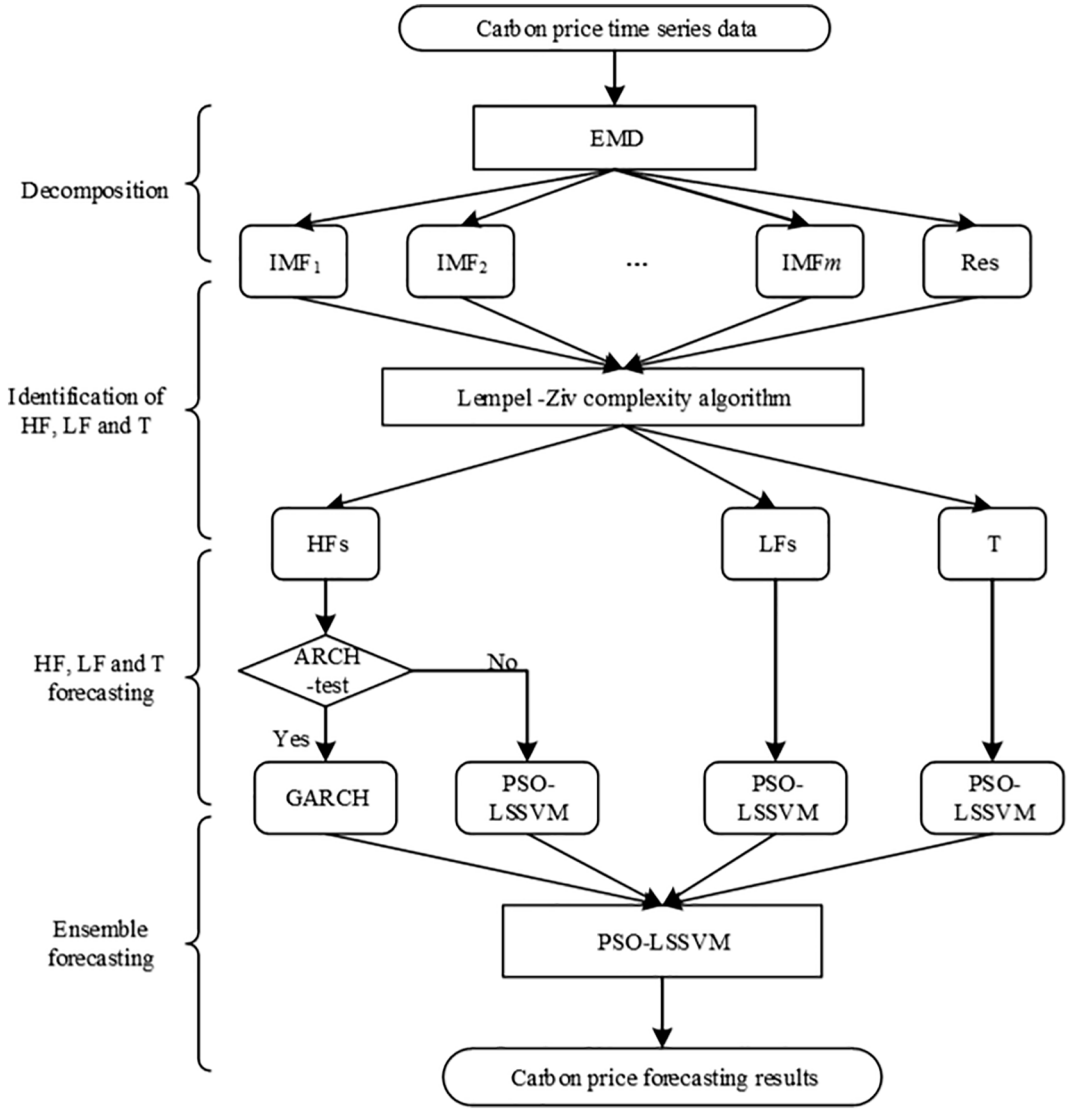
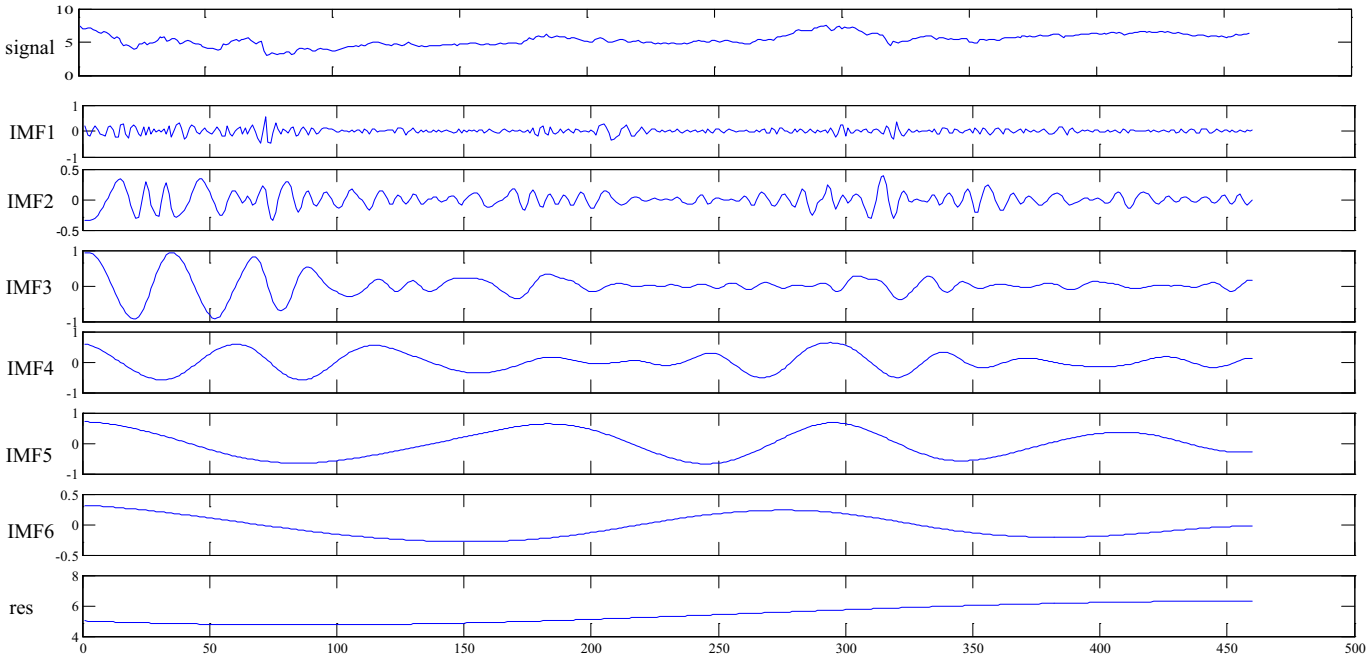
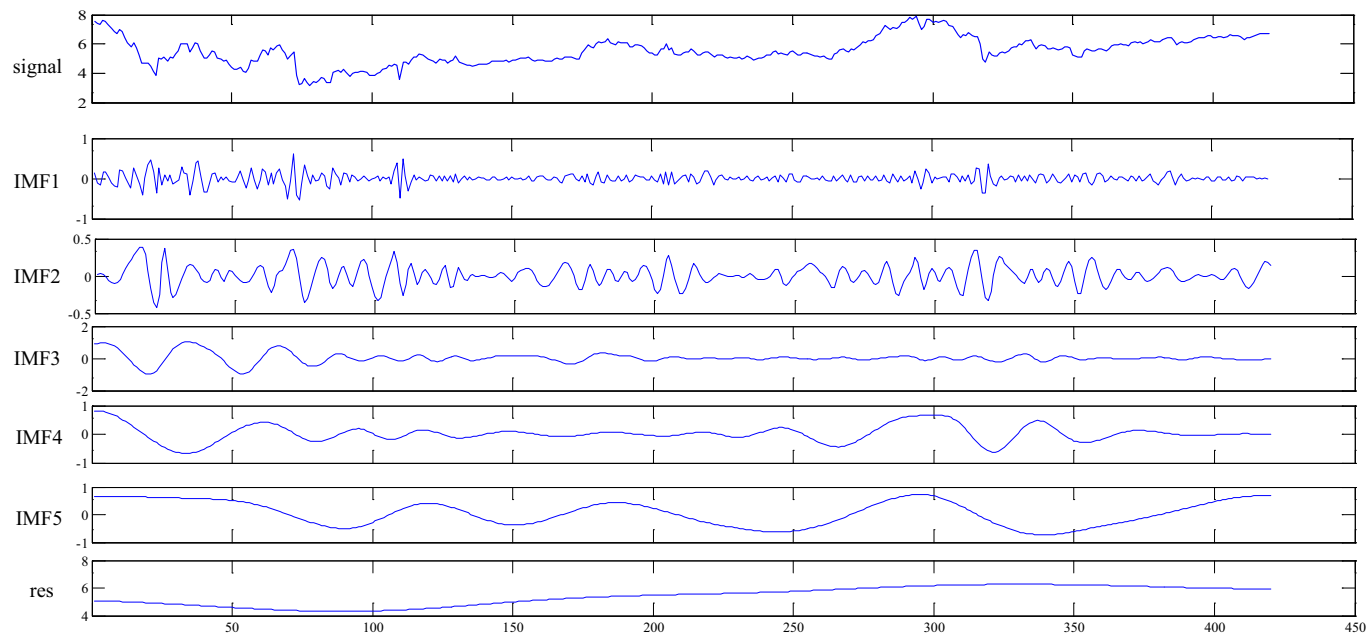

Carbon price time series data
EMD
Decomposition
IMF1
IMF2
...
IMFm
Res
Identification of HF, LF and T
Lempel-Ziv complexity algorithm
HFs
LFs
T
HF, LF and T forecasting
ARCH -test
No
Yes
GARCH
PSO-LSSVM
PSO-LSSVM
PSO-LSSVM
Ensemble forecasting
PSO-LSSVM
Carbon price forecasting results

Figure 2

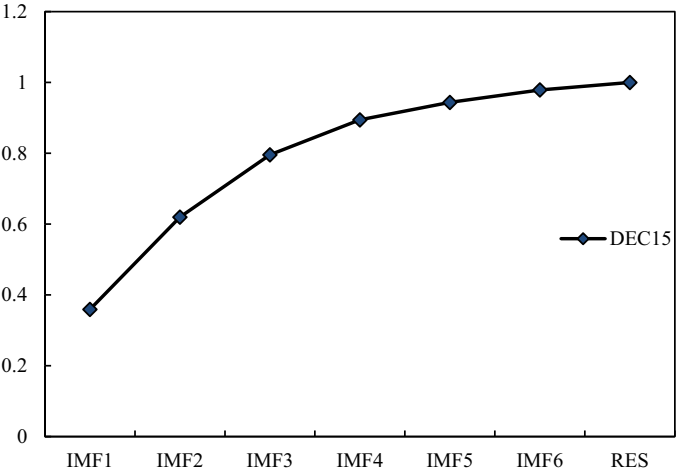


(a) The first decomposition result of DEC15 during January 02, 2013- October 16, 2014

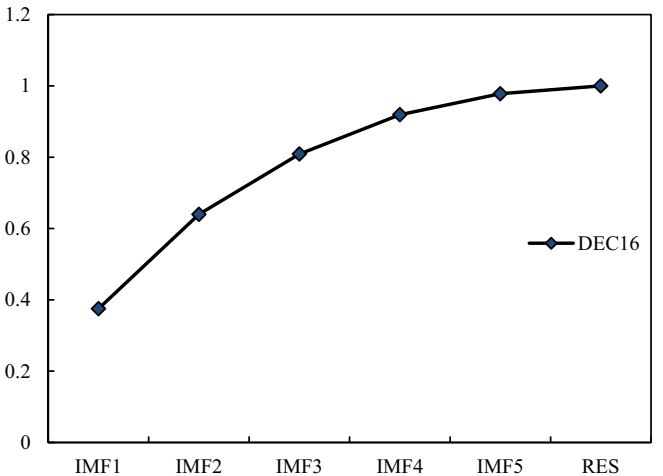


(b) The first decomposition result of DEC16 during January 02, 2013- August 21, 2014

Figure 3

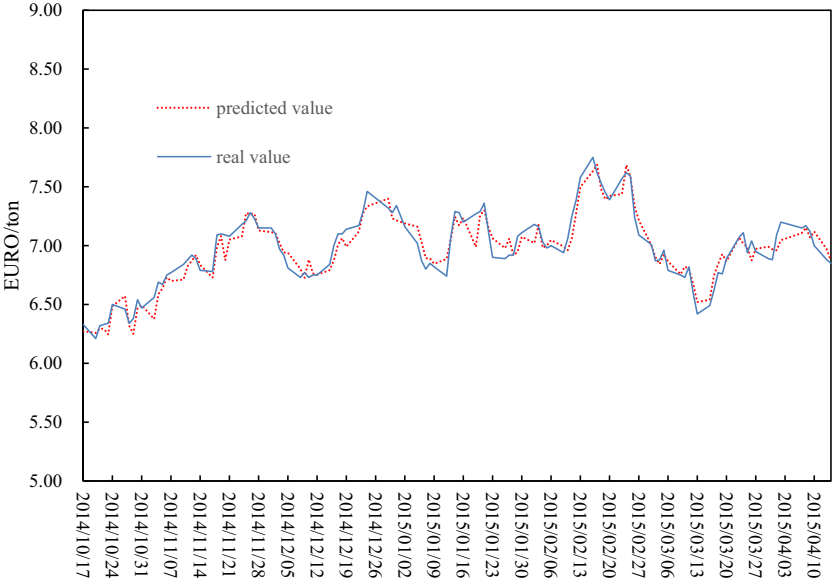


(a) The complexities of the first group of IMFs for DEC15 during January 02, 2013- October 16, 2014

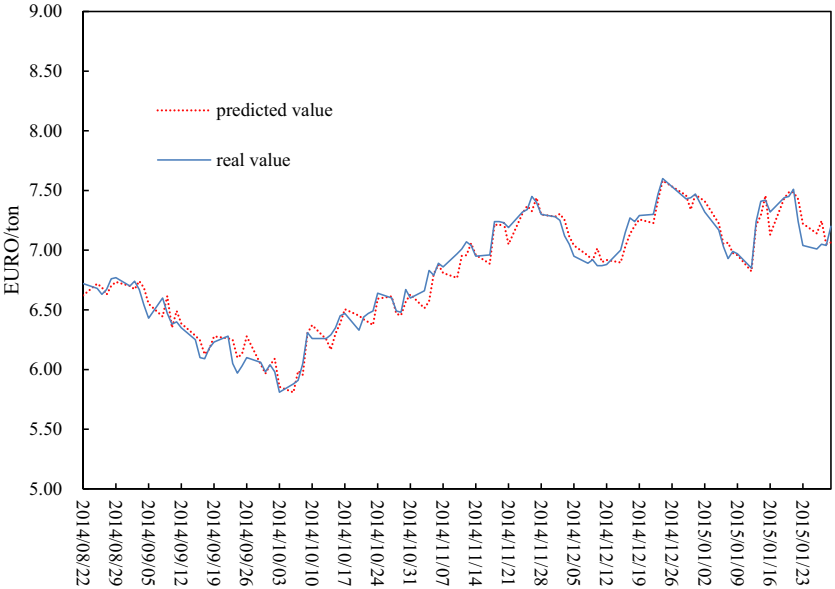


(b) The complexities of the first group of IMFs for DEC16 during January 02, 2013- August 21, 2014

Figure 4



(a) Out-of-sample prediction results of DEC15



(b) Out-of-sample prediction results of DEC16

Figure 5