Energy Conversion and Management 153 (2017) 589–602

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

Multi-step ahead wind speed forecasting using a hybrid model based on two-stage decomposition technique and AdaBoost-extreme learning machine

Tian Peng, Jianzhong Zhou*, Chu Zhang, Yang Zheng

School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
Hubei Key Laboratory of Digital Valley Science and Technology, Wuhan 430074, China

ARTICLE INFO

Keywords:
Wind speed forecasting
Extreme learning machine
Complementary ensemble empirical mode decomposition with adaptive noise
Variational mode decomposition
AdaBoost.RT

ABSTRACT

Accurate and reliable wind speed forecasting is of great importance to wind power generation and integration. This paper shows the development of a novel hybrid model for wind speed forecasting and demonstrates its efficiency. In the proposed hybrid model, a two-stage decomposition algorithm combining the complementary ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and the variational mode decomposition (VMD) is introduced to deal with the nonlinearity of wind speed time series. The CEEMDAN is exploited to decompose the original wind speed series into a series of intrinsic mode functions (IMFs) with different frequencies. The VMD is employed to re-decompose the IMF with the highest frequency using CEEMDAN into a number of modes successively. And then an improved AdaBoost.RT algorithm is coupled with extreme learning machine (ELM) to forecast all the decomposed modes using CEEMDAN and VMD. The AdaBoost.RT technique is improved by updating the threshold value self-adaptively. Finally, the forecasting value of the original wind speed series is obtained by adding up the forecasting results of all the decomposed modes. The proposed hybrid model has been applied to four datasets of wind speed observations for one-, two- and three-step ahead forecasting. The proposed model is compared with the non-denoising methods, namely the Bagging method, the partial least squares (PLS) model, the back propagation (BP) neural network, the support vector machine (SVM), the ELM and the AdaBoost-ELM models as well as several kinds of CEEMDAN-based and two-stage decomposition based methods. Results obtained from this study indicate that the proposed hybrid model can capture the nonlinear characteristics of wind speed time series and thus provide more accurate forecasting results compared with the other methods.

1. Introduction

Wind energy has been one of the most economically, abundant and environmental friendly renewable energy over the world. Integrating wind power to the power grid is difficult for its intermittent nature. According to [1], the fluctuations of wind speed are amplified by the power of three when generating wind power from a wind turbine. As a result, enhancing wind speed predictability is of great significance to various aspects of wind power generating system such as unit commitment decision, dispatch planning, maintenance scheduling and regulation.

Physically-based models, mainly including a few kinds of numeric weather prediction (NWP) models, have been widely used for wind speed forecasting in several studies [2–3]. Developing NWP models for wind speed forecasting requires collecting meteorological data obtained from meteorological stations or satellites. These data include humidity, terrain temperature, pressure, wind speed and direction etc., which makes preparing data for NWP models a difficult and time-consuming task. What's more, the modeling process of NWP models is operated by solving complex mathematical models, thus the NWP models are usually rendered on supercomputers as the modeling process needs lots of computation [2]. Al-Yahyai et al. [4] gave a general review of NWP models and how they overcome the limitations for classical wind energy measurement.

Unlike the physically-based models, statistical models and artificial intelligence (AI) models can obtain wind speed predictions based on the mapping relationship between the inputs and outputs in the fitting or training process. The modeling process is simpler. These models are usually constructed using less meteorological data and the model inputs are more accessible. Statistical models, including autoregressive moving average (ARMA) [5] models, auto regressive integrated moving average (ARIMA) models, seasonal- [6] or fractional-ARIMA [7] models, and ARIMA with exogenous terms (ARIMAX) [8] models have been widely exploited. These models are linear models and their ability to predict highly nonlinear or non-stationary time series is limited. AI

* Corresponding author at: School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China.
E-mail addresses: husthydropt@126.com (T. Peng), jz.zhou@mail.hust.edu.cn (J. Zhou).

http://dx.doi.org/10.1016/j.enconman.2017.10.021
Received 29 August 2017; Received in revised form 4 October 2017; Accepted 7 October 2017
Available online 18 October 2017

models, in the other way, are non-linear black-box models and can explain the underlying relationships between input data and output wind speed. In most cases, AI models can provide better prediction results than statistical models if the nonlinear characteristics of the time series are prominent [9]. Artificial neural networks (ANN), as a major AI tool, have been developed as an effective tool for wind speed forecasting over the last decades. Traditional ANN models mainly include back propagation neural network (BPNN), radial basis function (RBF) neural network, generalized regression neural network (GRNN), Elman neural network and multi-layer feed forward (MLF) network etc. The extreme learning machine (ELM), which was first introduced by Huang et al. [10], is a new type of single hidden layer feed-forward network (SLFN) of which the input weights and hidden layer biases do not need to be tuned. The ELM is famous for its fast speed and has been exploited for wind speed forecasting in various studies in recent years. Salcedo-Sanz et al. [11] presented a hybrid approach which combined the coral reefs optimization algorithm (CRO) with ELM for short-term wind speed forecasting. The CRO approach is used to select appropriate input variables from several meteorological predictive variables from a physical model. Wang et al. [6] combined ELM with Ljung-Box Q-test and SARIMA for mean monthly and mean daily wind speed forecasting. The SARIMA model is used to forecast the wind speed residue to be added to the forecasting result of the ELM model. The SARIMA model turned out to have improved the forecasting accuracy. Abdoos [12] calculated the forecasting errors of ANN, support vector machine (SVM), wavelet neural network (WNN) and ELM for 10-min wind power forecasting. It can be seen from the results that the ELM model can achieve better forecasting results compared with the other models.

In addition to the use of AI models alone, it can be seen from recent studies that data-preprocessing techniques have been a new tendency to deal with the nonlinear and nonstationary characteristics of wind speed signal and thus to enhance the forecasting accuracy. Liu et al. [13] presented a hybrid approach which combined wavelet transform (WT), genetic algorithm (GA) and SVM for 0.5 h wind speed forecasting. Meng et al. [14] came up with wavelet packet decomposition (WPD) to improve the forecasting accuracy of the BP models optimized by the crisscross optimization (CSO) algorithm for predicting short-term wind speed at 1 h interval up to 5 h. Ren et al. [15] compared the performance of empirical mode decomposition (EMD) and its improved versions including ensemble EMD (EEMD), complementary EEMD (CEEMD) and complete EEMD with adaptive noise (CEEMDAN) in decomposing a complex time series into several intrinsic mode functions (IMFs) and a residue. These decomposing techniques are coupled with SVR and ANN to construct the hybrid models for forecasting 1, 3 and 5 h ahead wind speed. The experimental prediction results showed that the proposed CEEMDAN-SVR model obtained the best results. Du et al. [16] introduced CEEMD to decompose the original wind speed series into a number of sub-layers. The Elman neural network model optimized by the multi-objective ant lion optimization algorithm (MOALO) is exploited as the predictor. Zhang et al. [17] developed a hybrid model for short-term wind speed forecasting by combining optimized variational mode decomposition (OVMD), hybrid backtracking search algorithm (HBSA) and ELM. In the proposed model, the OVMD method is employed to eliminate the redundant noises of the wind speed signal.

Although there have been a lot of achievements in the application of single decomposition techniques coupled with AI models for time series forecasting, while because the single decomposition techniques often cannot thoroughly deal with the nonlinear and nonstationary characteristics of wind speed signal, there still exists much room to improve in the process of modeling. Guo et al. [18] proposed a modified EMD-FNN (feed-forward neural network) model for wind speed forecasting. The IMF1 component is discarded because it is the most disorder and unsystematic part of the wind speed series and has little regularity. The developed model turned out to perform better than the basic FNN and the unmodified EMD-FNN models. Yu et al. [19] coupled singular spectrum analysis (SSA) with Elman neural network and EMD, EEMD and CEEMDAN, respectively to forecast 1 h-average wind speed series. The SSA method is exploited to extract the trend of IMF1 component. The results showed that through the retreatment of IMF1 using SSA, the performances of the hybrid models with single decomposition technique have been improved significantly. Liu et al. [20] proposed a new hybrid approach based on the secondary decomposition algorithm (SDA) and the Elman neural network for wind speed forecasting. The SDA combined WPD with FEEMD to decrease the non-stationarity of the wind speed signal. The FEEMD is employed to re-decompose the detailed components generated by WPD into a few IMFs. Wang et al. [21] presented a hybrid two-layer decomposition ensemble model for multi-step ahead electricity price forecasting. The VMD is applied to further decompose the high frequency IMFs generated by FEEMD.

Recent studies show that a new trend to improve the generalization ability of a single ANN model is to take advantage of ensemble techniques by utilizing multiple ANNs. In an ANN ensemble model, a number of ANNs trained for the same purpose as a single ANN are combined to generate a unique output [22]. Boosting is one of the most widely used method in regression which attempts to boost the performance of a weak learner. Boosting approach has been used successfully in the area of forecasting over the last decade such as pooled flood frequency analysis [22], estimation of ice thickness on lakes [23], drought prediction [24] and wind speed forecasting [25]. It has been shown that better results can be achieved by combining forecasts than by choosing the best one [23].

In this study, a combination of the improved AdaBoost.RT algorithm with the ELM is proposed for wind speed forecasting. In the AdaBoost-ELM (AELM) model for time series forecasting, the AdaBoost.RT algorithm is exploited to generate a sequence of ELM models and each subsequent ELM concentrates more on the points which are not well predicted by the previous one [23]. In addition, a two-stage decomposition algorithm combining CEEMDAN and VMD is introduced to filter the noises of the original wind speed time series. In the two-stage decomposition technique, the VMD is employed to re-decompose the component with the highest frequency generated by CEEMDAN. The reminder of the paper is arranged as follows: Section 2 reviews the CEEMDAN, VMD, ELM and AELM methodologies. Section 3 gives a detailed description of the proposed CEEMDAN-VMD-AELM (simplified as CVAELM) model. Section 4 describes the construction and development of the models for wind speed forecasting. Section 5 compares the performance of the proposed model with several other models on four different wind speed datasets. And finally Section 6 presents the conclusions obtained from the experimental results.

2. Methodologies

2.1. Complementary ensemble empirical mode decomposition with adaptive noise (CEEMDAN)

Empirical mode decomposition (EMD) developed by Huang et al. [26] is an adaptive data analysis method dealing with non-linear and non-stationary signal. The main idea of EMD is to decompose the complicated signal into a finite and often small number of intrinsic mode functions (IMFs) based on the local characteristics of the signal [26]. Based on the EMD method, a noise-assisted data analysis method called ensemble EMD (EEMD) was proposed by Wu and Huang [27] to overcome the drawback of mode mixing of EMD. In EEMD, the true IMF components are defined as the mean of an ensemble of trials to decompose the noise-added signals using EMD. The added white Gaussian noise signal will provide a relatively uniform reference scale distribution to facilitate the EMD process and thus alleviate the mode mixing of EMD. However, after a finite number of iterations, the reconstruction error caused by the white noise may not be eliminated. The accuracy of the reconstructed time series will be influenced and thus affect the forecasting accuracy. Though the reconstructed error can be decreased by increasing the number of iterations, the computational cost is high.

To lower the reconstruction error as well as the computational cost, a complete EEMD with adaptive noise (CEEMDAN) was developed by Torres et al. [28]. The decomposition process of CEEMDAN is given as follows:

Step 1: Generate a number of noise-added series:

$$X^i(t) = X(t) + p_0\omega^i(t) \tag{1}$$

where $X(t)$ denotes the original signal, $\omega^i(t)(i = 1,...,I)$ denotes different white Gaussian noise with $N(0,1)$ and p_0 is a noise coefficient which controls the signal-to-noise ratio.

Step 2: Decompose each $X^i(t)$ using EMD to get the corresponding first modes $IMF_1^i(t)$. Calculate the first mode of CEEMDAN by averaging all the modes:

$$\overline{IMF_1^i(t)} = \frac{1}{I}\sum_{i=1}^{I} IMF_1^i(t) \tag{2}$$

Step 3: Calculate the first residue $r_1(t) = X(t)-\overline{IMF_1^i(t)}$ and decompose the noise-added residue $r_1(t) + p_1E_1(w^i(t))$ to obtain the second mode:

$$\overline{IMF_2^i(t)} = \frac{1}{I}\sum_{i=1}^{I} E_1(r_1(t) + p_1E_1(w^i(t))) \tag{3}$$

where $E_1(\bullet)$ is a function to produce the first mode of EMD.

Step 4: Repeat to obtain the other modes until the residue component does not have at least two extreme values.

2.2. Variational mode decomposition (VMD)

Variational mode decomposition (VMD) developed by Dragomiretskiy and Zosso in 2014 is a newly non-recursive signal processing technique which attempts to decompose an one dimensional signal$f(t)$ into K independent modes$u_k(k = 1,2,...,K)$with limited bandwidth in spectral domain [29]. It is assumed that each variational mode u_k has a center pulsation (frequency) ω_k which is determined along with the decomposition process. To access the bandwidth of each u_k, Dragomiretskiy proposed the method below in [29]: (1) compute the corresponding analytical signal using Hilbert transform to obtain the unilateral frequency spectrum for each subseries u_k; (2) translate each mode's frequency spectrum to baseband regions using an exponential tuned to the respective estimated frequency; (3) estimate the bandwidth of each mode u_k through the H^1 Gaussian smoothness of the demodulated signal.

The constrained variational problem can be presented as follows:

$$\min_{u_k,w_k}\left\{\sum_k \left\| \partial_t\left[\left(\delta(t) + \frac{j}{\pi t}\right)*u_k(t)\right]e^{-jw_kt}\right\|_2^2\right\}$$
$$s.\ t.\quad \sum_k u_k = f(t) \tag{4}$$

where $\partial(t)$ denotes the Dirac distribution and * denotes convolution.

To translate the above optimization problem into an unconstraint one, the quadratic penalty term and Lagrangian multipliers are introduced:

$$L(u_k.\ w_k,\lambda) = \alpha\sum_k \left\| \partial_t\left[\left(\delta(t) + \frac{j}{\pi t}\right)*u_k(t)\right]e^{-jw_kt}\right\|_2^2 + \left\| f-\sum_k u_k\right\|_2^2 + \left\langle \lambda, f-\sum_k u_k\right\rangle \tag{5}$$

To solve the problem, the alternate direction method of multipliers (ADMM) is exploited [29]. The optimization of Eq. (5) can be illustrated as follows:

(1) Minimization of u_k:

$$\hat{u}_k^{n+1} = \frac{\hat{f}(\omega)-\sum_{i\neq k}\hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega-\omega_k)^2} \tag{6}$$

(2) Minimization of ω_k:

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega|\hat{u}_k(\omega)|^2d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2d\omega} \tag{7}$$

where $\hat{u}_i(\omega),\hat{u}_k^{n+1}\hat{f}(\omega),\hat{\lambda}(\omega)$ denote the Fourier transform of $u_i(\omega),u_k^{n+1}f(\omega),\lambda(\omega)$, respectively, and n denotes the number of iterations.

Readers may refer to [29] for more information about VMD.

2.3. Extreme learning Machine (ELM)

Extreme Learning Machine (ELM) introduced by Huang et al. [10] is a type of single hidden layer feed-forward network (SLFN) of which the input weights and hidden layer biases do not need to be tuned. After the input weights and hidden biases are randomly assigned, the output weights can be calculated by simply applying generalized inverse operation to the hidden layer output matrix [10]. Assuming a set of N samples, i.e. (x_t,y_t) for $t = 1,2,...,N$, where $\mathrm{x}_t = [x_{t1},x_{t2},...,x_{tn}]^T \in \mathrm{R}^n$ and $\mathrm{y}_t = [y_{t1},y_{t2},...,y_{tm}]^T \in \mathrm{R}^m$, the output of an ELM with L hidden neurons can be expressed as:

$$f_L = \sum_{i=1}^{L}\beta_i g(\mathrm{w}_i.\ \mathrm{x}_t +_i) = \mathrm{o}_t\ t = 1,...,N \tag{8}$$

where $\mathrm{w}_i = [w_{1i},w_{2i},...,w_{ni}]$ is the weight vector that connects the n input neurons with the ith hidden neuron; $\beta_i = [\beta_{i1},\beta_{i2},...,\beta_{im}]^T$ is the weight vector that connects the ith hidden neuron with the m output vectors, b_i is the bias, g is the activation function and $\mathrm{o}_t = [o_{t1,}o_{t2},...,o_{tm}]^T$ is the output vector of the network.

The purpose of ELM is to find the appropriate β_i,w_i,b_i such that $\sum_{t=1}^{N}\|\mathrm{o}_t-\mathrm{y}_t\| = 0$, i.e.,

$$f_L = \sum_{i=1}^{L}\beta_i g(\mathrm{w}_i.\ \mathrm{x}_t + b_i) = \mathrm{y}_t\quad t = 1,...,N \tag{9}$$

Eq. (9) can be simplified as:

$$\mathrm{H}\beta = \mathrm{T} \tag{10}$$

where

$$\mathrm{H}=\begin{bmatrix}\mathrm{h}(\mathrm{x}_1)\\ \vdots\\ \mathrm{h}(\mathrm{x}_N)\end{bmatrix} = \begin{bmatrix} g(w_1.\ \mathrm{x}_1 + b_1) & \cdots & g(w_L.\ \mathrm{x}_1 + b_L)\\ \vdots & \cdots & \vdots\\ g(w_1.\ \mathrm{x}_N + b_1) & \cdots & g(w_L.\ \mathrm{x}_N + b_L)\end{bmatrix}_{N\times L} \tag{11}$$

$$\beta = \begin{bmatrix}\beta_1^T\\ \vdots\\ \beta_L^T\end{bmatrix}_{L\times m}\quad \text{and}\quad \mathrm{T}=\begin{bmatrix}\mathrm{t}_1^T\\ \vdots\\ \mathrm{t}_N^T\end{bmatrix}_{N\times m} \tag{12}$$

and $\boldsymbol{H}$ is called the hidden layer output matrix of ELM. The coefficient β of the ELM can be obtained by finding the least-square solution of the following linear equation:

$$\|\mathrm{H}\hat{\beta}-\mathrm{T}\| = \|HH^\dagger T-T\| = \min_\beta\|\mathrm{H}\beta-\mathrm{T}\| \tag{13}$$

The special solution can be written as:

$$\hat{\beta} = \mathrm{H}^\dagger\mathrm{T} \tag{14}$$

where $\mathrm{H}^\dagger$ is the Moore-Penrose generalized inverse of the hidden layer output matrix [10].

2.4. The AdaBoost-ELM wind speed forecasting model

AdaBoost developed by Freund and Schapire is a machine learning meta-algorithm which was originally designed for classification problems [30] and then extended to regression ones [31]. The AdaBoost.RT developed by Solomatine [31] is a variation of AdaBoost.R [30] for regression problems. The AdaBoost.RT can be applied to many other types of learning algorithms (weak learners) to improve their performance. In this study, the AdaBoost.RT algorithm is applied as an ensemble method to enhance the forecasting ability of the ELM network. When in conjunction with ELM, AdaBoost.RT improve the performance of the algorithm by generating a sequence of ELM models such that each subsequent ELM model concentrates more on the training cases that do not perform well. The ELM is famous for its fast learning speed. However, due to the random initiation of the weights and biases, the result of the ELM model is instable. The combination of AdaBoost.RT and ELM can take advantage of the fast convergence speed of ELM as well as overcome the drawback of instability.

The specific process of the AdaBoost-ELM (AELM) model is presented as follows:

Step 1: Give a dataset with N training samples $[x_1,y_1],...,[x_N,y_N]$, x and y denote the input and output variables, respectively; Initialize the parameters of the AdaBoost.RT approach such as the threshold ϕ, the maximum iteration number K.

Step 2: Initialize the weight of each element: $D_k(i) = 1/N$; set the initial error $\varepsilon_k = 0$; let the iteration number $k = 1$.

Step 3: Train the ELM network based on the weight distribution such that $f_k(x) = y$.

Step 4: Compute the error of each sample according to:

$$E_k(i) = \left|\frac{f_k(x_i)-y_i}{y_i}\right| \quad \text{over all } i \tag{15}$$

where $f_k(x_i)$ denotes the output of the network for x_i.

And then compute the training error for the network according to:

$$\varepsilon_k = \sum_{i:E_k(i)>\phi} D_k(i) \tag{16}$$

Step 5: Compute $\beta^k = \varepsilon_k^n$, $n = 1,2$ *or* 3, n is adopted as 1 in this study.

Step 6: Update the weight distribution of the training samples according to:

$$D_{k+1}(i) = \frac{D_k(i)}{Z_k} \times \begin{cases} \beta_k, & E_k(i) \leqslant \phi \\ 1, & \text{otherwise} \end{cases} \tag{17}$$

where Z_k is a normalization factor such that $\sum_{i=1}^{N} D_k(x_i) = 1$.Step 7: Set $k = k + 1$ and repeat steps 3–6 till the maximum iteration number and output the final result $f_{fin}(x)$:

$$f_{fin}(x) = \sum_k \left(\log\frac{1}{\beta_k} f_k(x) / \sum_k \log\frac{1}{\beta_k}\right), \; k = 1,...,K \tag{18}$$

It can be seen from the above description that in the implementation of a boosted ELM ensemble model using AdaBoost.RT, the threshold ϕ needs to be optimally selected. Experiments with the AdaBoost.RT have shown that the performance of the committee machine is sensitive to ϕ [32]. If the ϕ is too low, then it is generally difficult to get a sufficient number of prediction samples. Too high value of ϕ will result in overfitting problems in the other hand. In this study, a self-adaptive method for updating the value of ϕ is employed by adjusting it according to the root mean square error (RMSE) of the training results [33]. The process of the self-adaptive method for updating ϕ is presented as follows:

Step 1: Calculate the RMSE of the training results in each iteration.

$$e_k = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (\hat{y}_i - y_i)^2} \tag{19}$$

Step 2: Updating ϕ_k according to the following equation such that the value of ϕ_k is increased along with the increasing of e_k.

$$\begin{cases} \phi_{k+1} = \phi_k(1-\lambda) & e_t < e_{t-1} \\ \phi_{k+1} = \phi_k(1+\lambda) & e_t > e_{t-1} \end{cases} \tag{20}$$

where λ is relative to the change of e_t:

$$\lambda = \frac{1}{2}\left|\frac{e_t - e_{t-1}}{e_t}\right| \tag{21}$$

The framework of the AELM method is presented in Fig.1.

2.5. Model performance evaluation

It is important to apply multiple error measure indices when evaluating the forecasting ability of the developed models. Three statistical error measures, the RMSE, the mean absolute error (MAE) and the

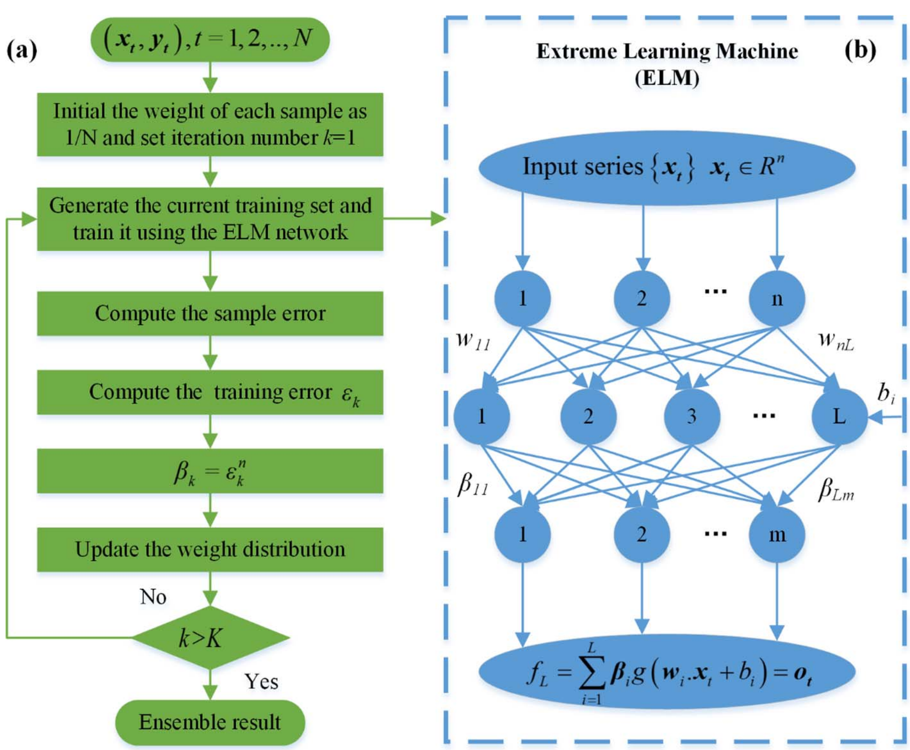

Fig. 1. The framework of the AELM method. (a) The flowchart of AdaBoost.RT algorithm. (b) The structure of ELM.

mean absolute scale error (MASE) have been exploited in this paper. Among the three statistical measures, RMSE is the square root of the average squared errors between the predicted values and the observed values and reflects the degree of variation, the MAE reflects the actual forecasting error in a more balanced perspective and the MASE is a measure of accuracy for the forecasts and is independent of scales. The RMSE, MAE and MASE are estimated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (q_p(i)-q_o(i))^2} \tag{22}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |q_p(i)-q_o(i)| \tag{23}$$

$$\mathrm{MASE} = \frac{\sum_{i=1}^{N} |q_p(i)-q_o(i)|}{\frac{N}{N-1}\sum_{i=2}^{N} |q_o(i)-q_o(i-1)|} \tag{24}$$

where $q_p(i)$ nand $q_o(i)$ are the predicted and observed wind speed data, respectively, and N is the length of the time series.

To further describe the improvement of a certain model, improved percentages indices including P_{RMSE}, P_{MAE} and P_{MASE} are employed to describe the improvement degree. The three indices between two different models named Model 1 and Model 2 can be described as:

$$P_{\mathrm{RMSE}} = \left(\frac{\mathrm{RMSE1}-\mathrm{RMSE2}}{\mathrm{RMSE1}}\right) \times 100 \tag{25}$$

$$P_{\mathrm{MAE}} = \left(\frac{\mathrm{MAE1}-\mathrm{MAE2}}{\mathrm{MAE1}}\right) \times 100 \tag{26}$$

$$P_{\mathrm{MASE}} = \left(\frac{\mathrm{MASE1}-\mathrm{MASE1MASE2}}{\mathrm{MASE1}}\right) \times 100 \tag{27}$$

3. Procedures of the proposed method

Fig. 2 illustrates the overall framework of the proposed CVAELM model constructed in this study. The proposed CEEMDAN-VMD-AELM (CVAELM) model can be implemented in four steps. Firstly, the original wind speed time series is denoised using the two-stage decomposition algorithm. In the denoising process, the wind speed series is first decomposed into several IMFs and a residue using CEEMDAN. And then the IMF with the highest frequency is re-decomposed into several modes using VMD. Secondly the AELM models are constructed and trained for all the modes decomposed using the two-stage decomposition technique. The AELM models are constructed by replacing the base learning algorithm in AdaBoost.RT with ELM. The input variables are selected according to the partial autocorrection function (PACF) values. For each mode, the antecedent time series of which the PACF values are outside the 95% confidence level are selected as input variables for forecasting, with the maximum time lag determined as 48. Thirdly, the well-trained AELM models are applied to the test datasets for each mode to obtain the predicted results. Finally, the final prediction is obtained by adding up the predicted results of all the modes.

4. Model construction and development

4.1. Data collection

Wind speed data gathered from a wind farm in Inner Mongolia, China is studied to test the efficiency of the wind speed forecasting models. And a total number of four weeks of 15 min wind speed data are collected. In the experiments, four different sets of wind speed data are targeted for one-, two- and three-step ahead forecasting. Each dataset contains seven days of 15-min wind speed data (672 observations) and is separated into two parts: the first 520 wind speed observations for training and the remaining 152 observations for testing. The four datasets are denoted as Dataset 1, Dataset 2, Dataset 3 and Dataset 4, respectively. The statistical information including mean value, median value (Med.), maximum (Max.) value, minimum (Min.) value, standard deviation (SD), skewness (Skew.) and kurtosis (Kurt.) of the four datasets are calculated and shown in Table 1. The four datasets of 15-min wind speed data are presented in Fig. 3.

4.2. Parameter settings

Two categories of wind speed forecasting models have been developed in this study: one is the individual regression methods without decomposition processes and the other is the decomposition-based regression models. The individual models include the Bagging method, the partial least squares (PLS) model, the BP neural network, the SVM, the ELM and the AELM models. The decomposition-based models including the CEEMDAN-ELM (CELM), the CEEMDAN-AELM (CAELM), the CEEMDAN-VMD-ELM (CVELM) and the CEEMDAN-VMD-AELM (CVAELM) models. All the models are developed in MATLAB environment in this study. Before sending the wind speed data into the forecasting models, the input-matrix needs to be constructed. The input variables plays an important role in ensuring the quality of the forecasting results. In this study, the PACF values of the time series are calculated and exploited to select the relevant and important variables. The antecedent wind speed series of which the PACF values are outside the 95% confidence level are selected as input variables for forecasting, with the maximum time lag adopted as 48. The input selection strategy is used in all the models constructed in this study to assure fair and valid comparisons.

The Matlab's "Bag" function is utilized to run the bagging decision trees model. The number of hidden neurons of the ELM and BP neural networks are selected using grid search (GS) algorithm. The search range is set as $[m,2n+20]$, where n denotes the number of input neurons, m is set as $2n-20$ when n is bigger than 10, otherwise m is set as 1. The searching step is set as 1. The parameters of SVM are also selected using the GS algorithm. The searching range of the penalty parameter C and the kernel parameter σ of SVM are both set as $[2^{-5},2^5]$. In the AELM forecasting models, the initial threshold ϕ is adopted as 0.2, the maximum iteration number is adopted as 20. In the CEEMDAN based models (CEEMDAN-ELM, CEEMDAN-AELM), the CEEMDAN decomposition technique is used to decompose the original wind speed series into different modes. The number of iterations for CEEMDAN is adopted as 20. The standard deviation of the Gaussian noise for CEEMDAN is 0.2. In the two-stage decomposition based models (CVELM, CVAELM), the number of modes to be decomposed using VMD needs to be predefined. The number of modes is selected according to the center frequencies ω_k of the decomposed modes.

4.3. Wind speed decomposition

In the proposed CVAELM model, the original series is decomposed into several modes using the two-stage decomposition algorithm to reduce the non-stationary and non-linear characteristics of the original wind speed series. The decomposition results of Dataset 1 using the two-stage decomposition technique are shown in Fig. 4. The parameters of CEEMDAN and VMD are determined according to the descriptions of Section 4.2. As has been illustrated in Fig. 4, the original series is decomposed into 9 IMFs and a residual which are named as IMF1, IMF2, …, IMF9 and Res via CEEMDAN in the first stage of the decomposition process. These IMFs decrease in frequency and increase in wavelength from IMF1 to IMF9. Among these decomposed subseries, IMF1 has the highest frequency and shows the detailed information of the wind speed series while the residual is a trend term which describes the general tendency. In the second stage of the decomposition process, the IMF1 with the highest frequency is re-decomposed into three modes via VMD

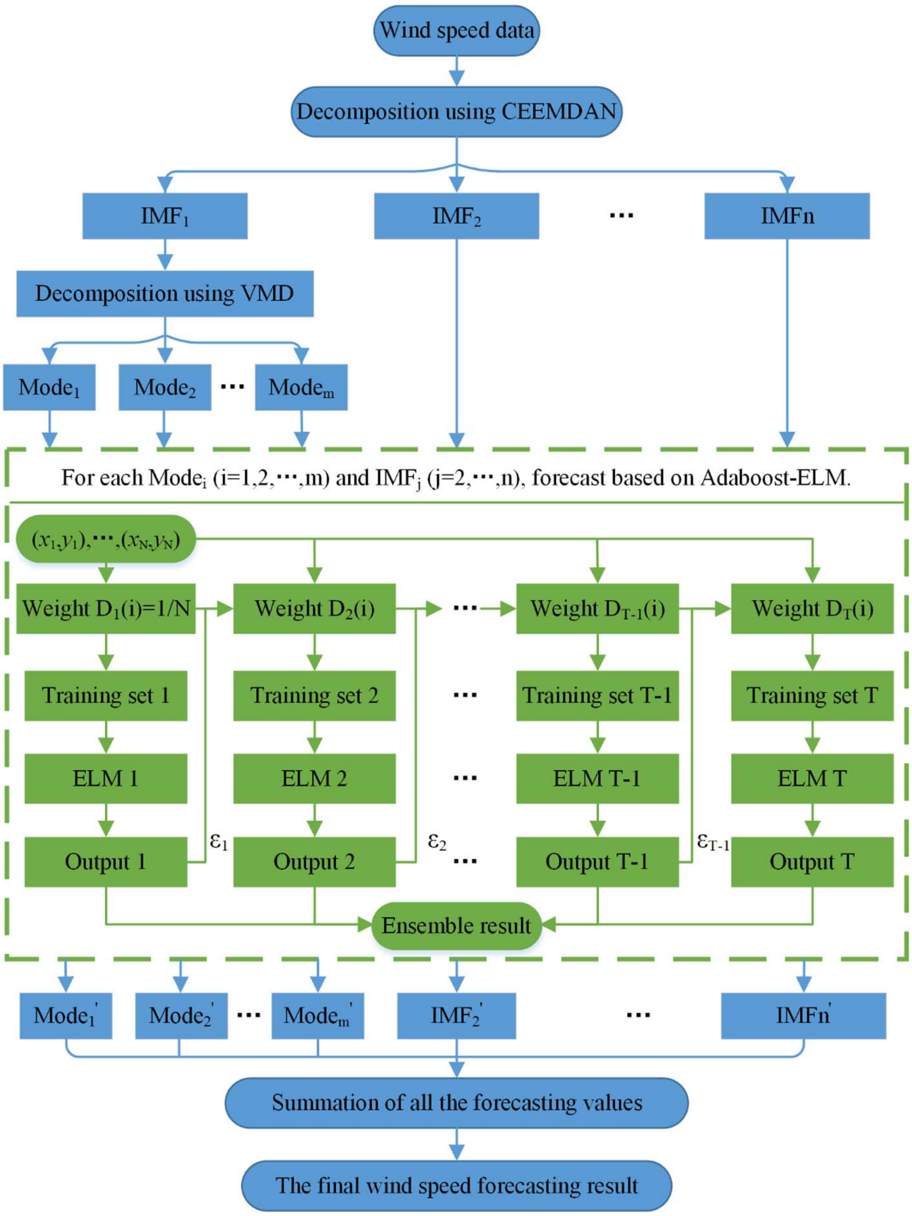

Fig. 2. The overall framework of the proposed CVAELM model.

Table 1
Statistical information for the four datasets from different periods.

Datasets	Mean(m/s)	Med. (m/s)	Max. (m/s)	Min. (m/s)	SD	Skew.	Kurt.	Data period
Dataset 1	7.12	7.20	17.85	0.33	4.08	0.42	2.62	20 th Mar. – 26 th Mar.
Dataset 2	6.59	5.79	17.01	0.31	4.33	0.44	2.01	1 st Apr. – 7 th Apr.
Dataset 3	9.06	8.93	18.83	0.74	4.75	-0.05	1.85	11 st Apr. – 17 th Apr.
Dataset 4	8.77	9.76	20.60	0.64	5.03	0.05	1.84	21 st Apr. – 27 th Apr.

which are named as Mode1, Mode2 and Mode3 to further decrease the non-stationarity of IMF1. The decomposition results are exhibited in Fig. 4, too. Unlike the IMFs decomposed via CEEMDAN, Mode3 shows the highest frequency among the three VMD modes.

5. Results and discussions

A total number of ten different models including Bagging, PLS, BP neural network, SVM, ELM, AELM, CELM, CAELM, CVELM and CVAELM models have been developed in this study. The performance of the six individual modes including Bagging, PLS model, BP neural network, SVM, ELM and AELM models are compared to highlight the effectiveness of the AELM model for wind speed forecasting. To demonstrate the efficiency of the two-stage decomposition technique (CEEMDAN and VMD) in improving the forecasting accuracy of the AELM model, the CVAELM model has been constructed to compare with the AELM model. The performances of the CVAELM and the CAELM models have also been compared. In the CAELM model, the AELM model is exploited to forecast the IMFs and the residual decomposed by CEEMDAN. The performances of the ELM, the CELM and the CVELM models have been discussed and compared to further reveal the effects of different decomposition techniques in enhancing the wind

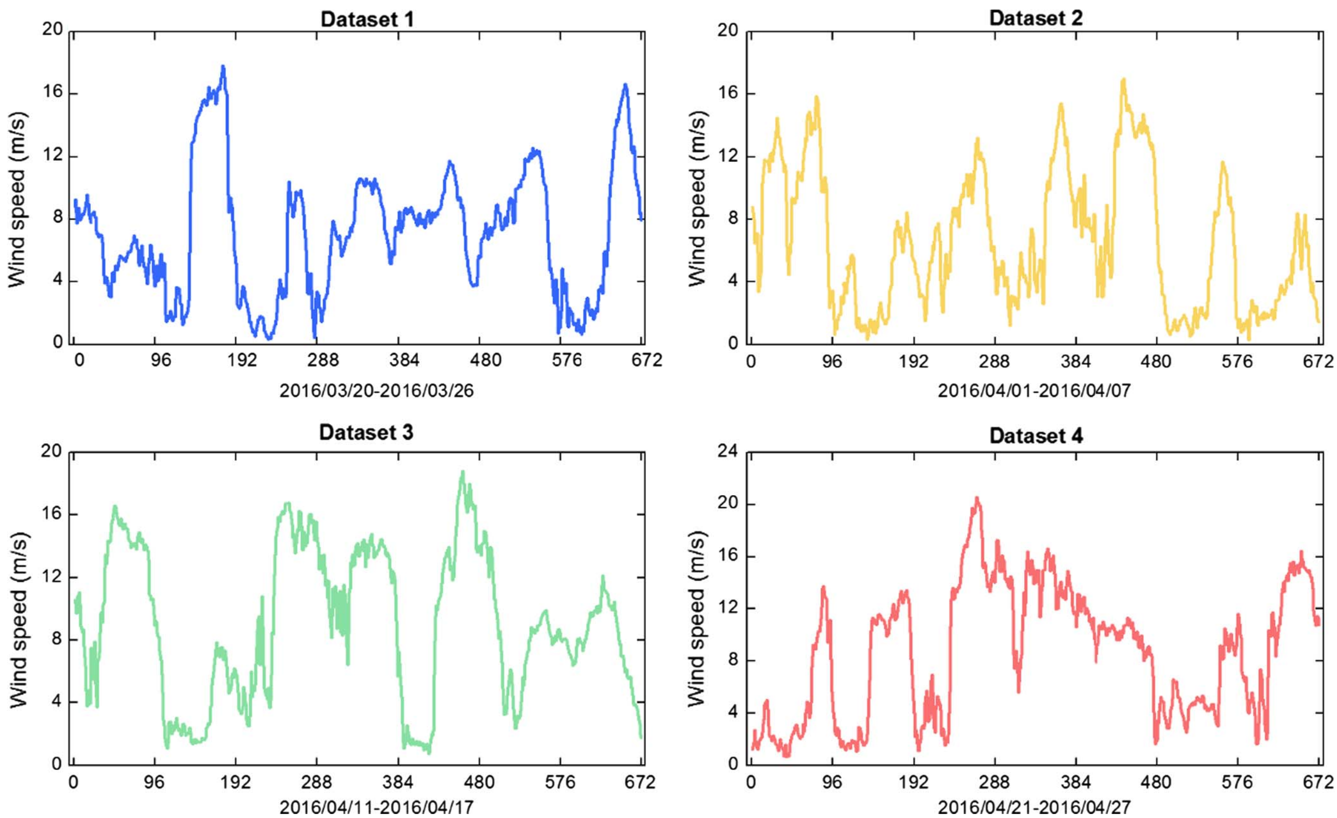

Fig. 3. Four datasets of 15 min wind speed data on different time periods.

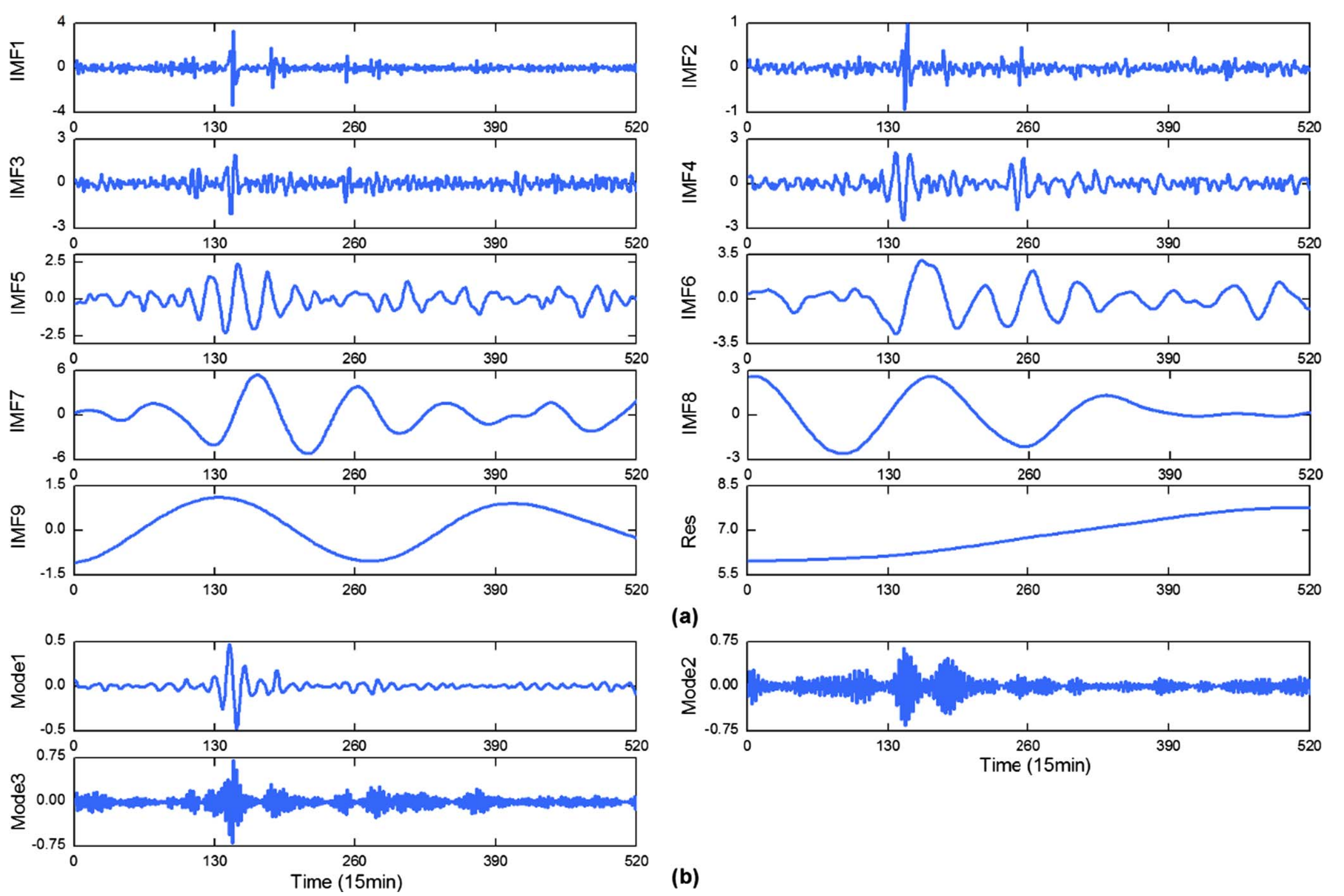

Fig. 4. Decomposition results of the original wind speed series by the two-stage decomposition technique for Dataset 1. (a) Decomposed IMFs using CEEMDAN; (b) Decomposed modes of IMF1 using VMD.

Table 2
Performance evaluations of different wind speed forecasting models for Dataset 1.

Horizons	One-step ahead			Two-step ahead			Three-step ahead		
Indices	RMSE (m/s)	MAE (m/s)	MASE	RMSE (m/s)	MAE (m/s)	MASE	RMSE (m/s)	MAE (m/s)	MASE
Bagging	0.78	0.54	1.32	1.06	0.75	1.83	1.35	1.00	2.44
PLS	0.64	0.44	1.06	0.97	0.70	1.70	1.26	0.96	2.33
BP	0.64	0.43	1.05	0.97	0.69	1.69	1.27	0.97	2.36
SVM	0.64	0.42	1.02	0.95	0.64	1.55	1.23	0.88	2.14
ELM	0.63	0.45	1.10	0.86	0.63	1.53	1.05	0.79	1.93
AELM	0.59	0.41	0.99	0.80	0.57	1.40	1.05	0.80	1.95
CELM	0.47	0.35	0.85	0.66	0.49	1.20	0.84	0.67	1.62
CVELM	0.45	0.32	0.78	0.63	0.47	1.14	0.81	0.62	1.51
CAELM	0.39	0.29	0.71	0.48	0.37	0.90	0.63	0.45	**1.10**
CVAELM	**0.36**	**0.28**	**0.68**	**0.45**	**0.35**	**0.85**	**0.57**	**0.45**	**1.10**

Values in bold means to highlight the methods with best performance.

speed forecasting accuracy.

5.1. Statistical error measures comparison

The forecasting errors including RMSE, MAE and MASE of the total ten models in one-, two- and three- step ahead for Dataset 1 are presented in Table 2. The forecasting results of the aforementioned models for Dataset 1 have been compared and discussed. From Table 2, it can be found that: (a) among the six individual regression methods, the AELM models performed best in one-step ahead forecasting. The RMSE, MAE and MASE of the AELM models are all the smallest. For example, the RMSE of the AELM model for Dataset 1 is 0.59 m/s while those of the Bagging, BP, PLS, SVM and ELM models are 0.78 m/s, 0.64 m/s, 0.64 m/s, 0.64 m/s and 0.63 m/s, respectively. The same conclusion has also been obtained in two-step ahead forecasting according to the forecasting results. In three-step ahead forecasting, the performance of the AELM model is similar to that of the ELM model according to the three indices and is much better than those of the other conventional methods. The well performance of the AELM models demonstrates that the proposed AELM model is an effective tool for accurate wind speed forecasting; (b) from the comparison of the ELM-based (ELM, CELM and CVELM) and the corresponding AELM-based (AELM, CAELM and CVAELM) models, it can be found that the RMSE, MAE and MASE of the AELM-based models are nearly all smaller than those of the corresponding ELM-based models for the three horizons, which verifies the contribution of the improved AdaBoost.RT method on boosting the performance of the ELM models. For example, the CVAELM models outperform the CVELM models and have lower RMSE values of 0.36 m/s, 0.45 m/s and 0.57 m/s in contrast to 0.45 m/s, 0.63 m/s and 0.81 m/s for the CVELM models in one-, two- and three-step ahead forecasting, respectively; (c) the models that embedded with two-stage decomposition technique outperform the models that embedded with CEEMDAN and the models did not use decomposition techniques. For example, compared with the AELM models, the CVAELM models lead to obvious reductions in terms of the three indices in 1–3 step ahead forecasting. And the CVAELM models are superior to the CAELM models in terms of RMSE, MAE and MASE in 1–2 step ahead forecasting. In three-step ahead forecasting, the MAE and the MASE of the two models are the same. The model comparisons demonstrate that the proposed two-stage decomposition approach is an effective data-preprocessing technique in enhancing the wind speed forecasting accuracy; (d) the proposed CVAELM model have the highest performance among the ten models in terms of the three statistical error measures, which demonstrates that the CVAELM model can be exploited as an effective tool in generating reliable and comparatively high-accuracy wind speed forecasting values.

To further exhibit the forecasting performance of different models, the forecasting results of the Bagging, ELM, AELM, CAELM and CVAELM models in one-step ahead forecasting for Dataset 1 have been illustrated in Fig.5. Comparisons between the forecasting values and the observed values are shown in the middle of the figure. The statistical error measures including RMSE, MAE and MASE are shown in the right side of the figure and the scatter plots with a linear regression curve are shown in the bottom of the figure. From the subgraph in the middle of Fig.5, it can be seen that the forecasting values of the five models approximate the observed values well except the Bagging method. From the subgraphs in the right side of Fig.5 it is obvious that the proposed CVAELM model have the lowest RMSE, MAE and MASE values. The rankings of the performance of the five models from the lowest to the highest are Bagging, ELM, AELM, CAELM and CVAELM. From the bottom subgraphs of Fig.5, it can be seen that the scatter points of the CAVELM model distribute the most uniformly around the regression line and also are the closest to the line, which demonstrates that the proposed CVAELM model can produce a better forecasting result compared with the other methods.

To further investigate the performance of the proposed model, the ten models developed in this study have been applied to Datasets 2–4 for 1–3 step ahead forecasting. The values of the statistical error measures of the ten models for 1–3 step ahead forecasting are presented in Tables 3–5 for Dataset 2, Dataset 3 and Dataset 4, respectively. And the corresponding figures to illustrate the one-step ahead forecasting performance of Bagging, ELM, AELM, CAELM and CVAELM models are presented in Figs. 6-8 for Dataset 2, Dataset 3 and Dataset 4, respectively. The conclusions obtained for Datasets 2–4 are similar to those of Dataset 1. It can be found from Tables 3–5 and Figs. 6–8 that the CVAELM models exhibit the best performance while the Bagging methods exhibit the worst. The AELM-based models generally performed better than the ELM-based models. The models used the two stage decomposition technique can lead to significant reductions in RMSE, MAE and MASE compared with the models that did not using data-preprocessing techniques.

5.2. Improved percentages comparison

Taken the Bagging method as the benchmark model, the values of the three improve percentages indices including P_{RMSE}, P_{MAE} and P_{MASE} for the ELM, AELM, CAELM and CVAELM models for the four datasets in 1–3 step ahead forecasting are presented in Table 6. And the improved percentages of different models compared with the Bagging method for the four datasets have been illustrated in Figs. 9 and 10. As can be seen from Table 6 and Figs. 9 and 10, the values of P_{RMSE}, P_{MAE} and P_{MASE} of the CVAELM models are the highest compared with the other models, which further testifies the suitability of the proposed model in both one-step and multi-step ahead wind speed forecasting. Taken Dataset 1 in one-step ahead forecasting for instance, the P_{RMSE} of the CVAELM model is 53.19% while those of the ELM, AELM and C-AELM models are 19.34%, 23.86% and 49.68%, respectively. Similarly, it has been further proved that the two-stage decomposition based

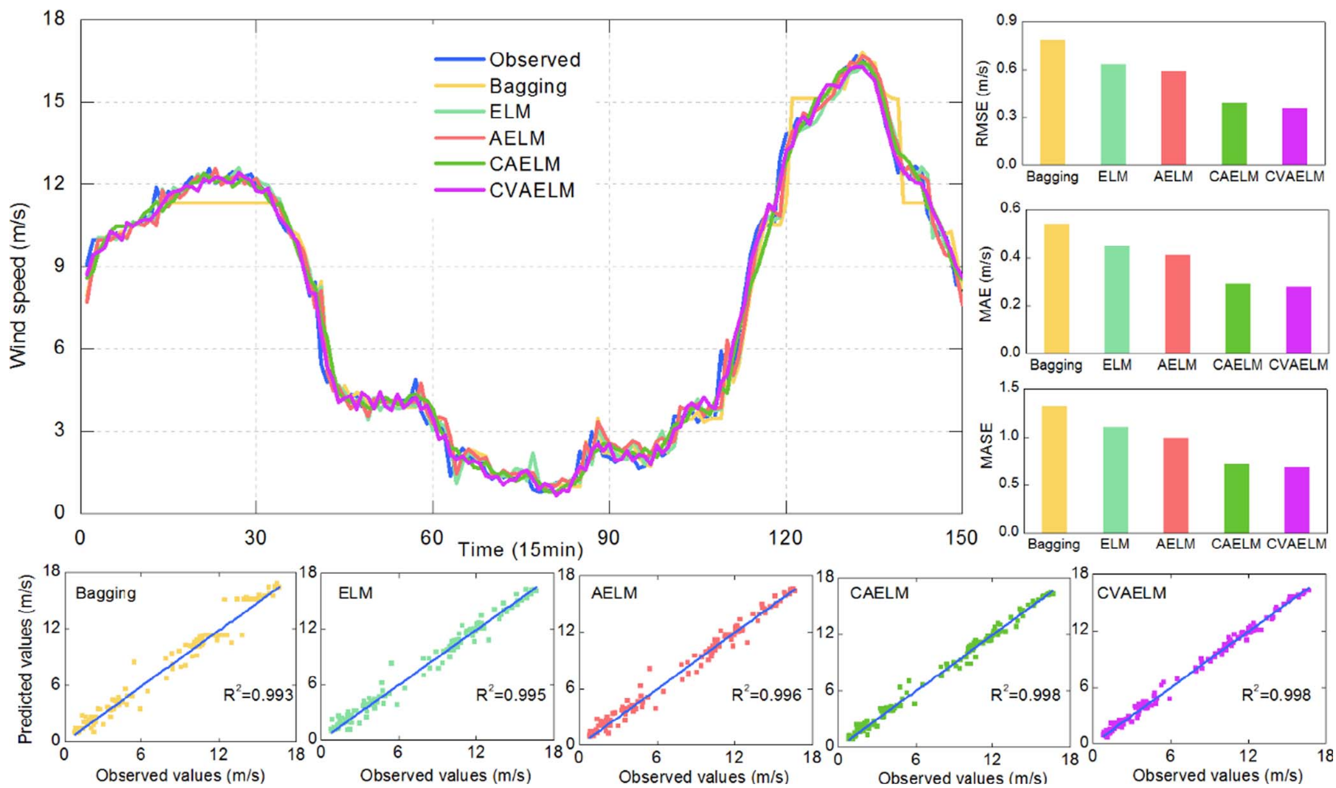

Fig. 5. Forecasting performance of the Bagging, ELM, AELM, CAELM and CVAELM models in one-step ahead forecasting for Dataset 1.

Table 3
Performance evaluations of different wind speed forecasting models for Dataset 2.

Horizons	One-step ahead			Two-step ahead			Three-step ahead		
Indices	RMSE (m/s)	MAE (m/s)	MASE	RMSE (m/s)	MAE (m/s)	MASE	RMSE (m/s)	MAE (m/s)	MASE
Bagging	0.80	0.57	1.26	1.10	0.77	1.70	1.43	1.00	2.24
PLS	0.69	0.45	1.01	1.06	0.69	1.52	1.37	0.95	2.13
BP	0.69	0.48	1.06	1.05	0.71	1.56	1.30	0.90	2.02
SVM	0.70	0.45	1.00	1.07	0.69	1.52	1.37	0.94	2.11
ELM	0.69	0.46	1.03	1.06	0.71	1.57	1.37	0.98	2.20
AELM	0.66	0.47	1.05	1.01	0.75	1.66	1.27	0.95	2.13
CELM	0.53	0.39	0.87	0.73	0.53	1.18	0.85	0.65	1.47
CVELM	0.53	0.39	0.87	0.72	0.56	1.23	0.81	0.63	1.43
CAELM	0.52	0.39	0.86	0.67	0.52	1.16	0.77	0.61	1.37
CVAELM	**0.50**	**0.36**	**0.80**	**0.64**	**0.47**	**1.04**	**0.72**	**0.54**	**1.21**

Values in bold means to highlight the methods with best performance.

Table 4
Performance evaluations of different wind speed forecasting models for Dataset 3.

Horizons	One-step ahead			Two-step ahead			Three-step ahead		
Indices	RMSE (m/s)	MAE (m/s)	MASE	RMSE (m/s)	MAE (m/s)	MASE	RMSE (m/s)	MAE (m/s)	MASE
Bagging	0.79	0.61	2.32	1.08	0.91	3.43	1.50	1.27	4.67
PLS	0.35	0.27	1.01	0.54	0.43	1.62	0.71	0.58	2.14
BP	0.36	0.28	1.07	0.56	0.43	1.64	0.73	0.60	2.20
SVM	0.41	0.32	1.20	0.66	0.52	1.97	0.93	0.77	2.84
ELM	0.34	0.27	1.01	0.52	0.41	1.55	0.68	0.53	1.96
AELM	0.32	0.25	0.95	0.49	0.39	1.46	0.62	0.50	1.84
CELM	0.29	0.23	0.88	0.40	0.32	1.20	0.52	0.40	1.46
CVELM	0.31	0.24	0.93	0.39	0.32	1.19	0.47	0.38	1.41
CAELM	0.27	0.21	0.80	0.35	0.28	1.06	**0.41**	**0.33**	**1.22**
CVAELM	**0.26**	**0.20**	**0.77**	**0.33**	**0.26**	**0.99**	**0.41**	**0.33**	1.23

Values in bold means to highlight the methods with best performance.

Table 5
Performance evaluations of different wind speed forecasting models for Dataset 4.

Horizons	One-step ahead			Two-step ahead			Three-step ahead		
Indices	RMSE (m/s)	MAE (m/s)	MASE	RMSE (m/s)	MAE (m/s)	MASE	RMSE (m/s)	MAE (m/s)	MASE
Bagging	1.12	0.61	1.03	1.54	0.97	1.62	1.86	1.31	2.17
PLS	1.11	0.60	1.01	1.47	0.91	1.52	1.82	1.21	2.01
BP	1.08	0.59	1.00	1.41	0.91	1.52	1.66	1.11	1.84
SVM	1.12	0.59	0.99	1.48	0.89	1.49	1.84	1.19	1.98
ELM	1.09	0.63	1.05	1.45	0.93	1.56	1.72	1.16	1.92
AELM	1.09	0.59	0.99	1.42	0.87	1.47	1.72	1.15	1.90
CELM	0.76	0.51	0.86	1.13	0.75	1.26	1.29	0.95	1.57
CVELM	0.70	0.50	0.84	1.13	0.75	1.25	1.27	0.96	1.60
CAELM	0.56	0.38	0.65	0.69	**0.47**	0.80	0.73	0.54	0.90
CVAELM	**0.47**	**0.34**	**0.58**	**0.64**	**0.47**	**0.79**	**0.70**	**0.52**	**0.86**

Values in bold means to highlight the methods with best performance.

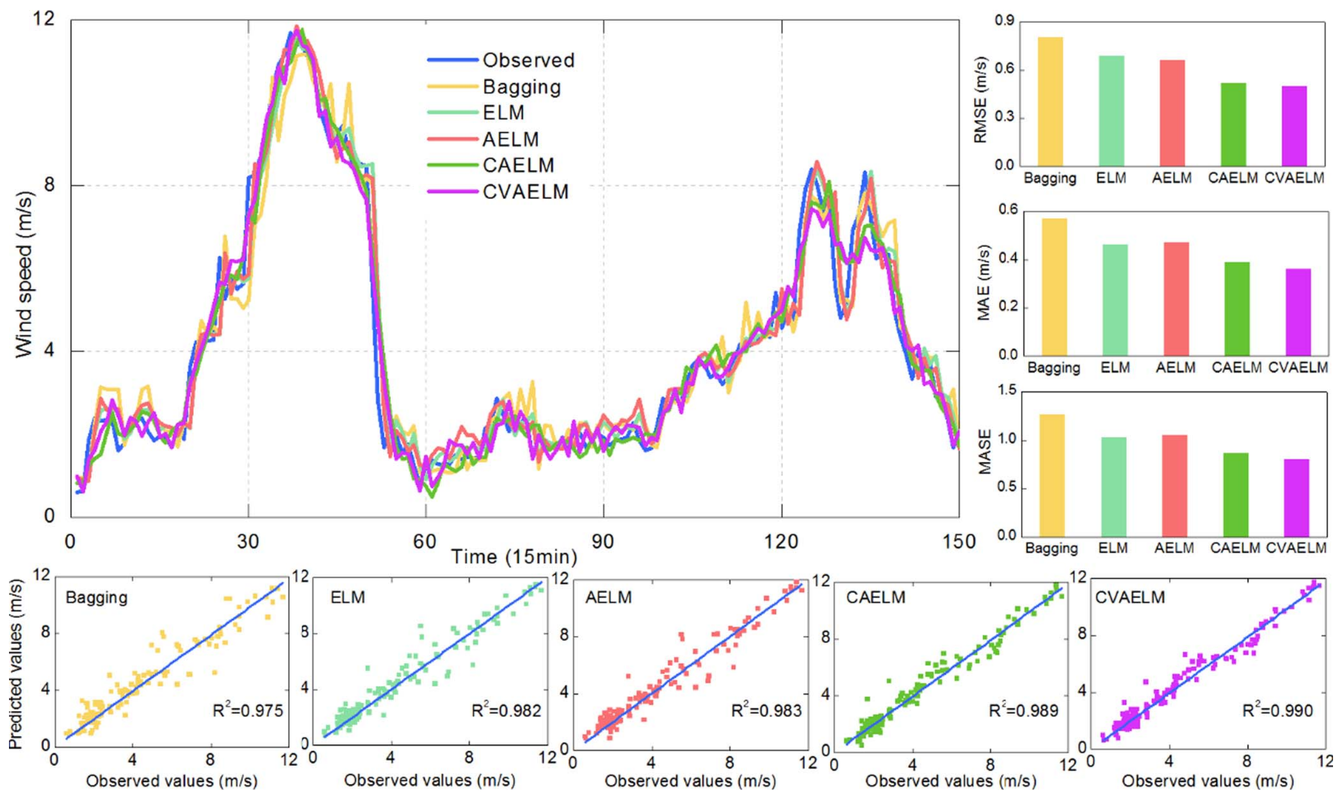

Fig. 6. Forecasting performance of the Bagging, ELM, AELM, CAELM and CVAELM models in one-step ahead forecasting for Dataset 2.

models (CVAELM) performed better than the single decomposition based models (CAELM). And the models with decomposition techniques perform better than the ones without. It has also been verified that the AdaBoost.RT algorithm has a positive effect on the ELM model.

5.3. Comparison of the number of error points

As high wind speed has a greater effect on the performance of the wind turbine generator for generating wind power, the number of points of which the relative error is bigger than 5% for high wind speed (> 10 m/s) for several models in the test stage have been compared. The number of error points is presented in Table 7. As can be seen from Table 7, the CVAELM models have the smallest number of error points for 1–3 step ahead forecasting for the four datasets in nearly all the forecasting cases. The performance rankings of the models in terms of the number of error points from the lowest to the highest are Bagging, SVR, ELM, AELM, CAELM and CVAELM, which agrees with the conclusions obtained from the above analysis in Section 5.1 and Section 5.2.

6. Conclusions

This paper investigates the performance of a hybrid model embedded with CEEMDAN, VMD, AdaBoost.RT and ELM for forecasting short-term wind speed in Inner Mongolia, China. The proposed hybrid model has been applied to four wind speed datasets for 1–3 step ahead forecasting to evaluate its effectiveness and applicability. The specific objectives are to develop and evaluate methods for improving wind speed forecasting accuracy by comparing the performance of different models based on various performance evaluation indices including three statistical error measures RMSE, MAE, MASE, three improved percentage measures P_{RMSE}, P_{MAE}, P_{MASE}, and the number of error points for high wind speed. The AELM-based models perform better than the corresponding ELM-based models according to the

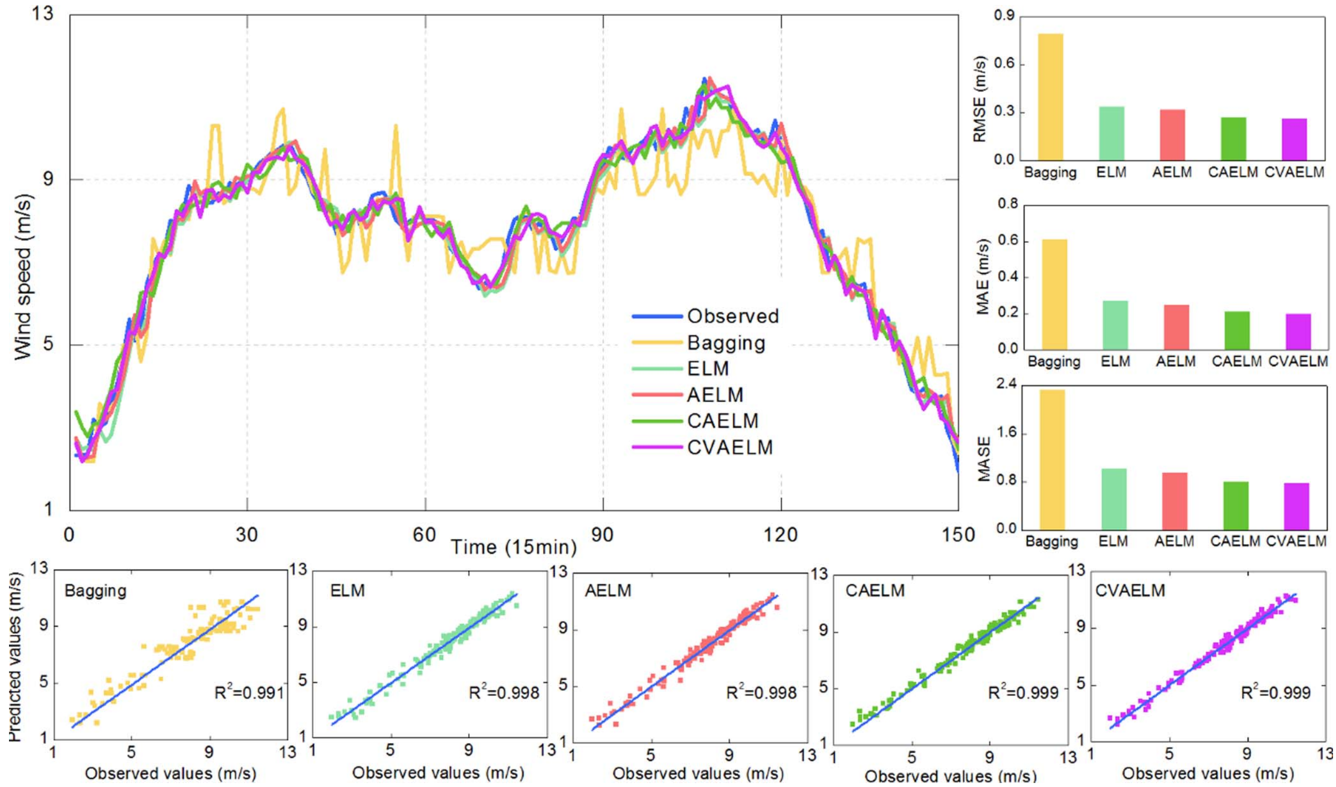

Fig. 7. Forecasting performance of the Bagging, ELM, AELM, CAELM and CVAELM models in one-step ahead forecasting for Dataset 3.

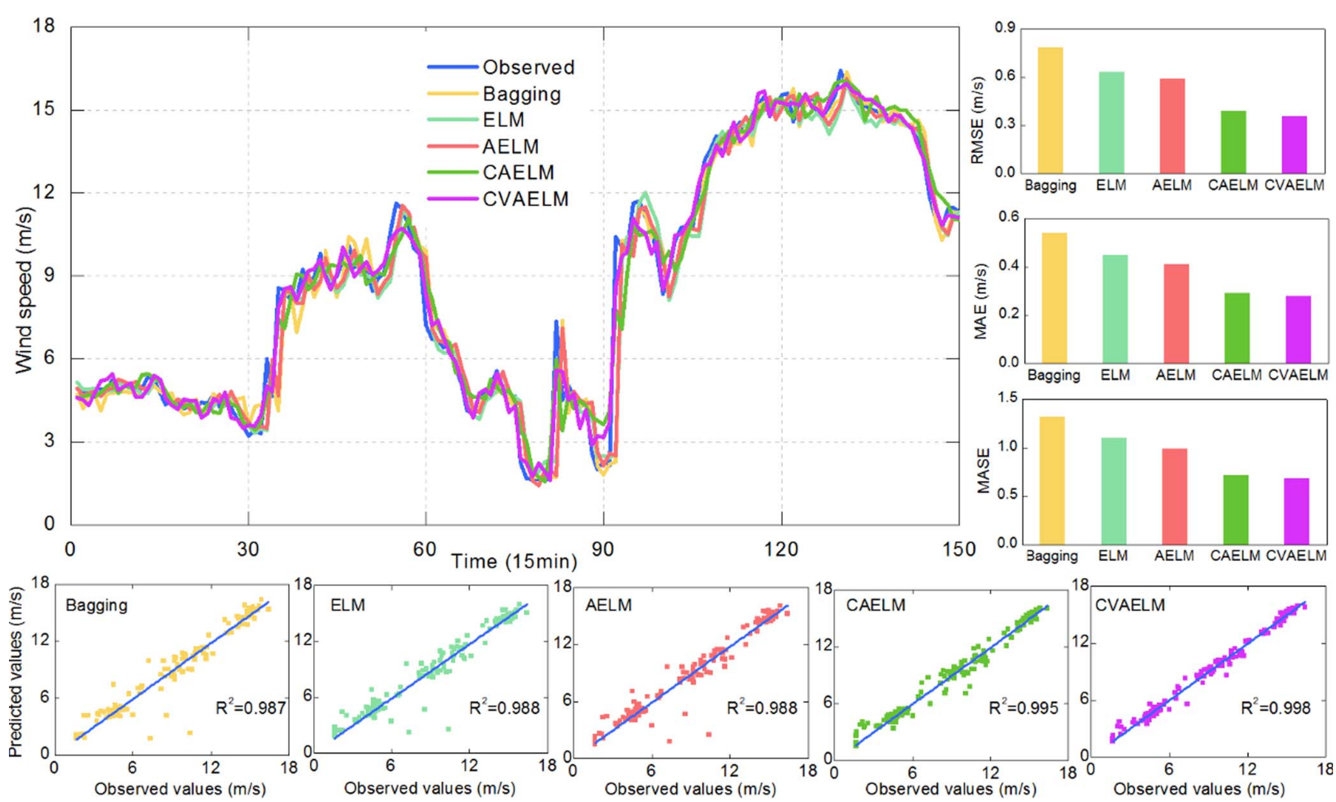

Fig. 8. Forecasting performance of the Bagging, ELM, AELM, CAELM and CVAELM models in one-step ahead forecasting for Dataset 4.

comparisons and analysis, which reveals the effectiveness of the AdaBoost.RT technique in boosting the performance of ELM models for wind speed forecasting. Comparisons between models embedded with decomposition techniques and models without decomposition techniques indicate that the conjunction of single decomposition technique with ELM and AELM models can significantly improve the models' accuracy in wind speed forecasting. And models with the two-stage decomposition technique achieve better forecasting results

Table 6
Improved percentages of the ELM, AELM, CAELM and CVAELM models compared with the Bagging method.

Horizons		One-step ahead			Two-step ahead			Three-step ahead		
Datasets	Models	P_{RMSE} (%)	P_{MAE} (%)	P_{MASE} (%)	P_{RMSE} (%)	P_{MAE} (%)	P_{MASE} (%)	P_{RMSE} (%)	P_{MAE} (%)	P_{MASE} (%)
Dataset 1	ELM	19.34	16.49	16.49	18.69	16.53	16.53	22.30	20.90	20.90
	AELM	23.86	24.49	24.49	24.59	23.68	23.68	22.66	20.20	20.20
	CAELM	49.68	45.82	45.82	54.87	50.77	50.77	53.69	54.90	54.90
	CVAELM	**53.19**	**48.23**	**48.23**	**57.29**	**53.49**	**53.49**	**57.79**	**54.98**	**54.98**
Dataset 2	ELM	13.70	18.37	18.37	3.74	7.60	7.60	3.86	1.73	1.73
	AELM	16.59	16.90	16.90	7.77	2.51	2.51	10.85	4.86	4.86
	CAELM	34.95	31.66	31.66	39.23	32.04	32.04	46.43	39.12	39.12
	CVAELM	**37.65**	**36.68**	**36.68**	**41.76**	**39.14**	**39.14**	**49.90**	**46.14**	**46.14**
Dataset 3	ELM	56.64	56.54	56.54	52.00	54.75	54.75	54.95	57.98	57.98
	AELM	58.97	59.34	59.34	54.96	57.43	57.43	58.79	60.47	60.47
	CAELM	66.19	65.44	65.44	67.94	69.18	69.18	**72.60**	**73.81**	**73.81**
	CVAELM	**67.33**	**66.95**	**66.95**	**69.37**	**71.26**	**71.26**	72.38	73.57	73.57
Dataset 4	ELM	3.07	-2.67	-2.67	5.90	3.73	3.73	7.64	11.76	11.76
	AELM	3.05	3.96	3.96	7.89	9.46	9.46	7.53	12.40	12.40
	CAELM	50.41	36.94	36.94	55.42	50.90	50.90	60.73	58.44	58.44
	CVAELM	**58.57**	**43.55**	**43.55**	**58.54**	**51.38**	**51.38**	**62.33**	**60.33**	**60.33**

Values in bold means to highlight the methods with best performance.

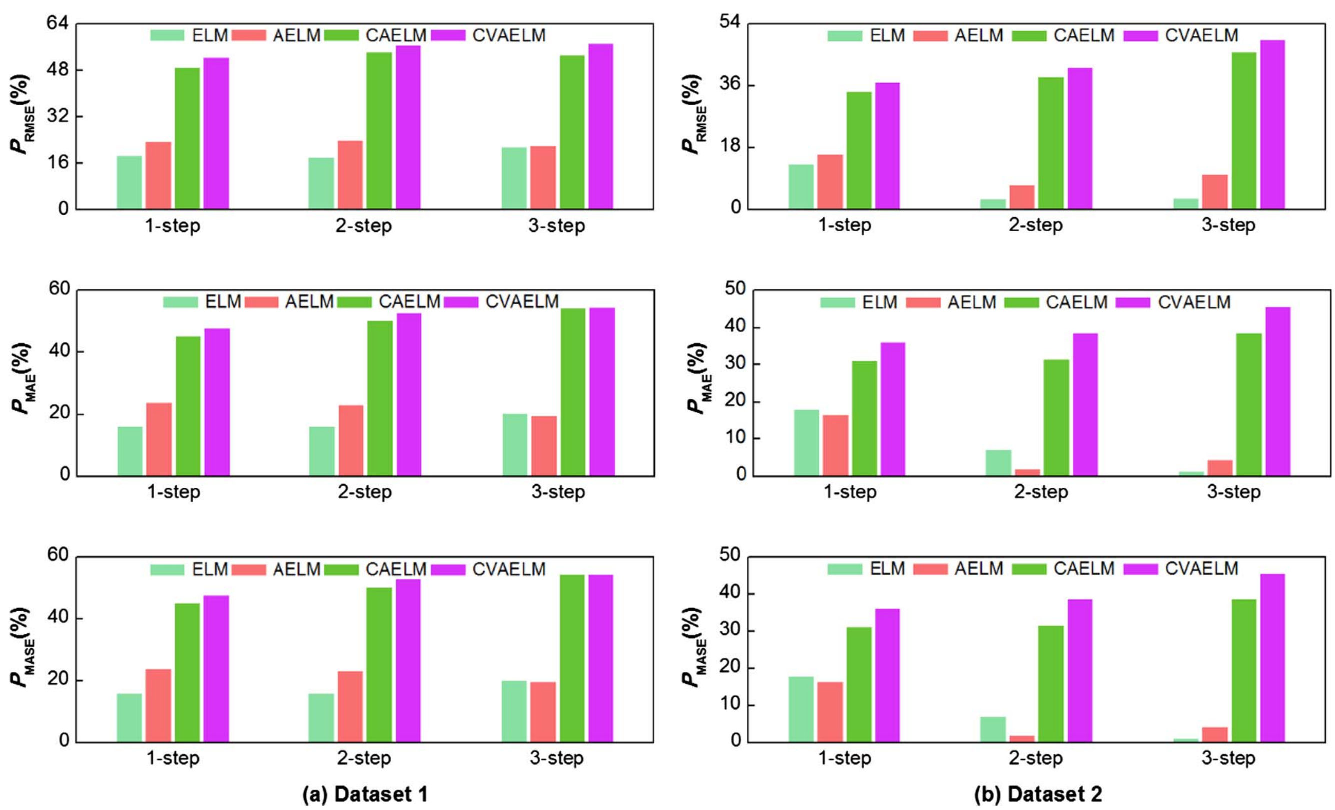

Fig. 9. Improved percentages of the ELM, AELM, CAELM and CVAELM models compared with the Bagging method for Dataset 1 and Dataset 2.

compared with models with the CEEMDAN technique, which verifies that using two successive single processing techniques is an effective tool to enhance the forecasting accuracy of the AELM models. Overall, it can be found from the results of this study that the proposed CVAELM model can be a successful tool for nonstationary wind speed forecasting. The proposed two-stage decomposition technique can decrease the unpredictability of the original wind speed time series considerably and the AdaBoost.RT algorithm can boost the forecasting performance of weak learners.

Acknowledgments

This work is supported by the National Key R & D Program of China (2016YFC0402205), the Key Program of the Major Research Plan of the National Natural Science Foundation of China (No. 91547208), the National Natural Science Foundation of China (No. 51579107) and the State Key Program of National Natural Science of China (No. 51239004) and special thanks are given to the anonymous reviewers and editors for their constructive comments.

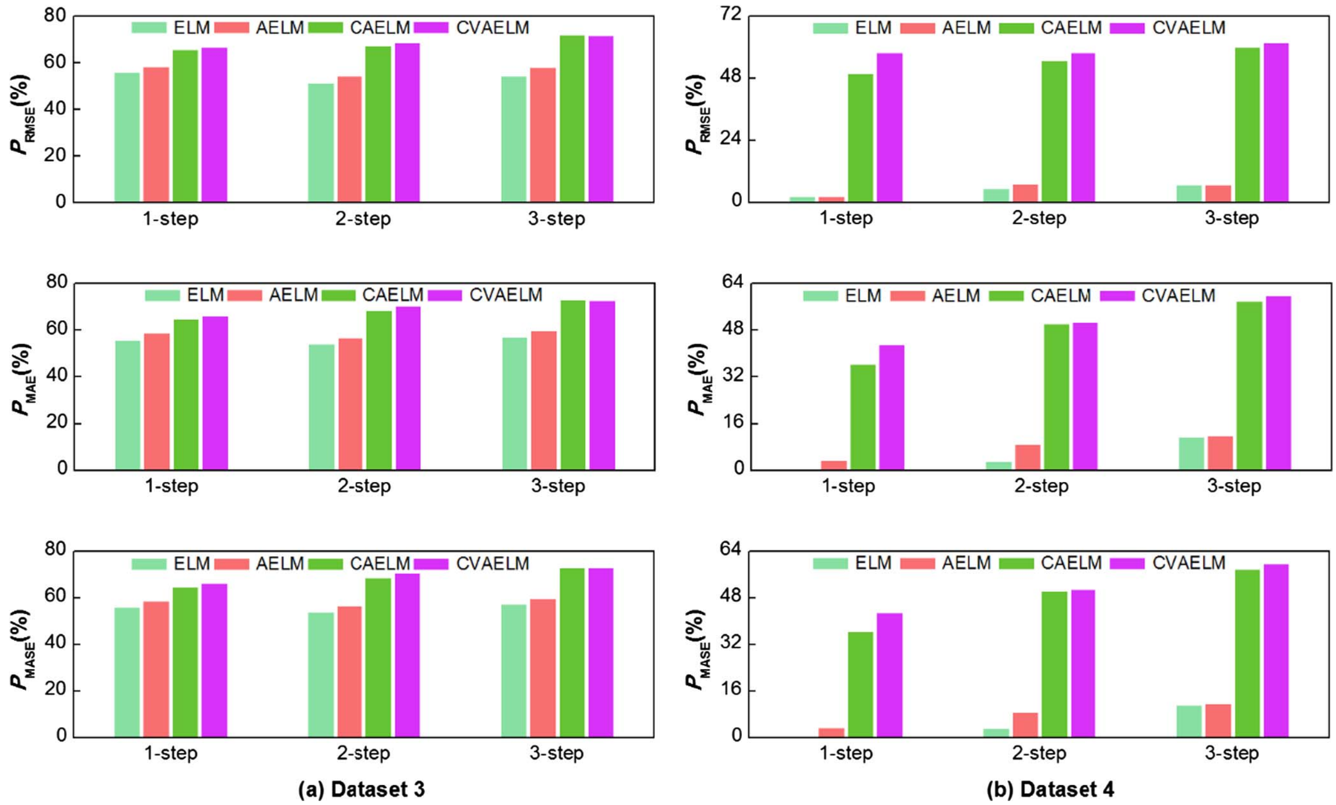

Fig. 10. Improved percentages of the ELM, AELM, CAELM and CVAELM models compared with the Bagging method for Dataset 3 and Dataset 4.

Table 7
Statistical number of the error points for high wind speed (> 10m/s).

Datasets	Models	Bagging	SVR	ELM	AELM	CAELM	CVAELM
Dataset 1	1-step	29	9	8	7	6	**4**
	2-step	37	21	22	16	18	**9**
	3-step	44	39	43	30	29	**18**
Dataset 2	1-step	3	3	3	**0**	1	1
	2-step	4	5	4	5	4	**0**
	3-step	6	6	6	7	7	**4**
Dataset 3	1-step	14	8	3	3	2	**1**
	2-step	12	11	4	5	5	**1**
	3-step	15	17	7	4	**2**	3
Dataset 4	1-step	16	18	21	15	16	**7**
	2-step	27	29	32	29	21	**15**
	3-step	35	33	31	28	29	**14**

Values in bold means to highlight the methods with best performance.

References

[1] Ren Y, Suganthan PN, Srikanth N. A novel empirical mode decomposition with support vector regression for wind speed forecasting. IEEE Trans Neural Networks Learn Syst 2016;27(8):1793–8.
[2] Soman S S, Zareipour H, Malik O, et al. A review of wind power and wind speed forecasting methods with different time horizons. North American Power Symposium (NAPS), IEEE; 2010. p. 1–8.
[3] Zhang J, Draxl C, Hopson T, et al. Comparison of numerical weather prediction based deterministic and probabilistic wind resource assessment methods. Appl Energy 2015;156:528–41.
[4] Al-Yahyai S, Charabi Y, Gastli A. Review of the use of Numerical Weather Prediction (NWP) Models for wind energy assessment. Renew Sustain Energy Rev 2010;14(9):3192–8.
[5] Han Q, Meng F, Hu T, et al. Non-parametric hybrid models for wind speed forecasting. Energy Convers Manage 2017;148:554–68.
[6] Wang J, Hu J, Ma K, et al. A self-adaptive hybrid approach for wind speed forecasting. Renewable Energy 2015;78:374–85.
[7] Yuan X, Tan Q, Lei X, et al. Wind power prediction using hybrid autoregressive fractionally integrated moving average and least square support vector machine. Energy 2017;129:122–37.
[8] Lydia M, Kumar SS, Selvakumar AI, et al. Linear and non-linear autoregressive models for short-term wind speed forecasting. Energy Convers Manage 2016;112:115–24.
[9] Jiang Y, Huang G. Short-term wind speed prediction: Hybrid of ensemble empirical mode decomposition, feature selection and error correction. Energy Convers Manage 2017;144:340–50.
[10] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: a new learning scheme of feedforward neural networks//Neural Networks, 2004. In: Proceedings. 2004 IEEE international joint conference on. IEEE, 2004; 2. p. 985–90.
[11] Salcedo-Sanz S, Pastor-Sánchez A, Prieto L, et al. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization–extreme learning machine approach. Energy Convers Manage 2014;87:10–8.
[12] Abdoos A. A new intelligent method based on combination of VMD and ELM for short term wind power forecasting. Neurocomputing 2016;203:111–20.
[13] Liu D, Niu D, Wang H, et al. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renewable Energy 2014;62:592–7.
[14] Meng A, Ge J, Yin H, et al. Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm. Energy Convers Manage 2016;114:75–88.
[15] 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.
[16] Du P, Wang J, Guo Z, et al. Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting. Energy Convers Manage 2017;150:90–107.
[17] Zhang C, Zhou J, Li C, et al. A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting. Energy Convers Manage 2017;143:360–76.
[18] Guo Z, Zhao W, Lu H, et al. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renewable Energy 2012;37(1):241–9.
[19] Yu C, Li Y, Zhang M. Comparative study on three new hybrid models using Elman Neural Network and Empirical Mode Decomposition based technologies improved by Singular Spectrum Analysis for hour-ahead wind speed forecasting. Energy Convers Manage 2017;147:75–85.
[20] Liu H, Tian H, Liang X, et al. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. Appl Energy 2015;157:183–94.
[21] Wang D, Luo H, Grunder O, et al. Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. Appl Energy 2017;190:390–407.
[22] Shu C, Burn DH. Artificial neural network ensembles and their application in pooled flood frequency analysis. Water Resour Res; 2004; 40(9).
[23] Zaier I, Shu C, Ouarda T, et al. Estimation of ice thickness on lakes using artificial neural network ensembles. J Hydrol 2010;383(3):330–40.

[24] Belayneh A, Adamowski J, Khalil B, et al. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. Atmos Res 2016;172:37–47.
[25] Ren Y, Qiu X, Suganthan PN. Empirical mode decomposition based adaboost-backpropagation neural network method for wind speed forecasting. In: Computational Intelligence in Ensemble Learning (CIEL), 2014 IEEE Symposium on. IEEE; 2014. p. 1–6.
[26] Huang NE, Shen Z, Long SR, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc Roy Soc Lond A Math Phys Eng Sci The Roy Soc 1971;1998(454):903–95.
[27] Wu Z, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv Adaptive Data Anal 2009;1(01):1–41.
[28] Torres M E, Colominas M A, Schlotthauer G, et al. A complete ensemble empirical mode decomposition with adaptive noise. In: Acoustics, speech and signal processing (ICASSP), 2011 IEEE international conference on. IEEE; 2011. p. 4144–47.
[29] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Trans Signal Process 2014;62(3):531–44.
[30] Freund Y, Schapire RE. Experiments with a new boosting algorithm. icml 1996;96:148–56.
[31] Solomatine D P, Shrestha D L. AdaBoost. RT: a boosting algorithm for regression problems. In: Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on. IEEE; 2004. 2: p. 1163–8.
[32] Shrestha DL, Solomatine DP. Experiments with AdaBoost. RT, an improved boosting scheme for regression. Neural Comput 2006;18(7):1678–710.
[33] Tian HX, Mao ZZ. An ensemble ELM based on modified AdaBoost. RT algorithm for predicting the temperature of molten steel in ladle furnace. IEEE Trans Autom Sci Eng 2010;7(1):73–80.