Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression

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

[a] 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

Abstract: Conventional methods are less robust in terms of accurately forecasting non-stationary and nonlineary carbon prices. In this study, we propose an empirical mode decomposition-based evolutionary least squares support vector regression multiscale ensemble forecasting model for carbon price forecasting. Firstly, each carbon price is disassembled into several simple modes with high stability and high regularity via empirical mode decomposition. Secondly, particle swarm optimization-based evolutionary least squares support vector regression is used to forecast each mode. Thirdly, the forecasted values of all the modes are composed into the ones of the original carbon price. Finally, using four different-matured carbon futures prices under the European Union Emissions Trading Scheme as samples, the empirical results show that the proposed model is more robust than the other popular forecasting methods in terms of statistical measures and trading performances.

Keywords: carbon price forecasting; empirical mode decomposition; least squares support vector regression; particle swarm optimization

1. Introduction

Global climate change, as a grand challenge faced by the human society, is attracting more and more attention around the world in the recent few decades. To address this challenge, the Kyoto Protocol, signed in 1997, came into effect on February 16, 2005. The protocol established the quantitative greenhouse gas emission reduction targets for the developed and industrialized countries. To achieve these targets effectively, the European Union Emissions Trading System (EU ETS) was initiated in January 2005. The EU ETS has been the biggest carbon trading market so far. It also provides an important demonstration of carbon market construction for other countries or regions, as well as a new investment choice for investors [1]. In light of this, it is important to improve the accuracy of carbon price forecasting. On the one hand, accurately forecasting carbon prices can contribute to a deep understanding on the characteristics of carbon prices so as to establish an effective and stable carbon pricing mechanism. On the other hand, it can provide a practical guidance for production operations and investment decisions, helping to avoid carbon price risks and maximize carbon assets. Therefore, carbon price prediction has become one of the most popular topics in energy research.

As we know, prediction technology generally can be classified into two categories: (i) time series forecasting, and (ii) multi-factor forecasting. Although multi-factor forecasting can consider the influences of exogenous variables, it is used to forecast the carbon price in the premise of forecasting the exogenous variables, which will inevitably lead to the problem of error accumulation so as to make the failure of carbon price prediction. Time series prediction can predict the future trend of carbon price by establishing a mathematical model to extend the trend of its own historical changeable law without the influences of exogenous variables, which can obtain a good prediction accuracy. Many studies have proven that time series prediction is applicable for energy and carbon price forecasting. Thereby, multi-factor forecasting is excluded, and time series forecasting is utilized to predict carbon price in this study. Recently, carbon price forecasting has attracted more and more research attentions [2-11]. The time series forecasting approaches

* Corresponding author: wpzbz@126.com (Bangzhu Zhu); wuzhanchi@sina.com(Zhanchi Wu).

used so far can be roughly divided into two broad categories: statistical and econometric models, and artificial intelligence (AI) models. The former includes the multiple linear regression [2], GARCH [3], MS-AR-GARCH [4], FIAPGARCH [5], HAR-RV [6], and nonparametric models [7]. The latter includes artificial neural networks (ANNs) [8,9] and least squares support vector regression (LSSVR) [10,11]. Although the existing methods can obtain good results when they are applied for stationary time series forecasting, they are not robust for forecasting accurately carbon price due to its highly non-stationary and nonlinear characteristics [12].

Empirical mode decomposition (EMD), proposed by Huang and his co-authors in 1998, is an effective approach for handling the nonlinear and non-stationary time series [13,14,15]. EMD can disassemble any carbon price into several intrinsic mode functions (IMFs) plus a residue with high stability and high regularity. When the IMFs and residue are used as the inputs of ANN or LSSVR, it can improve learning efficiency and forecasting accuracy by providing better understanding and feature-capturing [11,16]. Thereby, the accuracy of carbon price forecasting can be enhanced through EMD. During the past few years, the EMD-based ANN and/or LSSVR models have been applied for time series forecasting [17-26], including carbon price forecasting [11,16]. However, the traditional back-propagation ANNs, used as the predictors, can lead to the overfitting problems. Although LSSVR, built on the structural risk minimization, can effectively solve the overfitting problem [27], the performance of a LSSVR predictor is sensitive to its own model selection. Yet the hybrid EMD and LSSVR models have rarely been employed for carbon price forecasting. Thus, this study seeks to address this gap in carbon price forecasting methodology.

The aim of this study is to develop an EMD-based evolutionary LSSVR model to forecast carbon prices with high accuracy. The contributions of the study are two-fold. On the one hand, an EMD-based evolutionary LSSVR model (EMD–LSSVR–ADD) is constructed to forecast carbon prices: (1) each carbon price is decomposed into several IMFs plus a residue with high stability and high regularity via EMD; (2) all the IMFs and residues are respectively predicted via LSSVR trained by particle swarm optimization (PSO); (3) the forecasted values of all the IMFs and residues are aggregated into the ones of the original carbon price. On the other hand, using the empirical data from four different-matured carbon futures under the EU ETS, the study compares the forecasted results of the proposed model with the single ARIMA and LSSVR models, the hybrid ARIMA+LSSVR model, and a variation of the forecasting model (EMD-ARIMA-ADD) to demonstrate its robustness. Guo et al. (2012)[28] argued that it may be more suitable to integrate all IMFs without IMF_1 when forecasting wind speed. Thus the study adds two models by removing the IMF_1 from EMD–ARIMA–ADD and EMD–LSSVR–ADD to test whether this approach is feasible in the prediction of carbon prices, denoted as EMD–ARIMA–IMF_1–ADD and EMD–LSSVM–IMF_1–ADD models respectively. The study adopts the well-established evaluation criteria, including level forecasting, directional prediction, the Diebold–Mariano (DM) test, the Rate test, and trading performances including the Annualized return, Annualized volatility and Information ratio, to assess the robustness of the proposed EMD–LSSVR–ADD model.

The paper is organized as follows. Section 2 describes the EMD, LSSVR, and the proposed models. Section 3 reports the empirical analysis, and Section 4 concludes the study.

2. Methodology

2.1 EMD

EMD can decompose a carbon price into several IMFs and one residue by its local feature scales, as follows:

Step 1: Find out the local extreme points of carbon price $x(t)$;

Step 2: Shape the upper and lower envelopes, $e_{\max}(t)$ and $e_{\min}(t)$, respectively;

Step 3: Obtain the mean of $e_{\max}(t)$ and $e_{\min}(t)$:

$$a(t) = [e_{\max}(t) + e_{\min}(t)] / 2$$

Step 4: Get the difference between $x(t)$ and $a(t)$:

$$d(t) = x(t) - a(t)$$

Step 5: Check $d(t)$. When $d(t)$ cannot meet the two conditions of IMF, let $x(t) = d(t)$, return to the step 1, and cannot repeat unless $d(t)$ meets the two conditions. Otherwise, $d(t)$ is defined as an IMF, and let the residue $r(t) = x(t) - d(t)$;

Step 6: Perform the steps 1-5 only when the termination criterion is met. EMD cannot stop unless $\sigma(t) < \theta_1$ for a prescribed fraction $(1-\alpha)$ and $\sigma(t) < \theta_2$ for the remaining fraction, where θ_1 and θ_2 are two thresholds aimed to ensure mean globally small changes while locally big excursions.

In this study, we use the termination criterion by Rilling *et al.* [29], in which $\alpha = 0.05$, $\theta_1 = 0.05$, and $\theta_2 = 0.5$.

Finally, we can obtain: $x(t) = \sum_{i=1}^{m} IMF_i(t) + R_m(t)$, where *m* is the number of IMFs, and $R_m(t)$ is the final residue.

2.2 LSSVR

For data $\{x_i, y_i\}, i = 1,2,\cdots,n$, LSSVR is defined as [27]:

$$\min\left\{\frac{1}{2}\|\omega\|^2 + \frac{1}{2}C\sum_{i=1}^{n}\xi_i^2\right\}$$

$$\text{s.t. } y_i = \omega\cdot\varphi(x_i) + b + e_i, i = 1,2,\cdots,n$$

in which ω: the weight vector, *C*: the penalty parameter, ξ_i: the error, φ: mapping function, and *b* :the bias.

The Lagrange function is used to find out the solutions for ω and ξ_i:

$$L(\omega, b, \xi, \alpha) = \frac{1}{2}\|\omega\|^2 + \frac{1}{2}C\sum_{i=1}^{n}\xi_i^2 - \sum_{i=1}^{n}\alpha_i\{\omega\cdot\varphi(x_i) + b + \xi_i - y_i\}$$

in which $\{\alpha_i, i = 1,2,\cdots,n\}$ are a set of Lagrange multipliers. The optimal solutions are obtained from:

$$\begin{cases} \dfrac{\partial L}{\partial \omega} = 0 \to \omega = \sum_{i=1}^{n}\alpha_i\varphi(x_i) \\ \dfrac{\partial L}{\partial b} = 0 \to \sum_{i=1}^{n}\alpha_i = 0 \\ \dfrac{\partial L}{\partial e_i} = 0 \to \alpha_i = Ce_i \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \to \omega\cdot\varphi(x_i) + b + e_i - y_i = 0 \end{cases}$$

Using the least squares method to resolve the linear equations, LSSVR can be obtained as: $y(x) = \sum_{i=1}^{n}\alpha_i K(x, x_i) + b$, in which the kernel function, $K(x, x_i) = \varphi(x)\cdot\varphi(x_i)$, fulfills the Mercer's principle.

2.3 Hybriding LSSVR and PSO for carbon price forecasting

The model selection of LSSVR is concerned with two key issues [29,30]: how to select an appropriate kernel function, and how to determine the optimal parameters of LSSVR. For the former, radial basis function (RBF), $K(x, y) = \exp\left(-\|x - y\|^2 / 2\sigma^2\right)$, is selected to build the LSSVR model, because RBF can yield good results in general [31]. For the latter, we use the PSO algorithm [32] to seek the optimal parameters (C and σ) of LSSVR.

In the modeling of PSO, $x_i = (x_{i1}, x_{i2}, \cdots, x_{iM})$ and $v_i = (v_{i1}, v_{i2}, \cdots, v_{iM})$ are respectively defined as the position and velocity of particle i, $i = 1,2,\cdots,m$. $p_{best} = (p_{i1}, p_{i2}, \cdots, p_{iM})$ and $g_{best} = (p_{g1}, p_{g2}, \cdots, p_{gM})$ are respectively

defined as the optimum positions of particle *i* and *m* particles at the current iteration. x_i and v_i of each particle are updated as:

$$x_{id}(t+1)=\begin{cases} -p_{\max}, & x_{id} < -p_{\max} \\ x_{id}(t)+v_{id}(t+1), & -p_{\max} \le x_{id} \le p_{\max} \\ p_{\max}, & x_{id} > p_{\max} \end{cases} \tag{1}$$

$$v_{id}(t+1)=\begin{cases} -v_{\max}, & v_{id} < -v_{\max} \\ w(t)\cdot v_{id}(t)+c_1\cdot r_1\cdot[p_{id}(t)-x_{id}(t)]+c_2\cdot r_2\cdot[p_{gd}(t)-x_{id}(t)], & -v_{\max} \le v_{id} \le v_{\max} \\ v_{\max}, & v_{id} > v_{\max} \end{cases} \tag{2}$$

where $1\le i\le m$, $1\le d\le M$, $x_{id}(t)$ and $v_{id}(t)$ are respectively the position and velocity of particle *i* at iteration *t*, p_{id} is the optimal position of particle *i* at iteration *t*, p_{gd} is the global optimal position, and w is the inertia weight, defined as:

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

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

The study introduces the PSO algorithm to seek the optimal parameters (C and σ) of LSSVR, in order to improve searching efficiency and prediction accuracy [10], as presented in Fig. 1.

Insert Figure 1

Step 1: set up the training and test sets. The carbon price data is divided into a training set and a test set. The former is used for establishing the model, and the latter is used to test the forecasting performance of the proposed model.

Step 2: initialization. Randomly generating m particles with coding C and σ by real values, and setting the parameters of PSO: maximal iterations $t_{\max}$, maximal position $p_{\max}$, maximal velocity $v_{\max}$, $w\in[w_{\min}, w_{\max}]$, acceleration coefficients c_1 and c_2, $C\in[C_{\min}, C_{\max}]$, $\sigma\in[\sigma_{\min}, \sigma_{\max}]$. Let t = 0, and training begins.

Step 3: selecting the root mean square error (RMSE) as the fitness function:

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

in which n is the number of training sample, x_i and $\hat{x}_i$ are the real and predicted values.

Step 4: evaluating the fitness. Calculating the fitness value of each particle by Eq. (4), and obtain p_{best} and g_{best} at the current generation.

Step 5: updating the position and velocity of each particle by Eqs. (1)–(3).

Step 6: Checking the end condition. When the end condition, the maximum iterations here, is satisfied, the optimization process ends, and the optimal parameters are obtained to build the LSSVR. If not, move to step 7.

Step 7: Let $t=t+1$, and return to the step 4.

2.4 The proposed EMD-based LSSVR model for carbon price forecasting

For the carbon price x_t $(t=1,2,\cdots,T)$, a *h*-step forecasting in advance, $\hat{x}_{t+h}$, can be expressed as

$$\hat{x}_{t+h}=f(x_t, x_{t-1},\cdots,x_{t-m+1})$$

where x_t, $\hat{x}_t$ and m are the real, predicted values, and lag order respectively.

As shown in Fig. 2, we propose an EMD-based LSSVR model (EMD-LSSVR-ADD) for carbon price forecasting, generally comprised of the subsequent three key steps:

Step 1: Each carbon price is decomposed into a batch of IMFs and one residue with high stability and regularity via EMD.

Step2: LSSVR is employed in forecasting the IMFs and residue respectively, so as to obtain their forecasted values.

Step 3: The forecasted values of all the IMFs and residues are aggregated into the final predicted values of the original carbon price.

In short, the proposed EMD-LSSVR-ADD is in essence an EMD (multiscale decomposition)–LSSVR (component forecast)–ADD (multiscale ensemble forecast) model, which is a utilization of "decomposition and ensemble" tactics [11,34]. In the next section, four carbon futures prices are used for testing the robustness of the proposed multiscale prediction approach.

Insert Figure 2

3. Empirical analysis

3.1 Data

As the biggest carbon trading market in the EU ETS, the European Climate Exchange (ECX) is an indicator of the global carbon markets. Four futures prices matured in Decembers of 2013, 2014, 2015 and 2016 (denoted as DEC13, DEC14, DEC15 and DEC16 respectively) are selected as empirical samples. The daily data has been collected in Euros/ton (excluding public holidays from April 2008 to October 2016). For the convenience of modeling, the samples are divided into two subsets: the training set and the testing set. The training set is used to establish prediction models, and the testing set is employed to test the robustness of the established models. The divided samples of carbon prices are reported in Table 1. The data used are obtained from the website of ECX (http://www.theice.com).

Insert Table 1

3.2 Evaluation criteria

Forecasting performance is measured by two main criteria: level forecasting and directional prediction. Level forecasting is measured via the root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}[\hat{x}(t)-x(t)]^2}$$

On the other hand, directional prediction is measured with the directional prediction statistic (D_{stat}) [10, 36]:

$$D_{stat} = \frac{1}{n}\sum_{t=1}^{n} a_t \times 100\%$$

$$a_t = \begin{cases} 1, & if\,[x(t+1)-x(t)][(\hat{x}(t+1)-x(t)] \geq 0 \\ 0, & otherwise \end{cases}$$

where $\hat{x}(t)$ and $x(t)$ are the real and predicted values respectively, and n is the number of test samples.

The DM test is further used to statistically contrast the predicted performances of various predictive models [37].

In this study, mean square prediction error (MSPE) is chosen as the loss function. Thus, the DM statistic is defined as

$$DM = \frac{\bar{d}}{\sqrt{\hat{V}_{\bar{d}} / T}} \sim N(0,1), \quad T \to \infty$$

where $\bar{d} = \frac{1}{T}\sum_{t=1}^{T}\left[g(e_{te,t}) - g(e_{re,t}) \right]$, $g(e_{te,t}) = \sum_{t=1}^{T} e_{te,t}^2$, $g(e_{re,t}) = \sum_{t=1}^{T} e_{re,t}^2$, $e_{te,t} = x_t - \hat{x}_{te,t}$, $e_{re,t} = x_t - \hat{x}_{re,t}$, $\hat{V}_{\bar{d}} = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j$, and $\gamma_j = \mathrm{cov}(d_t, d_{t-j})$. $\hat{x}_{te,t}$ and $\hat{x}_{re,t}$ denote the forecasted values of x_t calculated using the test model (*te*) and reference model (*re*) at time *t*, respectively. A one-tailed test is generally employed in evaluating the DM statistic. In the DM test, the null hypothesis, i.e. the tested model is not worse than the reference model, is tested. Therefore, only if *p* is lower than a frequently-used level of significance 0.05, we should reject it; otherwise, we should accept it.

The statistics of RT test 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. Using the two-sided test, when the absolute value of z_{RT} exceeds1.96, the null hypothesis is rejected at the significance level of 5%.

A good statistical accuracy does not always mean a good trading performance. For investors, they usually care more about a model's practicability in trading. In this section, inspired by Sermpinis et al.(2016) [37], we design a pseudo trading strategy to test the trading performance as an investor chooses to buy or sell (or stay watching) carbon assets when the forecasted return is above or below (or equal) zero at the current carbon price respectively in real market. This can illustrate our model's application value of making production and investment more profit. We use the Annualized return, Annualized volatility and Information ratio to evaluate the trading performance. They defined as:

$$\text{Daily return: } R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

$$\text{Annualized return: } R^A = 252 * \frac{1}{k} * \sum_{t=1}^{n} R_t$$

$$\text{Annualized volatility: } \sigma^A = \sqrt{252} * \sqrt{\frac{1}{k-1} * \sum_{t}^{k}(R^t - R^m)}$$

$$\text{Information ratio: } \mathrm{IR} = \frac{R^A}{\sigma^A}$$

where, P_t is daily price of carbon future; k is the number of test set and R^m is mean value of R^t.

In order to evaluate the predictive performance of proposed EMD–LSSVR–ADD model with other popular forecasting models, the study compares its outputs with the outputs of the single ARIMA and LSSVR models, a hybrid ARIMA+LSSVR model, variants of the EMD–ARIMA–ADD model, EMD–ARIMA–IMF_1–ADD and EMD–LSSVM–IMF_1–ADD models. In the variant of the EMD–ARIMA–ADD model, all the IMFs and residues extracted by EMD are independently forecasted by the ARIMA model, and the predicted values are summed into the final predicted ones of the original carbon price.

3.3 Results and discussions

Forecasting experiments are carried out in terms of the steps outlined in the previous section. We set the thresholds and tolerance as $[\theta_1, \theta_2, \alpha] = [0.05, 0.5, 0.05]$ [29], and the decomposition results via EMD are reported in Fig.3. It is evident that DEC13, DEC14, DEC15 and DEC16 tend to be non-stationary and nonlinear due to the fact that their means change over time. DEC13, DEC14 and DEC15 are respectively disassembled into seven IMFs and one residue, while DEC16 is disassembled into six IMFs and one residue. At the same time, all the IMFs and residues have a higher stability and stronger regularity compared with the original series.

Insert Figure 3

We apply the one shot testing method [38], in which one single model applies over all test period. Thus, we apply the fixed time window to train the models by the training set, and forecast the testing set by the trained models. Meantime, we perform the one-step-ahead forecasting for DEC13 to DEC16. - EViews developed by Quantitative Micro Software is used for ARIMA modeling. The optimum model is found via the Akaike information criterion. By trial and error, both the best models derived from DEC14 to DEC16 are ARIMA (2,1,0) models, while the best models for DEC13 is ARIMA (1,1,1) model. Moreover, as previously mentioned, ARIMA is also used to model each IMF and residue decomposed by EMD. The predicted values of all IMFs and residues are then aggregated into the predicted values of EMD–ARIMA–ADD model.

All the LSSVR models are built by the LSSVMlab by Suykens and his colleagues on the platform of MATLAB 2016b. The input of each LSSVR model is determined using a partial autocorrelation function method [16]. The optimal parameters are searched with 100 particles and 5 generations. $C \in [1,1000], \sigma \in (0,50]$, $c_1 = c_2 = 2$, $w_{\max} = 0.9$, $w_{\min} = 0.1$, $p_{\max} = 0.05$, and $v_{\max} = 50$. Moreover, as mentioned above, LSSVR is also used to forecast each IMF and residue decomposed via EMD, and the predicted values of all IMFs and residues are aggregated into the forecasted values of EMD–LSSVR–ADD model.

The hybrid ARIMA+LSSVR model is built as discussed above, the predicted values of the original carbon price by ARIMA and LSSVR models are equally weighted sum of the final predicted ones of carbon prices. Consequently, Inspired by Tang et al. (2012)[18], Guo et al. (2012)[28], Zhu and Wei,(2013)[10] and Yu et al. (2015)[39], two single models (LSSVR and ARIMA), a hybrid ARIMA+LSSVR model, and four multiscale forecasting models (EMD–ARIMA–$\underline{\text{IMF}_1}$–ADD, EMD–ARIMA–ADD, EMD–LSSVR–$\underline{\text{IMF}_1}$–ADD and EMD–LSSVR–ADD) are applied to forecast carbon prices. The results of RMSE and D_{stat} for the different models are shown in Table 2. The DM and RT test results are listed in Tables 3 and 4. Furthermore, the comparison of the trading performances is concluded in Table 5. The out-of-sample forecasted results for DEC13, DEC14, DEC15 and DEC16 by the proposed EMD-LSSVR-ADD are presented in Fig.4.

Insert Table 2

Insert Table 3

Insert Table4

Insert Table 5

From the perspective of level forecasting measured by RMSE, it can be found that, firstly, the prediction accuracy

of LSSVR model is superior to that of ARIMA model for its strong nonlinear approximation ability and excellent self-learning ability. Meanwhile, the optimization of PSO improves the learning and prediction abilities of LSSVM. The hybrid process of ARIMA and LSSVR models only improves the predication accuracy slightly here. Secondly, all the multiscale ensemble prediction models including EMD–ARIMA–$\underline{IMF_1}$–ADD, EMD–ARIMA–ADD, EMD–LSSVR–$\underline{IMF_1}$–ADD and EMD–LSSVR–ADD obviously outperform each single prediction models such as ARIMA, LSSVR and their hybrid model. The main reason is that after EMD decomposition, both LSSVR and ARIMA can effectively forecast the simple and stable components so as to significantly improve the prediction accuracy. Thirdly, among the multiscale ensemble prediction models, the EMD–LSSVR–ADD and EMD–ARIMA–ADD show better results than EMD–LSSVR–$\underline{IMF_1}$–ADD and EMD–ARIMA–$\underline{IMF_1}$–ADD, which differs from the conclusion of Guo et al.(2012) and shows the necessity to take all the IMFs into consideration when forecasting carbon prices. Last but not least, the result of EMD–LSSVR–ADD is superior to that of EMD–ARIMA–ADD, which shows the strong predictive power of the proposed EMD-LSSVR-ADD model. Comparing all models here, the highest level of accuracy achieved by the proposed EMD–LSSVR–ADD model implies the advantage of "decomposition and ensemble" principle.

In terms of the level of directional prediction, the results of D_{stat} are similar in terms of RMSE. The models established via EMD decomposition and LSSVR show higher accuracy in directional prediction. Therefore, the proposed EMD–LSSVR–ADD model achieves the highest value of D_{stat} (or at least as high as) compared with other models in the contracts from DEC13, DEC14, DEC15 and DEC16. Concerning the improvement level of directional prediction, EMD makes the largest contribution. Due to the significant advantages of EMD, LSSVR just produces a less progress than ARIMA here.

Two findings are derived from the DM test results. Firstly, all the multiscale ensemble prediction models remarkably outperform than single scale models at the significance level of 5%. There is no obvious difference between ARIMA, LSSVM and hybrid models in level prediction. This confirms the power of EMD for capturing different characteristics of carbon prices. Secondly, in general, the proposed EMD–LSSVR–ADD model is significantly superior to EMD–ARIMA–$\underline{IMF_1}$–ADD and EMD–LSSVR–$\underline{IMF_1}$–ADD models, but it has no obvious advantage compared with EMD–ARIMA–ADD model except for DEC15. Furthermore, the RT test reveals the similar results as the DM test, i.e. the multiscale ensemble models can obtain high accuracy of directional prediction at the confidence level of 5% for all carbon prices. Although the proposed EMD–LSSVR–ADD model has the highest accuracy from D_{stat}, it does not show significant differences from other multiscale ensemble models in accuracy for direction.

In terms of the trading performance, the proposed EMD–LSSVR–ADD model produces the best trading performances in all the carbon prices for its highest Annualized return and smallest Annualized volatility. This implies that our proposed EMD–LSSVR–ADD model is capable of achieving good trading gains, and also suggests the advantages of EMD over single scale models.

To sum up, we can draw a few conclusions from the empirical analysis results: (1) in terms of level prediction, directional forecasting, the DM test, and the RT test, compared with ARIMA, LSSVR, ARIMA+LSSVR, EMD–ARIMA–$\underline{IMF_1}$–ADD, EMD–ARIMA–ADD and EMD–LSSVR–$\underline{IMF_1}$–ADD models, our proposed EMD–LSSVR–ADD model can obtain better statistical and trading performances; (2), four multiscale ensemble models can achieve more precise prediction results than ARIMA, LSSVR, and ARIMA+LSSVR models, implying that "decomposition and ensemble" tactics can significantly enhance predictive capability; (3) the nonlinear approach (LSSVR) is more appropriate to forecast carbon prices than the linear model (ARIMA). Therefore, the proposed EMD-based LSSVR model is a promising approach for carbon price forecasting.

Insert Figure 4

4. Conclusions and future work

In this study, we propose a new empirical mode decomposition-based evolutionary least squares support vector regression for carbon price forecasting. This model use empirical mode decomposition to disassemble each carbon price into a batch of more stationary and more regular components, which can be easily forecasted by the particle swarm optimization-based evolutionary least squares support vector regression. The final forecasted values of carbon prices are obtained via aggregating the forecasted values of all the components. Finally, using four carbon futures prices from the European Union Emissions Trading Scheme as samples, the proposed empirical mode decomposition-based evolutionary least squares support vector regression has been empirically tested, and its predictive performance has been compared with the predictive performance of single, hybrid and variational multiscale forecasting models, in terms of statistical measures and trading performances. The empirical results suggest that the proposed model can yield the optimal statistical measures. Moreover, the proposed model shows great trading performance as well, which indicates its value in practice. Our main future work is (1) to improve the accuracy of empirical mode decomposition, (2) to build the best forecasting for each component in terms of their own characteristics, and (3) to explore the nonlinear ensemble of all the components. Through these efforts, the accuracy of high non-stationary and nonlinear carbon price forecasting is expected to be further improved. Furthermore, Based on the proposed model, how to develop an intelligent forecasting and trading decision support system for carbon market so as to make production and investment a maximum profit is also our another next work.

Acknowledgments

We thank the chief editor’s and two reviewers' comments, which help us improve this paper. Our heartfelt thanks should also be given to the National Natural Science Foundation of China(NSFC) (71473180, 71201010, 71303174,71303076, and 71673083), National Philosophy and Social Science Foundation of China (14AZD068, 15ZDA054,16ZZD049), Guangdong Young Zhujiang Scholar (Yue Jiaoshi [2016]95), Natural Science Foundation for Distinguished Young Talents of Guangdong (2014A030306031), Soft Science Foundation of Guangdong (2014A070703062), Social Science Foundation of Guangdong (GD14XYJ21), Distinguished Young Teachers of Guangdong ([2014]145), High-level Personnel Project of Guangdong([2013]246), 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

[1] 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.

[2]Guobrandsdottir HN, Haraldsson HO. Predicting the price of EU ETS carbon credits. System Engineering Procedia 2011; 1:481-489.

[3] Paolella MS, Taschini L. An econometric analysis of emission allowance prices. Journal of Banking & Finance 2008; 32:2022-2032.

[4]Benz E, Truck S. Modeling the price dynamics of CO_2 emission allowances. Energy Economics 2009; 31(1):4-15.

[5] Conrad C, Rittler D, Rotfub W. Modeling and explaining the dynamics of European Union Allowance prices at the high-frequency. Energy Economics 2012; 34(1): 316-326.

[6]Chevallier J, Sevi B. On the realized volatility of the ECX emissions 2008 futures contract: distribution, dynamics and forecasting. Ann Finance 2011; 7:1-29.
[7]Chevallier J. Nonparametric modeling of carbon prices. Energy Economics 2011; 33(6):1267-1282.
[8]Fan XH, Li SS, Tian LX. Chaotic characteristic identification for carbon price and a multi-layer perceptron network prediction model. Expert Systems with Applications, 2015; 42: 3945-3952.
[9]Atsalakis G S. Using computational intelligence to forecast carbon prices. Applied Soft Computing, 2016; 43:107–116.
[10]Zhu BZ, Wei YM. Carbon price prediction with a hybrid ARIMA and least squares support vector machines methodology. Omega 2013; 41:517-524.
[11]Zhu BZ, Shi XT, Chevallier J, Wang P, Wei YM. An adaptive multiscale ensemble learning paradigm for non-stationary and nonlinear energy price time series forecasting. Journal of Forecasting (2016); DOI: 10.1002/for.2395.
[12] Feng ZH, Zou LL, Wei YM. Carbon price volatility: Evidence from EU ETS. Applied Energy 2011; 88: 590-598.
[13] Huang NE, Shen Z, Long SR. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Process of the Royal Society of London 1998; A454: 903-995.
[14]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.
[15]Yu, L., Li, J., Tang, L., Wang, S., Linear and nonlinear Granger causality investigation between carbon market and crude oil market: A multi-scale approach. Energy Economics 2015; 51, 300–311
[16]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.
[17]Yu LA, Wang SY, Lai KK. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. Energy Economics 2008; 30(5): 2623-2635.
[18]Tang L, Yu LA, Wang S, Li JP, Wang SY. A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting. Applied Energy 2012; 93:432-443.
[19]Chen CF, Lai MC, Yeh CC. Forecasting tourism demand based on empirical mode decomposition and neural network. Knowledge-Based Systems 2012; 26: 281-287.
[20]Lin CS, Chiu SH, Lin TY. Empirical mode decomposition–based least squares support vector regression for foreign exchange rate forecasting. Economic Modelling 2012; 29(6): 2583-2590.
[21]Wei Y, Chen MC. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. Transportation Research Part C: Emerging Technologies 2012; 21(1):148-162.
[22]An N, Zhao WG, Wang JZ, Shang D, Zhao ED. Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting. Energy 2013; 49(1): 279-288.
[23]Yu LA, Wang ZS, Tang L. A decomposition–ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. Applied Energy 2015; 156: 251–267.
[24]Liu H, Tian HQ, Liang XF, Li YF. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. Applied Energy 2015; 157:183–194.
[25]Li W, Lu C. 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.
[26]Ghasemi A, Shayeghi H, Moradzadeh M, Nooshyar M. A novel hybrid algorithm for electricity price and load forecasting in smart grids with demand-side management. Applied Energy 2016; 177: 40–59.
[27] Suykenns JAK, Vandewalle J. Least squares support vector machine. Neural Processing Letter 1999; 9

(3):293-300.
[28]Guo ZH, Zhao WG, Lu HY, Wang JZ. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renewable Energy 2012; 37(1): 241-249.
[29]Riling G, Flandrin P, Goncalves P. On Empirical Mode Decomposition and its algorithms. In Proceedings of the IEEE EURASIP Workshop on Nonlinear Signal and Image Processing, Grado, Italy, June 2003.
[30]Pai PF, Lin CS. A hybrid ARIMA and support vector machines model in stock price forecasting. Omega 2005;33: 497-505.
[31]Cagdas HA, Erol E, Cem K. Forecasting nonlinear time series with a hybrid methodology. Applied Mathematics Letters 2009; 22:1467-1470.
[32]Chen KY, Wang CH. A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. Expert Systems with Applications 2007; 32:254-264.
[33]Liu XY, Shao C, Ma HF, Liu RX. Optimal earth pressure balance control for shield tunneling based on LS-SVM and PSO. Automation in Construction 2011, 20:321–327.
[34]Liu H, Tian HQ, Pan DF, Li YF. Forecasting models for wind speed using wavelet, wavelet packet, time series and Artificial Neural Networks. Applied Energy 2013;107:191-208.
[35]Yu L, Wang SY, Lai KK. A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates. Computers & Operations Research 2005; 32:2523-2541.
[36]Diebold FX, Mariano RS. Comparing predictive accuracy. Journal of Business & Economic Statistic 1995;13(3):253–63.
[37]Sermpinis G, Stasinakis C, Rosillo R, Fuente D. European Exchange Trading Funds Trading with Locally Weighted Support Vector Regression. European Journal of Operational Research 2016; 258: 372–384.
[38]Luis T. Data mining with R: Learning with case studies. CRC Press, 2011.
[39]Yu L, Wang Z, Tang L. A decomposition–ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. Applied Energy 2015; 156:251-267.

Appendix A

Abbreviations: EMD, empirical mode decomposition; LSSVR, least squares support vector regression; IMF, intrinsic mode function; PSO, particle swarm optimization; EU ETS, European Union Emissions Trading System; GARCH, generalized autoregressive conditional heteroscedasticity; ANN, artificial neural networks; ARIMA, autoregressive integrated moving average; RBF, radial basis function; ECX, European Climate Exchange; RMSE, root mean squared error; D_{stat} directional prediction statistic; DM, Diebold–Mariano; RT, Rate test.

Figures

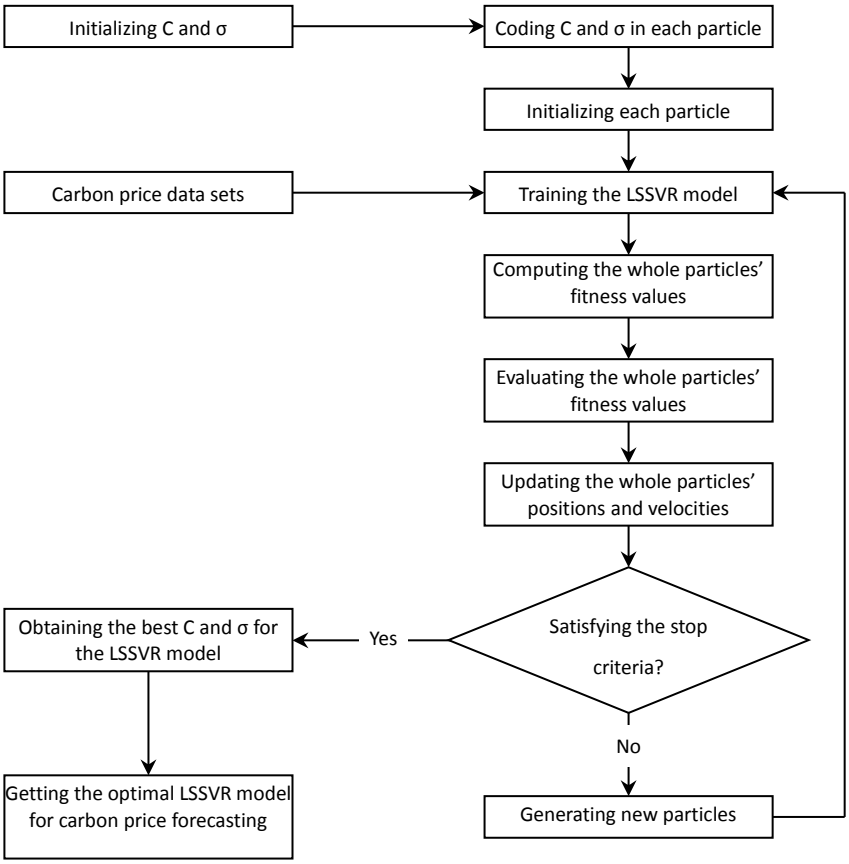


Fig. 1. The process of model selection for LSSVR using PSO.

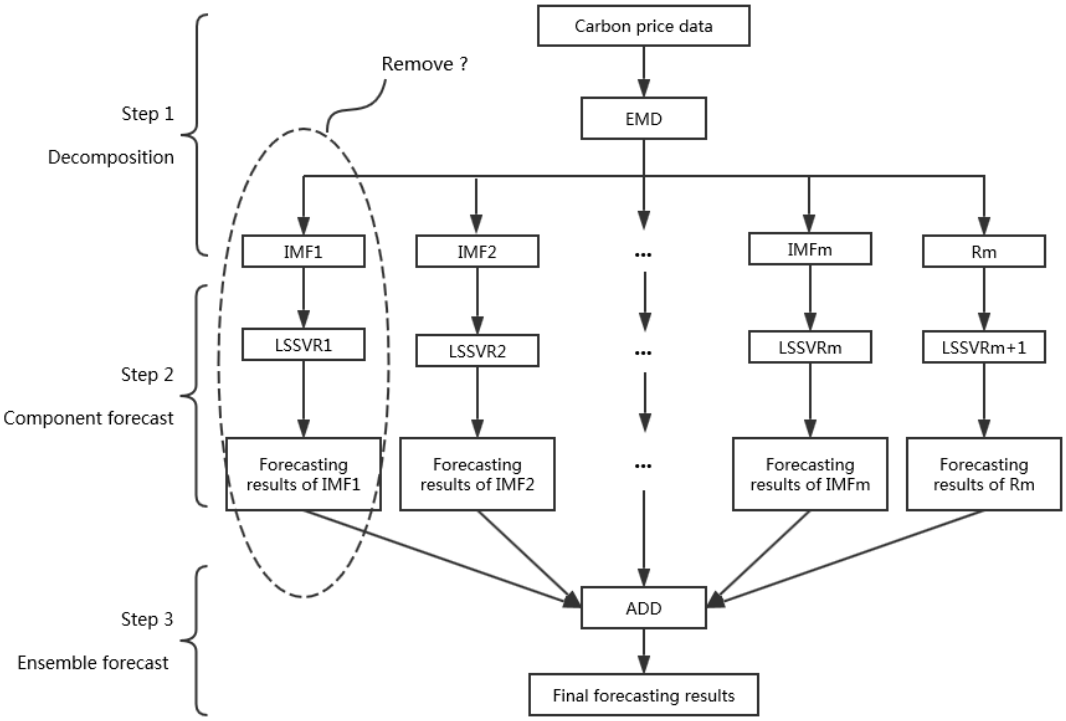
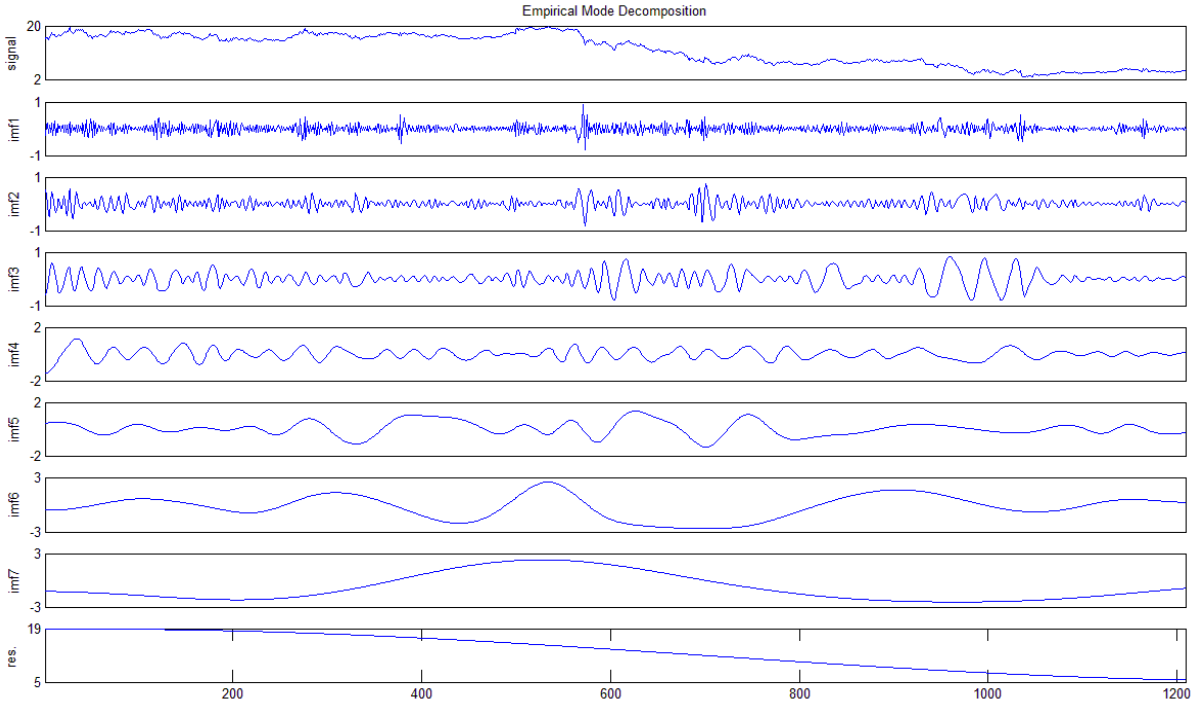
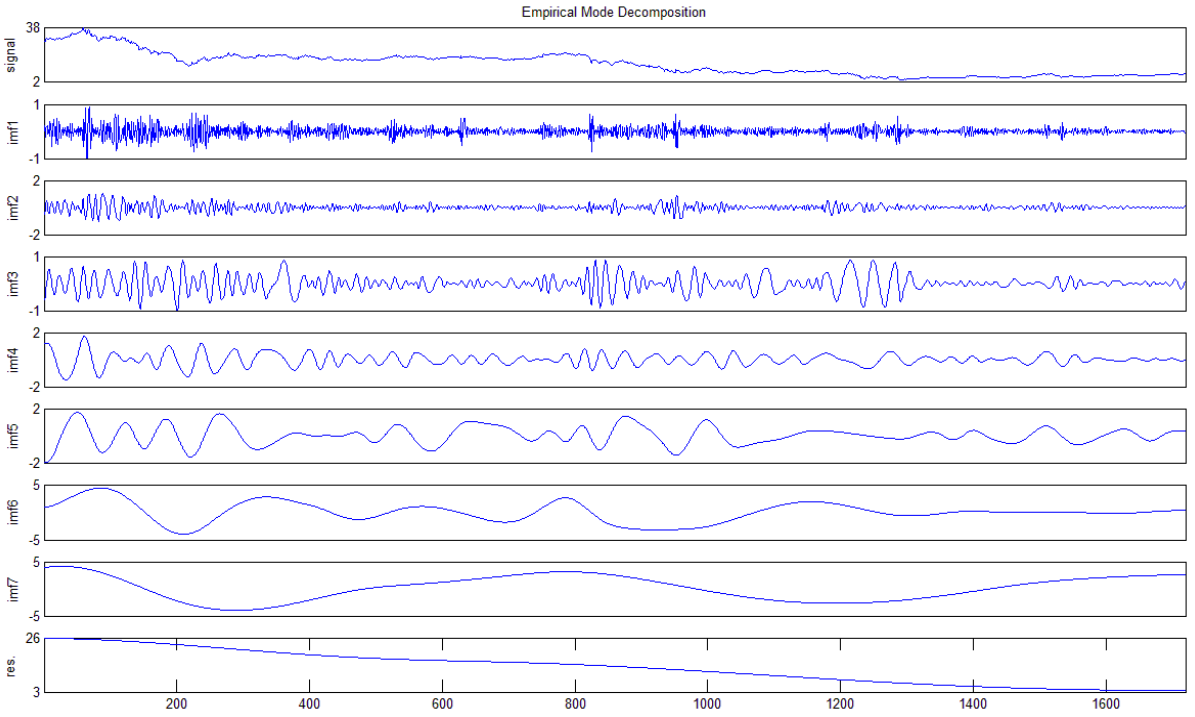


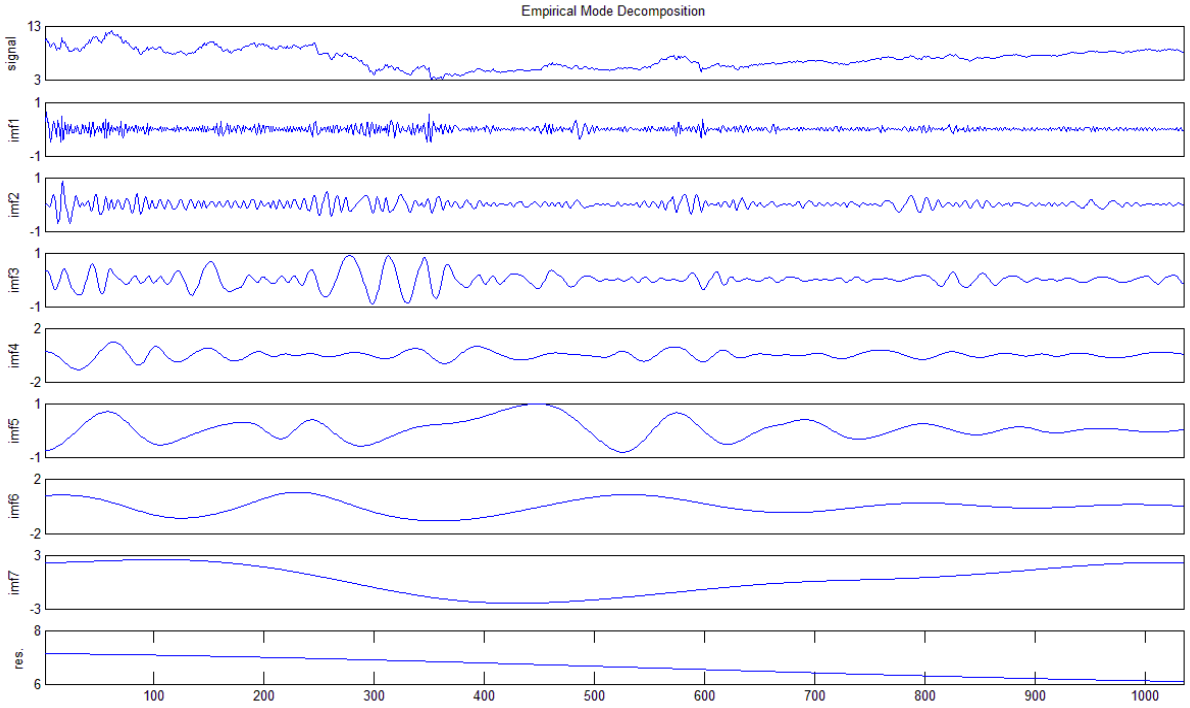
Fig. 2. The framework for the proposed multiscale prediction methodology.



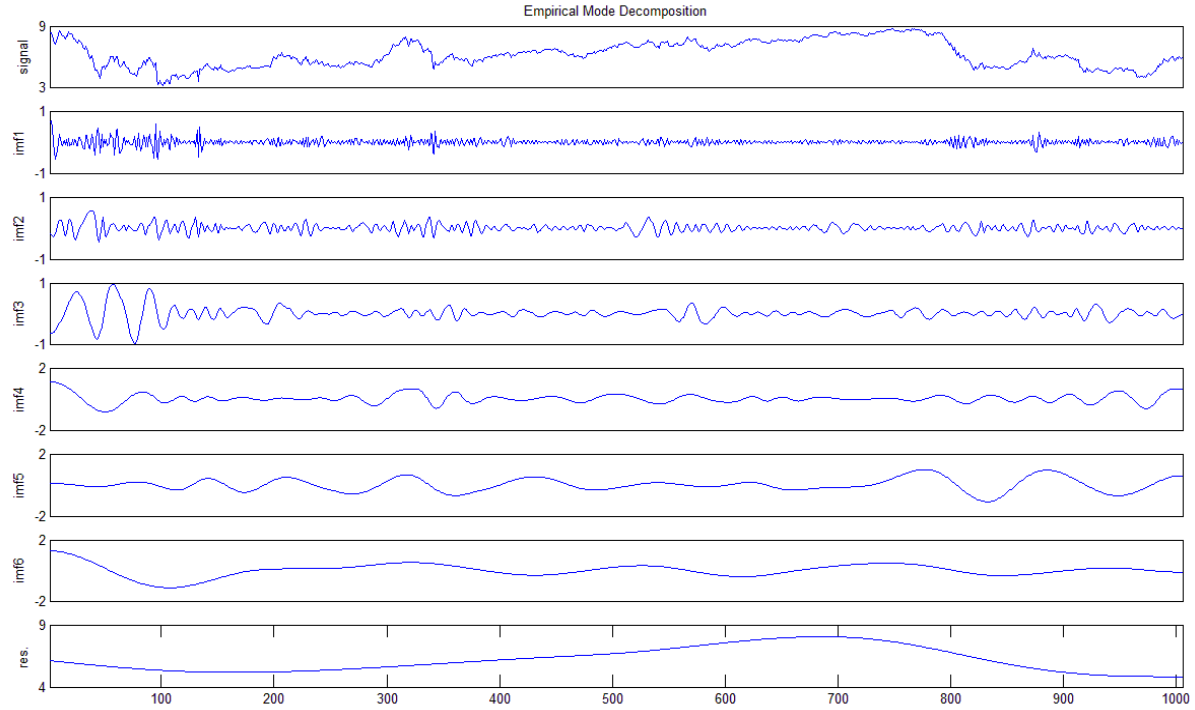
a. The decomposed results for DEC13.



b. The decomposed results for DEC14.

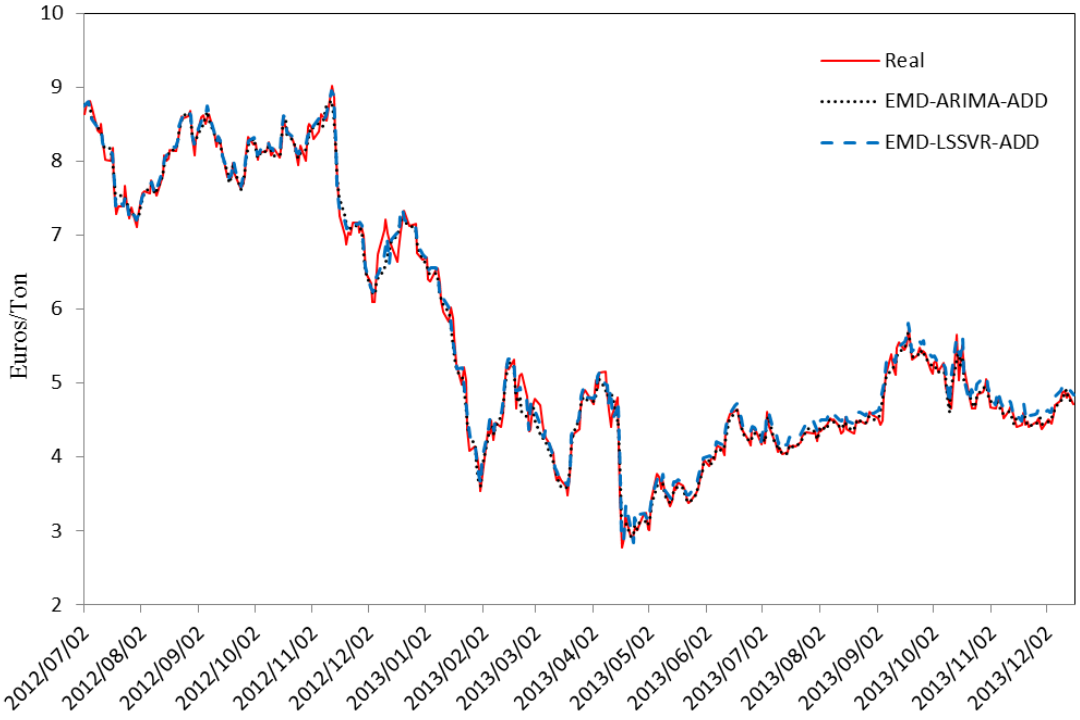


c. The decomposed results for DEC15.

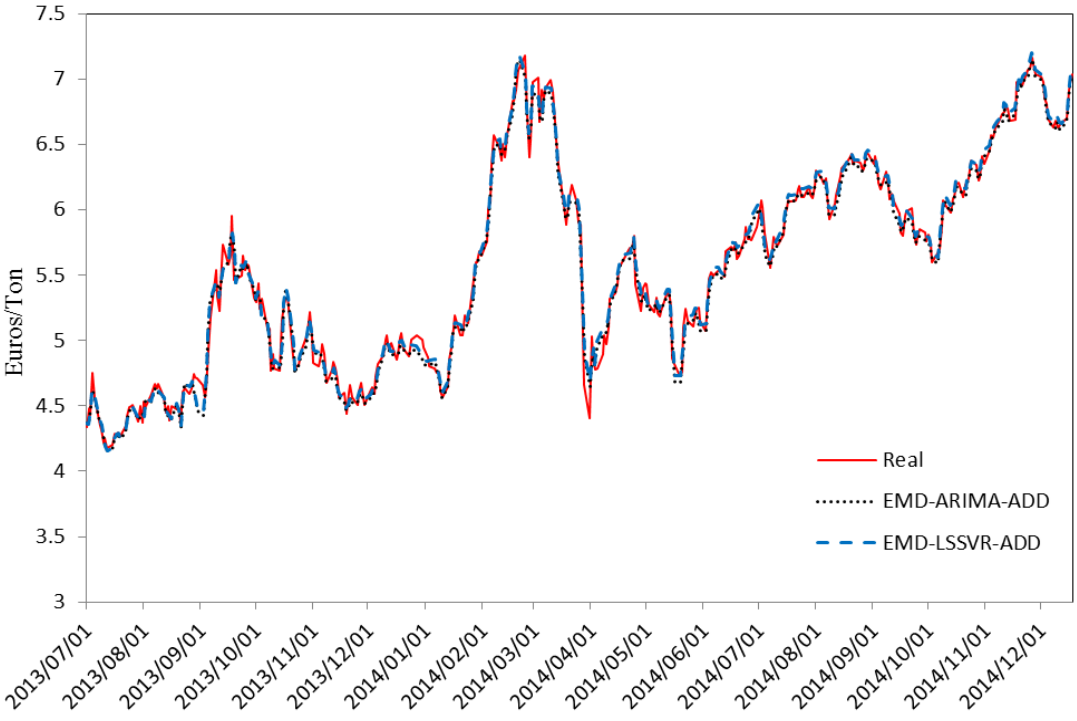


d. The decomposed results for DEC16.

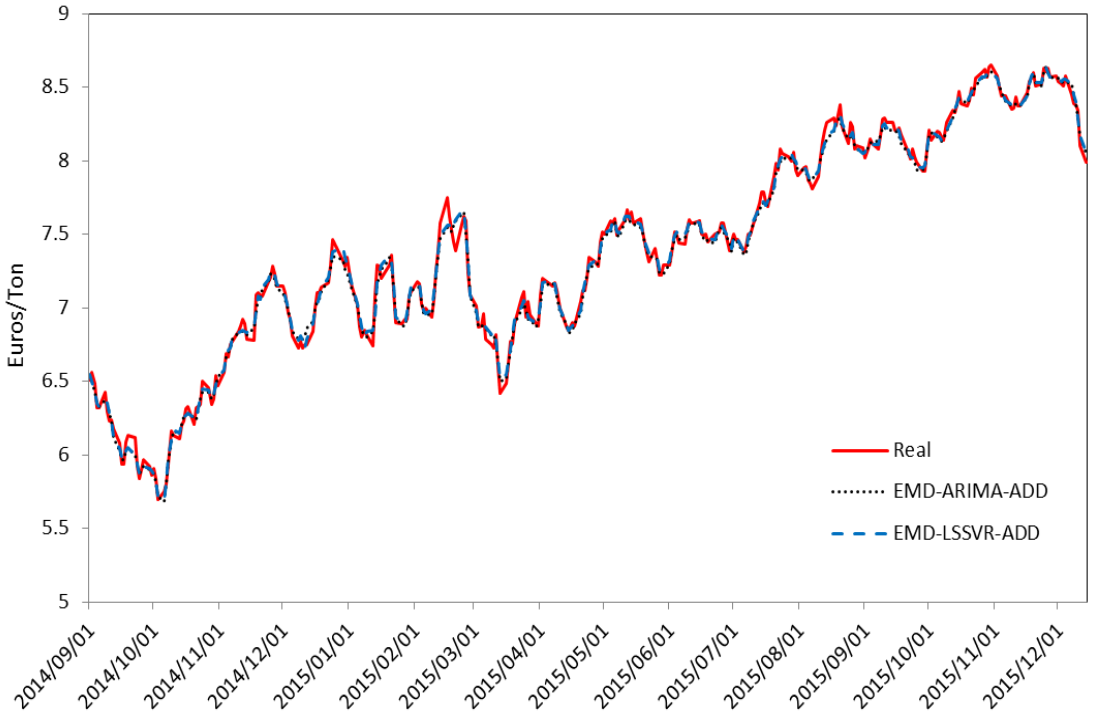
Fig.3.The decomposed results for carbon prices via EMD.



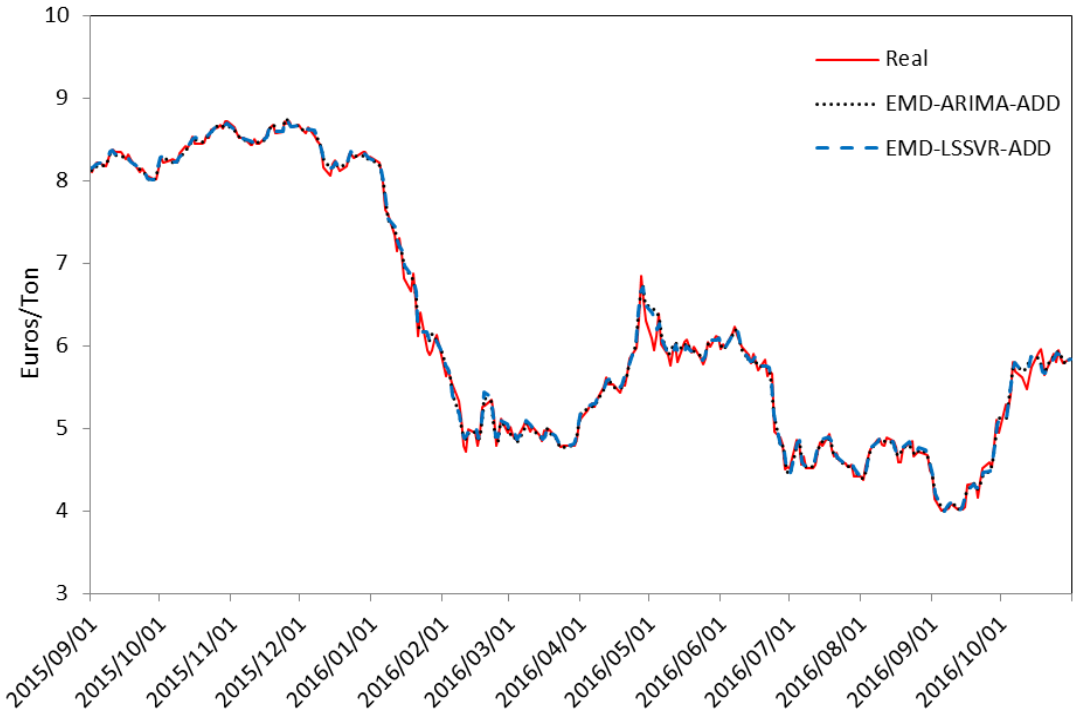
a. Out-of-sample forecasting results for DEC13.



b. Out-of-sample forecasting results for DEC14.



c. Out-of-sample forecasting results for DEC15.



d. Out-of-sample forecasting results for DEC16.

Fig.4. Out-of-sample forecasting results for carbon prices.

Tables

Table 1. Samples of carbon prices

Carbon price		Size	Date
DEC13	Sample set	1209	1 April 2009 - 16 December 2013
	Training set	833	1 April 2009 - 29 June 2012
	Testing set	376	2 July 2012 - 16 December 2013
DEC14	Sample set	1719	8 April 2008 - 18 December 2014
	Training set	1340	8 April 2008 - 28 June 2013
	Testing set	379	1 July 2013 - 18 December 2014
DEC15	Sample set	1035	29 November 2011 - 14 December 2015
	Training set	704	29 November 2011 - 29 August 2014
	Testing set	331	1 September 2014 - 14 December 2015
DEC16	Sample set	1006	27 November 2012 - 31 October 2016
	Training set	705	27 November 2012 - 31 August 2015
	Testing set	301	1 September 2015 - 31 October 2016

Table 2 Out-of –sample comparisons of the RMSE and the D_{stat} of each prediction model

Models		ARIMA	LSSVM	Hybrid	EMD-ARIMA-IMF_1-ADD	EMD-ARIMA-ADD	EMD-LSSVR-IMF_1-ADD	EMD-LSSVR-ADD
RMSE	DEC13	0.211	0.209	0.206	0.127	0.126	0.139	0.125
	DEC14	0.142	0.141	0.141	0.080	0.073	0.081	0.072
	DEC15	0.091	0.091	0.090	0.055	0.050	0.053	0.047
	DEC16	0.142	0.142	0.141	0.082	0.079	0.085	0.077
D_{stat}	DEC13	66.22	62.50	57.71	79.52	80.05	72.87	79.79
	DEC14	55.41	61.48	59.10	83.91	86.02	83.91	87.34
	DEC15	74.62	69.49	72.21	80.36	83.69	84.29	84.89
	DEC16	68.44	60.80	69.10	83.06	87.38	83.06	86.71

Table 3 Out-of –sample comparisons of DM test of each prediction model

	Test model	Reference model					
		ARIMA	LSSVM	Hybrid	EMD-ARIMA-IMF_1-ADD	EMD-ARIMA-ADD	EMD-LSSVR-IMF_1-ADD
DEC13	LSSVM	0.8276					
	Hybrid	0.2558	0.2250				
	EMD-ARIMA-IMF_1-ADD	0.0000 *	0.0000*	0.0000 *			
	EMD-ARIMA-ADD	0.0000 *	0.0000 *	0.0000 *	0.7428		
	EMD-LSSVR-IMF_1-ADD	0.0000 *	0.0000 *	0.0001 *	0.0069*	0.0222*	
	EMD-LSSVR-ADD	0.0000 *	0.0000 *	0.0000 *	0.6718	0.8389	0.0004 *
DEC14	LSSVM	0.5796					
	Hybrid	0.4803	0.7311				
	EMD-ARIMA-IMF_1-ADD	0.0000 *	0.0000*	0.0000 *			
	EMD-ARIMA-ADD	0.0000 *	0.0000 *	0.0000 *	0.0128 *		
	EMD-LSSVR-IMF_1-ADD	0.0000 *	0.0000 *	0.0000 *	0.4789	0.0119 *	
	EMD-LSSVR-ADD	0.0000 *	0.0000 *	0.0000 *	0.0079 *	0.4791	0.0005 *
DEC15	LSSVM	0.7598					
	Hybrid	0.2193	0.1131				
	EMD-ARIMA-IMF_1-ADD	0.0000 *	0.0000 *	0.0000 *			
	EMD-ARIMA-ADD	0.0000 *	0.0000 *	0.0000 *	0.0000 *		
	EMD-LSSVR-IMF_1-ADD	0.0000 *	0.0000 *	0.0000 *	0.0443 *	0.0164 *	
	EMD-LSSVR-ADD	0.0000 *	0.0000 *	0.0000 *	0.0000 *	0.0137 *	0.0000 *
DEC16	LSSVM	0.8957					
	Hybrid	0.2819	0.4421				
	EMD-ARIMA-IMF_1-ADD	0.0000 *	0.0000 *	0.0000 *			
	EMD-ARIMA-ADD	0.0000 *	0.0000 *	0.0000 *	0.5032		
	EMD-LSSVR-IMF_1-ADD	0.0000 *	0.0000 *	0.0000 *	0.0225 *	0.2506	
	EMD-LSSVR-ADD	0.0000 *	0.0000 *	0.0000 *	0.0865	0.4930	0.0175 *

Note: This table reports the P-values of DM test. * denotes that the null hypothesis is rejected at the significant level of 5%.

Table 4 Out-of –sample comparisons of Rate test of each prediction model

Test model		Reference model					
		ARIMA	LSSVM	Hybrid	EMD-ARIMA-IMF_1-ADD	EMD-ARIMA-ADD	EMD-LSSVR-IMF_1-ADD
DEC13	LSSVM	0.2872					
	Hybrid	0.0163 *	0.1801				
	EMD-ARIMA-IMF_1-ADD	0.0000 *	0.0000 *	0.0000 *			
	EMD-ARIMA-ADD	0.0000 *	0.0000 *	0.0000 *	0.8565		
	EMD-LSSVR-IMF_1-ADD	0.0477 *	0.0024 *	0.0000 *	0.0324 *	0.0204 *	
	EMD-LSSVR-ADD	0.0000 *	0.0000 *	0.0000 *	0.9268	0.9291	0.0257 *
DEC14	LSSVM	0.0902					
	Hybrid	0.3048	0.5034				
	EMD-ARIMA-IMF_1-ADD	0.0000 *	0.0000 *	0.0000 *			
	EMD-ARIMA-ADD	0.0000 *	0.0000 *	0.0000 *	0.4167		
	EMD-LSSVR-IMF_1-ADD	0.0000 *	0.0000 *	0.0000 *	1.0000 *	0.4167	
	EMD-LSSVR-ADD	0.0000 *	0.0000 *	0.0000 *	0.1786	0.5931	0.1786
DEC15	LSSVM	0.1417					
	Hybrid	0.4831	0.4417				
	EMD-ARIMA-IMF_1-ADD	0.0773	0.0013 *	0.0138 *			
	EMD-ARIMA-ADD	0.0041 *	0.0000 *	0.0004 *	0.2649		
	EMD-LSSVR-IMF_1-ADD	0.0021 *	0.0000 *	0.0002 *	0.1854	0.8334	
	EMD-LSSVR-ADD	0.0010 *	0.0000 *	0.0001 *	0.1243	0.6716	0.8308
DEC16	LSSVM	0.0502					
	Hybrid	0.8614	0.0330 *				
	EMD-ARIMA-IMF_1-ADD	0.0000 *	0.0000 *	0.0001 *			
	EMD-ARIMA-ADD	0.0000 *	0.0000 *	0.0000 *	0.1357		
	EMD-LSSVR-IMF_1-ADD	0.0000 *	0.0000 *	0.0001 *	1.0000	0.1357	
	EMD-LSSVR-ADD	0.0000 *	0.0000 *	0.0000 *	0.2116	0.8068	0.2116

Note: This table reports the P-values of Rate test. * denotes that the null hypothesis is rejected at the significant level of 5%.

Table 5 Out-of –sample comparisons of trading performances of each prediction model

Trading performances		ARIMA	LSSVM	Hybrid	EMD-ARIMA-IMF_1-ADD	EMD-ARIMA-ADD	EMD-LSSVR-IMF_1-ADD	EMD-LSSVR-ADD
DEC13	Annualized return(%)	-15.25	50.59	69.44	584.38	587.51	573.73	593.99
	Annualized volatility(%)	57.11	63.99	64.7	55.04	55.16	55.87	54.9
	Information ratio	-0.27	0.79	1.07	10.62	10.65	10.27	10.82
DEC14	Annualized return(%)	4.39	21.73	10.3	385.71	402.26	385.06	417.63
	Annualized volatility(%)	39.24	37.3	36.17	33.8	32.92	33.84	32.42
	Information ratio	0.11	0.58	0.28	11.41	12.22	11.38	12.88
DEC15	Annualized return(%)	56.28	40.15	51.99	194.27	206.55	203.93	211.6
	Annualized volatility(%)	17.48	17.11	17.51	16.64	16.05	15.85	15.75
	Information ratio	3.22	2.35	2.97	11.67	12.87	12.87	13.43
DEC16	Annualized return(%)	71	52.14	83.35	368.31	385.38	363.39	395
	Annualized volatility(%)	36.6	35.12	37.37	32.81	32.28	33.27	31.95
	Information ratio	1.94	1.48	2.23	11.23	11.94	10.92	12.36