



A randomized-algorithm-based decomposition-ensemble learning methodology for energy price forecasting

Ling Tang ^a, Yao Wu ^b, Lean Yu ^{b,*}

^a School of Economics and Management, Beihang University, Beijing, 100191, China

^b School of Economics and Management, Beijing University of Chemical Technology, Beijing, 100029, China

ARTICLE INFO

Article history:

Received 18 April 2017

Received in revised form

2 April 2018

Accepted 22 May 2018

Available online 23 May 2018

Keywords:

Decomposition-ensemble learning methodology

Randomized algorithm

Energy price forecasting

Extreme learning machine

Random vector functional link network

Random kitchen sinks

ABSTRACT

Inspired by the interesting idea of randomization, some powerful but time-consuming decomposition-ensemble learning paradigms can be extended into extremely efficient and fast variants by using randomized algorithms as individual forecasting tools. In the proposed methodology, Three major steps, (1) data decomposition via ensemble empirical mode decomposition, (2) individual prediction via a randomized algorithm (using randomization to mitigate training time and parameter sensitivity), and (3) results ensemble to produce final prediction, are included. Different from other existing decomposition-ensemble models using traditional econometric approaches or computational intelligence methods in individual prediction, this study employs some emerging randomized algorithms—extreme learning machine, random vector functional link network (using randomly fixed weights and bias in neural networks), and random kitchen sinks (using randomly mapping features to approximate kernels)—to dramatically save computational time and enhance prediction accuracy. With the Brent oil prices and the Henry Hub natural gas prices as studying samples, the empirical study statistically confirms that the proposed randomized-algorithm-based decomposition-ensemble learning models are proved to be excellently efficient and fast, relative to popular single techniques (including computational intelligence methods and randomized algorithms) and similar decomposition-ensemble counterparts (using the aforementioned single techniques as individual forecasting tools).

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

In the era of big data, energy price prediction, which is for capturing evolution laws of energy systems based on sufficient history observations (a typical case of big data) and thus providing a reliable future evaluation, has become an increasingly hot but challenging issue [1]. On the one hand, with the rapid development of the Internet and big data techniques, there exist a rich of data available concerning energy markets, which requires an urgent innovation of energy price forecasting techniques toward fast algorithms. Taking oil prices for example, besides historical time series data in different oil markets (e.g., the Brent and West Texas Intermediate), there are also a lot of myriad information with reference to the influencing factors like market factors (e.g., supplies and demands) [2] and external factors (e.g., substitutability with other energy resources, weather, stock levels, economic

growth, political changes, demographics, emergency events, and even psychological expectations) [3]. Accordingly, a fast learning algorithm is extremely desirable to effectively process these big data and produce prediction results rapidly. On the other hand, given that a high level of noise cannot be avoided within energy systems, how to capture the true information and enhance prediction accuracy still remains a key issue in the area of energy price prediction. For example, an accurate prediction for oil prices can help to improve the corresponding plans of production, marketing and investment, control potential risks and increase future profits in the oil-related sectors [4]. For this purpose, this paper will focus on improving the existing forecasting techniques toward efficient and fast algorithms, in the context of big data.

According to the existing studies, there are an abundance of forecasting techniques for energy prices, which generally fall into three major groups—econometric approaches, computational intelligences (CIs) and hybrid algorithms (integrating two or more single models in any aforementioned type(s)) [5]. In energy price prediction, for example, popular econometric models are auto-regressive integrated moving average (ARIMA) [6], generalized

* Corresponding author.

E-mail address: yulean@amss.ac.cn (L. Yu).

Nomenclature

ELM	extreme learning machine
RVFL	random vector functional links
RKS	random kitchen sinks
BPNN	back propagation neural network
LSSVR	least square support vector regression
ARIMA	autoregressive Integrated Moving Average Model
EEMD	ensemble empirical mode decomposition
IMF	Intrinsic Mode Function
x_t	the original time series
$c_{j,t}$	the j th IMF extracted in the j th extraction process at time t
r_t	the final residue
t	the period time

$g(x)$	the activation function
N	the number of samples
M	the number of hidden neuron nodes
x	the input vector
y	the output vector
ω	the weight vector connecting the input nodes with the hidden neuron node
β	the weight vector connecting the output nodes with the hidden neuron node
b	the threshold (or bias) of the hidden neuron node
H	the output matrix of hidden layer
X	the matrix of input
Y	the matrix of targets
D	the size of testing dataset

autoregressive conditional heteroskedasticity (GARCH) [7], random walk (RW) [8], vector auto-regression (VAR) [9] and error correction models (ECM) [10]. However, these models hold the data assumptions of stationarity and nonlinearity which contradict the real energy systems. In such a context, CIs have been become the most dominant approaches in energy price forecasting, such as artificial neural networks (ANNs) [5], support vector regression (SVR) [11], least squares support vector regression (LSSVR) [12] and various CI-based optimization tools. However, these conventional CI techniques have two intrinsic shortcomings—time consuming and parameter sensitivity [13]. For example, ANNs, using gradient descent methods for tuning parameters (such as weights and bias), take a long training time but frequently fall into local optimum [14]. Similarly, SVR and LSSVR, using iterative learning algorithms (such as the grid searching method or the trial-and-error method) to determine regularization and kernel parameters, cannot avoid the double problems of time consuming and parameter sensitivity [15]. Due to the respective disadvantages of the first two types, the third type, i.e., hybrid techniques combining two or more single algorithms, have emerged and offered an excellent performance in energy price prediction.

In particular, the decomposition-ensemble learning paradigms based on the promising concept of “decomposition and ensemble” has been widely considered as an excellent case among hybrid methods [16]. In a typical decomposition-ensemble model, three major steps are included—data decomposition to decompose the original complex data into relatively simple components for reducing data complexity, individual prediction to model each extracted component independently, and results ensemble to aggregate individual predictions to the final predictions [12,16]. The superiority of decomposition-ensemble techniques has been proved in terms of prediction accuracy in the forecasting of energy prices such as oil prices [6] and gas prices [17]. However, the “decomposition and ensemble” strategy poses a big challenge, i.e., a large computational burden for modeling all the decomposed components individually. Furthermore, most existing decomposition-ensemble models employed CI-based individual predictors with iterative tuning processes, such as ANN, SVR and LSSVR [5,11,12], which largely aggravates the time-consuming problem [14]. In addition, the performances of these CI algorithms are heavily dependent on the predesigned parameters concerning the iterative learning process, any one of which set inappropriately will make a great difference in the final prediction [15]. In this sense, the emerging decomposition-ensemble learning methodology severely suffers from the double problems of time consuming and parameter sensitivity. For this purpose, this study

will try to address the both issues of time consuming and parameter sensitivity.

Fortunately, the double issues of time consuming and parameter sensitivity can be nicely overcome by using the interesting idea of randomization, thereby an extremely efficient and fast decomposition-ensemble learning methodology can be developed. Based on randomization, some randomized algorithms have recently presented and shown excellent capabilities in terms of fast speed and prediction accuracy. In particular, the randomized algorithms employ randomly fixed parameters, randomly mapping features, randomly generated samples or randomly selected variables rather than iteratively tuned ones in conventional CIs, which effectively ensures an extremely fast learning speed and an excellent generalization performance [14]. Furthermore, without setting stopping criteria, learning rate, learning epochs and other parameters in learning processes, the problem of parameter sensitivity can be greatly solved. Typical cases include extreme learning machine (ELM) [18] and random vector functional links (RVFL) network [19] using randomly fixed input weights and hidden bias in neural networks, random kitchen sinks (RKS) [20] using randomly mapping features to approximate shift invariant kernels, and random forest (RF) [21] using randomly bootstrapped samples and randomly selected variables to grow a decision tree. These above randomized algorithms have extensively been applied to various complex systems such as electricity load [22,23], electricity production [24], pedestrian detection [25], wave energy flux [26], water demand [27], wind farm power ramp rates [28], mineral prospectivity [29], etc. Therefore, this paper try to introduce such emerging randomized algorithms to formulate some efficient and fast decomposition-ensemble learning models. To have a better overview of the conventional CIs and randomized algorithms, Table 1 synthetically presented the advantages and disadvantages of those intelligent approaches.

By using randomized algorithms, some randomized-algorithm-based decomposition-ensemble learning models have recently been developed and obtained satisfactory forecasting results. For example, Tang et al. [35] introduced ELM as the individual predictor into the “decomposition and ensemble” framework and observed the effectiveness of the proposed methodology in oil price prediction in terms of time-saving and accuracy. Wang et al. [36] used two-phase decomposition technique and modified extreme learning machine to forecast air quality index. Lu and Shao [37] developed an ensemble learning approach with ELM as the individual forecasting tool for computer products sales forecasting. Shrivastava [38] built a wavelet-based ELM decomposition-ensemble model for electricity price forecasting. Tang et al. [15]

Table 1

Advantages and disadvantages of different intelligent algorithms.

Models	Advantages	Disadvantages
RVFL	Universal approximator for a continuous function, direct links in RVFL improve the performance, less computational time via non-iterative learning [15,22,30]	Is randomness beneficial to general data-mining problems, or is there a way of constructing informative, functional, and cheap feature spaces without randomization techniques that can perform favorably? How randomness in an algorithm affects the resulting error bounds? [34]
ELM	Avoiding tedious and parameter tuning, good generalization performance, not very sensitive to the number of hidden nodes [31]	
RKS	Map the input data to a randomized low-dimensional feature space to approximate many popular kernels, the random features are powerful and economical for large scale learning [20]	
BPNN	Requiring less formal statistical training, ability to implicitly detect complex nonlinear relationships, ability to detect all possible interactions between predictor variables, and the availability of multiple training algorithms [32]	Local minima, slow learning speed and poor generalization performance [33]
LSSVR	Excellent generalization ability [14]	The kernel matrix scales poorly with the size of the training data leading to time-consuming [14]
ARIMA	Good ability to solve linear problem [15]	Lose power in complex time series and long-term predictions [15]

using RVFL developed a decomposition-ensemble learning paradigm for oil price forecasting. However, to the best of our knowledge, there were few decomposition-ensemble learning paradigms by using other probably more competitive randomized algorithms like RKS etc. Therefore, this study especially fills in such a literature gap by introducing various promising randomized algorithms and conducting a thorough comparison to explore whether the idea of randomization does improve the existing decomposition-ensemble learning paradigms in terms of speed and accuracy.

Generally speaking, this study aims to formulate some efficient and fast decomposition-ensemble learning models by using the emerging randomized algorithms in individual prediction for energy price forecasting, which well solves the double problems of time consuming and parameter sensitivity. The major contributions of this study can be summarized into two perspectives. First, by introducing various randomized algorithms, a series of randomized-algorithm-based decomposition-ensemble models are formulated. Second, a thorough comparison is conducted to check whether the idea of randomization does improve the existing decomposition-ensemble learning paradigms from the perspectives of speed and accuracy, and to explore the most efficient and fast one in energy price prediction.

The rest of this study is organized as follows. Section 2 describes the formulation process of the proposed methodology. For illustration and verification, the proposed methodology is performed to predict the Brent crude oil spot prices and the Henry Hub natural gas prices, as the results presented in Section 3. Finally, Section 4 concludes the major contributions of the paper, and discusses some interesting directions for future research.

2. Methodology formulation

This section presents the formulation process of the proposed methodology. In particular, Section 2.1 designs the general model framework, and Sections 2.2 and 2.3 describe the related techniques in detail.

2.1. Model framework

In general, three major steps are included in a typical decomposition-ensemble learning paradigm, i.e., data decomposition, individual prediction and results ensemble. However, given that a set of decomposed components are extracted in data decomposition and modeled in individual prediction, a challenge, i.e., large computational burden, arises. In such a context, this paper tries to employ the emerging randomized algorithms (e.g., ELM, RVFL and RKS) as the individual prediction tools to improve the existing models toward extremely efficient and fast variants for

energy price forecasting. In particular, the randomized algorithms use the promising idea of randomization, in terms of randomly fixed parameters, randomly generated samples or randomly mapping features rather than iteratively tuned ones, which effectively guarantees a much faster learning speed and significantly boost the performance or efficiency of neural networks [14]. Fig. 1 illustrates the general framework of the proposed decomposition-ensemble methodology using randomized algorithms.

Generally, the proposed methodology involves three major steps, data decomposition via EEMD, individual prediction via a randomized algorithm, and results ensemble via linear addition, which are elaborated below.

2.1.1. Data decomposition

In this step, an effective multi-scale analysis technique, EEMD [39], is employed to decompose the original time series $x_t, (t = 1, 2, \dots, T)$ into n intrinsic mode functions (IMFs) $c_{j,t}, (j = 1, 2, \dots, n)$ and one residue r_t . In particular, EEMD, an improved EMD, employs random white noises to address mode mixing in original EMD and to extract the true inner patterns, which can helpfully ensure the analysis capabilities of data decomposition.

2.1.2. Individual prediction

In the second step, some randomized algorithms, such as ELM, RVFL and RKS, are introduced to efficiently and fast model all the extracted IMFs and residue, and the corresponding individual prediction results $\hat{c}_{j,t}$ and \hat{r}_t can be generated. Particularly, such emerging randomized algorithms, using randomization (in terms of randomly fixed parameters, randomly generated samples or randomly mapping features) without a tuning process, can extremely save computational time and largely solve the problem of parameter sensitivity in conventional CI models.

2.1.3. Results ensemble

In the final step, all the individual prediction results are aggregated into the final ensemble prediction $\hat{x}_t = \sum_{j=1}^n \hat{c}_{j,t} + \hat{r}_t$, given that the sum of all the decomposed components exactly equals to the original data, i.e., $x_t = \sum_{j=1}^n c_{j,t} + r_t$.

In the proposed methodology, several related techniques, i.e., EEMD for data decomposition and the randomized forecasting algorithms (i.e., ELM, RVFL and RKS) for individual prediction, are briefly introduced in Subsections 2.2 and 2.3, respectively.

2.2. EEMD

EEMD, an improved variant in the EMD family, employs random

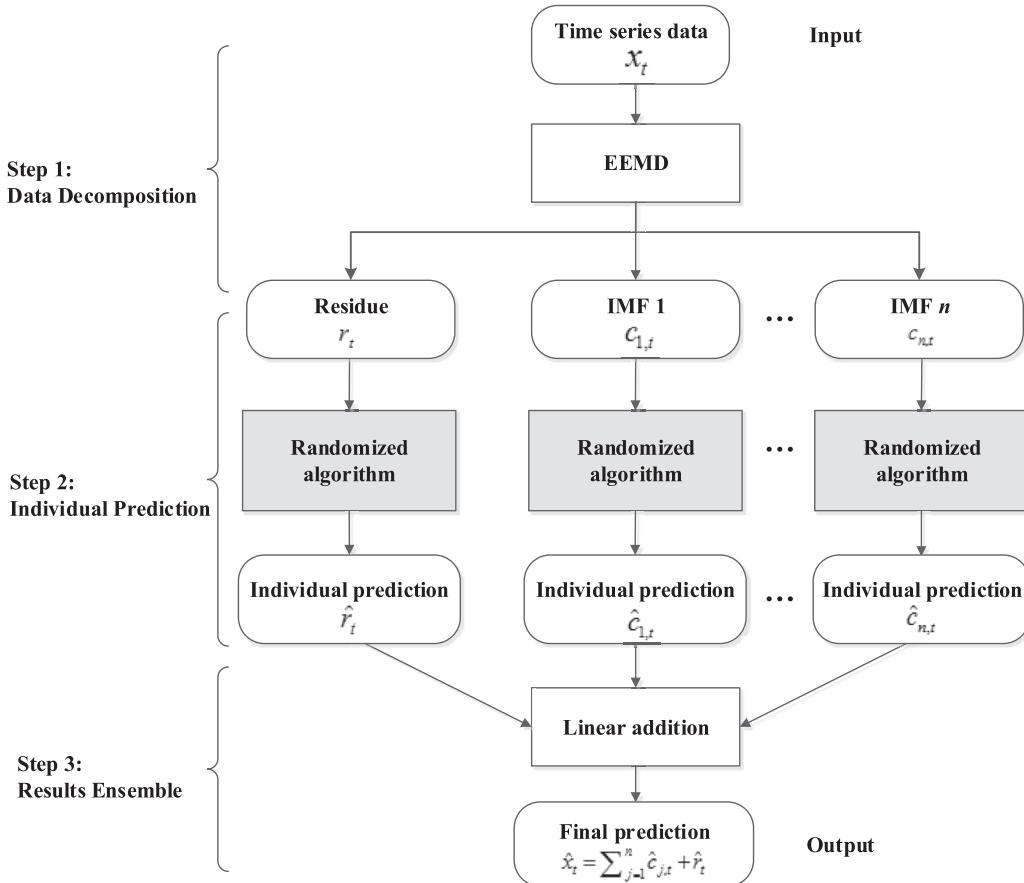


Fig. 1. General framework of the proposed randomized-algorithm-based decomposition-ensemble methodology.

noises to effectively address the mode mixing problem in the original EMD method and to capture the true inner patterns [40]. Different from other traditional decomposition methods using fixed bases (e.g., Fourier and wavelet analyses), EEMD is an empirical, intuitive, direct, self-adaptive data processing approach, which is designed for nonlinear and nonstationary data. Due to the above merits, EEMD has extensively been applied to diverse complex system analysis [15]. Therefore, this study employs this effective data decomposition technique to analyze energy systems.

The main purpose of EMD is to extract the coexisting oscillatory inner patterns, i.e., IMFs, from the original time series. Mathematically, each IMF must meet the following two conditions. First, in the whole IMF, the number of extrema (including maxima and minima) and the number of zero crossings are equal or different at most by one. Second, IMF must be symmetric with respect to a local zero mean. Through n extraction processes, the original data series $x_t (t = 1, 2, \dots, T)$ is decomposed into n IMFs and a final residue in a linear combination form.

$$x_t = \sum_{j=1}^n c_{j,t} + r_t, \quad (1)$$

where $c_{j,t} (j = 1, 2, \dots, n)$ is the j th IMF extracted in the j th extraction process at time t , r_t is the final residue, and $n = \log_2 T$ is the number of extraction processes.

To address the problem of mode mixing in EMD, EEMD especially adds random white noises to the original time series data, under the assumption that complex data are contaminated by noise, while the ensemble means of data with different levels of

random white noises are closer to the true information. Under such an assumption, B series of white noises are randomly produced with a similar standard deviation ε , and then employed to capture the true IMFs and offset themselves via ensemble averaging. The final errors ε_B can be well controlled [34].

$$\varepsilon_B = \frac{\varepsilon}{\sqrt{B}}. \quad (2)$$

2.3. Randomized algorithms

In the proposed methodology, the energy price series is decomposed into n IMFs and one final residue in the first step, thus a challenge, i.e., large computational burden, arises accordingly in modeling all of the decomposed components in the second step of individual prediction. In such a context, this paper tries to introduce the randomized algorithms to improve the existing decomposition-ensemble learning models into extremely efficient and fast variants. In particular, based on the promising idea of randomization, such emerging randomized algorithms use randomly fixed parameters, randomly generated samples, or randomly mapping features rather than iteratively tuned ones, which effectively ensures an extremely faster learning speed and a much better generalization performance. In this study, three popular randomized algorithms, i.e., ELM, RVFL and RKS are adopted.

2.3.1. ELM

Inspired by the interesting concept of randomization, the

randomized algorithm of ELM is actually a special case of the single hidden layer feedforward neural networks (SLFN) using randomly fixed weights and bias [18]. In standard SLFNs with M hidden neuron nodes and the activation function $g(x)$, for N samples (x_i, y_i) ($i = 1, 2, \dots, N$) (where $x_i = [x_{i,1}, x_{i,2}, \dots, x_{i,n}]^T \in \mathbb{R}^n$ is the input and $y_i = [y_{i,1}, y_{i,2}, \dots, y_{i,m}]^T \in \mathbb{R}^m$ is the output), the mathematical regression function [35] is expressed as

$$\sum_{i=1}^M \beta_i g_i(x_j) = \sum_{i=1}^M \beta_i g_i(\omega_i x_j + b_i) = t_j, j = 1, 2, \dots, N, \quad (3)$$

where $\omega_i = [\omega_{i,1}, \omega_{i,2}, \dots, \omega_{i,n}]^T$ denotes the weight vector connecting the input nodes with the i th hidden neuron node, $\beta_i = [\beta_{i,1}, \beta_{i,2}, \dots, \beta_{i,m}]^T$ is the weight vector connecting the output nodes with the i th hidden neuron node, and b_i is the threshold (or bias) of the i th hidden neuron node. In a conventional SLFN, the input weights ω_i and the hidden layer biases b_i are iteratively tuned by minimizing errors to approximate samples.

Different from typical SLFNs, ELM otherwise generates and fixes these parameters randomly based on an appropriate given distribution (e.g., the Gaussian distribution), and such a training process is transformed to evaluating a minimum norm least squares (LS) solution $\hat{\beta}$ for the linear system $H\beta = Y$, where H denotes the output matrix of hidden layer and Y indicates the matrix of targets.

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \dots & h_M(x_1) \\ \vdots & \ddots & \vdots \\ h_1(x_N) & \dots & h_M(x_N) \end{bmatrix} \lim_{x \rightarrow \infty}, \quad (4)$$

$$Y = [y_1^T, \dots, y_N^T]^T. \quad (5)$$

Accordingly, estimating the output weights β linking the hidden layer to the output layer is equivalent to exploring the LS solution $\hat{\beta} = [\beta_1^T, \dots, \beta_N^T]^T$ to the following linear system,

$$\begin{aligned} & \|H(\omega_1, \omega_2, \dots, \omega_M, b_1, \dots, b_M)\hat{\beta} - Y\| \\ &= \min_{\beta} \|H(\omega_1, \omega_2, \dots, \omega_M, b_1, \dots, b_M)\beta - Y\|. \end{aligned} \quad (6)$$

By using a pseudoinverse, the minimum norm LS solution to Eq. (6) can be obtained [41].

$$\hat{\beta} = H^+ Y, \quad (7)$$

where H^+ is the Moore-Penrose (MP) generalized inverse of matrix H . The minimum norm LS solution is unique and has the smallest norm among all LS solutions.

2.3.2. RVFL

Similar to ELM, RVFL randomly fixes the parameters of input weights and the hidden layer biases, rather than tune them via a time-consuming iterative process [16]. Different from ELM, RVFL considers a direct link from input layer to output layer in terms of input patterns [19]. Accordingly, there exist two kinds of nodes in RVFL, i.e., enhancement patterns $\sum_{i=1}^M \beta_i g_i(\omega_i x_j + b_i)$ (see Eq. (3)) and input patterns (i.e., the input-output link) $\sum_{i=M+1}^{M+n} \beta_i x_j$. Thus Eq. (3) can be extended into Eq. (8).

$$f(x_j) = \sum_{i=1}^M \beta_i g_i(\omega_i x_j + b_i) + \sum_{i=M+1}^{M+n} \beta_i x_j, j = 1, 2, \dots, N. \quad (8)$$

Accordingly, the input patterns $\sum_{i=M+1}^{M+n} \beta_i x_j$ serve as an input-output link, while there is no link in ELM. Based on Eq. (8), the original data $x_i = [x_{i,1}, x_{i,2}, \dots, x_{i,n}]^T \in \mathbb{R}^n$ can be mapped from a n -dimension space to a $(M+n)$ -dimension space.

Similar to ELM, RVFL randomly fixes the parameters ω_i and b_i , and sets the weight terms β_i by minimizing errors in the following system,

$$E = \frac{1}{2N} \sum_{i=1}^N (y_i - \beta d_i)^2, \quad (9)$$

Where $d = [g(\omega_1^T \mathbf{X} + b_1), \dots, g(\omega_M^T \mathbf{X} + b_M), x_1, \dots, x_n]$ is the input pattern and enhancement pattern vector. The minimum of system error in Eq. (9) can be found by a conjugate gradient approach [42]. However, if matrix inversion with the use of a pseudoinverse is feasible, the optimal weights β_i can be obtained by a single step without any iteration [19]. To save computational time, the pseudoinverse is used to obtain the weights β_i in this study.

2.3.3. RKS

RKS [20], a typical randomized method, employs randomly mapping features to approximate shift invariant kernels. The starting point of RKS is a celebrated result which is characteristic of the class of positive definite functions. According to Bochner's theorem [43], for any normalized continuous positive definite function, there exists a finite non-negative Borel measure $\mu \in \mathbb{R}^d$.

$$f(x) = \int_{\mathbb{R}^d} e^{-i\omega^T x} d\mu(\omega), \quad (10)$$

Where f is the Fourier transform of a finite non-negative Borel measure μ on \mathbb{R}^d . Accordingly, each entry of kernel matrix $K \in \mathbb{R}^{n \times n}$ can be approximated,

$$K_{i,j} = k(x_i - x_j) = \int p(z) e^{iz^T(x_i - x_j)} dz \approx \langle z(x_i), z(x_j) \rangle, \quad (11)$$

where $z(x)$ is the random feature function for a given shift-invariant kernel K (such as the Gaussian kernel here) based on a proper probability density function $p(z)$, which can be computed via the inverse Fourier transform of the kernel. Given the input data x_i ($i = 1, 2, \dots, N$), the training kernel methods can enhance speed by using linear algorithms incorporated with $z(x_i)$, and achieve higher accuracy than the linear algorithms directly operated on the original data x_i . Compared with kernel methods, linear algorithms using x_i run far more quickly especially in low-dimension data spaces, but with less accuracy. However, this disadvantage can be finely addressed by applying $z(x_i)$, which contains the information of kernels.

In this study, RKS is applied to least squares regression, by solving the least squares problem $\min_{\omega} \|z^T \omega - t\|_2^2 + \lambda \|\omega\|_2^2$ where t denotes the vector of targets, ω is the vector of weights, and z is the matrix of random features. By employing RKS, the prediction result $\omega^T z(x_i)$ can be fast obtained base on the input data x_i .

3. Empirical study

For illustration, the Brent crude oil spot prices and the Henry Hub natural gas prices are selected as the studying samples. The experiment is designed in Section 3.1, and Section 3.2 presents the empirical results and discusses whether the proposed model by using randomized algorithms statistically improves the energy

price prediction in terms of speed, accuracy and robustness.

3.1. Experimental design

As for sample data, two typical energy prices, the Brent crude oil spot prices and Henry Hub natural gas prices, are employed. For data availability, a total of 7370 daily observations of oil price and 4873 daily observations of natural gas are collected from the US energy information administration (EIA) (<http://www.eia.doe.gov/>), as shown in Fig. 2 where x and y coordinates represent date and the price of crude oil (dollars per barrel), natural gas prices (dollars per million btu) respectively. The sample data are further divided into training set accounting for the first 80% of the total samples and testing set accounting for the last 20%. Moreover, multi-step ahead predictions with the horizons of 1–5 days are performed to test the robustness of models.

Regarding benchmarking models, the most popular forecasting techniques especially for energy prices are employed, including the most typical CIs of back propagation neural network (BPNN) and LSSVR [5,12], and the emerging randomized algorithms of ELM, RVFL and RKS. It is worth noticing that ELM and RVFL using randomly fixed weights and bias are the improved variants of conventional neural networks (e.g., BPNN here), while RKS using randomly mapping features is an improved variant of traditional kernel methods (e.g., LSSVR). Furthermore, the above single conventional benchmarks of BPNN and LSSVR are also employed as individual forecasting tools to formulate two decomposition-ensemble benchmarks. Generally speaking, for the proposed learning paradigms using randomized algorithms (i.e., EEMD-ELM, EEMD-RVFL and EEMD-RKS), five single models (i.e., ELM, RVFL, RKS, BPNN and LSSVR) and two ensemble models (i.e., EEMD-BPNN and EEMD-LSSVR) are formulated as benchmarking methods. For the above abbreviations for ensemble models, the first part represents for decomposition tool, and the second part for single forecasting tools.

As for evaluation criteria, the mean absolute percent error (MAPE) is selected to evaluate level accuracy [44].

$$MAPE = \frac{1}{D} \sum_{t=1}^D \left| \frac{x_t - \hat{x}_t}{x_t} \right|, \quad (12)$$

where D is the size of testing dataset. For directional accuracy, a directional statistic ($Dstat$) [15] is utilized:

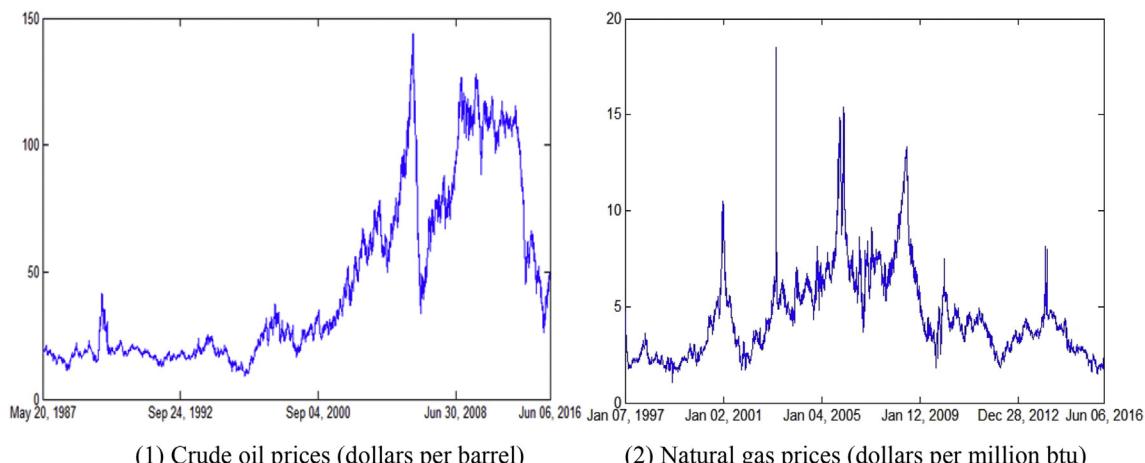


Fig. 2. Sample data in the experiments.

$$Dstat = \frac{1}{D} \sum_{t=1}^D a_t \times 100\%, \quad (13)$$

where $a_t = 1$ if $(\hat{x}_{t+1} - x_t)(x_{t+1} - x_t) \geq 0$, or otherwise $a_t = 0$. To measure the speed of models, the computational time covering both training and testing processes are computed. All the models are conducted in MATLAB environment running on a desktop with CPU 3.3 GHz and 4 GB RAM.

In performance comparison, the improvement rate (IR) is a useful index to directly show the improvement of a target model A on the benchmark B in terms of a given criterion C .

$$IR_C = \frac{C_A - C_B}{C_B} \times 100\%, \quad (14)$$

where C_i is the performance measurements (in terms of MAPE, $-Dstat$ or computational time here) produced by model i ($i = A, B$). If IR_C is positive, the target model A appears more powerful than B in terms of criterion C . The larger IR_C is, the more superior the target model A can be tested.

To statistically test the superiority of the proposed models, a unilateral t -test is conducted on the null hypothesis that the prediction performance of the target model A is inferior to the benchmarking model B . Accordingly, the t -test statistic [45] can be designed as

$$t_C = \frac{\bar{C}_A - \bar{C}_B}{S_{\bar{C}_A - \bar{C}_B}}, \quad (15)$$

where \bar{C}_i is the average value of performance measurements of model i ($i = A, B$) during the testing period (in terms of MAPE, $-Dstat$ or computational time), and $S_{\bar{C}_A - \bar{C}_B} = \sqrt{(s_A^2 + s_B^2)/D}$ where s_i^2 is the unbiased estimator for the variance of performance measurements.

3.2. Empirical results

A comprehensive comparison is conducted between the three proposed decomposition-ensemble learning algorithms (using randomized algorithms as individual forecasting tools) and the benchmarking models (including both single techniques and similar ensemble counterparts with conventional individual forecasting tools), in order to (1) check out whether the proposed

methodology with randomized algorithms does improve prediction, and (2) explore the most efficient and fast method(s) for energy prices. For a clear discussion, the comparison results are analyzed from the following three perspectives. First, the prediction performances of the five single benchmarks are compared in Section 3.2.1. Second, the proposed methods are compared with all the benchmarks for testifying the superiority, as discussed in Section 3.2.2. Third, Section 3.3 concludes the major findings of the empirical study.

3.2.1. Performance comparison of single methods

The five benchmarking models, ELM, RVFL, RKS, BPNN and LSSVR, are first employed to predict the Brent crude oil prices and the Henry Hub natural gas prices. As for the three neural networks, the hidden nodes in both RVFL and ELM is determined via the grid searching method [15]. Accordingly, the number of enhancement nodes in RVFL is set to 10 for crude oil prices and 8 for natural gas prices by the grid search method, while the number of hidden neurons in ELM is 20 for crude oil prices and 8 for natural gas prices,

respectively. The input weights and hidden biases in RVFL and ELM are randomly fixed based on the Gaussian distribution, and the activation function is the sigmoidal function $g(x) = 1/(1 + e^{-x})$. In BPNN, a typical three-layer network is employed, in which the number of hidden nodes is set to 7 and the number of iterations is 10,000. In LSSVR, the Gaussian RBF kernel function is selected, and the regularization and kernel parameters are set via the grid search method [12]. In RKS, the number of random features is set to 100.

The prediction performances of five single models are compared in terms of *MAPE*, *Dstat* and computational time, as the results reported in Figs. 3–5 and the corresponding *t*-test results in Tables S1–S6 in the Supplementary, respectively. From these comparison results, one important conclusion can be obviously seen that the three randomized algorithms, ELM, RVFL and RKS, can be shown to be not only effective in producing stable results in terms of level accuracy (see Fig. 3), but also extremely fast with the least computational time (Fig. 5). The *t*-test statistically proves the above conclusion, given that all the *p*-values are far below the significance level of 1% when testing the three randomized

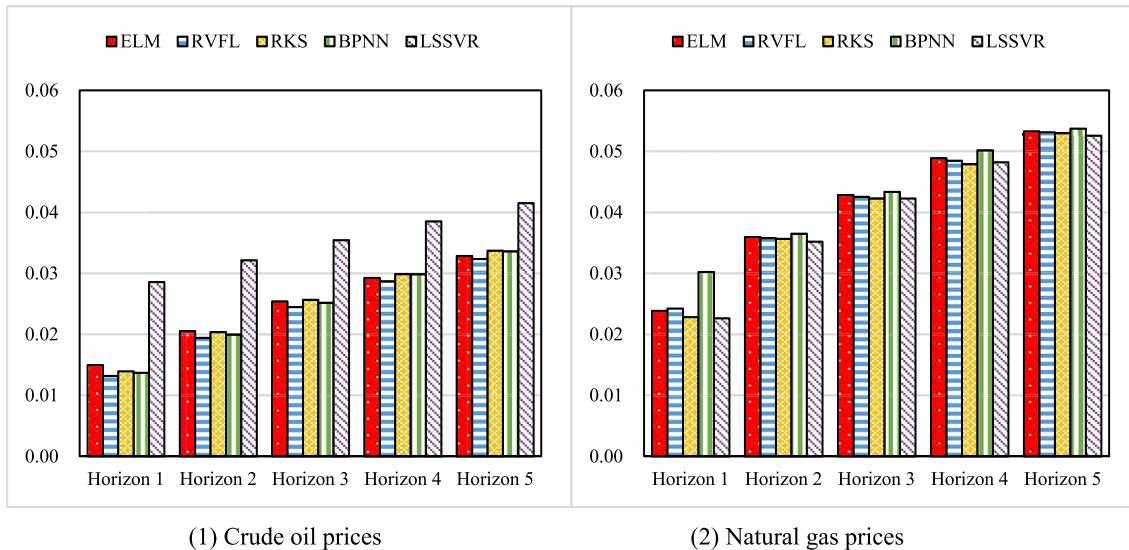


Fig. 3. Performance comparison of single methods in terms of *MAPE* criterion.

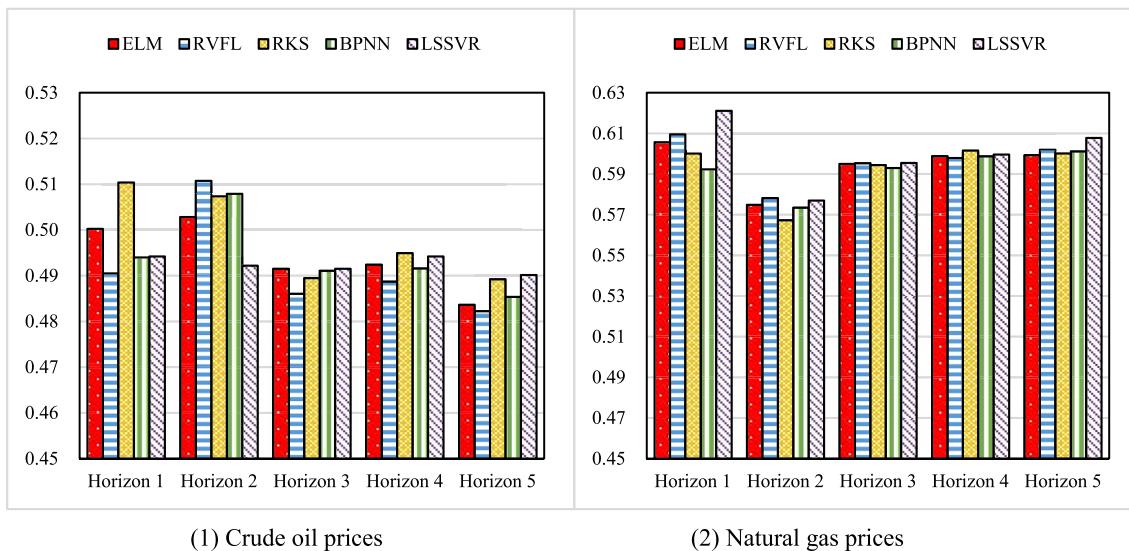


Fig. 4. Performance comparison of single methods in terms of *D_{stat}* criterion.

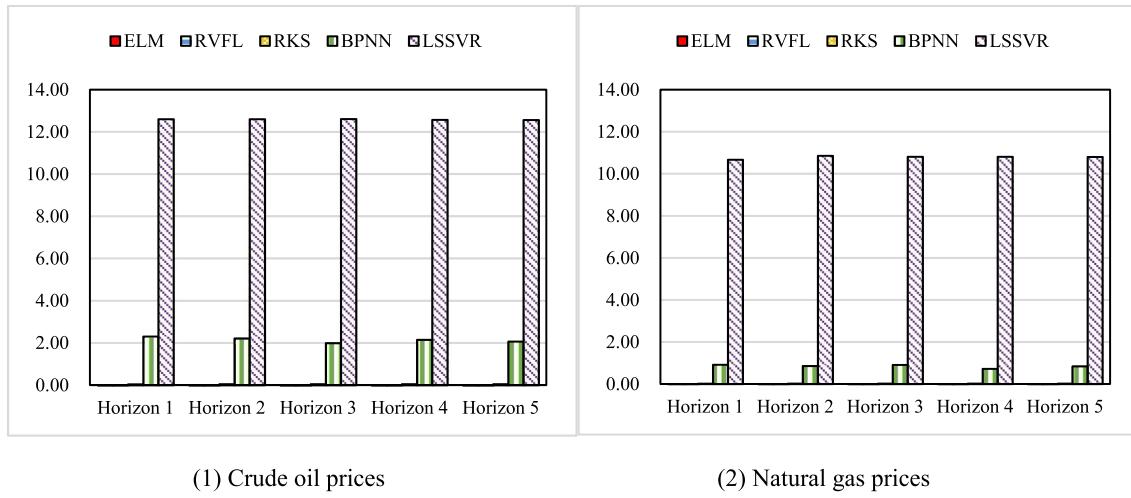


Fig. 5. Performance comparison of single methods in terms of computational time (seconds).

algorithms on computational time (Tables S3 and S6, which are shown in Supplementary).

As for level accuracy in terms of MAPE, it is obvious that the three randomized algorithms, ELM, RVFL and RKS, can be shown to be robust forecasting methods without an instable result in the both cases of oil prices and gas prices. In particular, ELM, RVFL, RKS and BPNN perform excellently compared with LSSVR in crude oil price prediction, while except for BPNN all the single models appear a similar performance for natural gas prices. In the same way, it is not hard to find some other models, such as BPNN and LSSVR, are observed to lose power in some cases. In the two conventional CI models, LSSVR produces extremely poor forecasting results across all the horizons for crude oil prices, and BPNN becomes instable in the one-step-ahead prediction for natural gas prices. The main reasons can be attributed to the super-sensitivity problem in the traditional CI models [15]. Nevertheless, the randomized algorithms with randomization can largely address such a problem. From the above results, the idea of randomization can prove to be effective in improving energy price forecasting because ELM, RVFL (using randomly fixed parameters) and RKS (using randomly mapping features) consistently produce stable satisfactory prediction results for both crude oil prices and natural gas prices.

Regarding to directional accuracy measured by D_{stat} , it can be seen that all the five single models reveal a similar poor performance, and none of them can be observed to be better than the others in all the cases. Furthermore, the directional accuracy of all the five single models is at an extremely low level, with the maximal D_{stat} values approximately 0.53 for crude oil prices and 0.63 for natural gas prices, respectively. The potential reasons mainly lie in the limitation of single models, while the promising concept of “decomposition and ensemble” can effectively improve energy price forecasting, especially from the perspective of directional accuracy [5].

In terms of speed, Fig. 5 indicates that the three randomized algorithms, ELM, RVFL and RKS, are significantly faster than the other single methods—LSSVR (via the grid searching method for parameter tuning) and BPNN (via the gradient descent method). This result finely verifies to the effectiveness of the randomization idea in addressing the problems of time-consuming and slow convergence in the conventional forecasting techniques. When comparing the three randomized algorithms, ELM and RVFL are slightly faster than RKS. Coupling the results of accuracy and speed, the randomized algorithms, ELM, RVFL and RKS can be proved to be

efficient and fast.

To statistically test the differences among the five single benchmarks, the unilateral *t*-test is conducted, as the results reported in Tables S1–S3 for crude oil prices and Tables S4–S6 for natural gas prices. According to the testing results, the main conclusions deduced above can be statistically confirmed. As for level accuracy (see Tables S1 and S4), the three randomized algorithms, ELM, RVFL and RKS, can be tested to be the most powerful algorithms in crude oil price prediction, given that all the corresponding *p*-values are far below the significance level of 1% in most cases. However, none of the single models can be consistently shown to be more powerful than the others for natural gas prices. Regarding to directional accuracy (Tables S2 and S5), it can be obviously seen that none of five single models can be consistently proved to be better than the others, at the significance level of 1%. As for speed, when testing ELM, RVFL and RKS against the other single benchmarks, all the *p*-values are far below the significance level of 1% without any exception, which statistically supports the effectiveness of randomization in saving computational time. Therefore, the *t*-test statistically confirms that the randomized algorithms are efficient and fast techniques, compared with conventional CI approaches. Thus, this study utilizes them to improve the existing decomposition-ensemble learning paradigms severely suffering from computational burden and parameter sensitivity.

3.2.2. Performance comparison of decomposition-ensemble learning paradigms

The five decomposition-ensemble learning models using the above single algorithms as individual forecasting tools are then performed. First, EEMD is employed to decompose the original series into relatively simple components, by using 100 ensemble members with different random white noises of a standard deviation 0.01. Fig. 6 shows the decomposition results, in which the IMFs are listed in the order from the highest to the lowest frequencies, and the last one is the residue. The x and y coordinates represent date and the price of crude oil (dollars per barrel), natural gas prices (dollars per million btu) respectively. In the second step, the five forecasting models, ELM, RVFL, RKS, BPNN and LSSVR are respectively performed to model the extracted components. For consistency purpose, parameter specification of each individual forecasting tool follows a similar way to that when performing as a single benchmark. In the third step, the individual predicted results are aggregated into the one ensemble result as the final prediction.

First, the decomposition-ensemble learning models are

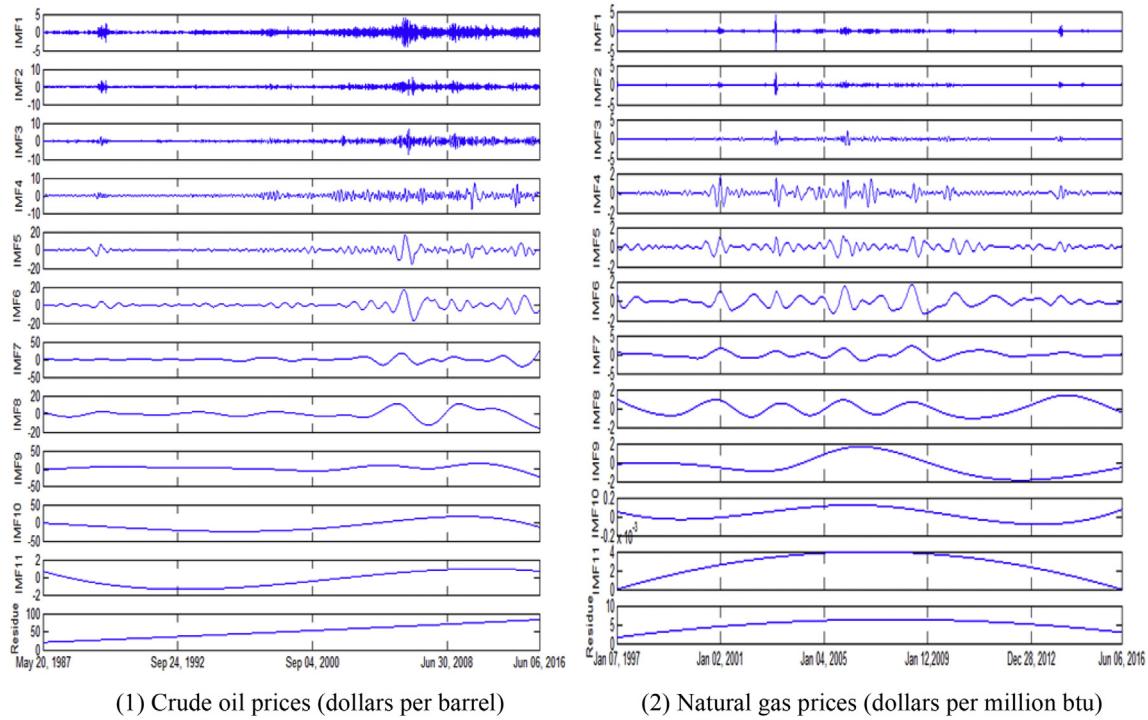


Fig. 6. Decomposition results via EEMD.

compared with their respective single models, to check out whether the concept of “decomposition and ensemble” improves prediction, as the *IR* results listed in Table 2. And to show the observed and predicted results of single models and ensemble models clearly, ten-point data were showed in Figs. 7 and 8, while all the predicted result of test data was showed in the supplementary (see Figs. S1–S2). Two important findings can be obtained from the comparison results. On the one hand, all the proposed decomposition-ensemble learning models using randomized individual forecasting algorithms outperform their respective original single forms without the decomposition strategy. In particular, the *IR* values of EEMD-ELM, EEMD-RVFL and EEMD-RKS over ELM, RVFL and RKS are all above 36.20%, which well verifies to the effectiveness of the promising “decomposition and ensemble” concept. On the other hand, in the cases of BPNN, a number of exceptions with negative *IRs* (in bold) exist. The main reasons can be attributed to the problem of parameter sensitivity, i.e., an

inappropriate parameter in any individual member may make a great difference in the final prediction. Therefore, the stable randomized algorithms, largely addressing the problem of parameter sensitivity, are strongly recommended to formulate decomposition-ensemble learning paradigms.

Then, the five models are compared with each other in terms of *MAPE*, *Dstat* and computational time, as the results presented in Figs. 9–11, respectively. Furthermore, Tables S7–S12 in Supplementary represent the corresponding results of the *t*-test. From these comparison results, one important conclusion can be obviously obtained that the proposed decomposition-ensemble models, EEMD-ELM, EEMD-RVFL and EEMD-RKS, are extremely efficient and fast techniques, in terms of both accuracy (see Figs. 9 and 10) and speed (see Fig. 11), and the *t*-test statistically confirms their superiority in most cases at the confidence level of 99% (see Tables S7–S12). The result finely supports that the idea of randomization, in terms of randomly fixed parameters or randomly

Table 2
IR of ensemble methods over the respective single benchmarks.

Benchmark	<i>IR</i> in <i>MAPE</i> for crude oil price forecasting					<i>IR</i> in <i>MAPE</i> for natural price forecasting				
	One	Two	Three	Four	Five	One	Two	Three	Four	Five
ELM	63.48%	64.50%	63.43%	66.77%	66.62%	33.32%	49.84%	48.72%	53.53%	53.44%
RVFL	58.59%	62.92%	62.63%	67.65%	68.35%	42.57%	55.11%	53.64%	58.32%	58.50%
RKS	36.23%	47.84%	52.24%	57.31%	58.02%	26.09%	47.29%	45.40%	50.16%	51.83%
BPNN	−137.86%	−49.21%	−26.05%	−7.89%	8.26%	32.21%	39.38%	43.88%	46.94%	45.25%
LSSVR	26.63%	30.22%	32.73%	34.42%	34.71%	28.90%	47.08%	48.48%	52.38%	51.71%
Horizon										
Benchmark	<i>IR</i> in <i>Dstat</i> for crude oil prices forecasting					<i>IR</i> in <i>Dstat</i> for natural price forecasting				
	One	Two	Three	Four	Five	One	Two	Three	Four	Five
ELM	71.51%	64.97%	55.44%	54.10%	50.28%	41.26%	42.38%	28.50%	26.25%	21.97%
RVFL	75.05%	62.62%	58.27%	57.03%	54.20%	46.13%	49.55%	34.92%	32.64%	28.53%
RKS	64.45%	56.97%	54.62%	51.88%	48.18%	49.04%	50.30%	34.75%	30.88%	28.30%
BPNN	27.61%	18.28%	21.06%	18.93%	21.00%	35.14%	35.33%	26.13%	22.85%	18.26%
LSSVR	14.56%	9.38%	9.12%	6.73%	7.06%	37.02%	42.17%	31.38%	28.77%	22.80%

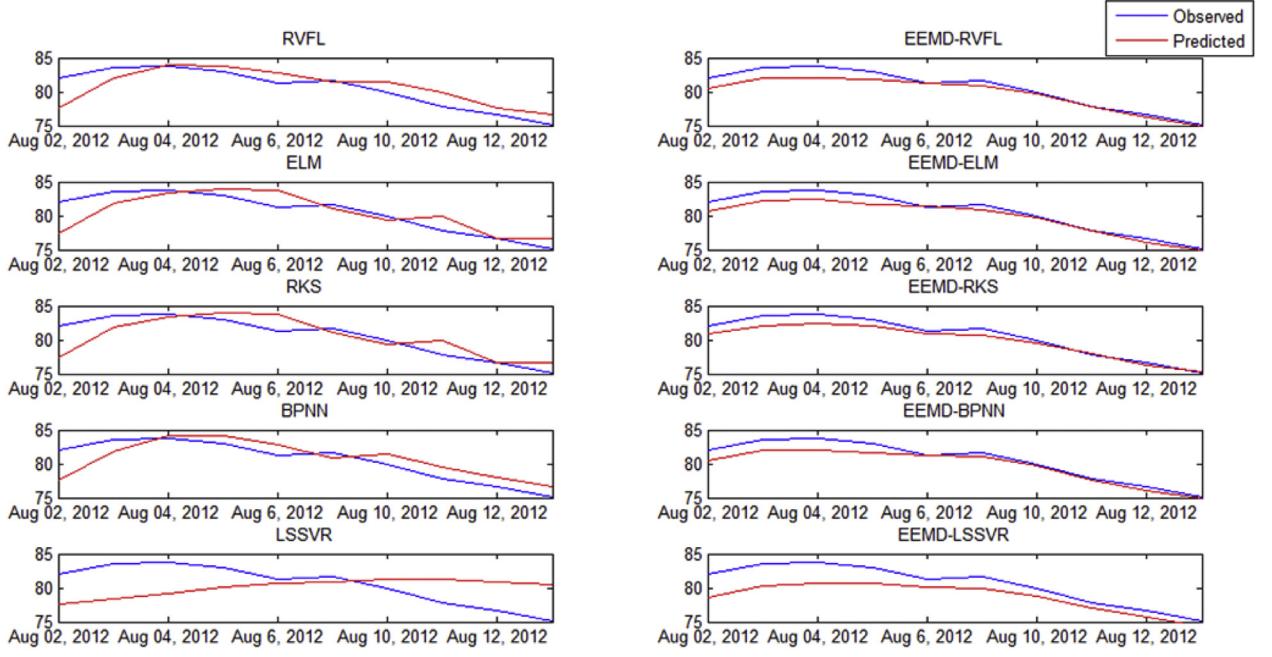


Fig. 7. Observed and predicted crude oil prices (dollars per barrel).

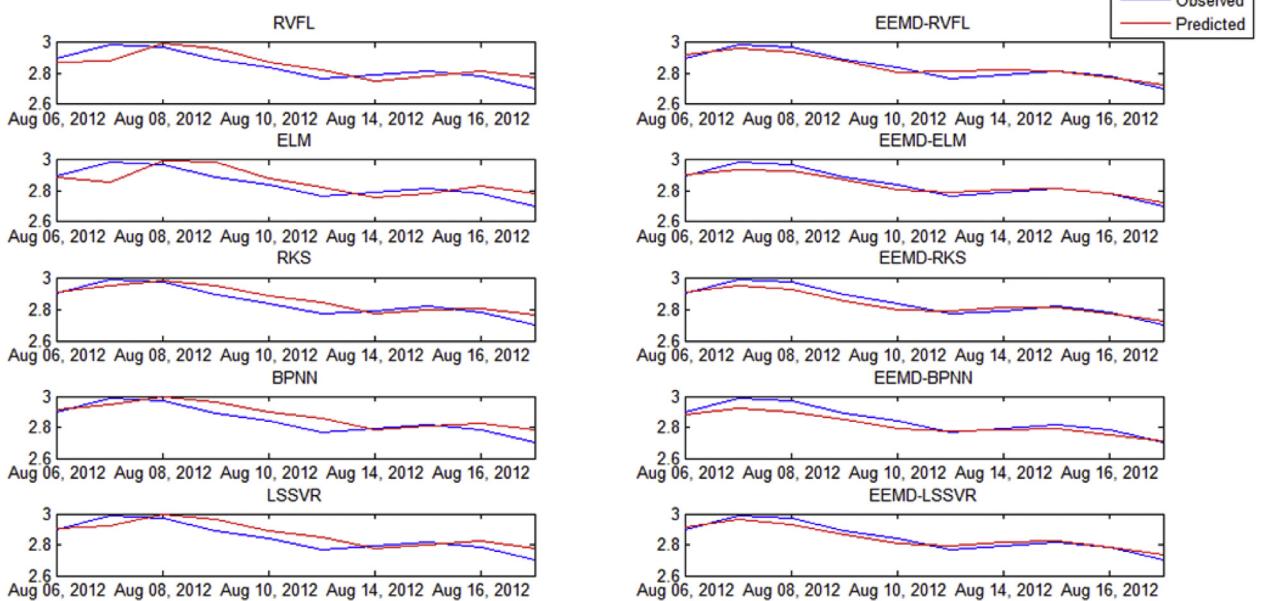


Fig. 8. Observed and predicted natural gas prices (dollars per million btu).

mapping features, can effectively address the double challenging problems (large computational burden and parameter sensitivity) in the existing decomposition-ensemble learning algorithms.

As for level accuracy in terms of *MAPE*, four important conclusions can be drawn from Fig. 9. First, the three proposed models, EEMD-ELM, EEMD-RVFL and EEMD-RKS, outperform all the listed ensemble benchmarks for oil price forecasting, while EEMD-RVFL ranks the first for natural gas prices, in terms of *MAPE*. The results imply the effectiveness of the randomization idea in improving the decomposition-ensemble methodology. Second, among the three randomized models, i.e., EEMD-ELM, EEMD-RVFL and EEMD-RKS, the former two neural networks-based models

defeat the last in all the cases, mainly due to the elaborative designs in neural networks. Third, among the three neural networks-based models, EEMD-ELM and EEMD-RVFL with randomization significantly beat EEMD-BPNN with iterative learning, repeatedly testifying to the effectiveness of randomization in improving decomposition-ensemble methodology. Fourth, as for the most effective model, EEMD-RVFL significantly excels in level accuracy, due to both the randomization idea (in terms of randomly fixed parameters) and the special design (neural networks with the input-output links) in RVFL.

For directional accuracy (measured by D_{stat}), the three proposed randomized models, EEMD-ELM, EEMD-RVFL and EEMD-RKS,

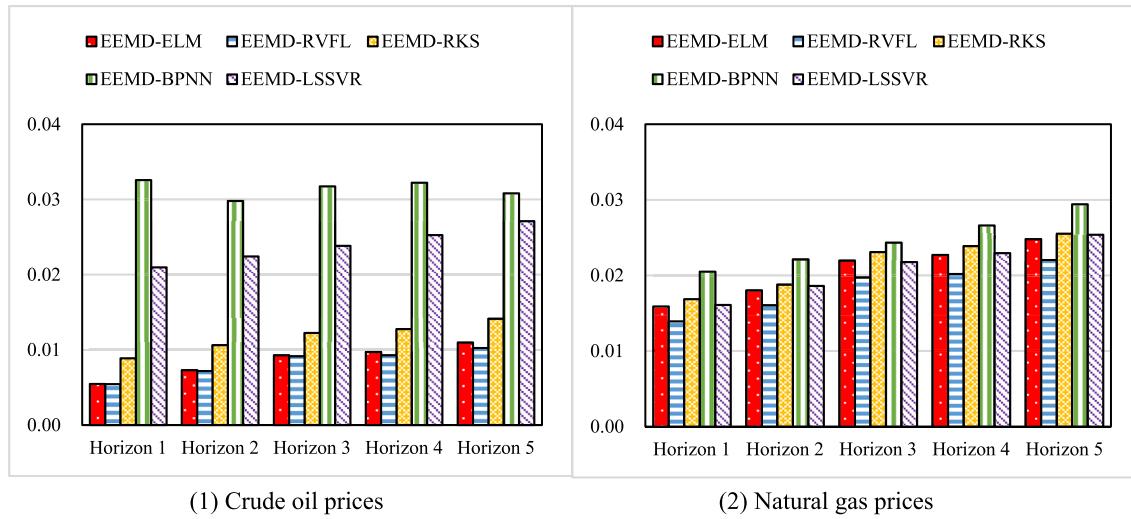


Fig. 9. Performance comparison of ensemble methods in terms of *MAPE* criterion.

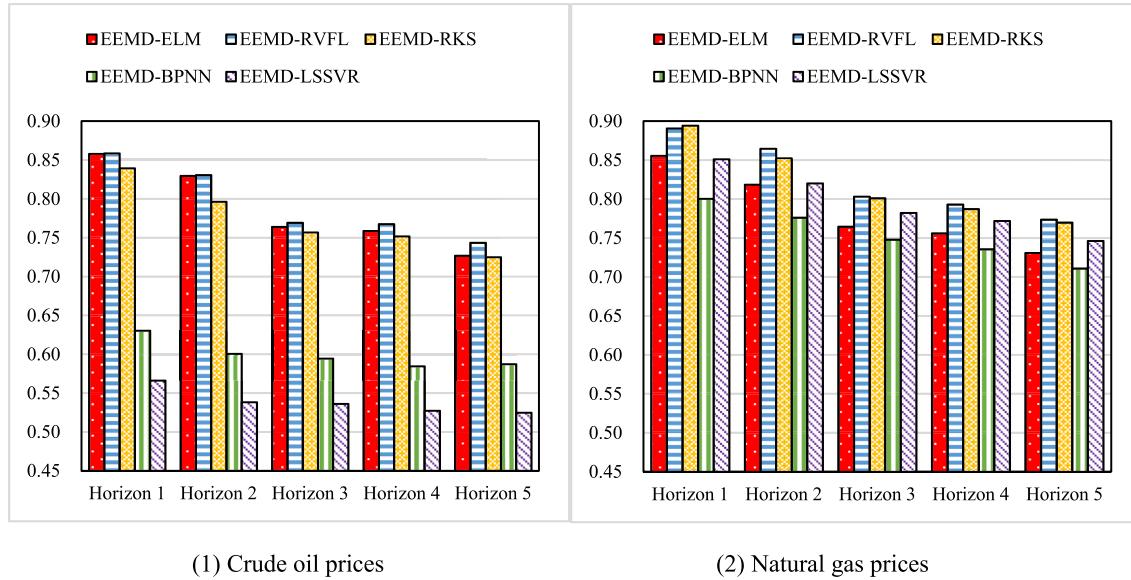


Fig. 10. Performance comparison of ensemble methods in terms of *Dstat* criterion.

produce satisfactory results with high directional hit ratios. Furthermore, three interesting findings can be obtained as follows. First, the “decomposition and ensemble” concept can be proved to improve directional accuracy, given that the decomposition-ensemble models excel their respective single models in most cases. Second, among the three proposed models, EEMD-RVFL performs the best in the predictions for crude oil prices and natural gas prices, which can be attributed to the special designs of neural networks with the input-output links. Therefore, with the randomization idea and special designs, EEMD-RVFL is proved to be the appropriate forecasting method in terms of directional accuracy. Third, EEMD-RKS significantly outperforms EEMD-LSSVR in all the cases, which shows that the randomly mapping features in RKS can remarkable improve the conventional decomposition-ensemble models with LSSVR in addressing sensitivity to kernel parameters. Similarly, the two randomized neural networks-based ensemble algorithms, EEMD-ELM and EEMD-RVFL (using randomly fixed parameters), significantly outperform the BPNN-based counterpart (using iteratively tuned ones), which repeatedly

supports the promising randomization idea.

Regarding for speed, three important results can be found from Fig. 11. First, the proposed methods, EEMD-ELM, EEMD-RVFL and EEMD-RKS (using randomized individual forecasting tools), are extremely faster than the conventional models of EEMD-BPNN and EEMD-LSSVR in terms of computational time. The result indicates that with randomization, in terms of randomly fixed parameters (for EEMD-ELM and EEMD-RVFL) and randomly mapping features (for EEMD-RKS), the major challenging problem of time-consuming in the existing conventional decomposition-ensemble learning paradigms can be utterly overcome. Second, when comparing the three randomized models, EEMD-ELM and EEMD-RVFL defeat EEMD-RKS with a much shorter computational time. Third, when comparing ensemble models with single models, the single models appear much faster, given that the EEMD decomposes the original data (one time series) into 12 individual series for both crude oil price and natural gas price, an additional large computational burden of conducting 12–1=11 more predictions arose. And all the steps of data decomposition, individual prediction for all the

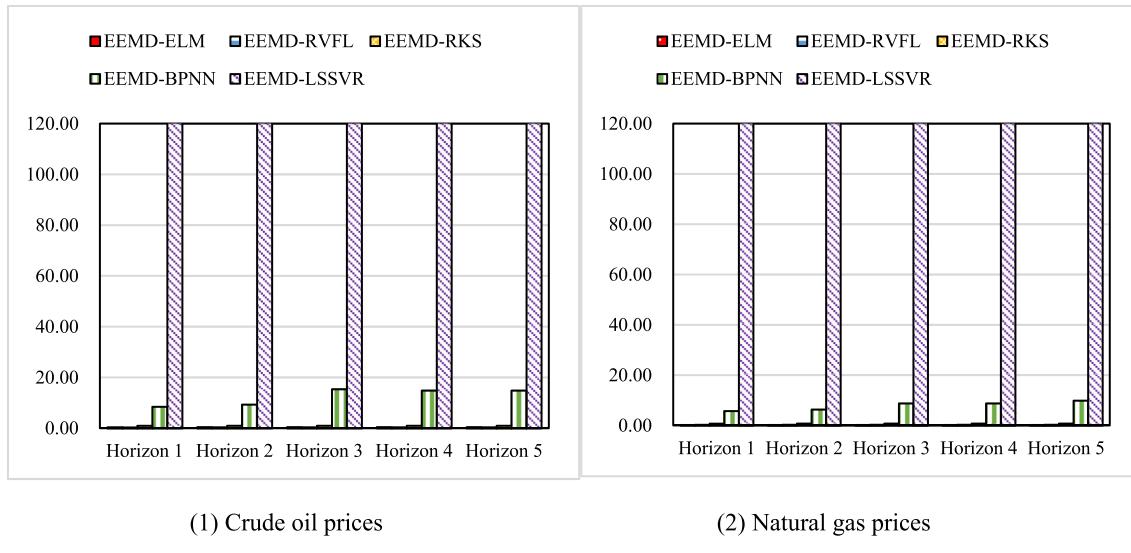


Fig. 11. Performance comparison of ensemble methods in terms of computational time (seconds).

extracted components and results ensemble need to consume additional computing time. Even under such a situation, the proposed three randomized decomposition-ensemble learning models are still much faster than most single benchmarks, especially LSSVR.

From Tables S7–S12, the *t*-test statistically confirms the two important conclusions. On the one hand, as for speed, when the proposed models (EEMD-ELM, EEMD-RVFL and EEMD-RKS) are treated as target models against the other ensemble counterparts, the *p*-values are all far smaller than the significance level of 1% (see Tables S9 and S12), statistically verifying that they are all extremely fast forecasting techniques. This implies that the time-consuming problem can well solved by the promising idea of randomization. On the other hand, as for accuracy, EEMD-RVFL can be shown to be the most powerful model in terms of both MAPE (Tables S7 and S10) and *Dstat* (Tables S8 and S11), demonstrating the effectiveness of the randomization idea (in terms of randomly fixed weights and bias in neural networks) and the special design (neural networks with the input-output links) in RVFL. Notably, the importance of the input-output links in effectively capturing the relationship between inputs and outputs and further improving prediction in terms of accuracy and speed has been previously observed in Refs. [15] and [22].

3.3. Summarizations

From the above discussions, we can deduce the following five interesting conclusions from the above empirical analysis.

(1) The proposed ensemble models, i.e., EEMD-ELM, EEMD-RVFL and EEMD-RKS (using randomized algorithms as individual forecasting tools), can be statistically proved to be extremely efficient and fast techniques, in terms of speed and accuracy.

(2) Among single benchmarks, the randomized algorithms, ELM, RVFL and RKS, can be shown to be not only effective in producing stable results but also extremely fast compared with traditional CI approaches.

(3) Decomposition-ensemble models significantly defeat their respective single models, which supports that the promising concept of “decomposition and ensemble” can substantially improve prediction.

(4) As for the most appropriate model, EEMD-RVFL, with both randomization (in terms of randomized fixed parameters) and

elaborative design (neural networks with input-output links), is strongly recommended for energy price forecasting.

The above conclusions further lead to (5) a typical decomposition-ensemble learning model, EEMD-RVFL, combining the powerful framework of “decomposition and ensemble” with the promising randomized neural networks of RVFL, can be used as an extremely efficient and fast technique for energy prices forecasting.

4. Conclusions

To solve the time-consuming issue in existing decomposition-ensemble learning paradigms, this study introduces the emerging randomized algorithms to develop an efficient and fast decomposition-ensemble learning methodology for energy price forecasting. In particular, the randomized algorithms use randomization, in terms of randomly fixed parameters, randomly mapping features, randomly generated samples or randomly selected variables rather than iteratively tuned ones, which effectively ensures an extremely fast learning speed and a better generalization performance. Compared with the existing studies, this study makes major contributions to the literature from the following two perspectives. First, to the best of our knowledge, besides ELM and RVFL, there were few decomposition-ensemble models using other potentially effective randomized algorithms as individual forecasting techniques. Accordingly, this study introduces the most popular randomized algorithms to formulate a set of decomposition-ensemble learning algorithms. Second, a thorough comparison is conducted to check whether the idea of randomization does improve the existing decomposition-ensemble learning methods, and to explore the most efficient and fast method for energy price forecasting.

For illustration and verification, the Brent crude oil spot price and the Henry Hub natural gas price are selected as the studying sample, and three important findings can be obtained. First, the proposed decomposition-ensemble learning models (using randomized algorithms as individual forecasting tools) are considered as a set of extremely efficient and fast techniques compared with other listed benchmarking models, in terms of speed and accuracy. Second, by using randomized individual forecasting tool, the double challenges of time-consuming and parameter sensitivity in the conventional decomposition-ensemble models (with CI individual

predictors) can be excellently overcome. Third, as for the most efficient and fast decomposition-ensemble learning methods, EEMD-RVFL, with both randomization (in terms of randomized fixed parameters) and elaborate designs (neural networks with input-output links), is strongly recommended for energy price forecasting.

Besides energy markets, the proposed decomposition-ensemble methodology using randomized algorithms can be also extended to other complex systems, to thoroughly test its generalization and universality. Furthermore, this study only considers univariate time series analysis, and extending into multivariate analysis covering various external factors is also an interesting direction for future research. Finally, besides the three considered randomized algorithms, other even more efficient and fast forecasting techniques, such as radial basis function network, recurrent neural networks, deep neural networks and convolutional neural networks [14] can be employed. It is worth noticing that the technical novelty of this work is to introduce randomized algorithm such as RVFL network, ELM, RKS into the “decomposition and ensemble” framework and propose a fast decomposition-ensemble learning model for energy price forecasting. However, making a substantial innovation over randomized techniques remains a tough but important task. We will look into these interesting issues in the near future.

Acknowledgements

This work is supported by grants from the National Natural Science Foundation of China (NSFC Nos. 71622011, 71433001 and 71301006), the National Program for Support of Top Notch Young Professionals, and Beijing Advanced Innovation Center for Soft Matter Science and Engineering.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.energy.2018.05.146>.

References

- [1] Yu L, Zhao Y, Tang L. Ensemble forecasting for complex time series using sparse representation and neural networks. *J Forecast* 2017;36(2):122–38.
- [2] Monge M, Gil-Alana LA, de Gracia FP. Crude oil price behaviour before and after military conflicts and geopolitical events. *Energy* 2017;120:79–91.
- [3] Jones DW, Leiby PN, Paik IK. Oil price shocks and the macroeconomy: what has been learned since 1996. *Energy J* 2004;25(2):1–32.
- [4] Zhao Y, Li J, Yu L. A deep learning ensemble approach for crude oil price forecasting. *Energy Econ* 2017;66:9–16.
- [5] Lahmiri S. Comparing variational and empirical mode decomposition in forecasting day-ahead energy prices. *IEEE Syst J* 2017;11(3):1907–10.
- [6] Jianwei E, Bao Y, Ye J. Crude oil price analysis and forecasting based on variational mode decomposition and independent component analysis. *Phys Stat Mech Appl* 2017;484:412–27.
- [7] Van Goor H, Scholtens B. Modeling natural gas price volatility: the case of the UK gas market. *Energy* 2014;72:126–34.
- [8] Murat A, Tokat E. Forecasting oil price movements with crack spread futures. *Energy Econ* 2009;31(1):85–90.
- [9] Chiu FP, Hsu CS, Ho A, Chen CC. Modeling the price relationships between crude oil, energy crops and biofuels. *Energy* 2016;109:845–57.
- [10] Lanza A, Manera M, Giovannini M. Modeling and forecasting cointegrated relationships among heavy oil and product prices. *Energy Econ* 2005;27(6):831–48.
- [11] Ren Y, Suganthan PN, Srikanth N. A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods. *IEEE Trans Sustain Energy* 2015;6(1):236–44.
- [12] Zhu B, Han D, Wang P, Wu Z. Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. *Appl Energy* 2017;191:521–30.
- [13] Shayeghi H, Ghasemi A, Moradzadeh M, Nooshyar M. Day-ahead electricity price forecasting using WPT, GMI and modified LSSVM-based S-OLABC algorithm. *Soft Comput* 2017;21(2):525–41.
- [14] Zhang L, Suganthan P. A survey of randomized algorithms for training neural networks. *Inf Sci* 2016;364:146–55.
- [15] Tang L, Wu Y, Yu L. A non-iterative decomposition-ensemble learning paradigm using RVFL network for crude oil price forecasting. *Appl Soft Comput* 2017. <https://doi.org/10.1016/j.asoc.2017.02.013>.
- [16] Qiu X, Suganthan PN, Amaralunga GAJ. Short-term electricity price forecasting with empirical mode decomposition based ensemble kernel machines. *Procedia Comput Sci* 2017;108:1308–17.
- [17] Nguyen HT, Nabney IT. Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. *Energy* 2010;35(9):3674–85.
- [18] Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. *Neurocomputing* 2006;70(1):489–501.
- [19] Pao YH, Park GH, Sobajic DJ. Learning and generalization characteristics of the random vector functional-link net. *Neurocomputing* 1994;6(2):163–80.
- [20] Rahimi A, Recht B. Random features for large-scale kernel machines. In: Proc. 21th annual neural information processing systems conference, Vancouver, B.C., Canada; 2007. p. 1177–84.
- [21] Breiman L. Random forests. *Mach Learn* 2001;45(1):5–32.
- [22] Ren Y, Suganthan PN, Srikanth N, Amaralunga G. Random vector functional link network for short-term electricity load demand forecasting. *Inf Sci* 2016;367:1078–93.
- [23] Nedellec R, Cugliari J, Goude Y. GEFCom2012: electric load forecasting and backcasting with semi-parametric models. *Int J Forecast* 2014;30(2):375–81.
- [24] Zamo M, Mestr O, Arbogast P, Pannekoucke O. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: deterministic forecast of hourly production. *Sol Energy* 2014;105:792–803.
- [25] Wang Z, Yoon S, Xie S, Lu Y, Park D. A high accuracy pedestrian detection system combining a cascade AdaBoost detector and random vector functional-link net. *Sci World J* 2014;2014:1–7.
- [26] Ibarra-Berastegi G, Saénz J, Esnaola G, Ezcurra A, Ulatzia A. Short-term forecasting of the wave energy flux: analogues, random forests, and physics-based models. *Ocean Eng* 2015;104:530–9.
- [27] Herrera M, Torgo L, Izquierdo J, Pérez-García R. Predictive models for forecasting hourly urban water demand. *J Hydrol* 2010;387(1):141–50.
- [28] Zheng H, Kusiak A. Prediction of wind farm power ramp rates: a data-mining approach. *J Sol Energy Eng* 2009;131(3). 031011.
- [29] Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, Chica-Rivas M. Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol Rev* 2015;71:804–18.
- [30] Igelnik B, Pao YH. Stochastic choice of basis functions in adaptive function approximation and the functional-link net. *IEEE Trans Neural Network* 1995;6(6):1320–9.
- [31] Huang GB, Wang DH, Lan Y. Extreme learning machines: a survey. *Inter J Machine Learn Cyber* 2011;2(2):107–22.
- [32] Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 1996;49(11):1225–31.
- [33] Zhang JR, Zhang J, Lok TM, Lyu MR. A hybrid particle swarm optimization—back-propagation algorithm for feedforward neural network training. *Appl Math Comput* 2007;185(2):1026–37.
- [34] Scardapane S, Wang D. Randomness in neural networks: an overview, wiley interdisciplinary reviews. *Data Min Knowl Discov* 2017;7(2). <https://doi.org/10.1002/widm.1200>.
- [35] Tang L, Dai W, Yu L, Wang S. A novel CEEMD-based EELM ensemble learning paradigm for crude oil price forecasting. *Int J Inf Technol Decis Making* 2015;14(01):141–69.
- [36] Wang D, Wei S, Luo H, Yue C, Grunder O. A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine. *Sci Total Environ* 2017;580:719–33.
- [37] Lu C, Shao Y. Forecasting computer products sales by integrating ensemble empirical mode decomposition and extreme learning machine. *Math Probl Eng* 2012;2012:1–15.
- [38] Shrivastava N, Panigrahi B. A hybrid wavelet-ELM based short term price forecasting for electricity markets. *Int J Electr Power Energy Syst* 2014;55:41–50.
- [39] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, in: proceeding of Royal Society of London A: Mathematical, Phys Eng Sci 1998;454(1971):903–95.
- [40] Wu Z, Huang N. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv Adapt Data Anal* 2009;1(01):1–41.
- [41] Banerjee K. Generalized inverse of matrices and its applications. *Technometrics* 1973;15(1):197.
- [42] Fletcher R, Reeves C. Function minimization by conjugate gradients. *Comput J* 1964;7(2):149–54.
- [43] Rudin W. Fourier analysis on groups. New York, USA: John Wiley & Sons; 2011.
- [44] Xie G, Wang S, Zhao Y, Lai KK. Hybrid approaches based on LSSVR model for container throughput forecasting: a comparative study. *Appl Soft Comput* 2013;13(5):2232–41.
- [45] Cressie N, Whitford H. How to use the two sample t-Test. *Biom J* 1986;28(2):131–48.