

Article

Short-Term Wind Speed Forecasting Based on Hybrid Variational Mode Decomposition and Least Squares Support Vector Machine Optimized by Bat Algorithm Model

Qunli Wu ^{1,2} and Huaxing Lin ^{1,*}

¹ Department of Economics and Management, North China Electric Power University, 689 Huadian Road, Baoding 071003, China; 51851220@ncepu.edu.cn

² Beijing Key Laboratory of New Energy and Low-Carbon Development, North China Electric Power University, Beijing 102206, China

* Correspondence: 2172218066@ncepu.edu.cn

Received: 24 December 2018; Accepted: 23 January 2019; Published: 26 January 2019

Abstract: With the integration of wind energy into electricity grids, wind speed forecasting plays an important role in energy generation planning, power grid integration and turbine maintenance scheduling. This study proposes a hybrid wind speed forecasting model to enhance prediction performance. Variational mode decomposition (VMD) was applied to decompose the original wind speed series into different sub-series with various frequencies. A least squares support vector machine (LSSVM) model with the pertinent parameters being optimized by a bat algorithm (BA) was established to forecast those sub-series extracted from VMD. The ultimate forecast of wind speed can be obtained by accumulating the prediction values of each sub-series. The results show that: (a) VMD-BA-LSSVM displays better capacity for the prediction of ultra short-term (15min) and short-term (1h) wind speed forecasting; (b) the proposed forecasting model was compared with wavelet decomposition (WD) and ensemble empirical mode decomposition (EEMD), and the results indicate that VMD has stronger decomposition ability than WD and EEMD, thus, significant improvements in forecasting accuracy were obtained with the proposed forecasting models compared with other forecasting methods.

Keywords: Wind speed forecasting; variational mode decomposition (VMD); least squares support vector machine (LSSVM); bat algorithm (BA)

1. Introduction

Wind power has been recognized as one of the most major and efficient renewable energy and has been extensively applied throughout the world [1]. With the rapid development of wind power generation, wind speed forecasting has become a hot issue in the field of power generation research, due to its important role in energy generation planning, power grid integration and turbine maintenance scheduling [2]. For example, a 10% deviation in the expected wind speed leads to an approximate 30% deviation in the expected wind power generation because the power potential is proportional to the cubic power of the wind speed [3]. Therefore, it is of great importance to develop relatively accurate wind speed forecasting models.

In the technical literature, the short-term wind speed forecasting methods can be classified into four categories: (a) the physical model; (b) the traditional statistical model; (c) the artificial intelligent (AI) model; and (d) the hybrid model [4].

The physical model relies on the information stated in the numerical weather forecast. [5]. This model was established with many complicated factors such as pressure, temperature, obstacles and roughness, which are usually difficult to utilize in practical applications [6]. The traditional statistical

model based on the mature statistical equations to obtain the potential evolution rule [7,8]. The most commonly used traditional statistical models for wind speed forecasting include the autoregressive model (AR) [9], autoregressive moving average (ARMA) [10] and autoregressive integrated moving average model (ARIMA) [11]. Liebl [12] proposed a new statistical perspective using a functional factor model for modeling and forecasting electricity spot prices that accounts for the merit order model. Statistical models have simple principles and high efficiencies. However, the prediction accuracy of low-order statistical models is relatively low, while high-order model parameters are tremendously difficult to obtain. As for the AI methods, artificial neural networks (ANN) [13], support vector regression (SVR) [14], regularized extreme learning machine (RELM) [15] and Least square support vector machines (LSSVM) [16] might be the most frequently used models for wind speed forecasting, and empirical analysis shows that they are superior to traditional linear models. Yeh [17] proposed a parameter-free simplified swarm optimization for ANN training for time-series prediction and demonstrated its robustness and efficiency. Santamaría-Bonfil et al. [18] employed the SVR model and their results showed that the proposed model was more accurate than the persistence and auto-regressive models in medium short-term wind speed and wind power forecasting. Zhou et al. [16] built a LSSVM based model for one-step ahead wind speed forecasting, and obtained relatively accurate results compared to the persistence approach.

However, up until now, using only an artificial intelligence model does not provide satisfactory prediction accuracy due to the non-stationary nature of the original wind speed series for training the forecasting model [19]. To overcome this non-stationary problem and further improve the forecasting accuracy, many studies have proposed a hybrid model. One approach is to combine multiple intelligent algorithms to form a new hybrid model. The parameter optimization algorithms can improve the performance of the prediction algorithms by searching for the optimal parameters [20]. Amjady and Keynia [21] constructed a hybrid method composed of cascaded forecasters where each forecaster consists of a neural network (NN) and an evolutionary algorithms (EA). Wang et al. [22] proposed the LSSVM model whose parameters are tuned by an artificial intelligence (PSOSA) model, and it was built to make forecasts. Liu et al. [23] developed a hybrid model for wind speed forecasting, where the CSO algorithm was used to optimize the parameters of the SVM model. It is recognized that these heuristic optimization algorithms can search the global optimal and obtain the optimal parameters [24]. Another approach is to add signal decomposition techniques to the hybrid model, which aims to further decompose the non-linear wind speed time series into more stationary and regular subseries. The signal decomposition techniques used in wind speed prediction mainly include wavelet decomposition (WD), wavelet packet decomposition (WPD), empirical mode decomposition (EMD), ensemble empirical model decomposition (EEMD), and variational mode decomposition (VMD). Niu et al. [25] developed a hybrid model based on the WD-SVM optimized by a genetic algorithm (GA). The WD was applied to reduce the high-frequency components, the GA was incorporated into SVM for parameter optimization. The results indicated that proposed method is more efficient than a SVM-GA model without WT. Wang et al. [26] proposed a wind speed forecasting method based upon EEMD and an optimized BP neural network (GA-BP) for on-line short term (1 h) and ultra-short term (10 min) wind speed forecasting, and computational results have shown good performance of EEMD. Zhou et al. [27] employed a new decomposition-optimization model created by integrating VMD, the backtracking search algorithm (BSA), and RELM to enhance forecasting accuracy. Liu et al. [28] built up a model which applied both a WT and EMD decomposition method. Wang et al. [29] presented a hybrid model based on the FEEMD (fast ensemble empirical mode decomposition), VMD, BP and FA (firefly algorithm). The FEEMD was used to decompose the original series into several sub-series, while the VMD was used to further decompose high frequency sub-series. The results showed that the proposed model excelled compared to the FEEMD-FA-BP model and the VMD-FA-BP model.

The application of signal processing technology in wind speed prediction enhances the prediction performance. WD has good time-frequency localization characteristics, but the decomposition effect depends on the choice of the basis function, and the adaptability is poor. Besides, there are some problems in EMD, such as endpoint effect and over envelope. EEMD is an

improved method of EMD, which reduces the phenomenon of mode aliasing. Recently, VMD, which is a new signal decomposition technology has been proposed by [30]. Compared with the recursive element screening mode of EMD and EEMD, VMD decomposes the signal into non-recursive and variational mode. It has been stated that VMD has a better theoretical foundation compared to the sequential iterative sifting of EMD. It was also demonstrated that VMD has some advantages in tone separation and is less sensitive to noise and sampling [31].

The principal purpose of this study was to investigate a more accurate forecasting method for wind speed. A hybrid model based on VMD-bat algorithm (BA)-LSSVM was employed to forecast wind speed. Noticeably, in this work the two parameters of LSSVM were fine-tuned by the BA to ensure the generalization and the learning ability of LSSVM. In addition, different models (EEMD-BA-LSSVM, WD-BA-LSSVM, BA-LSSVM, PSO-LSSVM and LSSVM) were developed for comparative analysis. The proposed model is composed of three steps: (a) VMD is adopted to decompose the raw wind speed series into a discrete number of components with different frequencies; (b) LSSVM optimized by BA is employed to forecast each component; (c) the ultimate forecast result of wind speed can be obtained by accumulating the prediction values of each components.

The main contributions of this study are as follows: (a) VMD, as a competitive signal decomposition method to decompose the nonlinear features of the wind speed signals, is firstly combined with LSSVM optimized by BA to forecast wind speed; and (b) The proposed forecasting model was compared with EEMD-BA-LSSVM and WD-BA-LSSVM, and VMD was found to be more thorough and stable than WD and EEMD in the high frequency decomposition of wind speed series, which enhanced the accuracy of wind speed prediction to a certain extent. Few papers have compared and analyzed the application of various signal processing techniques in wind speed prediction. This paper fills the gap.

The rest of this paper is organized as follows: Section 2 describes the modelling approaches. In Section 3 a hybrid model is constructed that is designed to forecast wind speed. Then, in Section 4 the proposed model is examined by experimental and comparative analysis. Finally, Section 5 provides some conclusions of the entire research.

2. Methods

The research methodology used in this paper includes: variational mode decomposition, ensemble empirical mode decomposition, wavelet decomposition, least squares support vector machine and the bat algorithm. A brief description of those methods is outlined as follows.

2.1. Variational Mode Decomposition

VMD was proposed by Dragomiretskiy and Zosso in 2014 [30], and it is a newly developed multi-resolution for non-recursive signal processing. The VMD can adaptively decompose a real-valued signal $f(t)$ into a discrete number of band-limited intrinsic mode function (BIMF) u_k with specific sparsity properties. Each BIMF u_k is compact around a center pulsation ω_k which is determined along with the process of decomposition and its bandwidth is estimated by using the H^1 Gaussian smoothness of the shifted signal. Thus, the process of decomposition is implemented by settling a constrained variational problem:

$$\left\{ \begin{array}{l} \min \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ s. t. \sum_{k=1}^K u_k = f(t) \end{array} \right. \quad (1)$$

Making use of both a quadratic penalty term and Lagrangian multipliers λ , the above constrained problem can be converted to the unconstrained one which is easier to address. The augmented Lagrangian is described as follows:

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 \quad (2)$$

$$+ \langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \rangle$$

where α denotes the balancing parameter of the data-fidelity constraint. The alternate direction method of multipliers (ADMM) can be used to solve Equation (1). Therefore, it is implied that updating u_k , ω_k and λ_k in two directions is conducive for realizing the analysis process of VMD, and the solutions of u_k , ω_k and λ_k can be calculated as follows:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i^n(\omega) + (\hat{\lambda}(\omega)/2)}{1 + 2\alpha(\omega - \omega_k)^2} \quad (3)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \quad (4)$$

$$\hat{\lambda}_k^{n+1}(\omega) = \hat{\lambda}_k^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i^{n+1}(\omega) \right) \quad (5)$$

where $\hat{f}(\omega)$, $\hat{u}_k^n(\omega)$, $\hat{\lambda}(\omega)$ and $\hat{u}_k^{n+1}(\omega)$ represent the Fourier transforms of $f(t)$, and n denotes the number of iterations.

The termination condition of the VMD algorithm is presented as follows:

$$\frac{\sum_k \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \epsilon \quad (6)$$

where ϵ is tolerance of convergence criterion.

BIMF u_k can be obtained from the entire decomposition process for VMD according to the following steps:

Step 1: Initialize parameters for VMD method including $\{u_k^1\}$, $\{\omega_k^1\}$, and λ^1 , and set iteration number $n = 1$.

Step 2: Calculate $\hat{u}_k^{n+1}(\omega)$ and ω_k^{n+1} using the Equations (3) and (4).

Step 3: Update the Lagrangian multiplier λ_k in terms of Equation (5).

Step 4: Given the tolerance of the convergence criterion $\epsilon > 0$, if the convergence condition of Equations (6) is satisfied, the iteration is stopped, otherwise n increases to $n + 1$ and returns to step 2. Then, the final BIMF can be obtained. The main variables of VMD are listed in Table 1.

Table 1. Main variables involved in variational mode decomposition (VMD).

Variable	Meaning	Variable	Meaning
$f(t)$	real-valued signal	ω_k	center pulsation
u_k	band-limited intrinsic mode function	λ	Lagrangian multipliers

2.2. Ensemble Empirical Mode Decomposition

EMD, originally proposed by Huang [32], is a powerful signal decomposition technology that aims to decompose complicated signals into several intrinsic mode function (IMF) components. However, sometimes EMD cannot correctly decompose the raw data sequences. These IMFs extracted by EMD have lost their physical meanings and weaken the regularity. Compared with EMD, EEMD has good performance in non-stationary signal decomposition. EEMD adds a white noise series to the raw signal $f(t)$ to eliminate the mode mixing, obtaining the IMFs through the EMD procedures. The computation steps of the EEMD algorithm are described as follows:

Step 1: Calculate $f^i(t) = x(t) + n^i(t)$, where $n^i(t)$ ($i = 1, 2, 3, \dots, N$) represent the random white Gaussian noise series.

Step 2: Decompose the series $x^i(t)$ using the EMD technology to obtain IMF modes imf_m^i ($m = 1, 2, 3, \dots, N$).

Step 3: Compute the mean of the corresponding series $imf_m^i(t)$ as follows:

$$\overline{imf_m(t)} = \frac{1}{N} \sum_{i=1}^N imf_m^i(t) \quad (7)$$

Step 4: Repeat the above mean procedure to complete the process of EEMD. The decomposed results of the original signal series $x(t)$ will be obtained as follows:

$$x(t) = \sum_{m=1}^k \overline{imf_m(t)} + r_k(t) \quad (8)$$

where $\overline{imf_m(t)}$, ($m = 1, 2, 3, \dots, k$) are the IMFs decomposed by EEMD, $r_k(t)$ denotes the corresponding residue. The main variables of EEMD are listed in Table 2.

Table 2. Main variables involved in ensemble empirical mode decomposition (EEMD).

Variable	Meaning	Variable	Meaning
$f(t)$	raw signal	$r_k(t)$	residue
imf_m^i	intrinsic mode function	$n^i(t)$	random white Gaussian noise series

2.3. Wavelet Decomposition

WD is a signal decomposition technique with more applications. The basic principle is to decompose the non-stationary discrete wind speed sequence $f(t)$ into a high frequency detail sequence (d_1, d_2, \dots, d_J) with different frequencies and a low frequency approximation sequence a_J according to the multi-resolution idea proposed by Mallat [33]. J is the maximum number of decomposition layers. A 4-layer decomposition is usually performed using the db4 wavelet base. The decomposition process is:

$$\begin{cases} a_{j+1} = H(a_j) \\ d_{j+1} = G(d_j) \end{cases} \quad (9)$$

where a_j , d_j are the low frequency signal and high frequency signal, respectively, of the original signal at the resolution 2^{-j} , which are the components of the original signal on different adjacent frequency segments; H is the low pass filter; G is the high pass filter. The decomposition process utilizes two decimations so that each layer of the decomposition signal is half the pre-decomposition signal data, and two interpolation reconstructions are required to restore the signal length, as in equation (10).

$$\begin{cases} A_J = (H^*)^J a_j \\ D_J = (H^*)^{j-1} G^* d_j \end{cases} \quad (10)$$

H^* and G^* are the dual operators of H and G , respectively. After reconstructing (d_1, d_2, \dots, d_J) and a_J , the detail sequence D_1, D_2, \dots, D_J and the approximate sequence A_J are obtained. The main variables of WD are listed in Table 3.

Table 3. Main variables involved in wavelet decomposition (WD).

Variable	Meaning	Variable	Meaning
$f(t)$	raw signal	A_J	approximate sequence
D_J	detail sequence		

2.4. Least Squares Support Vector Machine

The LSSVM, put forward by Suykens [34], is a variation of the standard support vector machine (SVM), adopting the loss function different from SVM and minimizing the square error. A quadratic

programming problem can be transformed into linear equations by replacing inequality constraints with equality constraints, greatly reducing the computational complexity. In the LSSVM model, for a given training sample set $S = \{(x_i, y_i) | i = 1, 2, 3, \dots, t\}$, where x_i is the i^{th} input of sample space R^{input} , y_i is the i^{th} output of sample space R^{output} , t is the size of the training sample. Then, the optimal decision function is framed by using the high dimensional feature space. The decision function can be expressed as follows:

$$f(x) = \omega^T \varphi(x) + b \quad (11)$$

where $\varphi(x)$ represents the nonlinear mapping function from input space to high dimensional feature space, ω is weight, b is bias, and $f(x)$ is the prediction value.

The structural risk minimization can be described as follows:

$$R = \frac{1}{2} \|\omega\|^2 + c R_{emp} \quad (12)$$

where $\|\omega\|^2$ suggests the complex degree of the model, c is the regularization parameter, controlling the degree of punishment beyond the error samples, R_{emp} is the empirical risk function, the objective function of LSSVM is obtained as follows:

$$\min Z(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^t \xi_i^2 \quad (13)$$

$$s.t. \quad y_i = \omega \varphi(x_i) + b + \xi_i \quad i = 1, 2, 3, \dots, t$$

where ξ_i is the error, the Lagrange function can be defined as follows:

$$L(\omega, b, \xi, \lambda) = \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^t \xi_i^2 - \sum_{i=1}^t \lambda_i [\omega \varphi(x_i) + b + \xi_i - y_i] \quad (14)$$

where $\lambda_i (1, 2, 3, \dots, t)$ are the Lagrange multipliers.

According to the Karush-Kuhn-Tucker (KKT) conditions, Equation (11) is shown as follows:

$$\begin{cases} \omega - \sum_{i=1}^t \lambda_i \xi_i^2 = 0 \\ \sum_{i=1}^t \lambda_i = 0, \lambda_i - c \xi_i = 0 \\ \omega \varphi(x_i) + b + \xi_i - y_i = 0 \end{cases} \quad (15)$$

In the light of Equation (11), the optimization problem can be converted into the process of solving linear equations, which is presented as follows:

$$\begin{bmatrix} 0 & I^T \\ I & J + \frac{1}{c} \end{bmatrix} \begin{bmatrix} b \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (16)$$

where $I = [1, 1, \dots, 1]^T$ is a $t \times 1$ dimensional column vector, $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_t]^T$, $y = [y_1, y_2, \dots, y_t]^T$, $J_{ij} = \varphi(x_i)^T \varphi(x_j) = G(x_i, x_j)$, K is the kernel function which satisfies the condition of Mercer, the final form of the LSSVM model emerges as follows:

$$f(x) = \sum_{i=1}^t \lambda_i K(x_i, x_j) + b \quad (17)$$

In this research, the radial basis function (RBF) is selected as the kernel function, as shown in Equation (18):

$$G(x_k, x_j) = \exp\left(-\frac{\|x_k - x_j\|^2}{2\sigma^2}\right) \quad (18)$$

where σ^2 is the parameter of the kernel function.

Thus, there are two parameters, the regularization parameter c and the kernel parameter σ^2 , determining the LSSVM model. In previous studies, experimental comparison, grid searching methods and cross validation methods were applied to optimize the two parameters, but they are all time-consuming and inefficient. Therefore, this paper adopts a BA to optimize the two parameters, which can enhance and further the adaptability of the model and effectively improve the forecasting accuracy. The main variables of the LSSVM are listed in Table 4.

Table 4. Main variables involved in the least squares support vector machine (LSSVM).

Variable	Meaning	Variable	Meaning
$f(x)$	prediction value	ξ_i	Error
x_i	Input	λ_i	Lagrange multipliers
y_i	Output	$\ \omega\ ^2$	complex degree
c	regularization parameter	σ^2	parameter of the kernel function

2.5. The Bat Algorithm (BA)

The BA is a novel meta-heuristic algorithm inspired by the echolocation behavior of bats. The BA offers an excellent way to optimize and classify a selection of complicated problems [35]. The basic flow of the BA can be generalized by the pseudo code listed in Algorithm 1.

Algorithm 1. Pseudo code of the Bat Algorithm.

- (1) Initialize the position of bat population $x_i = (1, 2, \dots, n)$ and v_i
- (2) Initialize pulse frequency f_i at x_i , pulse rates r_i and the loudness A_i
- (3) While ($t < \text{maximum number of iterations}$)
- (4) Generate new solutions by adjusting frequency
- (5) Update the velocities and solutions
- (6) If ($\text{rand} > r_i$)
- (7) Select a solution among the best solutions
- (8) Generate a local solution around the selected best solution
- (9) End if
- (10) Generate a new solution by flying randomly
- (11) If ($\text{rand} < A_i \& f(x_i) < f(x^*)$)
- (12) Accept the new solutions
- (13) Increase r_i and reduce A_i
- (14) End if
- (15) Rank the bats and find the current best x^*
- (16) End while

3. Wind Speed Forecasting Models

In this section, the proposed model (VMD-BA-LSSVM) is described in detail. The flowchart of the presented model is given in Figure 1. The following three parts constitute the hybrid model.

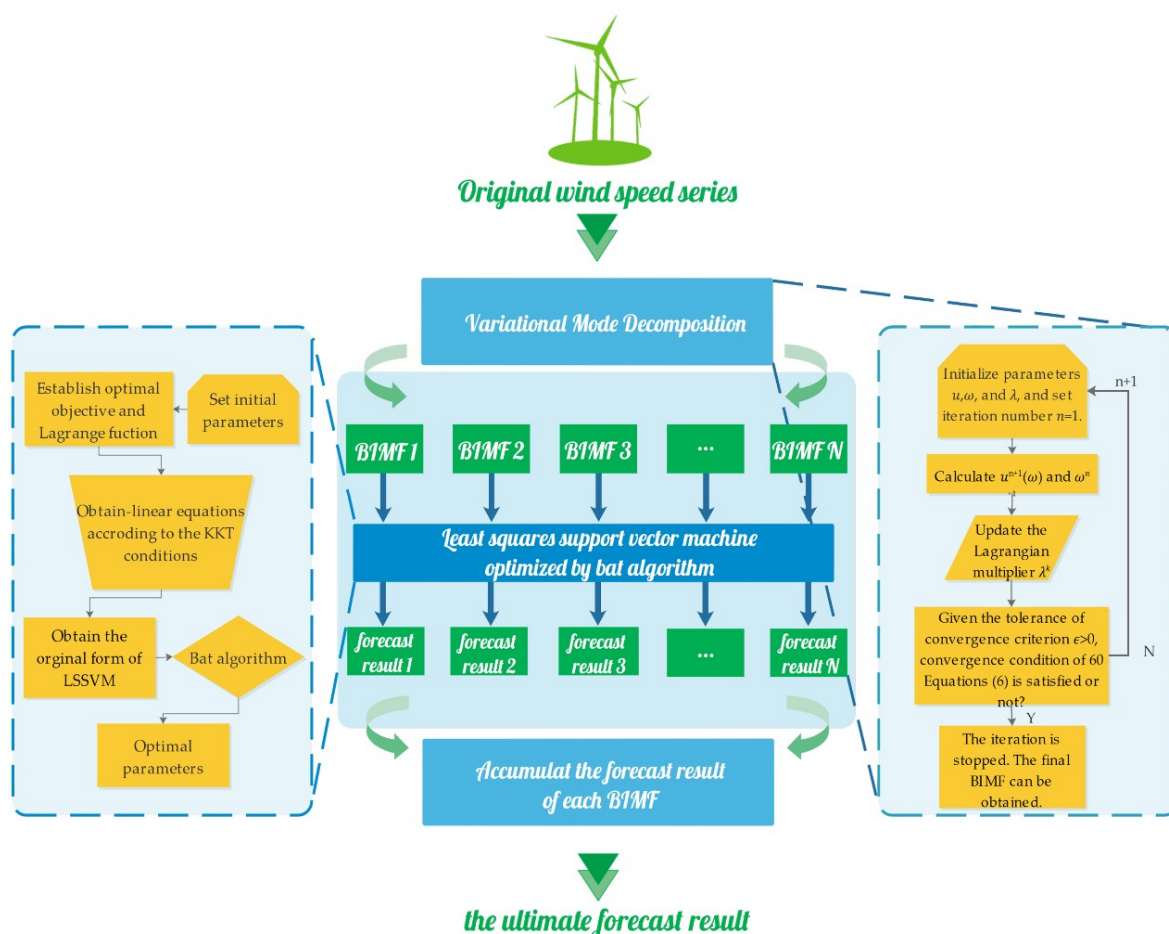


Figure 1. The structure of the VMD-BA-LSSVM model.

Part one: Data preprocessing. The VMD approach is employed to decompose the original wind speed series into a discrete number of components with different frequencies that are respectively denoted by BIMF1, BIMF2, ..., BIMFN. The aim of this technique is to diminish the non-stationary character of the series for the high precision short-term forecast.

Part two: Training and validation of the model. In this study, forecasting for each component is in the light of LSSVM-BA model, the basic steps can be described as follows:

Step 1: Parameter setting

The main parameters of BA are initial population size n , maximum iteration number N , original loudness A , pulse rate r , location vector x , and speed vector v .

Step 2: Initialize population

Initialize the bat population's position, each bat location strategy is a component of (γ, σ^2) , which can be defined as follows:

$$x = x_{min} + rand(1, d) \times (x_{max} - x_{min}) \quad (19)$$

where the dimension of the bat population: $d = 2$.

Step 3: Update parameters

Calculate the fitness value of population, find the current optimal solution and update the pulse frequency, velocity and position of bats as follows:

$$f_i = f_{min} + (f_{max} - f_{min}) \times \beta \quad (20)$$

$$v_i^t = v_i^{t-1} + (x_i^t - x^*) \times f_i \quad (21)$$

$$x_i^t = x_i^{t-1} + v_i^t \quad (22)$$

where β denotes uniformly random numbers, $\beta \in [0,1]$; f_i is the search pulse frequency of the bat i , $f_i \in [f_{min}, f_{max}]$; v_i^t and v_i^{t-1} are the velocities of the bat i at time t and $t - 1$, respectively; further, x_i^t and x_i^{t-1} represent the location of the bat i at time t and $t - 1$, respectively; x^* is the present optimal solution for all bats.

Step 4: Update loudness and pulse frequency

Produce a uniformly random number $rand$, if $rand > r_i$, disturb the optimal strategy randomly and acquire a new strategy; if $rand < A_i$ and $f(x) > f(x^*)$, then the new strategy can be accepted, the r_i and A_i of the bat are updated as follows:

$$A_i^{t+1} = \alpha A_i^t \quad (23)$$

$$r_i^{t+1} = r_i^0 [1 - \exp(-\gamma t)] \quad (24)$$

where α and γ are constants.

Step 5: Output the global optimal solution

The current optimal solution can be obtained by relying on the rank of all fitness values of the bat population. Repeat the steps from Equation (20) to Equation (22) until the maximum iterations are completed and output the global optimal solution. Therefore, a wind speed forecasting model can be generated. In addition, the LSSVM approach is adopted to model the training set, and the mean square errors of the true values and forecasting values are adopted as the fitness functions of the BA. Then, the group of parameters of LSSVM is optimized by BA for the minimum fitness value. Finally, the LSSVM model with optimal parameters can be developed to predict the wind speed.

Part three: Wind speed forecasting. In this part, the LSSVM approach with the parameters optimized by the BA is employed to predict each BIMF decomposed by VMD. Then, the ultimate forecast result of wind speed can be obtained by accumulating the prediction values of each BIMF.

4. Experimental Results and Comparative Analysis

Experimental results are analyzed in this section to illustrate the effectiveness of the proposed method in a comparative analysis. Case 1 utilizes the proposed method for ultra-short term (15 min) wind speed forecasting, and in Case 2, a short term (1 h) wind speed forecasting is discussed to verify the generalization ability of the model.

4.1. Study Area and Data Set

Two case studies are analyzed in this section to illustrate the effectiveness of the proposed method. The actual wind speed series of a real wind farm with installed capacity of 33.25 MW in Jiangsu Province of China were chosen as the research object. Fifteen-minute wind speed data from 00:00 November 1, 2013 to 23:45 November 7, 2013 were chosen as the Case I study. One-hour wind speed data from 1:00 December 3, 2013 to 00:00 December 31, 2013 were chosen as the Case 2 study to discuss its generalization ability. In each Case, the 1st–576th observation and 577st–672th observations are respectively adopted as the training and test data. Training data corresponds to input x in Equation (1), the predicted value $f(x)$ obtained in Equation (1) was compared with the test data. The original wind speed series for the two cases are shown in Figure 2. Table 5 demonstrates the descriptive statistics for the wind speed series.



Figure 2. Two original wind speed series. (a) Wind speed series for 15min; (b) Wind speed series for 1h.

Table 5. Descriptive statistics of wind speed series.

Case	Data set	Statistics				
		Minimum (m/s)	Maximum (m/s)	Mean (m/s)	Median (m/s)	Standard Deviation (m/s)
15 min	Training data	0.11	8.8	4.278	4.35	1.978
	Test data	1.85	7.73	5.016	5.335	1.455
1 hour	Training data	0	12.31	5.594	5.476	2.5
	Test data	3.4	11.36	6.81	6.902	1.617

4.2. Performance Criteria of Prediction Accuracy

In this paper, the root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) were employed as evaluation criteria to quantitatively assess the forecasting performance of the proposed model:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2} \quad (25)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i| \quad (26)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100\% \quad (27)$$

where x_i is the actual data at i , and \hat{x}_i is the corresponding predictive data. $i = 1, 2, \dots, n$.

4.3. Original Wind Speed Series Decomposition Results

In order to improve the forecasting performance of wind speed series, this paper firstly adopts signal decomposition technique to decompose the original wind speed series into several components, and the data decomposition results of these two wind speed series (15 min and 1 h) are listed in Figure 3 and Figure 4. Figure 3(a) and Figure 4(a) are wind speed sequences decomposed by VMD; it is obvious that each wind speed series is decomposed into 8 components, which are respectively denoted by BIMF1, BIMF2, ..., BIMF8. $\alpha = 2000$ and $\tau = 0.3$ to ensure the fidelity of the data decomposition. For the comparative analysis, EEMD and WD were also used to decompose the wind speed series. The number of decompositions is automatically generated in the recursive

process with the decomposition process using EEMD, and wind speed series is decomposed into 8 intrinsic mode functions (IMF) components and 1 residual (RES). The results of EEMD are shown in Figure 3(b) and Figure 4(b). Figure 3(c) and Figure 4(c) show the results of WD with 4-layer decomposition and reconstruction using db 4 wavelet base. The wind speed series are decomposed into an approximation component A1 and 4 detail components, which are denoted by D1, D2, D3, D4.

Observing the component characteristics after signal decomposition, the components of VMD, EEMD and WD range in order from higher frequency to lower frequency. It is generally believed that the high frequency component is the random part of wind speed. Some lower frequency components have strong sinusoidal fluctuation characteristics and can be considered as periodic components of wind speed. The last low frequency part is the trend term of wind speed, which reflects the trend of the wind speed series. From Figure 3, we can see that the amplitude fluctuation of the high frequency component obtained by VMD is approximately between $[-1,1]$. The amplitude of the high frequency component of EEMD fluctuates greatly, and it ranges from $[-2,2]$. The fluctuation of the high frequency component of WD is the largest, which is $[-5.5]$. This shows that VMD decomposes the high frequency part more thoroughly than EEMD and WD.

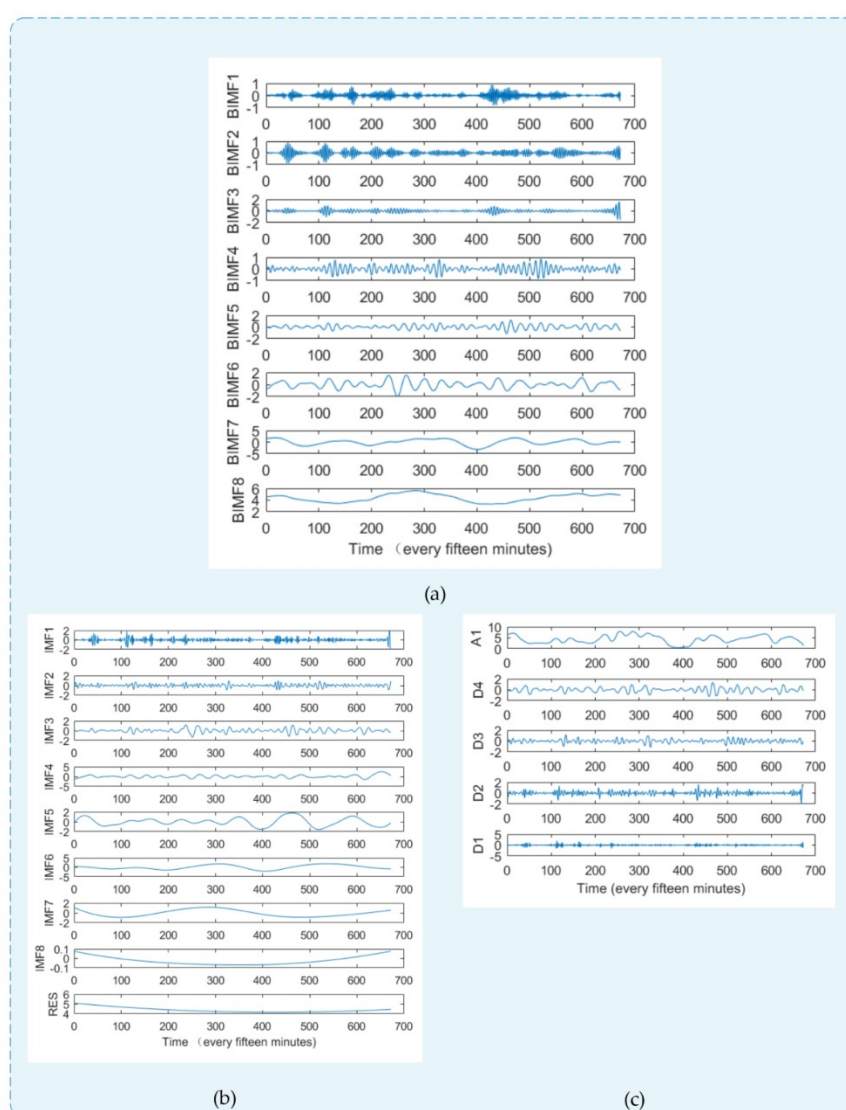


Figure 3. VMD, EEMD and WD for wind speed series for 15 min. (a) Wind speed sequences decomposed by VMD; (b) Wind speed sequences decomposed by EEMD; (c) Wind speed sequences decomposed by WD.

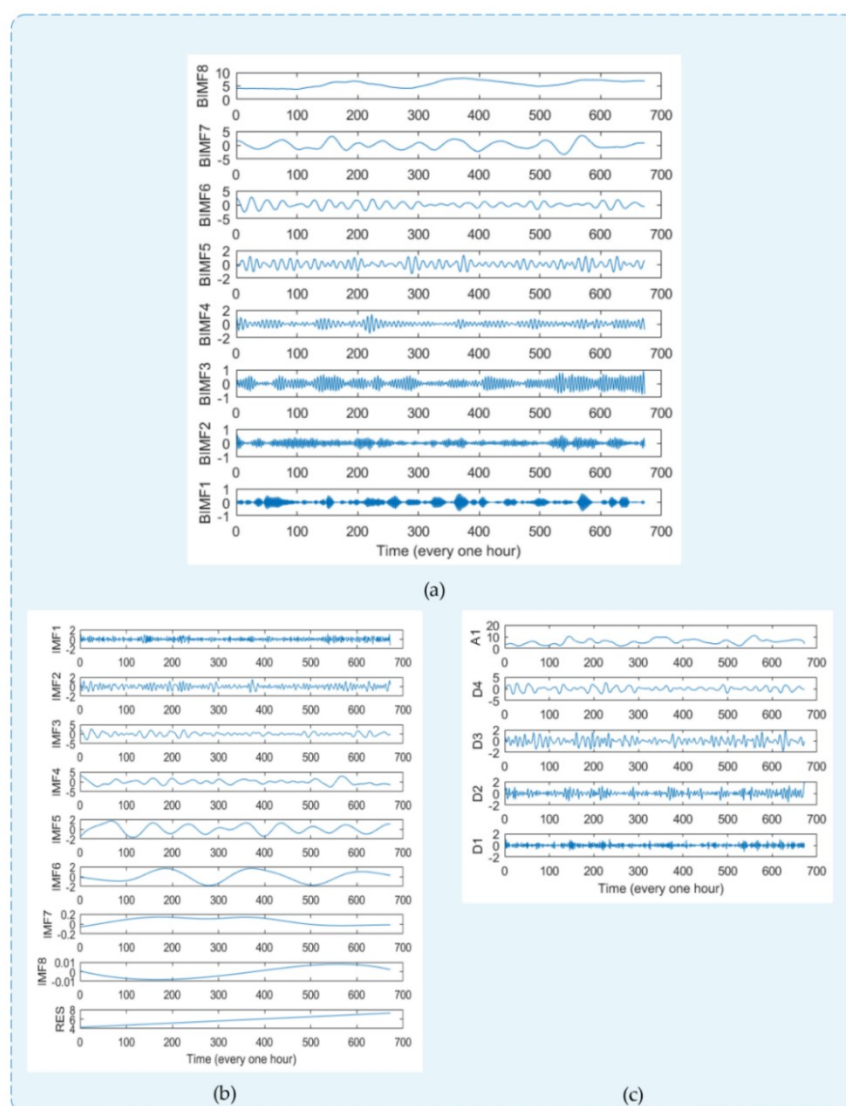


Figure 4. VMD, EEMD and WD for wind speed series for 1h. (a) Wind speed sequences decomposed by VMD; (b) Wind speed sequences decomposed by EEMD; (c) Wind speed sequences decomposed by WD.

4.4. Parameter Settings

Previous studies based on the LSSVM model for forecasting have indicated that the performance of the LSSVM approach depends on its parameters and the kernel function. The BA is a population intelligent optimization algorithm that is used to search the optimal parameters of LSSVM. In this paper, RBF was chosen as the kernel function of the LSSVM algorithm, which decreases the complexity of the model and improves the training speed. Thus, the regularization parameter c and kernel parameter σ^2 can obtain the optimal values using the reliable, automatic searching ability of BA. The main parameters of the BA are listed in Table 6. Table 7 shows the optimal parameters (c, σ^2) of each sub-series in the LSSVM models obtained using the BA approach for VMD, EEMD, and WD.

Table 6. Main parameters of the bat algorithm (BA).

Parameters	Values	Parameters	Values
Initial population size	10	Minimum frequency	0
Initial loudness	0.25	Maximum frequency	5
Pulse rate	0.5	Max-iteration number	50

Table 7. The optimal parameters in the LSSVM model for VMD, EEMD and WD.

VMD			EEMD			WD		
Components	c	σ^2	Components	c	σ^2	Components	c	σ^2
BIMF1	0.1898	14.8432	IMF1	0.098	9.5004	A1	0.1696	10.1624
BIMF2	12.3084	1.0987	IMF2	0.2118	0.0239	D1	0.0014	8.1789
BIMF3	1.3475	0.4603	IMF3	0.2942	0.0532	D2	13.6794	10.1307
BIMF4	5.6934	2.2004	IMF4	24.51	2.4311	D3	4.5363	0.4641
BIMF5	8.4486	4.7532	IMF5	1983.8	0.1635	D4	0.0092	11.1173
BIMF6	0.811	6.835	IMF6	28.3727	0.303			
BIMF7	9.1664	0.8689	IMF7	80.9834	0.3148			
BIMF8	8.8698	9.4069	IMF8	3.87	0.5182			
			RES	0.728	0.0892			

4.5. Comparative Analysis of Different Models

In order to demonstrate the advantages of the VMD-BA-LSSVM model, LSSVM, PSO-LSSVM, BA-LSSVM, WD-BA-LSSVM, EEMD-BA-LSSVM models are taken as the comparison models.

4.5.1. Case 1: Ultra Short-Term (15 min) Wind Speed Forecasting

The values for MAPE, MAE and RMSE of the proposed and various comparison models are presented in Table 8. Compared with other forecasting models, the proposed model displays better prediction of wind speed, and achieves good forecasting performance. This conclusion can be further verified by the results shown in Figure 5, which present the fit and absolute error between the predicted wind speed and actual wind speed. To further analyze the performance and differences between the models, the specific analysis is shown below.

Table 8. Error comparison among different forecasting models in Case 1I.

Case 1	Error	Model					
		LSSVM	PSO-LSSVM	BA-LSSVM	WD-BA-LSSVM	EEMD-BA-LSSVM	VMD-BA-LSSVM
Fifteen Minutes	MAPE	20.99%	19.69%	15.44%	14.93%	3.42%	1.03%
	MAE	0.92	0.8708	0.6873	0.6972	0.1538	0.0427
	RMSE	1.0866	0.9858	0.8764	0.8094	0.2035	0.0543

(1) Optimal parameters. Figure 5 (a) (b) (c) shows that the prediction accuracy of BA-LSSVM and PSO-LSSVM is higher than LSSVM, and the predicted wind speed series is consistent with the variation trend of the original wind speed. It shows that the prediction accuracy of the LSSVM model can be improved by optimizing the kernel parameters. Table 8 shows that the prediction accuracy of BA-LSSVM is higher than that of PSO-LSSVM. It shows that the BA algorithm has better search ability and convergence speed than the PSO algorithm, and the prediction accuracy of LSSVM kernel parameters optimized by BA algorithm is higher. The improved LSSVM models have better performance than a single LSSVM approach. The primary reason for this may be that the process of automatic searching is added to the improved LSSVM model, which equips the LSSVM model with better learning and generalization ability so that it easily acquires the global optimal solution.

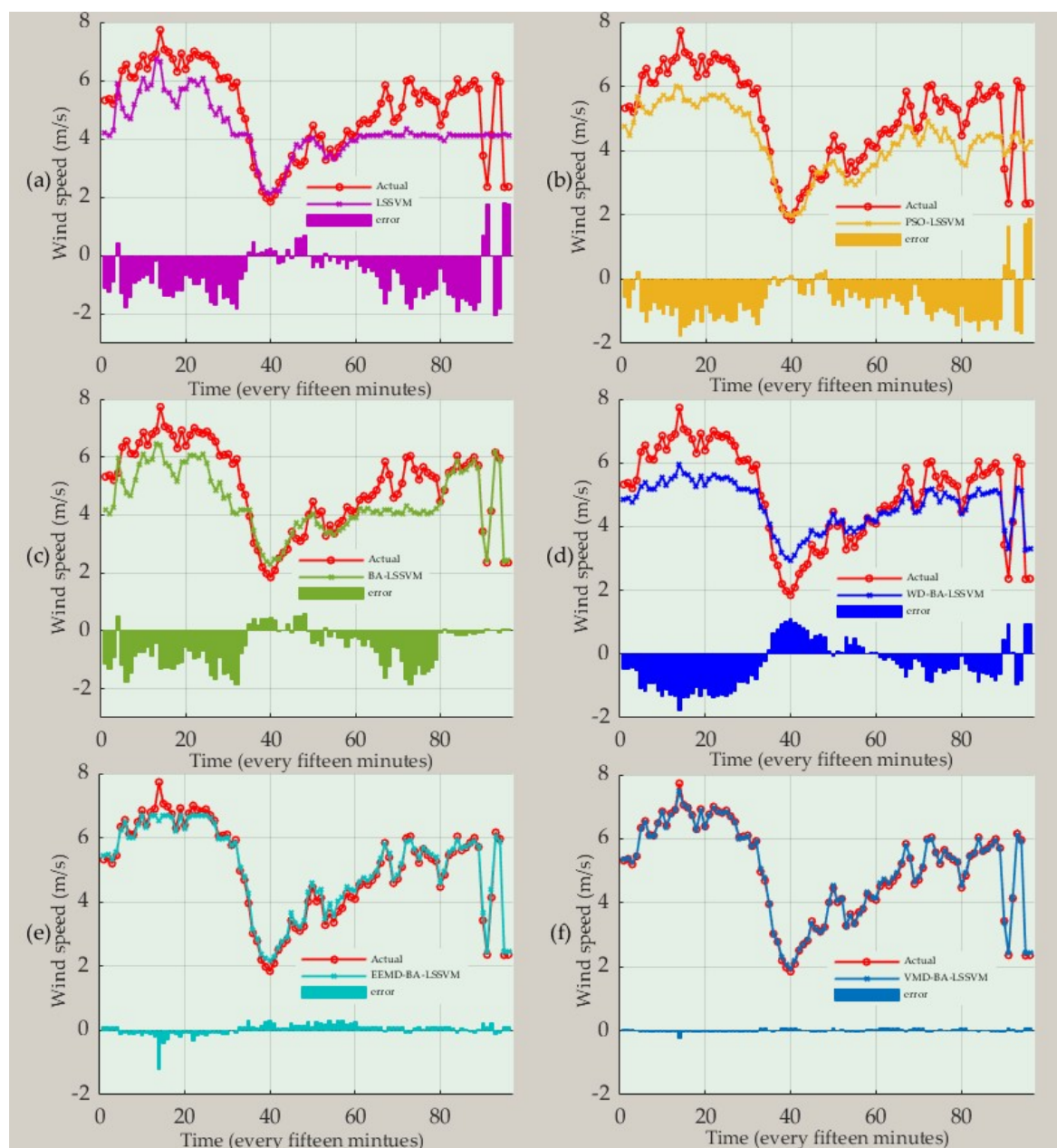


Figure 5. The forecasting results from various models (15 min). (a) LSSVM model; (b) PSO-LSSVM model; (c) BA-LSSVM model; (d) WD-BA-LSSVM model; (e) EEMD-BA-LSSVM; (f) VMD-BA-LSSVM.

(2) Absolute error. From the error distribution in Figure 5, the error value of VMD-BA-LSSVM fluctuates very little near zero. Although EEMD-BA-LSSVM fluctuates less than WD-BA-LSSVM, BA-LSSVM, PSO-LSSVM and LSSVM, and gets better prediction results, it was found that VMD-BA-LSSVM has better follow up to the original wind speed series than EEMD-BA-LSSVM, and the error value is also the smallest among all the models. This fully shows that the proposed algorithm has high accuracy and great advantages in wind speed prediction.

(3) Signal decomposition technique. From Table 8 and Figure 5(c–f), the prediction accuracy of VMD-BA-LSSVM, EEMD-BA-LSSVM and WD-BA-LSSVM based on signal decomposition is better than that of traditional prediction models without signal decomposition. It shows that signal decomposition technology can effectively reduce the non-stationary characteristics of wind speed series. Among the three signal decomposition models, VMD-BA-LSSVM is better than WD-BA-LSSVM and EEMD-BA-LSSVM, and the prediction accuracy evaluation index is the best in each comparison model. This is because the decomposition ability of the three signal decomposition techniques differs when the high frequency part of the wind speed series is decomposed. In WD-BA-

LSSVM, the high-frequency variation of each component is the largest, which affects the prediction accuracy of the model to a certain extent. The MAPE value is 14.93%. The high frequency range of each component in EEMD-BA-LSSVM is smoother than that in WD, thus, the prediction accuracy of EEMD-BA-LSSVM is higher than that of WD-BA-LSSVM, and the MAPE value is 3.42%. The wind speed sequence in VMD-BA-LSSVM changes steadily after decomposition, which makes the prediction accuracy of each component better, thereby, the integrated results can further improve the prediction accuracy of the model. The MAPE value of VMD-BA-LSSVM is only 1.03%.

The above analysis shows the superiority of the proposed method in ultra short-term wind speed forecasting.

4.5.2. Case 2: Short-Term (1h) Wind Speed Forecasting

In order to further verify the generalization ability of the forecasting method proposed in this paper, short-term (1 h) wind speed forecasting was conducted and is discussed in this section. On the basis of Case 1, BA-LSSVM, WD-BA-LSSVM, EEMD-BA-LSSVM and VMD-BA-LSSVM were used to predict wind speed. By comparing Figure 6(a–d), the forecasting results of VMD-BA-LSSVM are better than that of EEMD-BA-LSSVM, WD-BA-LSSVM and BA-LSSVM. It is obvious that signal processing technology can significantly improve the accuracy of wind speed prediction. The results of the proposed hybrid VMD and BA-LSSVM model is much better than that of the hybrid EEMD or WD and BA-LSSVM model most of the time. The improved BA-LSSVM model is able to forecast the trend in the wind speed series, but the prediction accuracy is unsatisfactory. Table 9 presents the results of the four models, and similar conclusion as that in Case 1 can be made here, that is, the proposed model is also suitable for short-term (1h) wind speed forecasting.

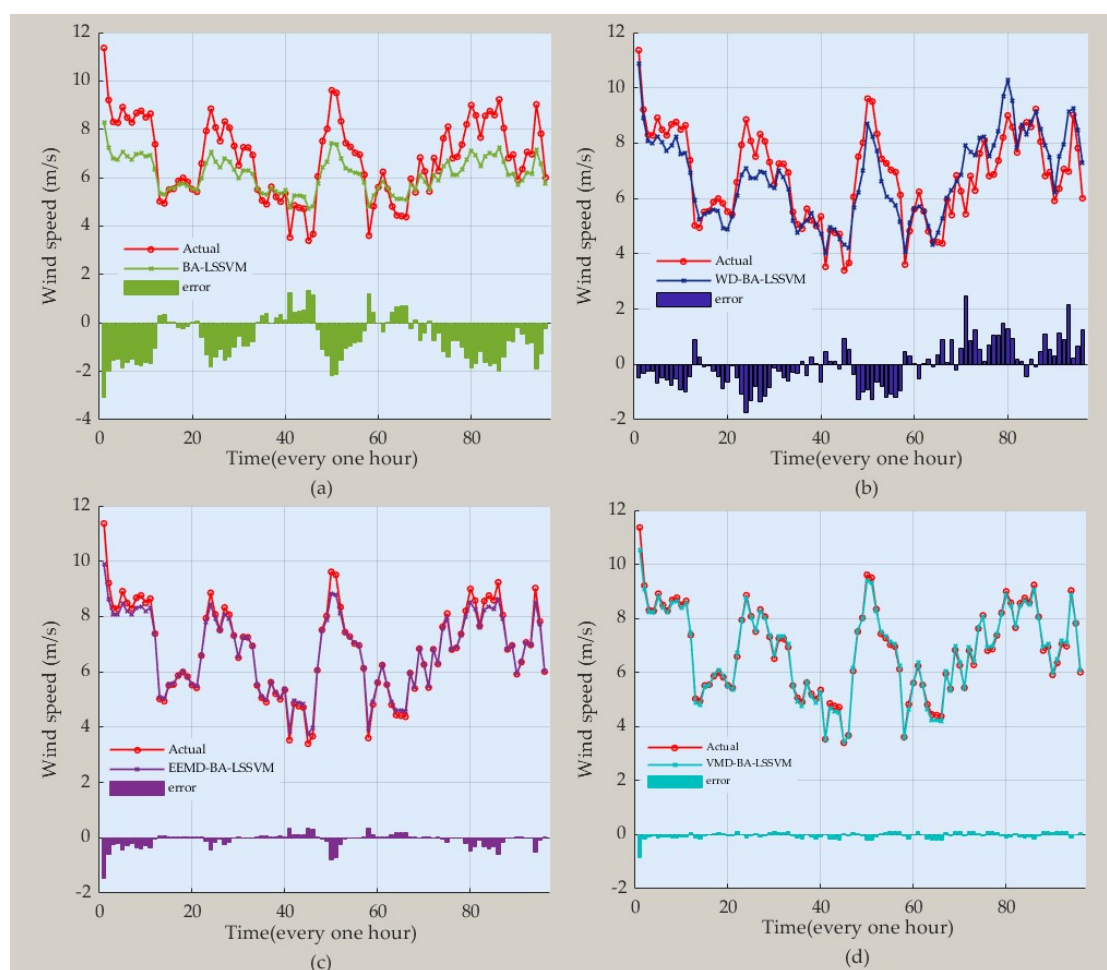


Figure 6. The forecasting results from various models (1h). (a) BA-LSSVM model; (b) WD-BA-LSSVM model; (c) EEMD-BA-LSSVM; (d) VMD-BA-LSSVM.

Table 9. Error comparison among different forecasting models in Case 2.

Case 2	Error	Model			
		BA-LSSVM	WD-BA-LSSVM	EEMD-BA-LSSVM	VMD-BA-LSSVM
One hour	MAPE	14.42%	9.50%	2.23%	1.56%
	MAE	0.9283	0.6463	0.1585	0.1015
	RMSE	1.1309	0.8026	0.2715	0.1367

5. Conclusion

In order to enhance the efficient and accurate prediction of wind speed, a hybrid model is proposed in this paper. First, the VMD technique was employed to decompose the original wind speed series. Then, the relevant parameters of the proposed model were optimized by a BA. Finally, the hybrid VMD and BA-LSSVM model with excellent learning and generating abilities was developed to forecast wind speed. EEMD and WD were also employed to compare the application of various signal processing techniques in wind speed forecasting models. In Case 1, the MAPE, MAE and RMSE of the proposed model were 1.03%, 0.0427 and 0.0543, respectively. In Case 2, the MAPE, MAE and RMSE of the proposed model were 1.56%, 0.1015 and 0.1367, respectively. The hybrid VMD and BA-LSSVM model is much better than the traditional LSSVM, PSO-LSSVM, BA-LSSVM models and the hybrid EEMD, WD and BA-LSSVM model.

The superiority of the proposed hybrid model over other models may be accounted for by the following aspects: (a) Signal decomposition technique plays an essential role in wind speed forecasting. Therefore, VMD was employed to decompose the original wind speed series. Besides, the decomposition effects of VMD, EEMD and WD were also analyzed in this paper, and we found that the forecasting performance of wind speed series can be greatly augmented by using an VMD technique; (b) The parameters of the LSSVM models play an important role in wind speed forecasting. Therefore, the BA algorithm was employed to optimize the parameters of the LSSVM model, and it was concluded that BA-LSSVM is better than PSO-LSSVM and LSSVM in learning and generalization ability; (c) The hybrid model comprehensively captures the characteristics of the original wind speed series, whilst the single models only reflect the trend of the wind speed series with limited prediction accuracy.

Therefore,, the proposed hybrid model clearly performs better than the other single or hybrid models as shown by the MAE, RMSE and MAPE criteria, and it is suitable for ultra short-term (15 min) and short-term (1 h) wind speed forecasting.

Author Contributions: Q.W. designed this paper and provided overall guidance; H.L. wrote the entire manuscript.

Acknowledgements: The authors gratefully acknowledge the financial support from the National Social Science Fund of China (Grant No. 17BGL252).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wu, Q.; Peng, C. Wind power generation forecasting using least squares support vector machine combined with ensemble empirical mode decomposition, principal component analysis and a bat algorithm. *Energies* **2016**, *9*, 261.
2. Kusiak, A.; Zhang, Z.J.; Verma, A. Prediction, operations, and condition monitoring in wind energy. *Energy* **2013**, *60*, 1–12.
3. Ackermann, T.; Söder, L. Wind energy technology and current status: A review. *Renew. Sustain. Energy Rev.* **2000**, *4*, 315–374.
4. Jung, J.; Broadwater, R.P. Current status and future advances for wind speed and power forecasting. *Renew. Sustain. Energy Rev.* **2014**, *31*, 762–777.
5. Barbounis, T.G.; Theocharis, J.B. A locally recurrent fuzzy neural network with application to the wind speed prediction using spatial correlation. *Neurocomputing* **2007**, *70*, 1525–1542.

6. Cassola, F.; Burlando, M. Wind speed and wind energy forecast through kalman filtering of numerical weather prediction model output. *Appl. Energy* **2012**, *99*, 154–166.
7. Erdem, E.; Shi, J. Arma based approaches for forecasting the tuple of wind speed and direction. *Appl. Energy* **2011**, *88*, 1405–1414.
8. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **2014**, *30*, 1030–1081.
9. Maatallah, O.A.; Achuthan, A.; Janoyan, K.; Marzocca, P. Recursive wind speed forecasting based on hammerstein auto-regressive model. *Appl. Energy* **2015**, *145*, 191–197.
10. Liu, H.; Erdem, E.; Shi, J. Comprehensive evaluation of arma-garch(-m) approaches for modeling the mean and volatility of wind speed. *Appl. Energy* **2011**, *88*, 724–732.
11. Liu, H.; Tian, H.Q.; Chen, C.; Li, Y.F. A hybrid statistical method to predict wind speed and wind power. *Renew. Energy* **2010**, *35*, 1857–1861.
12. Liebl, D. Modeling and forecasting electricity spot prices: A functional data perspective. *Ann. Appl. Stat.* **2013**, *7*, 1562–1592.
13. Kani, S.A.P.; Riahy, G.H. A new ANN-based methodology for very short-term wind speed prediction using markov chain approach. In Proceedings of the IEEE Canada Electric Power Conference (EPEC 2008), Vancouver, BC, Canada, 6–7 October 2008; pp 1–6.
14. Wang, J.; Qin, S.; Zhou, Q.; Jiang, H. Medium-term wind speeds forecasting utilizing hybrid models for three different sites in xinjiang, china. *Renew. Energy* **2015**, *76*, 91–101.
15. Wei, S.; Liu, M. Wind speed forecasting using FEEMD echo state networks with RELM in Hebei, China. *Energy Convers. Manag.* **2016**, *114*, 197–208.
16. Zhou, J.; Jing, S.; Gong, L. Fine tuning support vector machines for short-term wind speed forecasting. *Energy Convers. Manag.* **2011**, *52*, 1990–1998.
17. Yeh, W.C. New parameter-free simplified swarm optimization for artificial neural network training and its application in the prediction of time series. *IEEE Trans. Neural Netw. Learn. Syst.* **2013**, *24*, 661–665.
18. Santamaria-Bonfil, G.; Reyes-Ballesteros, A.; Gershenson, C. Wind speed forecasting for wind farms: A method based on support vector regression. *Renew. Energy* **2016**, *85*, 790–809.
19. Gendeel, M.; Zhang, Y.X.; Han, A.Q. Performance comparison of ANNs model with VMD for short-term wind speed forecasting. *IET Renew. Power Gener.* **2018**, *12*, 1424–1430.
20. Zhang, W.; Qu, Z.; Zhang, K.; Mao, W.; Ma, Y.; Fan, X. A combined model based on ceemdan and modified flower pollination algorithm for wind speed forecasting. *Energy Convers. Manag.* **2017**, *136*, 439–451.
21. Amjady, N.; Keynia, F. Day ahead price forecasting of electricity markets by a mixed data model and hybrid forecast method. *Int. J. Electr. Power Energy Syst.* **2008**, *30*, 533–546.
22. Wang, J.Z.; Wang, Y.; Jiang, P. The study and application of a novel hybrid forecasting model—A case study of wind speed forecasting in china. *Appl. Energy* **2015**, *143*, 472–488.
23. Liu, T.X.; Liu, S.Z.; Heng, J.N.; Gao, Y.Y. A new hybrid approach for wind speed forecasting applying support vector machine with ensemble empirical mode decomposition and cuckoo search algorithm. *Appl. Sci. I* **2018**, *8*, 22.
24. Zahraee, S.M.; Assadi, M.K.; Saidur, R. Application of artificial intelligence methods for hybrid energy system optimization. *Renew. Sustain. Energy Rev.* **2016**, *66*, 617–630.
25. Liu, D.; Niu, D.; Wang, H.; Fan, L. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renew. Energy* **2014**, *62*, 592–597.
26. Wang, S.; Zhang, N.; Wu, L.; Wang, Y. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and ga-bp neural network method. *Renew. Energy* **2016**, *94*, 629–636.
27. Zhou, J.; Sun, N.; Jia, B.; Peng, T.; Sciubba, E. A novel decomposition-optimization model for short-term wind speed forecasting. *Energies* **2018**, *11*, 1752.
28. Liu, D.; Li, H.; Ma, Z. One hour ahead prediction of wind speed based on data mining. In Proceedings of the International Conference on Advanced Computer Control, Shenyang, China, 27–29 March 2010; pp. 199–203.
29. Wang, D.Y.; Luo, H.Y.; Grunder, O.; Lin, Y.B.; Guo, H.X. 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.
30. Zosso, D.; Dragomiretskiy, K. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544.

31. Wang, Y.; Markert, R.; Xiang, J.; Zheng, W. Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system. *Mech. Syst. Signal Process.* **2015**, *60–61*, 243–251.
32. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995.
33. Mallat, S.G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693.
34. Suykens, J.A.K.; Vandewalle, J. Recurrent least squares support vector machines. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **2008**, *47*, 1109–1114.
35. Cheng, L.I. A new metaheuristic bat-inspired algorithm. *Comput. Knowl. Technol.* **2010**, *284*, 65–74.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).