Energy Conversion and Management 144 (2017) 340–350

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Short-term wind speed prediction: Hybrid of ensemble empirical mode decomposition, feature selection and error correction

Yan Jiang, Guoqing Huang *

Key Laboratory of High-speed Railway Engineering (MOE), School of Civil Engineering, Southwest Jiaotong University, Chengdu, Sichuan 610031, China
Research Center for Wind Engineering, School of Civil Engineering, Southwest Jiaotong University, Chengdu, Sichuan 610031, China

ARTICLE INFO

Article history:
Received 25 January 2017
Received in revised form 16 April 2017
Accepted 19 April 2017
Available online 2 May 2017

Keywords:
Wind speed prediction
Real time decomposition
Ensemble empirical mode decomposition
Feature selection
Kernel density estimation
Kullback-Leibler divergence
Energy measure
Least squares support vector machine
Error correction

ABSTRACT

Accurately forecasting wind speed is a critical mission for the exploitation and utilization of wind power. To improve the prediction accuracy, the nonlinearity and nonstationarity embedded in wind speed time series should be reduced. Because the subseries has less nonlinearity and nonstationarity after decomposition, the decomposition-based forecasting methods are widely adopted to provide the higher predictive accuracy. However, latest studies showed the real time decomposition-based forecasting methods could be worse than the single forecasting models. The aim of this study is to improve the performance of the real time decomposition-based forecasting method after the factors attributed to its unsatisfactory performance are uncovered. In this paper, the feature selection and error correction are adopted in the real time decomposition-based forecasting method to enhance the prediction accuracy. In the proposed method, the raw wind speed time series is decomposed into a number of different subseries by ensemble empirical mode decomposition; then two feature selection methods including kernel density estimation-based Kullback-Leibler divergence and energy measure are used to reduce the disturbance of illusive components; further the least squares support vector machine is adopted to establish the one-step ahead forecasting models for the remaining subseries; finally, the hybrid of least squares support vector machine and generalized auto-regressive conditionally heteroscedastic model is introduced to correct resulting error component if its inherent correlation and heteroscedasticity cannot be neglected. Based on two sets of measured data, the results of this study show that: (1) the real time decomposition-based method may be ineffective in practice; (2) both the feature selection and error correction can improve forecasting performance in comparison with the real time decomposition-based method; (3) compared with other involved methods, the proposed hybrid method has the satisfactory performance in both accuracy and stability.

1. Introduction

As a renewable and environmentally friendly energy resource, wind energy is of vital importance among the low-carbon energy technologies and has attracted the global attention. According to the World Wind Energy Report [1], the total installed wind power capacity has reached to 160,000 MW in 2009, and it is expected to double every 3 years. Along with the rapid development and utilization of wind energy, the green gas emission and the traditional fossil fuels utilization are drastically reduced. However, due to the stochastic and intermittent characteristics of the wind source [2], the integration efficiency of wind power into a multisource energy network poses many challenges on a series of tasks, such as the energy generation planning and turbine maintenance scheduling [3]. These challenges have severely hindered the exploitation of wind energy. One of the methods to mitigate these challenges is the improvement of short-term wind speed forecasting accuracy and the reduction of its uncertainty [4].

Abbreviations: ARMA, Auto-Regressive Moving Average; ARIMA, Auto-Regressive Integrated Moving Average; FARIMA, Fractional Auto-Regressive Integrated Moving Average; MTK, Modified Taylor Kriging; ANN, Artificial Neural Networks; SVM, Support Vector Machine; MLP, Multi Layer Perceptron; LSSVM, Least Squares Support Vector Machine; SSA, Singular Spectrum Analysis; RBFNN, Radial Basis Function Neural Network; GSA, Gravitational Search Algorithm; ANFIS, Adaptive Network-based Fuzzy Inference System; PSO, Particle Swarm Optimization; GARCH, Generalized Autoregressive Conditionally Heteroscedastic; ELM, Extreme Learning Machine; EMD, Empirical Mode Decomposition; EEMD, Ensemble Empirical Mode Decomposition; FEEMD, Fast Ensemble Empirical Mode Decomposition; WD, Wavelet Decomposition; WPD, Wavelet Packet Decomposition; GA-BP, Genetic Algorithm-Back Propagation; IMF, Intrinsic Mode Function; KDE, Kernel Density Estimation; KLD, Kullback-Leibler Divergence; MEMD, Multivariate Empirical Mode Decomposition; PDF, Probability Density Function; ACF, Autocorrelation Function; PACF, Partial Autocorrelation Function; MAE, Mean Absolute Error; MRPE, Mean Relative Percentage Error; RMSE, Root Mean Square Error; RMSRE, Root Mean Square Relative Error; KS, Kolmogorov-Smirnov; CDF, Cumulative Distribution Function.

* Corresponding author at: Key Laboratory of High-speed Railway Engineering (MOE), School of Civil Engineering, Southwest Jiaotong University, Chengdu, Sichuan 610031, China.

E-mail addresses: xnjtjiangyan@163.com (Y. Jiang), ghuang1001@gmail.com (G. Huang).

http://dx.doi.org/10.1016/j.enconman.2017.04.064
0196-8904/

In the past few decades, numerous methods have been proposed to enhance the prediction accuracy including physical approaches and statistical models. Every kind of methods has their own advantages and disadvantages. The physical approaches take into account the meteorological factors and are generally used for long-term wind speed prediction [5]. On the other hand, the statistical models are usually suitable for short-term wind speed prediction based on the historical data. These statistical models mainly include linear time series and nonlinear intelligent models, which will be reviewed below.

Time series models, such as auto-regressive moving average (ARMA) and auto-regressive integrated moving average (ARIMA), have been widely applied in short-term wind speed prediction. In Ref. [6], ARMA model was adopted to forecast the tuple of wind speed and direction, and the forecasting results validated the effectiveness of this method. Cadenas and Rivera [7] proposed an ARIMA model to forecast the wind speed and the forecasting results were better than the persistence model. In recent years, a number of new time series models have been developed. In Ref. [8], a fractional-ARIMA (FARIMA) was used to forecast wind speed on the 24 h and 48 h horizons and showed significant improvements compared to ARIMA model. Liu et al. [9] proposed a modified Taylor Kriging model for wind speed prediction and the forecasting precision was found to be higher than ARIMA model. These models can explicitly reveal the linear relationship in the time series. However, they usually have unsatisfactory forecasting performance for the wind speed data with strong nonlinear features.

Different from the time series models, the nonlinear intelligent models can explain the nonlinear relationship between the input and output. Thus, they can provide better prediction results if the nonlinear characteristics of wind speed time series are prominent. For example, Cadenas and Rivera [10] proposed an artificial neural networks (ANN) for wind speed prediction and the forecasting results showed the proposed method could improve the prediction accuracy effectively; Mohandes et al. [11] introduced a support vector machine (SVM) to conduct the wind speed prediction and the prediction results were better than the multilayer perceptron (MLP); Zhou et al. [12] presented an optimal least squares SVM (LSSVM) based on better tuning LSSVM model parameters for wind speed forecasting and the prediction results were satisfactory. Every coin has two sides. Compared with the linear time series model, these nonlinear models may suffer from inefficient or over-fitted training and need more manual intervention for model parameters tuning [13].

Recently, the uncertainty of wind power prediction has been incorporated. For example, Qadrdan et al. [14] proposed an efficient method to generate probabilistic wind power forecast scenarios using singular spectrum analysis (SSA), Monte Carlo simulation and a scenario reduction algorithm. The final results validated the effectiveness of the approach. Bramati et al. [15] developed a new method for wind power prediction using the dynamic system of equations, which considered the uncertainty from the climate and the functioning of turbines. The prediction results showed the proposed method outperformed the benchmark model.

To further enhance the accuracy of the wind speed or power prediction, various hybrid approaches have been developed. For example, Liu et al. [16] proposed four different hybrid methods for high-precision multi-step wind speed predictions based on the adaptive boosting algorithm and MLP neural networks and the results showed these hybrid methods were effective. Meng et al. [17] adopted a new hybrid model for wind speed prediction based on wavelet packet decomposition (WPD), crisscross optimization algorithm and ANN, and the results showed the proposed method had significant advantage over the other reference models. These hybrid models mainly include weighting-based approaches, parameter optimization-based approaches, error correction-based approaches and decomposition-based approaches and could generally utilize the strengths of individual methods [5]. For instance, Han and Liu [18] proposed a weighting-based forecasting method where the maximum entropy principle was utilized to obtain the weight coefficient of persistence, ARIMA and four ANN-based models. The forecasting results showed that this hybrid method outperformed those single models. Shi et al. [19] proposed another weight-based hybrid model combining LSSVM and radial basis function neural network (RBFNN) based on grey relational analysis and wind speed distribution features. The results showed that the proposed model had significantly improved the forecasting accuracy. Ref. [2] addressed a parameter optimization-based forecasting method where the parameters of LSSVM model were optimized by gravitational search algorithm (GSA) and the results validated the effectiveness of the approach. Pousinho et al. [20] presented another parameter optimization-based forecasting method based on adaptive network-based fuzzy inference system (ANFIS). The forecasting results showed that particle swarm optimization (PSO)-based hybrid model could improve the prediction accuracy by comprehensive parameter selection. Liu et al. [21] presented an error correction-based forecasting method which used generalized auto-regressive conditionally heteroscedastic (GARCH) model to modify the ARMA model forecasting results. The results showed the performance of the ARMA-GARCH was satisfactory. Liang et al. [4] proposed another error correction-based method for multi-step ahead wind speed prediction where SVM or extreme learning machine (ELM) was established to forecast the error component and the simulation results demonstrated the effectiveness of the proposed model. Liu et al. [22] applied decomposition-based forecasting method which used the recursive ARIMA model to forecast individual subseries of wind speed time series based on empirical mode decomposition (EMD). The forecasting results showed the hybrid model had superior performance over the single ARIMA model. Hu et al. [23] investigated the possibility of improving the quality of wind speed forecasting by combining ensemble EMD (EEMD) and SVM. This model showed the better prediction capacity compared with other models. Sun and Liu [24] developed a hybrid model which combines fast EEMD (FEEMD) with regularized ELM for wind speed forecasting and the simulation results showed that the built model was effective and practicable. Liu et al. [25] had developed another four different hybrid models by combining four mainstream signal decomposing algorithm [e.g., wavelet decomposition (WD), WPD, EMD and FEEMD] and ELM for multi-step wind speed forecasting. This hybrid method integrated the advantages of individual models and improved the forecasting accuracy. Clearly, although the hybrid methods may increase the prediction accuracy, they introduce the complexity of the algorithm. Hence, it is necessary to balance the prediction accuracy and the complexity before the hybrid methods are used.

Focusing on the decomposition-based forecasting methods, Wang and Wu [13] pointed out that many of these methods based on the pre-processing scheme that all data, including the known data (the training data) and the unknown data (the forecasting data), were decomposed only once before prediction. This pre-processing scheme was unreasonable and violated the purpose of the wind speed prediction. In order to address this problem, they used another pre-processing scheme that the original data should be divided into training and forecasting parts, and the decomposition for training part should be real time. With newly obtained data, the training data should be updated and re-decomposed. However, the results showed that this real time decomposition-based forecasting methods may be ineffective in comparison with the single SVM or genetic algorithm-back propagation (GA-BP) ANN model.

Typically, the forecasting results of these decomposition-based methods are equal to the summation of all subseries forecasting results. However, recent studies show not all subseries are beneficial in forecasting the final prediction results. In Ref. [26], a hybrid of EMD and ANN was proposed where the highest frequency Intrinsic Mode Function (IMF) was removed and the remaining subseries were adopted to train ANN model in the prediction. The results showed the summation of the remaining subseries forecasting results had better forecasting accuracy. Zhang et al. [27] proposed an EMD and feature selection-based forecasting method where level numbers of selected subseries were fixed in every forecasting process. They found the forecasting results were satisfactory. According to Ref. [13], in real time decomposition-based forecasting method even though the updated training parts differ by one sample point, the decomposed subseries could change obviously. Therefore, the subseries from feature selection may change with the updated training part.

Additionally, the errors in decomposition-based methods are always neglected due to the hypothesis of white noise. However, it may be unreasonable in practice. In Ref. [28] an error prediction approach was proposed to revise the forecasting results of WD-LSSVM and the final results validated the effectiveness of error prediction approach.

This paper focuses on real time decomposition-based forecasting methods. The existence of illusive components introduced by decomposition method itself, and the interdependence structure and heteroscedasticity of errors are both considered in this study. First, the mean value is extracted from original time series and this zero-mean time series is decomposed by EEMD into a set of individual subseries. Second, a novel feature selection method based on kernel density estimation (KDE)-based Kullback-Leibler divergence (KLD) and energy measure is proposed to reduce the influence of the illusive components. After the illusive subseries are identified, the LSSVM model are established to conduct the one-step ahead prediction for the remaining subseries. Third, based on the analysis of the resulting error component, an error correction model is established to correct the above forecasting results. Fourth, the comparison of the proposed method with other forecasting methods including ARIMA, LSSVM, EEMD-LSSVM, feature selection-based EEMD-LSSVM and error correction-based EEMD-LSSVM is conducted, and prediction results show that the proposed approach has better performance in both accuracy and stability. Finally, the conclusions are provided.

2. Ensemble empirical mode decomposition and forecasting models

As discussed in introduction, only real time decomposition to manipulate data in prediction is reasonable. Therefore, this study focuses on the real time decomposition-based forecasting method. In this section, a brief review of EEMD, LSSVM and GARCH models will be presented, respectively.

2.1. Ensemble empirical mode decomposition

EMD is a data-driven and adaptive algorithm to analyze nonlinear and nonstationary signals. It can be regarded as an approximate quasi-dyadic filter bank. Then, the signal $x(t)$ can be decomposed into a number of amplitude-frequency modulated IMFs and a residue component via an iterative sifting process. The summation of decomposition results is shown in Eq. (1)

$$x(t) = \sum_{j=1}^{M+1} c_j(t) \tag{1}$$

where $c_j(t)(j = 1, 2, \ldots M)$ is IMF at level j, and $c_{M+1}(t)$ represents the residue. Detailed discussions about the derivation of the residue or trend can be found in Refs. [29,30].

Because EMD has the drawback of the mode mixing, a new noise-assisted EEMD followed by the same iterative sifting process as in EMD has been proposed by Wu and Huang [31]. In this algorithm, a number of Gaussian white noise samples are added to original signal to reinforce the signal in all frequency bands, thus the mode mixing can be curbed. The IMFs are then obtained by removing the white noise with the help of the statistical ensemble average. In this study, the ensemble size and the parameter for estimating the magnitude of the added white noise are 100 and 0.2, respectively. Additionally, the end effect exists in both EMD and EEMD, which can be reduced by the mirror symmetry method. However, this cannot be curbed completely. Recently, some variants of EMD have also been developed, such as FEEMD [24] and multivariate EMD (MEMD) [32]. The latter was adopted to decompose multivariate data.

2.2. Least squares support vector machine

After the original data is decomposed into different subseries by EEMD, a popular LSSVM algorithm is adopted to build the forecasting model for each subseries [2]. Suppose that the training part of each subseries consists of n-m data sets, $\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_{n-m}, y_{n-m})\}$, where $\boldsymbol{x}_i \in R^m$ (R denotes real number) is input vector and $y_i \in R$ is the corresponding output, i.e.,

$$\begin{aligned} &\boldsymbol{x_i} = [x(i) \quad x(i+1) \quad \cdots \quad x(i+m-1)]; \\ &y_i = x(i+m), \quad i = 1, 2, \ldots, n-m \end{aligned} \tag{2}$$

where n denotes the sample number of each subseries; m is the dimension of $\boldsymbol{x}_i$ and can be determined by minimizing the root mean square error using the output of training part according to Ref. [12]. The regression formula can be written as follows:

$$y = \boldsymbol{w}^T \phi(\boldsymbol{x_i}) + b, \boldsymbol{w} \in G, b \in R \tag{3}$$

where $\phi(\boldsymbol{x}_i)$ denotes the nonlinear mapping function transferring input space to high dimensional feature space G; y is the fitted value; $\boldsymbol{w}$ and b are the parameters that can be obtained by optimizing the following function:

$$\begin{aligned} &\min \ \mathrm{J}(\boldsymbol{w}, \xi) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{\gamma}{2}\sum_{i=1}^{n-m}\xi_i^2 \\ &\textit{Subject to}: \ y_i = \boldsymbol{w} \cdot \phi(\boldsymbol{x}_i) + b + \xi_i, \ i = 1, 2, \ldots, n-m \end{aligned} \tag{4}$$

where γ is the penalty factor to balance complexity and precision of the model; ξ_i is a residual error between the true output value at the time point i and its fitted value.

To solve the above optimization problem, the corresponding Lagrange function is constructed as:

$$L(\boldsymbol{w}, b, \xi, \boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{\gamma}{2}\sum_{i=1}^{n-m}\xi_i^2 - \sum_{i=1}^{n-m}\alpha_i[\boldsymbol{w} \cdot \phi(\boldsymbol{x}_i) + b + \xi_i - y_i] \tag{5}$$

where α_i (i = 1, 2, ..., n-m) is the Lagrange multipliers.

After computing the partial derivatives of $L(\boldsymbol{w}, b, \xi, \boldsymbol{\alpha})$ with respect to $\boldsymbol{w}$, b, ξ_i and α_i, the nonlinear regression prediction function can be expressed as:

$$\hat{y} = \sum_{i=1}^{n-m}\alpha_i \cdot K(\boldsymbol{x}, \boldsymbol{x}_i) + b \tag{6}$$

where $\hat{y}$ is the predicted value; $\boldsymbol{x}$ denotes the latest input vector ($\boldsymbol{x} \in R^m$); $K(\boldsymbol{x}, \boldsymbol{x}_i) = \phi(\boldsymbol{x})^T \times \phi(\boldsymbol{x}_i)$ (i = 1, 2, ... , n-m) is kernel function, satisfying Mercer's conditions [33]. More details of LSSVM can be found in Ref. [12].

2.3. Generalized auto-regressive conditionally heteroscedastic model

GARCH model is a predominant model for modeling and estimating volatility in Ref. [21]. Based on the theory of LSSVM, the true value at time i can be given by:

$$y_i = \hat{y} + \xi_i \tag{7}$$

in which ξ_i is usually assumed to be independently identically distribution [34]. If ξ_i has the time-varying variance (i.e., heteroscedasticity), it can be modeled as:

$$\xi_i = \sqrt{h_i} \upsilon_i \tag{8}$$

where υ_i is a white noise sequence with zero mean and unit variance; h_i is the conditional variance and can be expressed as:

$$h_i = \varsigma_0 + \sum_{k=1}^{p} \varsigma_k h_{i-k} + \sum_{l=1}^{q} \eta_l \xi_{i-l}^2$$
$$\begin{cases} \varsigma_0 > 0 & \\ \varsigma_k \geqslant 0 & k = 1, 2 \ldots p \\ \eta_l \geqslant 0 & l = 1, 2 \ldots q \\ \sum_{k=1}^{p} \varsigma_k + \sum_{l=1}^{q} \eta_l < 1 & \end{cases} \tag{9}$$

where η_l and ς_k are nonnegative varying coefficients. In this notation, the current conditional variance depends on the previous conditional variance. That is, the error ξ_i follows GARCH(p,q) model. Specially, when p and q are both equal to one, GARCH(1,1) stands for the standard GARCH model and is adopted in this study for its simplicity.

Before implementing GARCH model, the Lagrange Multiplier (LM) statistic is used to test the heteroscedasticity of error component [35], which can be given as:

$$\mathrm{LM} = nr^2 \tag{10}$$

where r^2 is the fitting goodness. If LM statistic is larger than $\chi^2(q)$, it means that the heteroscedasticity exists in the error component, and GARCH model should be adopted to modify the forecasting results of LSSVM model.

3. Proposed method

As presented in introduction, the decomposition-based forecasting models mainly focus on different decomposition methods and forecasting models, and the final forecasting result is generally equal to the summation of all subseries forecasting results. However, some recent studies show that the summation of some subseries forecasting results may have better prediction accuracy [26,27]. This could be attributed to the existence of illusive components introduced by the decomposition methods. In this study, a pre-processing feature selection is used to conduct illusive components identification among all decomposed subseries and reduce their influence on wind speed prediction. To further improve the forecasting accuracy, a post-processing error forecast correction is adopted. The flowchart of proposed model is shown in Fig. 1 and the corresponding procedure is described as follows:

Step 1: Divide the original data into two parts, including the training part ($\{x'(1), \ldots, x'(n)\}$) and the forecasting part ($\{x'(n+1), \ldots, x'(n+N)\}$).

Step 2: Extract the mean value $\bar{x}[\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x'(t)]$ from the training part and obtain zero-mean training part, denoted by $\{x(1), \ldots, x(n)\}$. Then, decompose this zero-mean time series by EEMD into a number of individual subseries ($\{c_j(1), \ldots, c_j(n)\}$, $j = 1 \ldots M+1$) while the forecasting part is assumed to be unavailable.

Step 3: Apply the KLD and energy-based criteria for feature selection among all subseries. Eliminate the illusive subseries and reserve the remaining subseries.

Step 4: Build LSSVM to forecast the $n+1$th data, $\hat{c}_j(n+1)$, from the 1st-nth data of the remaining subseries. Obtain the resulting error component ($\{e_1, e_2, \ldots, e_n\}$). Analyze this error component and establish the suitable model to forecast the $n+1$th error, $\hat{e}_{n+1}$. Summarize these predicting results and mean value $\bar{x}$, and obtain the final forecasting result, $\hat{x}(n+1)$. See details in Section 4.2.

Step 5: Obtain actual one-data point and update the training data to $\{x'(2), \ldots, x'(n+1)\}$. Repeat the Steps 2–4, the corresponding forecasting result, $\hat{x}(n+2)$, can be obtained. By analogy, continue the one-step ahead prediction until all of the forecasting tasks are completed.

Step 6: Calculate the forecasting error via the forecasting data set.

3.1. Feature selection

3.1.1. Kullback-Leibler divergence-based feature selection

In this section, a KLD-based feature selection method is proposed to reduce the influence of illusive components [36]. Suppose that the probability density functions (PDFs) of original zero-mean wind speed time series $\{x(t)\}$ and the jth subseries $\{c_j(t)\}$ are $f(x)$ and $g(c_j)$, respectively. KLD known as the relative entropy between these two PDFs is given by [36]:

$$\mathrm{KLD}(f \| g) = \sum_{x \in \{x(t)\}} f(x) \log \frac{f(x)}{g(x)} \tag{11}$$

$$\mathrm{KLD}(g \| f) = \sum_{c_j \in \{c_j(t)\}} g(c_j) \log \frac{g(c_j)}{f(c_j)} \tag{12}$$

where the KLD satisfies two properties: (i) if the value of KLD is large, the difference between two signals is large, otherwise the difference is small; (ii) KLD = 0 denotes these two signals are the same.

In this study, a symmetric version of KLD is adopted because the symmetry property is very crucial in the classification issue [36], which is shown as follows:

$$\mathrm{KLD} = \frac{1}{2}\left(\sum_{x \in \{x(t)\}} f(x) \log \frac{f(x)}{g(x)} + \sum_{c_j \in \{c_j(t)\}} g(c_j) \log \frac{g(c_j)}{f(c_j)} \right) \tag{13}$$

Here, a non-parametric kernel smoothing method is proposed to determine $f(x)$ and $g(c_j)$. This approach attempts to estimate the PDF from data without assuming a particular distribution form. Taking an example of the time series $\{x(t)\}$, the KDE is [36]

$$f(x) \approx \frac{1}{nh} \sum_{t=1}^{n} K\left(\frac{x - x(t)}{h} \right) \tag{14}$$

where $K(\cdot)$ is a symmetric kernel function with integration equal to one and has many possible choices [37]. In this study, Gaussian kernel function is commonly used due to its computational advantage and defined as

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2} \right) \tag{15}$$

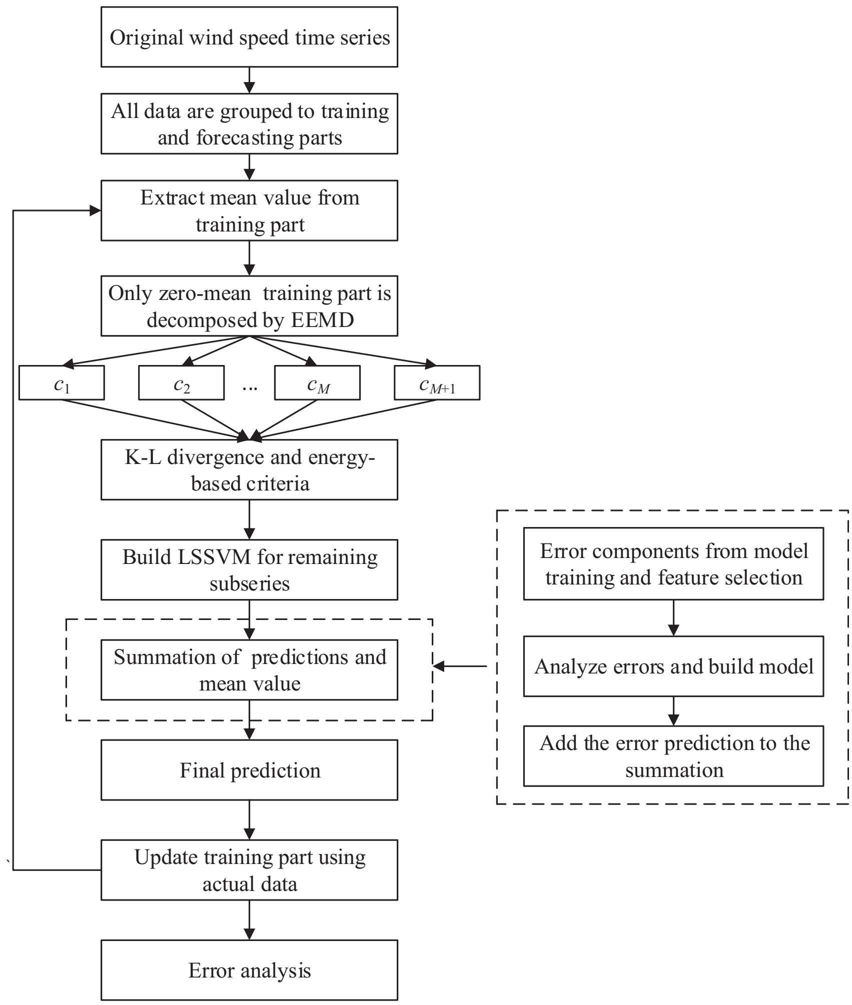


Fig. 1. Flowchart of proposed method.

and $h > 0$ is a smoothing parameter called the band width, which is given as [36]:

$$h = \left(\frac{4\sigma^5}{3n}\right)^{1/5} \tag{16}$$

in which σ represents the estimate of sample standard deviation. Thus, $f(x)$ will be acquired by substituting Eqs. (15) and (16) into Eq. (14). Similarly, $g(c_j)$ can also be obtained. With $f(x)$ and $g(c_j)$, the symmetric KLD between time series $\{x(t)\}$ and each subseries can be gained by Eq. (13). Here, the maximum among the KLDs is selected and the corresponding subseries should be eliminated as the illusive component.

3.1.2. Energy-based feature selection

In order to further eliminate the illusive components which may not be detected by KLD, a direct measure for signal strength (i.e., energy measure) is proposed [38]. The energy of each subseries is given by

$$E_{c_j(t)} = \sum_{t=1}^{n} [c_j(t)]^2 \tag{17}$$

Here, the subseries containing the least energy should be eliminated as the illusive component. Based on the above two feature selection criteria, the influence of illusive components can be reduced.

3.2. Error forecast correction

After feature selection is completed, the LSSVM model will be adopted to conduct one-step ahead prediction for the remaining subseries. Here, the error component comes from illusive subseries identification and model training.

To obtain higher forecasting accuracy, an error correction method is proposed. First, the interdependence structure of error component is analyzed by autocorrelation (ACF) and partial ACF (PACF) to determine whether it is necessary to build forecasting model for this error component [39]. Then, the heteroscedasticity of the error component is examined by LM statistic test to determine whether it is necessary to build GARCH model. Based on the above analysis results, four situations may exist in this study and the corresponding error prediction models are: (i) when the correlation and heteroscedasticity of the error component are obvious, the LSVVM-GARCH model is adopted; (ii) when both characteristics are negligible, it may be unnecessary to conduct error correction; (iii) if only correlation exists in the error component, the LSSVM model is adopted; (iv) if only heteroscedasticity exists apparently, the GARCH model is used.

4. Case study

To examine the effectiveness and reliability of the proposed method, a collection of wind speed data (data collection 1) measured from Colorado, USA (including 300 samples) is adopted, as shown in Fig. 2. Here, the original time series is divided into two parts: the 1st-225th sampling points for training and the subsequent 226th-300th data for testing. The training part is used to establish the forecasting model, and the remaining part is utilized to verify the performance of proposed model.

4.1. Evaluation criteria

In order to quantitatively evaluate the accuracy and stability of proposed model, the mean absolute error (MAE), mean relative percentage error (MRPE), root mean square error (RMSE) and root mean square relative error (RMSRE) are utilized in this study, which are given as:

$$\begin{aligned} \mathrm{MAE} &= \frac{1}{N'}\sum_{t=n+1}^{n+N'}|x'(t)-\hat{x}(t)| \\ \mathrm{MRPE} &= \frac{1}{N'}\sum_{t=n+1}^{n+N'}\left|\frac{x'(t)-\hat{x}(t)}{x'(t)}\right| \\ \mathrm{RMSE} &= \sqrt{\frac{1}{N'}\sum\nolimits_{t=n+1}^{n+N'}(x'(t)-\hat{x}(t))^2} \\ \mathrm{RMSRE} &= \sqrt{\frac{1}{N'}\sum\nolimits_{t=n+1}^{n+N'}\left(\frac{x'(t)-\hat{x}(t)}{x'(t)}\right)^2} \end{aligned} \tag{18}$$

where $\{x'(t)\}$ and $\{\hat{x}(t)\}$ represent the measured data and the predicted data at the time t, respectively; N' is the number of the data for performance evaluation (for one-step ahead prediction, N' is equal to the number of the data in forecasting part, i.e., $N' = N$).

Additionally, in order to confirm whether the above four evaluative criteria are statistically significant, the one-sided Kolmogorov-Smirnov (KS) test is adopted [40]. This test is to determine whether the prediction results from the proposed method possess stochastically smaller errors in comparison with the reference model. Specifically, if the cumulative distribution function (CDF) of forecasting errors from proposed model lies at the upper of that from the reference model, it means the proposed method has a lower stochastic error than the reference model. Then, the hypothesis on the one-sided KS test can be approximately shown as follows:

$$\begin{aligned} H_0 &: F(z) \leqslant F_v(z), \\ H_1 &: F(z) > F_v(z) \end{aligned} \tag{19}$$

where the $F(z)$ and $F_v(z)$ denote the empirical CDFs for the forecasting errors from proposed method and the reference model, respectively. For KS test, the decision making process relies on the calculation of probability value (p-value) [40]. If the p-value of one sided KS test is smaller than the significant level, we reject the null hypothesis (i.e., the prediction accuracy has statistically significant difference and the proposed model reports stochastically smaller errors).

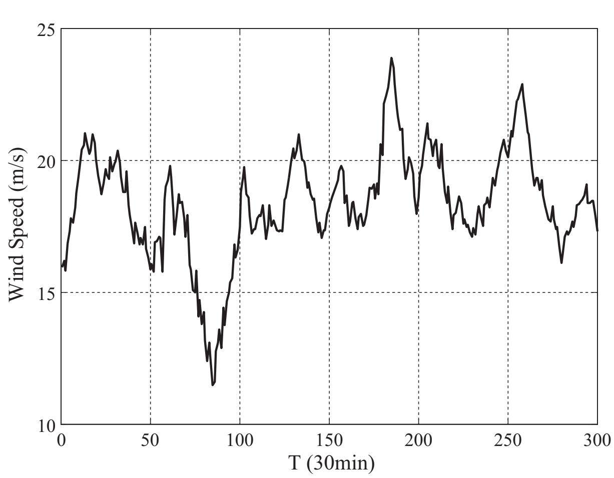


Fig. 2. Wind speed data collection 1.

In this study, the significant level is selected as 10% and the forecasting errors adopt the absolute errors and relative errors, respectively. According to Eq. (18), the KS test results of MAE and RMSRE are the same with these of RMSE and MRPE, respectively. Therefore, only the KS test results of MAE and RMSRE will be shown hereafter. Detailed information about KS test can be found in Ref. [40].

4.2. Wind speed prediction

Fig. 3 shows the zero-mean data and its decomposition results from the 1st-225th data points in data collection 1. Fig. 4 illustrates all the subseries arranged according to their respective magnitude in KLD and energy. As shown in Fig. 4, two IMFs (c_1–c_2) are regarded as illusive components based on the above two feature selection criteria. Then, the LSSVM models should be established for the remaining subseries (c_3–c_6) to conduct one-step ahead prediction.

According to Section 2.2, the dimension m for c_3-c_6 is 4, 6, 7 and 8, respectively. Here, only the m + 1th-225th data points in each subseries have training results. Thus, only the 9th-225th samples in all errors due to feature selection and model training are effective, which are denoted as error component I and shown in Fig. 5.

Fig. 6 shows the analysis results of the error component I by ACF and PACF. It is obvious that a significant correlation relationship exists in error component I. Then, the LSSVM model is adopted to model this error component. The dimension m for this error component is 5 by minimizing RMSE of the output of model training. Therefore, only the 6th to the last data points in error component I have effective training results and the corresponding model training errors are shown in Fig. 7, denoted as error component II.

Fig. 8 shows the ACF and PACF of the error component II. As seen from Fig. 8, there exists no significant correlation relationship in error component II. Therefore, under the homoscedastic hypothesis, the LSSVM model is enough for error component I. However, according to Section 2.3, the value of LM statistic (the selected test order is 2) for the error component II is 30.7, which is larger than $\chi^2_{0.05}(2) \cong 6.0$ [34]. That is, the heteroscedasticity obviously exists in error component II.

Based on the above analysis results, the correlation and heteroscedasticity of the error component cannot be neglected. Therefore, the LSSVM-GARCH model should be adopted in this situation to conduct the error prediction. The final forecasting result is equal to the summation of the mean value, remaining subseries forecasting results and the error correction result. Similarly, the above procedures are performed for other training parts and the corresponding forecasting results will be obtained.

4.3. Forecasting results and discussions

To illustrate the performance of the proposed method, six forecasting models are constructed: ARIMA model, LSSVM model, proposed model and other three models as shown in Table 1. One-step ahead predictions are carried out based on these six models and the estimated error results of these predictions are given in Table 2. Compared with the other five models, the improvements by the proposed method are summarized in Table 3. Fig. 9 shows the forecasting results at 226th-300th data based on the proposed method,

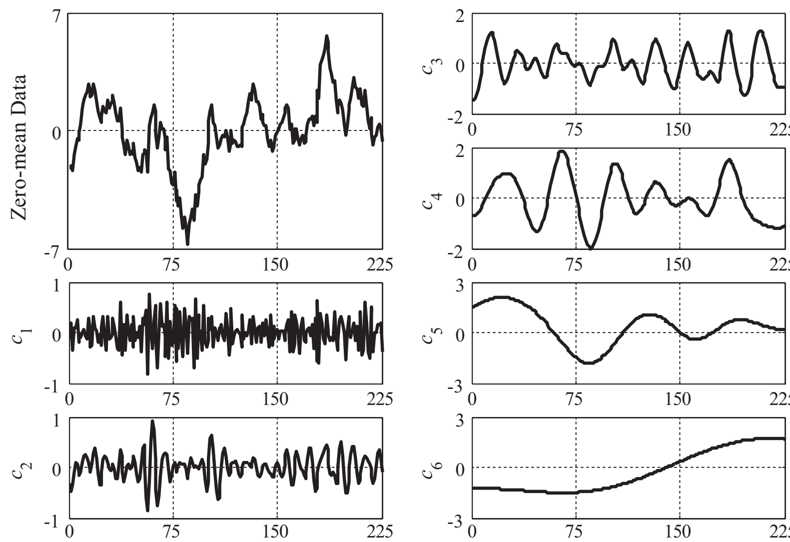
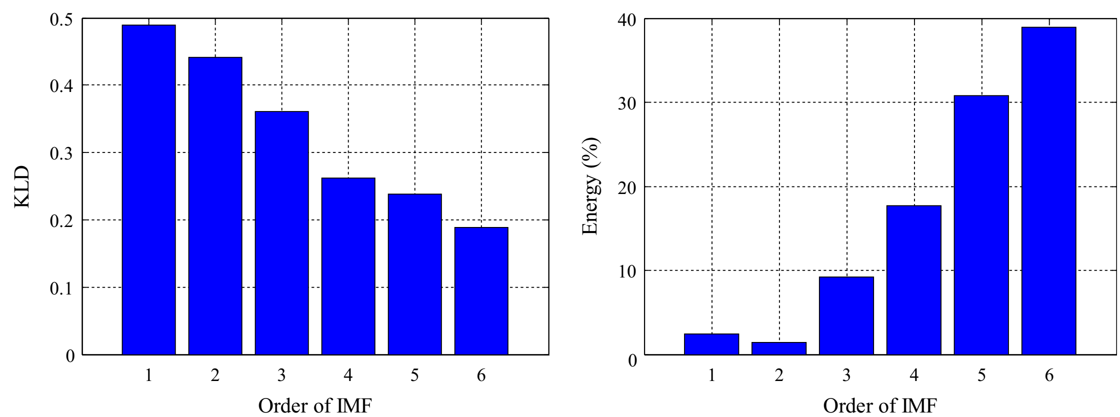


Fig. 3. Subseries decomposed from the first 225 samples of data collection 1.



(a) KLD of each subseries

(b) Energy of each subseries

Fig. 4. KLD and energy for each subseries.

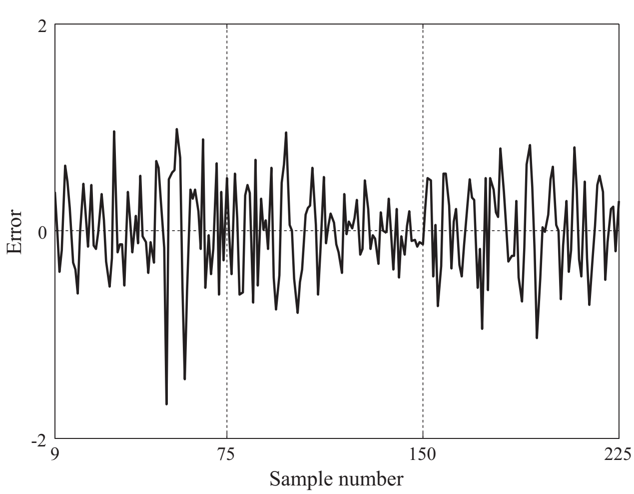


Fig. 5. Error component I.

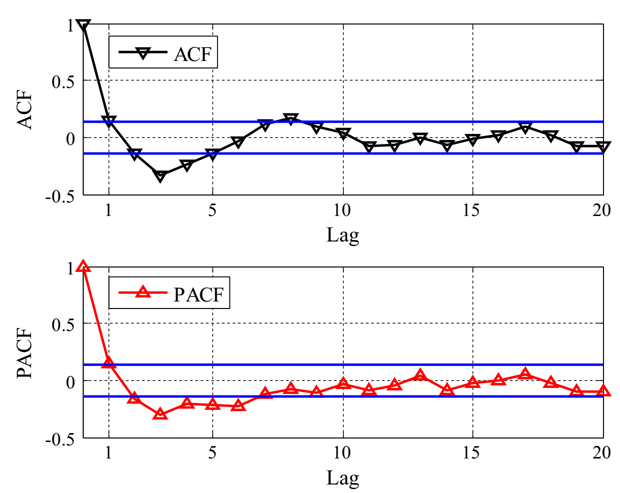


Fig. 6. The ACF and PACF of error component I with 95% confidence interval.

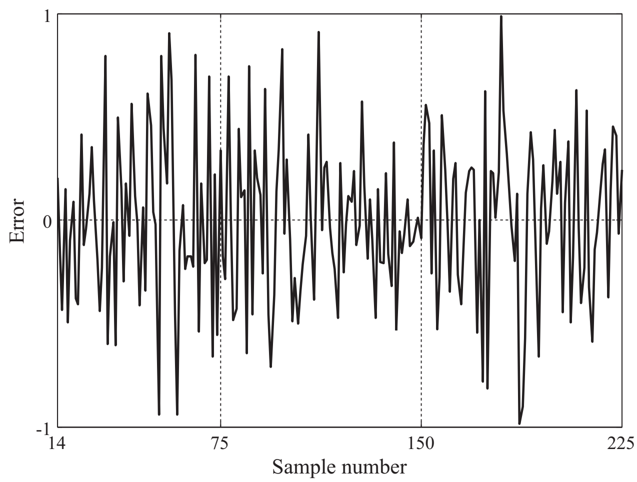


Fig. 7. Error component II.

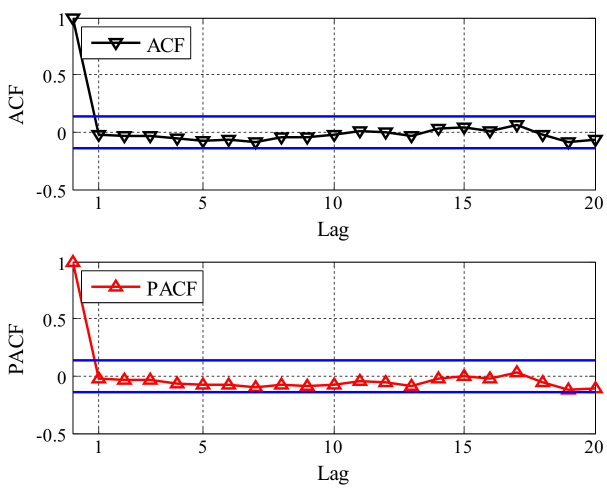


Fig. 8. The ACF and PACF of error component II with 95% confidence interval.

Table 1
Other three forecasting models.

Model 1	EEMD-LSSVM
Model 2	Error correction-based EEMD-LSSVM
Model 3	Feature selection-based EEMD-LSSVM

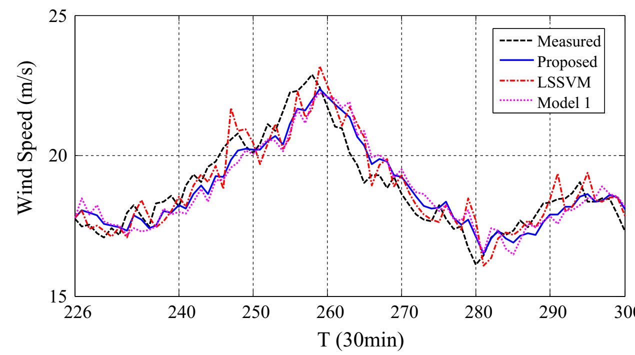


Fig. 9. Forecasting results by proposed method, LSSVM and model 1 (data collection 1).

the single LSSVM model and model 1. As seen from Tables 2–3 and Fig. 9, it can be observed as follows:

(1) Compared with the single ARIMA model, the single LSSVM model has better forecasting performance. For example, the MAE and MRPE of the single LSSVM model are 0.577 and 0.030, respectively. The MAE and MRPE of the single ARIMA model are 0.618 and 0.033, respectively. This can be attributed to the fact that there exists the nonlinearity in the data.
(2) Compared with the other five involved models, model 1 even perform worse overall. For instance, the four indexes (MAE, RMSE, MRPE and RMSRE) of the single LSSVM model are 0.577, 0.703, 0.030 and 0.037, respectively, while those of model 1 are 0.645, 0.772, 0.034 and 0.040, respectively. Therefore, considering the prediction accuracy and computation time, the real time decomposition-based method may be ineffective in practice. Clearly, although the nonstationarity of the original data can be reduced apparently, the real time decomposition may increase the other forecasting difficulties. One could be the appearance of illusive components introduced by the decomposition method.
(3) Models 2 and 3 both can improve the forecasting performance in comparison with model 1. It means that the incorporations of feature selection and error correction are all beneficial to the prediction. For example, when comparing model 3 with model 1, the former one has a little improvement in the overall performance. MAE, RMSE, MRPE and RMSRE decrease by 0.068, 0.131, 0.003 and 0.006, respectively.

Table 2
Estimated results of six forecasting models (data collection 1).

Indices	LSSVM	Proposed	Model 1	Model 2	Model 3	ARIMA
MAE	0.577	0.500	0.645	0.624	0.577	0.618
RMSE	0.703	0.605	0.772	0.731	0.641	0.706
MRPE	0.030	0.026	0.034	0.032	0.031	0.033
RMSRE	0.037	0.032	0.040	0.037	0.034	0.037

Table 3
Improvements by proposed method (data collection 1).

Indices	Proposed vs. LSSVM	Proposed vs. Model 1	Proposed vs. Model 2	Proposed vs. Model 3	Proposed vs. ARIMA
MAE (%)	13.269	22.491	19.905	13.297	19.010
RMSE (%)	13.931	21.660	17.305	5.663	14.281
MRPE (%)	13.499	22.662	17.723	14.015	19.558
RMSRE (%)	14.526	21.727	15.385	6.798	15.362

Table 4
Evaluation MAE and RMSRE of different models from statistical significance (data collection 1).

One-sided (p-value)	Proposed vs. LSSVM	Proposed vs. Model 1	Proposed vs. Model 2	Proposed vs. Model 3	Proposed vs. ARIMA
MAE	0.070*	0.018*	0.050*	0.065*	0.011*
RMSRE	0.047*	0.018*	0.036*	0.043*	0.018*

* Note: indicates results are statistically significant based on a significant level of 0.1.

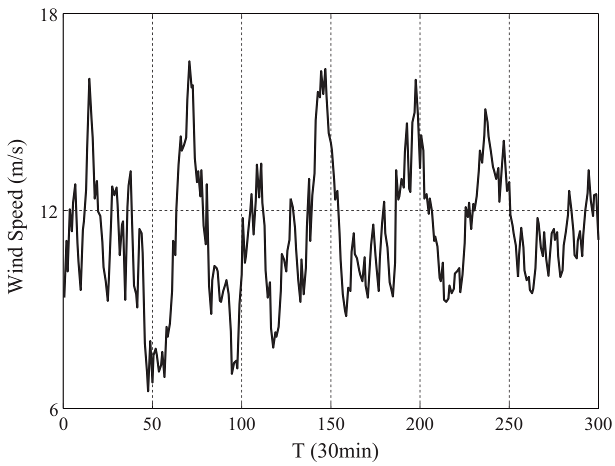


Fig. 10. Wind speed data collection 2.

(4) Compared with the single LSSVM model, both models 2 and 3 may not provide satisfactory forecasting results. It means only one of modifications may be not enough to provide satisfactory forecasting accuracy.
(5) Model 3 is superior to model 2. This indicates that the reduction of the disturbance of illusive components may be more effective than error correction in improving the forecasting accuracy.
(6) The proposed method outperforms the other five models noticeably, which takes advantage of the strengths of aforementioned models 2 and 3. For instance, compared with model 3, the improvements by proposed method in terms of MAE, RMSE, MRPE and RMSRE are 13.297%, 5.663%, 14.015% and 6.798%, respectively. The reason could be that the combination of error correction and feature selection is more effective in enhancing the forecasting accuracy.

Table 4 shows the KS test results of MAE and RMSRE. From Table 4, it can be found that the one-sided KS test provides sufficient evidence to conclude that the proposed method reports a stochastically lower errors in comparison with the other five models. Therefore, the KS tests for MAE and RMSRE illustrate that the evaluative criteria in this study have statistical significance and also show the proposed method has better forecasting accuracy from the statistical perspective.

5. Additional case

In order to examine the stability of the proposed method, another data (data collection 2) is used to make one-step ahead prediction. Fig. 10 shows the wind speed data measured from Minnesota in USA. For simplicity, only the evaluations are given in Tables 5–7. The forecasting results at 226th-300th data are shown in Fig. 11.

From Tables 5–7 and Fig. 11, the main results are similar to those from data collection 1 and can be given that: (1) the real time EEMD-LSSVM forecasting method may be ineffective in comparison with other models. (2) compared with model 1, both models 2 and 3 can promote the forecasting accuracy in some extent, but their results are unsatisfactory in comparison with LSSVM model; (3) LSSVM model is better than ARIMA model; (4) the pro-

Table 5
Estimated results of six forecasting models (data collection 2).

Indices	LSSVM	Proposed	Model 1	Model 2	Model 3	ARIMA
MAE	0.818	0.732	0.907	0.885	0.835	0.859
RMSE	1.008	0.881	1.084	1.067	0.979	1.038
MRPE	0.070	0.063	0.077	0.075	0.072	0.074
RMSRE	0.085	0.075	0.091	0.091	0.085	0.089

Table 6
Improvements by proposed method (data collection 2).

Indices	Proposed vs. LSSVM	Proposed vs. Model 1	Proposed vs. Model 2	Proposed vs. Model 3	Proposed vs. ARIMA
MAE (%)	10.533	19.286	17.286	12.324	17.369
RMSE (%)	12.547	18.728	17.438	9.992	17.759
MRPE (%)	10.072	18.653	17.035	12.663	18.350
RMSRE (%)	12.436	18.147	17.432	11.878	19.611

Table 7
Evaluation MAE and RMSRE of different models from statistical significance (data collection 2).

One-sided (p-value)	Proposed vs. LSSVM	Proposed vs. Model 1	Proposed vs. Model 2	Proposed vs. Model 3	Proposed vs. ARIMA
MAE	0.075*	0.043*	0.057*	0.067*	0.018*
RMSRE	0.083*	0.054*	0.065*	0.077*	0.023*

* Note: indicates results are statistically significant based on a significant level of 0.1.

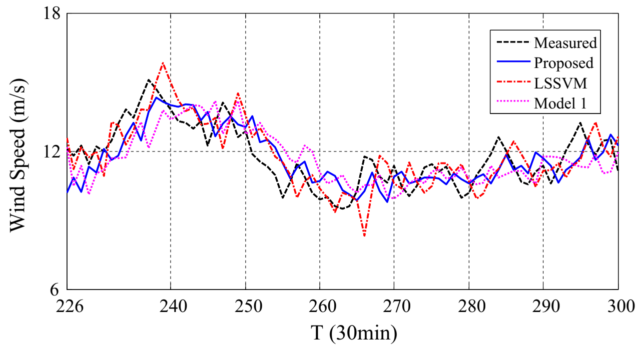


Fig. 11. Forecasting results by proposed method, LSSVM and model 1 (data collection 2).

posed method shows advantages over LSSVM model, ARIMA model and models 1–3; (5) the evaluative criteria in this study have statistical significance and the proposed method shows advantages over other models from the statistical perspective.

6. Conclusion

Although the real time decomposition-based forecasting methods are reasonable, they may be ineffective in practice. To address this difficulty, a novel hybrid method is proposed, which combines EEMD, feature selection and error correction. Two case studies based on measured data are provided in this paper and some conclusions can be summarized as follows:

(1) As pointed by Wang and Wu [13], real time decomposition-based forecasting methods may be ineffective in practical application. Although the nonstationarity of the original data can be reduced apparently, the real time decomposition may increase another difficulty. This could be the appearance of the illusive components caused by the decomposition.
(2) The incorporation of either the feature selection or the error correction can promote the forecasting accuracy to some extent for real time decomposition-based forecasting methods. However, the forecasting results may be unsatisfactory in comparison with the single LSSVM.
(3) In the proposed method, the KDE-based KLD method and energy measure are adopted for the feature selection in order to decrease the influence of illusive components among subseries, and a LSSVM-GARCH model is used to model the corresponding error component depending on the interdependence structure and heteroscedasticity of errors. Results show that the proposed method has the satisfactory performance in both accuracy and stability, compared with other methods, including ARIMA, LSSVM, EEMD-LSSVM, feature selection-based EEMD-LSSVM and error correction-based EEMD-LSSVM.

The major contribution of this study is that the novel hybrid method with the satisfactory performance in both accuracy and stability is proposed after the factors attributed to the unsatisfactory performance of the real time decomposition-based forecasting method have been inferred. Similar to other prediction methods, the proposed method needs further efforts to improve its performance. A few future tasks should be carried out and are given below.

(1) The existence of mode mixing and end effect from the decomposition method should be curbed more effectively, which deserves the in-depth study.
(2) In the current analysis for the error component, only the simple GARCH(1,1) model is adopted. The more suitable model should be analyzed in the future.
(3) A probabilistic wind speed forecasting algorithm should be developed to further consider the uncertainties of the wind speed.

Acknowledgements

Supports from the National Natural Science Foundation of China (Grant Nos. U1334201 and 51578471) are greatly acknowledged. In addition, the authors would like to thank the editor and reviewers for precious comments and suggestions.

References

[1] World Wing Energy Association. World Wind Energy Report; 2009.
[2] Yuan XH, Chen C, Yuan YB, Huang YH, Tan QX. Short-term wind power prediction based on LSSVM–GSA model. Energy Convers Manage 2015;101:393–401.
[3] Hu JM, Wang JZ. Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression. Energy 2015;93:1456–66.
[4] Liang ZT, Liang J, Wang CF, Dong XM, Miao XF. Short-term wind power combined forecasting based on error forecast correction. Energy Convers Manage 2016;119:215–26.
[5] Tascikaraoglu A, Uzunoglu M. A review of combined approaches for prediction of short-term wind speed and power. Renew Sustain Energy Rev 2014;34:243–54.
[6] Erdem E, Shi J. ARMA based approaches for forecasting the tuple of wind speed and direction. Appl Energy 2011;88(4):1405–14.
[7] Cadenas E, Rivera W. Wind speed forecasting in the south coast of Oaxaca, Mexico. Renewable Energy 2007;32(12):2116–28.
[8] Katris C, Daskalaki S. Comparing forecasting approaches for Internet traffic. Expert Syst Appl 2015;42(21):8172–83.
[9] Liu HP, Shi J, Erdem E. Prediction of wind speed time series using modified Taylor Kriging method. Energy 2010;35(12):4870–9.
[10] Cadenas E, Rivera W. Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks. Renew Energy 2009;34(1):274–8.
[11] Mohandes MA, Halawani TO, Rehman S, Hussain AA. Support vector machines for wind speed prediction. Renew Energy 2004;29(6):939–47.
[12] Zhou JY, Jing S, Gong L. Fine tuning support vector machines for short-term wind speed forecasting. Energy Convers Manage 2011;52(4):1990–8.
[13] Wang YM, Wu L. On practical challenges of decomposition-based hybrid forecasting algorithms for wind speed and solar irradiation. Energy 2016;112:208–20.
[14] Qadrdan M, Ghodsi M, Wu JZ. Probabilistic wind power forecasting using a single forecast. Int J Energy Stat 2013;1(02):99–111.
[15] Bramati MC, Arezzo MF, Pellegrini G. Short-term wind power forecasting based on dynamic system of equations. Int J Energy Stat 2016;4(03):1650012.
[16] Liu H, Tian HQ, Li YF, Zhang L. Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions. Energy Convers Manage 2015;92:67–81.
[17] Meng AB, Ge JF, Yin H, Chen SZ. Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm. Energy Convers Manage 2016;114:75–88.
[18] Han S, Liu YQ. The study of wind power combination prediction. In: Proceedings of the Asia-Pacific power and energy engineering conference (APPEEC).
[19] Shi J, Ding ZH, Lee WJ, Yang YP, Liu YQ, Zhang MM. Hybrid forecasting model for very-short term wind power forecasting based on grey relational analysis and wind speed distribution features. IEEE Trans Smart Grid 2014;5(1):521–6.
[20] Pousinho HMI, Mendes VMF, Catalão JPDS. A hybrid PSO–ANFIS approach for short-term wind power prediction in Portugal. Energy Convers Manage 2011;52(1):397–402.
[21] Liu HP, Erdem E, Shi J. Comprehensive evaluation of ARMA–GARCH(-M) approaches for modeling the mean and volatility of wind speed. Appl Energy 2011;88(3):724–32.
[22] Liu H, Tian HQ, Li YF. An EMD-recursive ARIMA method to predict wind speed for railway strong wind warning system. J Wind Eng Ind Aerodyn 2015;141:27–38.
[23] Hu JM, Wang JZ, Zeng GW. A hybrid forecasting approach applied to wind speed time series. Renewable Energy 2013;60:185–94.
[24] Sun W, Liu MH. Wind speed forecasting using FEEMD echo state networks with RELM in Hebei, China. Energy Convers Manage 2016;114:197–208.
[25] Liu H, Tian HQ, Li YF. Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms. Energy Convers Manage 2015;100:16–22.
[26] Guo ZH, Zhao WG, Lu HY, Wang JZ. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renew Energy 2012;37(1):241–9.

[27] Zhang C, Wei HK, Zhao JS, Liu TH, Zhu TT, Zhang KJ. Short-term wind speed forecasting using empirical mode decomposition and feature selection. Renew Energy 2016;96:727–37.
[28] Duan XW, Li QQ, Wang RQ, Zhang Y, Sun SM, Lang CY. An error-revision-based method for very short-term wind speed prediction using wavelet transform and support vector machine. In: International Conference on Control, Automation and Information Sciences (ICCAIS), 2015. p. 197–202.
[29] Breaker LC. Energy production trend extraction using ensemble empirical mode decomposition. Int J Energy Stat 2013;1(03):195–204.
[30] Su YW, Huang GQ, Xu YL. Derivation of time-varying mean for non-stationary downburst winds. J Wind Eng Ind Aerodyn 2015;141:39–48.
[31] Wu ZH, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv Adaptive Data Anal 2009;1(1):1–41.
[32] Huang GQ, Su YW, Kareem A, Liao HL. Time-Frequency analysis of nonstationary process based on multivariate empirical mode decomposition. J Eng Mech 2015;142(1):04015065.
[33] Vapnik VN, Chervonenkis AJ. Theory of pattern recognition; 1974.
[34] Zhu SM, Yang M, Han XS. Short-term generation forecast of wind farm using SVM-GARCH approach. In: IEEE International Conference on Power System Technology (POWERCON); 2012. p. 1–6.
[35] Bollerslev T, Wooldridge JM. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econ Rev 1992;11(2):143–72.
[36] Zhang F, Liu Y, Chen CJ, Li YF, Huang HZ. Fault diagnosis of rotating machinery based on kernel density estimation and Kullback-Leibler divergence. J Mech Sci Technol 2014;28(11):4441–54.
[37] Epanechnikov VA. Non-parametric estimation of a multivariate probability density. Theory Probab Appl 1969;14(1):153–8.
[38] Yan RQ, Gao RX. Rotary machine health diagnosis based on empirical mode decomposition. J Vib Acoust 2008;130(2):21007.
[39] Schwarz G. Estimating the dimension of a model. Ann Stat 1978;6(2):461–4.
[40] Hassani H, Silva ES. A Kolmogorov-Smirnov based test for comparing the predictive accuracy of two sets of forecasts. Econometrics 2015;3(3):590–609.