Hybrid Short Term Wind Speed Forecasting Using Variational Mode Decomposition and a Weighted Regularized Extreme Learning Machine

Nantian Huang *, Chong Yuan, Guowei Cai and Enkai Xing

School of Electrical Engineering, Northeast Electric Power University, Jilin 132012, China;
yuan_chong@126.com (C.Y.); caigw@mail.nedu.edu.cn (G.C.); xingenkai@126.com (E.X.)
* Correspondence: huangnantian@126.com; Tel.: +86-135-7855-5676

Abstract: Accurate wind speed forecasting is a fundamental element of wind power prediction. Thus, a new hybrid wind speed forecasting model, using variational mode decomposition (VMD), the partial autocorrelation function (PACF), and weighted regularized extreme learning machine (WRELM), is proposed to improve the accuracy of wind speed forecasting. First, the historic wind speed time series is decomposed into several intrinsic mode functions (IMFs). Second, the partial correlation of each IMF sequence is analyzed using PACF to select the optimal subfeature set for particular predictors of each IMF. Then, the predictors of each IMF are constructed in order to enhance its strength using WRELM. Finally, wind speed is obtained by adding up all the predictors. The experiment, using real wind speed data, verified the effectiveness and advancement of the new approach.

---

1. Introduction

Wind energy has been gaining more and more attention from all over the world as a renewable and clean energy source due to an increasing consumption of fossil energy and the pressure of environmental protection [1–4]. However, wind power is severely intermittent, volatile, and stochastic. It is not conducive to the safe and stable operation of the traditional electrical grid, power system generation planning, and economic electric power dispatch. Therefore, large-scale grid-connected operations should be based on accurate wind speed forecasting results in order to reduce wind power fluctuations in power systems. Accurate wind speed forecasting can improve the accuracy of wind power prediction; therefore, research on wind speed forecasting has significance and application value [5,6].

Existing wind speed forecasting methods can be divided into ultra-short-term, short-term, medium-term, and long-term forecasts from the forecast time scale [7]. Short-term wind speed forecasting is an important basis for the economic electric power dispatch of wind power grids. It helps to improve power quality and to maintain the reliability and stability of power grids. The short-term wind speed forecasting model can be divided into three types: statistical, physical, and intelligent models [8–13].

The statistical model is built through analysis of the correlation between the wind speed data of each time point in the wind speed series. Linear mathematical models, such as the auto regressive (AR) method [14] and auto regressive integrated moving average (ARIMA) [15], are established using the functional relationship between historical wind speed data and output wind speeds. Statistical

models have simple principles and high efficiencies. However, the prediction accuracy of low-order statistical models is not high, while high-order model parameters are extremely difficult to determine.

A physical model relies on information from numerical weather reports (NWPs). This model is constructed on numerous meteorological and geographic properties, including air pressure, temperature, humidity, surface roughness, and contour [16,17]. The input dimension of the physical model is extremely high and the operation is complex due to the large number of factors. These are not conducive to the improvement of prediction accuracies. Moreover, the influence of numerical weather report (NWP) error on wind speed prediction is difficult to quantitatively estimate.

An intelligent model can build the nonlinear intelligent prediction model, which can compensate for the deficiencies in the linear prediction model. The forecasting result is close to that of the real wind speed. The neural networks (NNs) in the existing intelligent model have excellent nonlinear learning and generalization abilities; however, the setting of neural network (NN) parameters is complex. The training time of the intelligent predictor is long and requires a large number of training samples. The limitation of NNs is that they are not conducive to the improvement of prediction accuracy [13,18]. An extreme learning machine (ELM) [19–22] is proposed, based on the single hidden layer feedforward network (SLFN) [19–22]. ELM has the advantages of a simple structure, high learning efficiency, and strong generalization ability. ELM has been proven capable of obtaining accurate prediction results using a small training set. However, the modeling process of ELM is considered a structural risk. Therefore, the prediction model based on traditional ELM is not the best model. Moreover, the prediction accuracy of the ELM model is vulnerable to outlier interference [23–25]. Weighted regularization ELM (WRELM) considers structure and empirical risks in the building of an optimal predictive model, and integrates weight in order to avoid interference from outliers in the training process [26,27]. WRELM can adjust the correlation weight of ELM automatically, according to the training errors during the training process. According to the characteristics of the training samples, an optimal forecasting model with fewer training samples and a high forecasting accuracy, based on WRELM, can be automatically built [26]. Thus, WRELM is more suitable for short-term wind speed forecasting.

The fluctuations and the randomness of wind speeds are the main factors that influence the accuracy of wind speed forecasting. The signal processing methods used to decompose the signal reduce the fluctuations and the randomness of the data in the wind speed sequence, and obtain a more regular sub-series [28–32]. Subsequently, a hybrid wind speed forecasting model is constructed using the sub-predictors of the sub-series. Thus, the fluctuations and randomness of the original signals are reduced and the classification of wind speed forecasting is improved. Existing hybrid methods often use wavelet transforms (WTs) and empirical mode decomposition (EMD) in order to decompose the original wind speed sequence [28–31]. WTs have very good decomposition ability for non-singular signals; however, obtaining satisfactory results according to specific circumstances with appropriate basis functions and decomposition scales is difficult [28]. As an adaptive decomposition method, intrinsic mode functions (IMFs), decomposed using EMD, could maintain the regularity of a wind speed signal in different frequency domains. Although constructing the hybrid wind speed forecasting model, based on EMD, is convenient, some defects, such as mode mixing and false modes, still exist. These defects lead to the uncertainty in the center frequency and bandwidth of the signal frequency band of the IMF component and affect the forecasting accuracy of the hybrid model. Variational mode decomposition (VMD) can adaptively decompose the signal in the frequency domain and transform the original signals into several IMFs with strong continuity and correlation [33–37]. Unlike EMD, VMD reduces the nonlinearity and fluctuations of the wind speed time series, and avoids the influence of mode mixing defects on wind speed predictions.

The results of the wind speed time series decomposition differ in different periods when the wind speed time series is pretreated with the signal processing method. Therefore, feature selection, according to the correlation of the elements, should be performed in each sub-sequence in order to construct the optimal prediction model for each. The partial autocorrelation function (PACF) is often

used to measure the degree of a time series and to remove autocorrelation interference [35,38]. The IMF sequence generated by the VMD treatment has a certain correlation and stability; therefore, the optimal input vector of the predictor for each IMF can be separately determined by PACF.

A new hybrid method of short-term wind speed prediction, using VMD, PACF, and WRELM, is proposed. First, VMD is used to decompose the wind speed sequence and to obtain relatively stable IMFs. This step reduces the influence of the random and fluctuating wind speed series on the prediction accuracy of wind speed forecasting. Second, PACF is used to determine the maximum input order for each IMF sequence in order to obtain the optimal feature vector of each predictor based on the WRELM for different IMFs. Subsequently, the parameters of the forecast model are automatically adjusted, and the WRELM-based optimal forecasting model is obtained in the training process. Finally, each forecasting model is used to predict the corresponding IMF sequence, and the prediction results of the WRELM model are added to obtain the final wind speed forecasting result. The experiment is performed using real wind speed data from the Measurement and Instrumentation Data Center (MIDC) of the National Renewable Energy Laboratory (NREL) in America in order to verify the effectiveness of the new method.

2. Structure and Methodology of the New Hybrid Model

The new method includes VMD wind speed decomposition, PACF feature selection, and modular WRELM prediction. Figure 1 shows the forecasting process of the new approach.
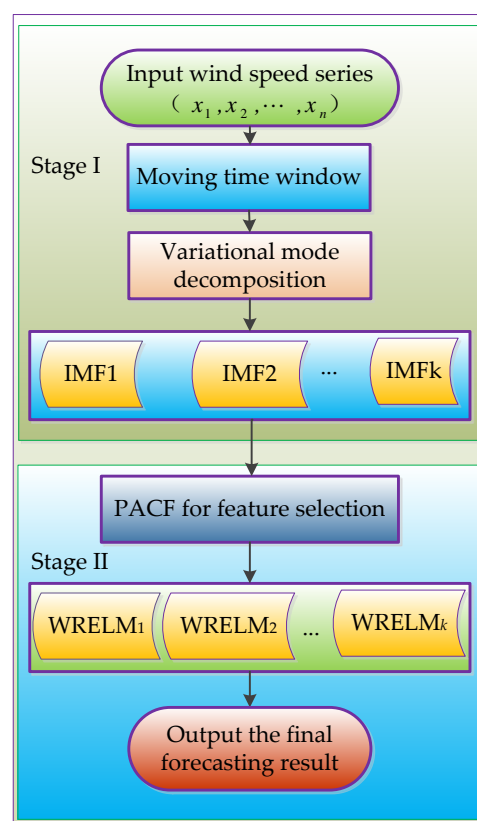

Figure 1. Structure of the new hybrid wind speed forecasting model.

As shown in Figure 1, the hybrid wind speed forecasting method mainly consists of two stages. The main contents of the two stages are described as follows.

Stage I: Data Processing

As shown in Figure 1, the new method first uses the VMD decomposition method to analyze the wind speed forecast time series in a specific time window. This method also obtains relatively stable IMF sub-sequences. The optimal variational mode layers are determined by the variation of the center frequency of the IMF subsequence. Finally, the optimal IMF sub-sequence is obtained by using the optimal decomposition layer number *k*.

Stage II: Hybrid Forecasting Model Construction

For every IMF sub-sequence, PACF is used to calculate the partial autocorrelation after the removal of the autocorrelation. Then, the input variable for the corresponding prediction model is selected, based on the partial autocorrelation measured by the PACF value. Subsequently, the WRELM prediction network is constructed for each IMF subsequence layer, with the output variable of wind speed in the forecasting point. A cross-validation method is used to determine the network parameters of WRELM, including the activation function and node number of the hidden layer. The final prediction results can be obtained by combining the predicted results of WRELMs corresponding to different IMF sequences.

2.1. Variational Mode Decomposition (VMD)

VMD decomposes the wind speed time series in the variational framework, which is different from EMD, which uses cyclic screening. The decomposition process of VMD is the construction and solution of the variational problem. Assuming that each mode has a limited bandwidth, which is compacted around a center frequency, the variational problem is decomposing the wind speed series, *f*, into $\{u_k(t)\}$, where *k* is number of the modes.

Wind speed time series *f* is assumed decomposed into *k* modes, and each mode has a finite bandwidth and a center frequency. The objective function seeks the *k* modes of wind speed decomposed by VMD, of which the sum value of the estimated bandwidth is the minimum. In addition, the sum of the IMFs must be equal to the input signal, *f*, which is the constraint condition. The estimation procedure of the IMF bandwidth is as follows:

(a) Hilbert transform is used to decompose the wind speed time series, *f*, and the analytic signal of each mode function, $u_k(t)$, is calculated in order to obtain the corresponding unilateral spectrum.
(b) Estimated pulsation is obtained by mixing $e^{-j\omega_k t}$. The frequency spectrum of each mode is modulated to the corresponding baseband.
(c) The L^2 norm of the demodulation signal gradient is calculated and the bandwidth of each mode is estimated.

Subsequently, the variational problem is described as follows:

$$\begin{cases} \min\limits_{\{u_k\},\{\omega_k\}} \left\{ \sum\limits_k \| \partial_t [(\delta(t) + \frac{j}{\pi t}) * u_k(t)]e^{-j\omega_k t} \|_2^2 \right\} \\ \text{s.t.} \sum\limits_k u_k = f \end{cases} \tag{1}$$

where $\{u_k\} = \{u_1, u_2, \cdots, u_k\}$, $\sum\limits_k = \sum\limits_{k-1}^{k}$, $\{\omega_k\} = \{\omega_1, \omega_2, \cdots, \omega_k\}$, and ω_k are the center frequency of the *k*th mode.

To obtain the optimal solution of the constrained variable problem, the issue is addressed as an unconstrained variable problem by introducing a quadratic penalty factor, *α*, and the Lagrange multiplication operator, $\lambda(t)$ [33,34]. The augmented Lagrange equation is defined as:

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

Then, the alternate direction method of multipliers (ADMM) is used to find the iterative sub-optimizations for the augmented Lagrange equation of Equation (2). The solutions for Equation (2) are expressed as:

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

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

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

2.2. Partial Autocorrelation Function (PACF)

The data correlation of the mode decomposed by the VMD in each period is different because of the different wind speed fluctuations in each period. Thus, in the design of the prediction model corresponding to various IMF sequences, the correlation between the data in each IMF needs to be analyzed based on the current decomposition results, and the optimal feature vector of each predictor can be selected by the correlation. PACF is used to evaluate the correlation and the result.

For IMF sub-sequence $1 \leq s \leq L$, let x_t be the forecasting variable of the predictor. If the partial autocorrelation at lag s is outside of the 95% confidence interval, which is calculated as $[-1.96/\sqrt{L}, \ 1.96/\sqrt{L}]$, x_{t-s} is one of the selected features for the predictor [35,38]. If all the PACF coefficients are within the 95% confidence interval, x_{t-1} will be selected as the only input variable of the predictor. The PACF process is as follows [35,38]; the covariance at lag s (if $s = 0$, it is the variance) is described as:

$$\hat{\gamma}_s = \frac{1}{L} \sum_{t=1}^{N-s} (x_t - \overline{x})(x_{t+s} - \overline{x}), \ s = 1, 2, \cdots, M \tag{5}$$

where $\overline{x}$ is the mean value of IMF and M is the maximum lag. Then, the autocorrelation function (ACF) at lag s is:

$$\hat{\rho}_s = \frac{\hat{\gamma}_s}{\hat{\gamma}_0} \tag{6}$$

According to Equations (5) and (6), PACF at lag s can be calculated as:

$$\left. \begin{array}{l} \hat{\alpha}_{1,1} = \hat{\rho}_1 \\ \hat{\alpha}_{s+1,s+1} = \frac{\hat{\rho}_{s+1} - \sum_{j=1}^s \hat{\rho}_{s+1-j}\hat{\alpha}_{sj}}{1 - \sum_{j=1}^s \hat{\rho}_j\hat{\alpha}_{sj}} \\ \hat{\alpha}_{s+1,j} = \hat{\alpha}_{sj} - \hat{\alpha}_{s+1,s+1} \cdot \hat{\alpha}_{s,s-j+1} \\ \qquad\qquad (j = 1, 2, \cdots, s) \end{array} \right\} \tag{7}$$

where $1 \leq s \leq M$.

2.3. Weighted Regularization Extreme Learning Machine (WRELM)

ELM is proposed based on SLFNs. Compared with other intelligent methods, it has the advantages of having a simple structure, fast calculation speeds, a high forecasting accuracy, and fewer training sample requirements [15–18]. According to the wind speed forcasting process, the ELM for regression is constructed in this paper using multiple inputs and a single output.

Assuming a data set with N training samples, $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, where the input vector is $\mathbf{x}_i \in \mathbb{R}^n$ and the output is $y_i \in \mathbb{R}$, the ELM predictor between $\mathbf{x}_i$ and y_i with L hidden nodes can be described as [24]:

$$\sum_{j=1}^L \beta_j g(\mathbf{w}_j \cdot \mathbf{x}_i + b_j) = y_i, \qquad i = 1, 2, \ldots, N \tag{8}$$

where $g(x)$ is the activation function, $\mathbf{w}_j$ is the randomly selected input weight vector between the ith hidden neuron and the input neuron, b_j is the randomly selected bias of the ith hidden node, and β_j is the weight that connects the ith hidden neuron and the output neuron. Equations can be described as a single linear system:

$$\mathbf{H}\beta = \mathbf{y} \tag{9}$$

where $\quad \mathbf{H}(\mathbf{w}_1, \ldots, \mathbf{w}_L, \mathbf{x}_1, \ldots, \mathbf{x}_N, b_1, \ldots, b_L) \quad = \quad \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix}_{N \times L},$

$\beta = [\beta_1, \beta_2, \ldots, \beta_L]^T$, and $y = [y_1, y_2, \ldots, y_N]^T$. The structure of the traditional ELM is shown in Figure 2.
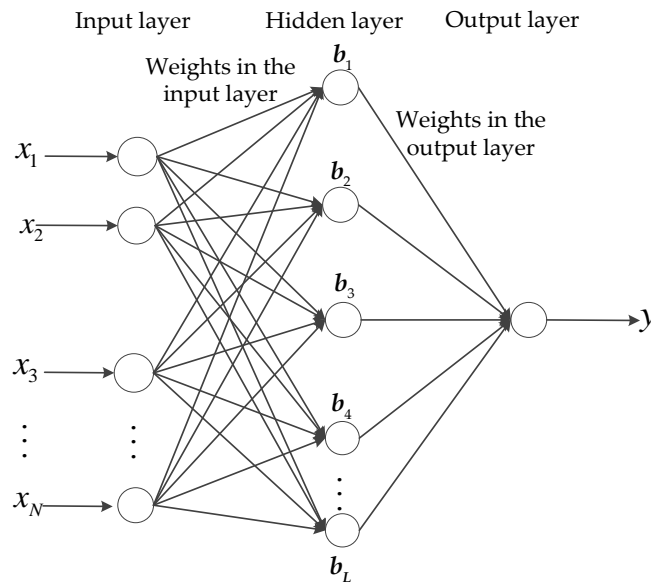

Figure 2. Structure of the extreme learning machine (ELM).

The goal of the ELM is to find the optimal least squares solution, β, by minimizing the associated loss function:

$$\min \| \mathbf{y} - \mathbf{H}\beta \|_2^2 \tag{10}$$

In the process of actual regression application, the training sample is more than the testing sample. Thus, the optimal solution can be written as:

$$\hat{\beta} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y} \tag{11}$$

Though ELM has many advantages, the regression accuracy of ELM still suffers from outliers in the training set. Otherwise, the goal of minimizing the training error might cause an overfitting limitation, which could affect the forcasting accuracy of ELM. To overcome the limitation of ELM, WRELM is employed in order to construct the wind speed predictor. WRELM [24] is designed using a regularized extreme learning machine (RELM) [27]. In RELM, a regularization parameter is used to balance the training error and the norm of output weight in order to avoid the overfitting limitation. WRELM further improves RELM by reducing the influence of outliers. The mathematical model of WRELM can be written as [24]:

$$\min_{\beta} C \parallel \mathbf{We} \parallel_2^2 + \parallel \beta \parallel_2^2 \quad subject\ to\ \mathbf{y} - \mathbf{H}\beta = \mathbf{e} \tag{12}$$

where $\mathbf{W} = diag\{w_1, w_2, \ldots, w_N\}$ is the extension form of $\parallel \mathbf{e} \parallel_2^2$, which is obtained from the error variable weight, e_i, using weighing factor w_i, $\mathbf{e} = [e_1, e_2, \ldots, e_N]^T$ $\mathbf{W} = diag\{w_1, w_2, \ldots, w_N\}$; C is a regularization parameter; and $\mathbf{e}$ is a training error vector with N variables, where $\mathbf{e} = [e_1, e_2, \ldots, e_N]^T$.

The optimal solution of β for WRELM is as follows:

$$\hat{\beta} = (\mathbf{H}^T \mathbf{W}^2 \mathbf{H} + \frac{\mathbf{I}}{C})^{-1} \mathbf{H}^T \mathbf{W}^2 \mathbf{y} \tag{13}$$

where $\mathbf{I}$ is a unit matrix.

Existing research certifies that WRELM has better regression accuracy than ELM or RELM. Moreover, WRELM effectively avoids the influence of overfitting and outliers on the forcasting accuracy [24]. The details of the derivation procedure and the parameter setting of WRELM can be found in Reference [24].

3. Case Study

In this section, the effectiveness of the proposed hybrid approach on historical wind speed observations in order to predict the short-term wind speed is verified using real wind speed data. The observations are obtained from NREL. All the experiments are performed in Matlab 8.5 (2015a) running on an Intel® Core™ i7-6700 processor operating at 3.40 GHz. The WRELM and VMD toolboxes are from Zhang [24] and Dragomiretskiy [33], respectively.

3.1. Data Sets and Evaluation Criteria

The change of wind speed data is uncontrollable because wind speed data is influenced by meteorological factors. The correlation between the wind speed data in the forecast period and the historic wind speed data with a long time interval is low. Therefore, the data quantity for model verification is set as in Reference [32]. The mean half-hour wind speed data for two days (i.e., 96 training samples) are used to train the forecasting model under a new approach. The model will then be constructed between the historical wind speed data and the current wind speed [32]. Considering the fluctuation of wind speeds and weather differences between day and night, the correlation between the 12-hour historical data and the forecasting data is analyzed using PACF in order to construct the optimal input vector of each predictor. Hence, length L of the wind speed time series for correlation analysis is 120, which means that a wind speed time series with 120 samples is decomposed by VMD. The maximum lag, M, is 24, and the number of training samples, N, for WRELM is 96.

In practice, two-thirds of the entire data set is used to train the model and one-third is used to test the forecasting accuracy. The testing set for the new approach contains 48 samples (i.e., one day) because the training set contains 96 samples. The data used for training and testing the effectiveness of the new approach for forecasting the wind speed on 20 February 2004 is shown in Figure 3.

To assess the forecasting accuracy of the new approach and the comparative methods, the mean absolute percentage error (MAPE) and root mean square error (RMSE) are used as evaluation criteria.

$$\text{MAPE} = \frac{1}{K} \sum_{i=1}^{K} \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100\% \tag{14}$$

$$\text{RMSE} = \sqrt{\frac{1}{K} \sum_{i=1}^{K} (x_i - \hat{x}_i)^2} \tag{15}$$

where K is the sample number of the testing set, x_i is the real value of wind speed at i, and $\hat{x}_i$ is the corresponding forecasting value.
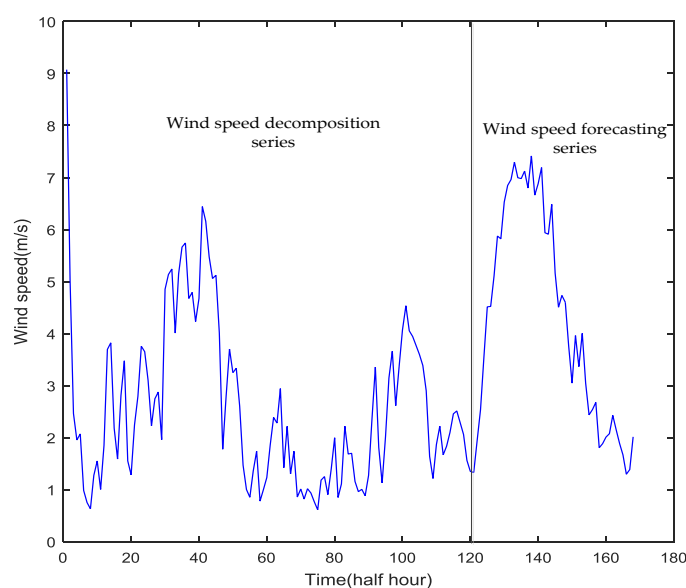
Figure 3. Wind speed series used to verify the effectiveness of the new approach (from 18 February 2004 to 20 February 2004).

3.2. Decomposition of the Wind Speed Series Based on VMD

Before the wind speed sequence is decomposed by VMD, the decomposition layers, *k*, need to be set first. If *k* is too small, the signal is not fully decomposed. Accurately characterizing the signal into different frequency ranges is difficult. If the signal is over-decomposed, the mode in the high frequency is excessive. Reducing the difference between each mode will increase the complexity and reduce the accuracy of the hybrid model.

The optimal mode number can be selected by the difference between the center frequencies (Δf) of IMF *k* and IMF ($k - 1$), as the center frequency is closely related to the decomposition results of VMD [35]. Table 1 shows the center frequency with different modes, *k*, of the decomposing results of the wind speed decomposition series in Figure 3.

As shown in Table 1, the value of x_{t-s} is significantly reduced when the value of *k* is from one to seven, while the value of Δf tends to be stable. Thus, we consider seven as the optimal value of *k*.

Table 1. Center frequency corresponding to different mode number, *k*.

k	Normalized Center Frequency										Δf
	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	IMF10	
1	0.0011	-	-	-	-	-	-	-	-	-	-
2	0.0011	0.2249	-	-	-	-	-	-	-	-	0.2238
3	0.0011	0.0536	0.2254	-	-	-	-	-	-	-	0.1718
4	0.0011	0.0535	0.2254	0.3867	-	-	-	-	-	-	0.1613
5	0.0011	0.0533	0.2245	0.2867	0.3902	-	-	-	-	-	0.1035
6	0.0011	0.0185	0.0682	0.2248	0.2873	0.4412	-	-	-	-	0.1584
7	0.0011	0.0185	0.0681	0.2248	0.2869	0.3845	0.4443	-	-	-	0.0598
8	0.0011	0.0181	0.0640	0.1420	0.2260	0.2878	0.3850	0.4446	-	-	0.0596
9	0.0011	0.0181	0.0639	0.1418	0.2259	0.2864	0.3250	0.3868	0.4449	-	0.0581
10	0.0011	0.0181	0.0548	0.0907	0.1598	0.2265	0.2869	0.3260	0.3871	0.4450	0.0579

To further validate the effectiveness of the method for *k* optimization, wind data from 20 days, randomly selected from 2004, are used in order to verify the forecasting accuracy of the new hybrid wind speed forecasting method with different *k* values, from one to 10. The optimal number of decomposition layers is evaluated by the prediction error (MAPE and RMSE) of one-step

forecasting using the new approach with different decomposition layers. In those 20 days, the optimal decomposition layer with a minimum error of 17 days is seven. The prediction error of seven decomposition layers of other days, with different optimal decomposition layers, is very close to that of the minimum error of the predicted result with the optimal layer number. The mean errors of the different decomposition layers of the one-step forecasting are shown in Table 2.

Table 2. Mean error of one-step-ahead forecasting using the new method with different *k* values.

k	Mean Error	
	MAPE (%)	RMSE (m/s)
1	20.7629	0.4742
2	15.8809	0.3963
3	14.7369	0.3545
4	13.0698	0.3154
5	11.9313	0.2968
6	11.3299	0.2851
7	11.0996	0.2810
8	11.1432	0.2832
9	11.6413	0.2913
10	11.9311	0.3031

As shown in Table 2, when the decomposing layer of VMD is seven ($k = 7$), the mean MAPE and RMSE are the minimum values, 11.0996 and 0.2810, respectively. To obtain optimal prediction results in general, the optimal decomposition layer with the minimum mean error of seven is selected in the VMD decomposition, according to the mean error of the statistical experiment.

On the basis of determining the decomposition level of VMD, a comparative experiment on decomposing a normalized wind speed series is proposed, based on VMD and EMD. The other parameters of VMD are set according to References [34–37], and the parameters of EMD are set according to Reference [30]. The experimental results of series decomposition are shown in Figures 4 and 5.
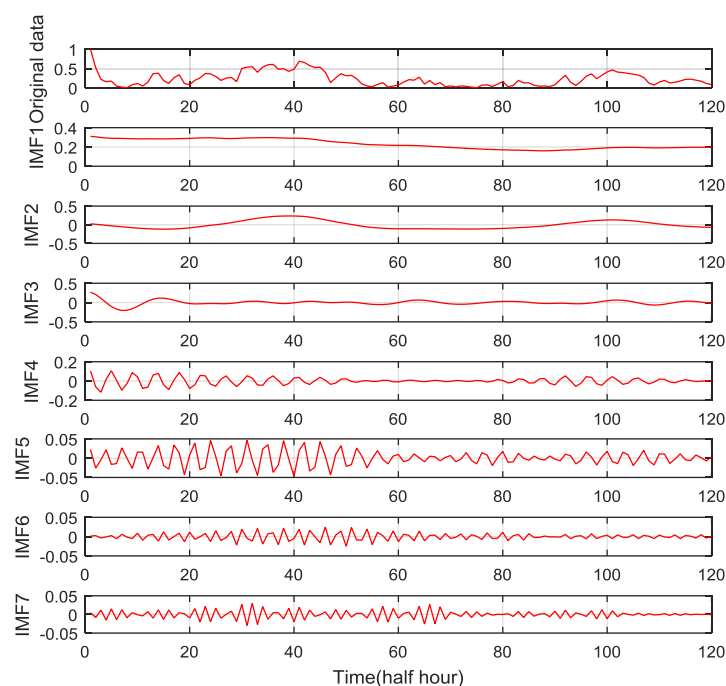

Figure 4. Decomposition results of variational mode decomposition (VMD).
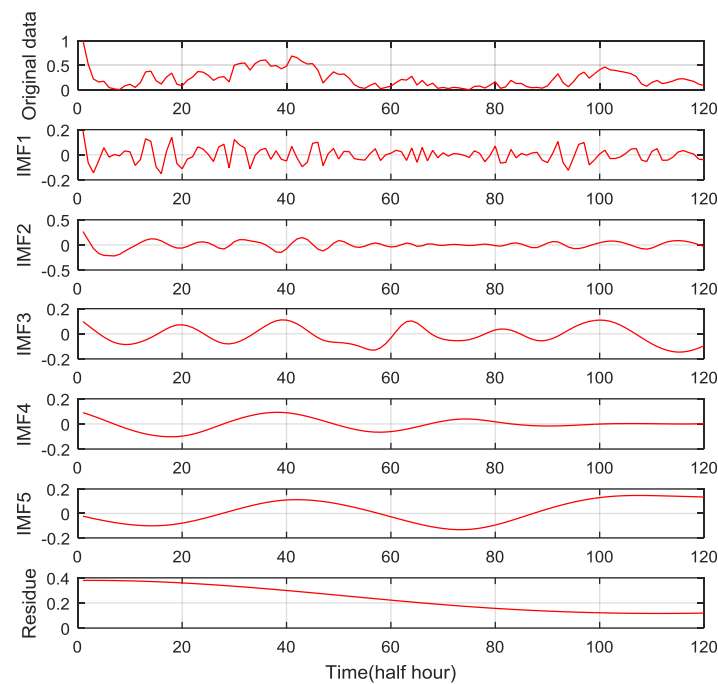
Figure 5. Decomposition results of empirical mode decomposition (EMD).

As shown in Figures 4 and 5, the original wind speed signal can be decomposed by VMD and EMD for seven and five IMFs, respectively. The results of the EMD method also include a residue component.

The IMF sequence generated by VMD in the low-frequency area with a large amplitude had fewer fluctuations. The change trends of the data with a large amplitude, such as IMF1, IMF2, and IMF3 of VMD, were stable. For these types of data, any predictor can achieve a better prediction effect. The predictors corresponding to the IMFs with large amplitudes will have fewer forecasting errors. The IMFs in the high-frequency area with a small amplitude changed drastically. The predictors corresponding to IMFs with small amplitudes will have larger errors. However, the error of the high-frequency predictor has a limited effect on the prediction accuracy of the hybrid model because of the small magnitude of the predicted object. The characteristics of the VMD decomposition results are conducive to the improvement of the prediction accuracy of wind speed forecasting because the final prediction result, based on VMD, is the sum of the prediction results of each IMF.

However, IMFs decomposed by EMD have different characteristics. IMFs with a large amplitude change drastically. The forecasting errors of IMFs with large amplitudes are also sizable because of the intense fluctuations in the EMD decomposition results. The violent fluctuation trends of IMFs with large amplitudes will lead to a larger error in the final prediction results of the EMD-based predictor. Furthermore, the results of the EMD are not conducive to the improvement of the prediction accuracy. IMFs decomposed by VMD are significantly more suitable for the hybrid wind speed forecasting model construction than those decomposed by EMD.

3.3. Optimal Feature Vector Construction Based on PACF

The change trends and the correlation between data elements are different in IMFs decomposed by VMD at different periods. To obtain the optimal predictors, the input variables of the predictors should be selected according to the decomposition results of VMD. In this paper, the length of the wind series, L, is 120, and its confidence interval of 95% is $[-0.179, +0.179]$. The partial autocorrelation of the original wind speed series and the IMFs decomposed by VMD are shown in Figure 6. According to the partial autocorrelograms, the variable with lag s, denoted as x_{t-s} with the partial autocorrelation value out of the confidence interval, will be selected to construct the input vector of the predictor corresponding to the specific IMF.
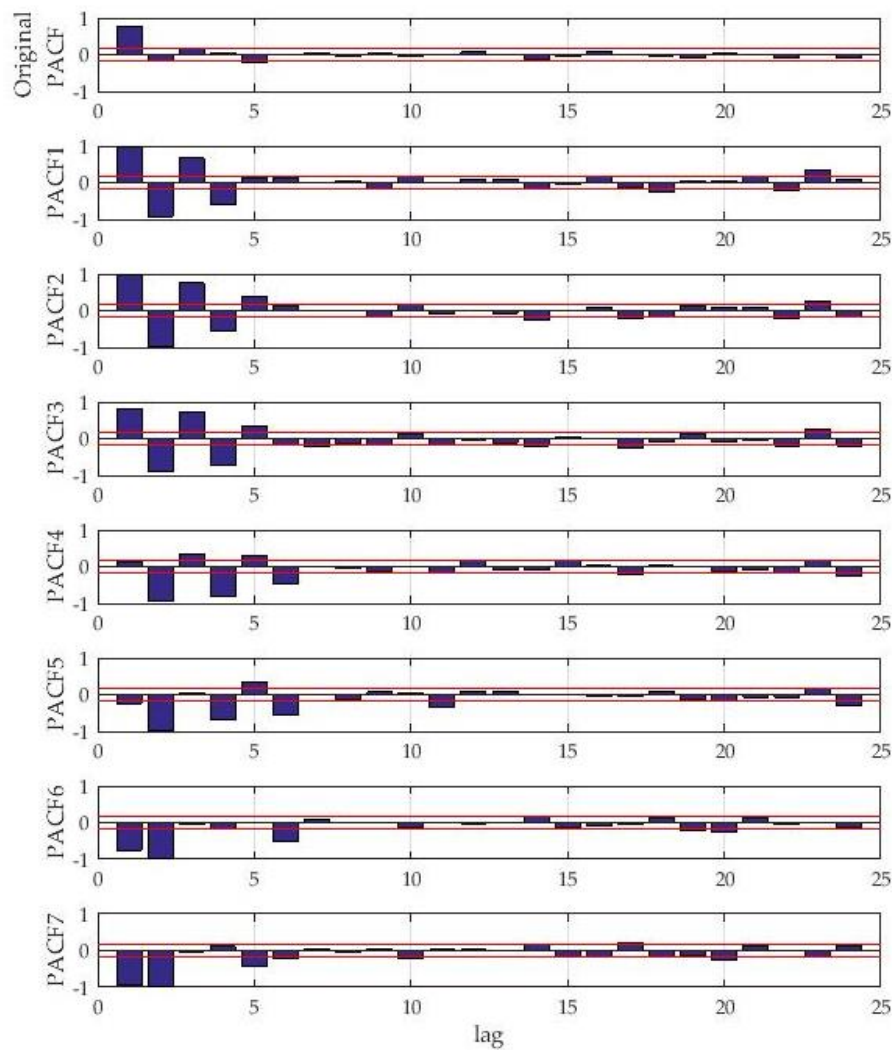
Figure 6. PACFs of the original data and the IMFs extracted from VMD.

Figure 6 shows the characteristics of the partial autocorrelation of the original wind speed and the IMFs. Only two variables can be selected by PACF because of the large fluctuation in the original wind speed. However, IMFs decomposed by VMD have a certain stability and regularity; thus, more related variables can be selected by PACF. Compared to the original wind speed series, the IMFs of VMD can easily establish a prediction model. The compositions of the input vectors of different predictors determined by PACF are shown in Table 3.

Table 3. Input variables determined by PACF.

Time Series	Input Variables
Original	x_{t-1}, x_{t-4}
IMF1	$x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-9}, x_{t-10}, x_{t-14}, x_{t-18}, x_{t-22}, x_{t-23}$
IMF2	$x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-14}, x_{t-17}, x_{t-22}, x_{t-23}$
IMF3	$x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-7}, x_{t-14}, x_{t-17}, x_{t-22}, x_{t-23}, x_{t-24}$
IMF4	$x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-15}, x_{t-17}, x_{t-24}$
IMF5	$x_{t-1}, x_{t-2}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-11}, x_{t-24}$
IMF6	$x_{t-1}, x_{t-2}, x_{t-6}, x_{t-19}, x_{t-20}$
IMF7	$x_{t-1}, x_{t-2}, x_{t-5}, x_{t-6}, x_{t-10}, x_{t-17}, x_{t-17}, x_{t-20}, x_{t-23}$

3.4. Forecasting Model for Wind Speed Based on WRELM

After determining the input vector, WRELMs are used to predict the IMFs decomposed by VMD. Using MAPE as the index, and randomly selecting 10 days as the test set, the activation function and the number of hidden layer neurons of WRELM are determined using the cross-validation method [25,28]. The forecasting accuracies with different characteristics of WRELM are shown in Table 4. As Table 4 shows, the WRELM with the "sine" activation function and 20 hidden layer neurons has the highest forecasting accuracy. Thus, the characteristics of WRELM will be selected by the result.

Table 4. Forecasting accuracy with different characters of WRELM.

Activation Function	MAPE(%) of Different Hidden Layer Numbers					
	5	10	15	20	25	30
Sine	11.7280	10.7585	10.7009	10.6955	10.7110	10.7631
Sigmoid	12.0360	10.7611	10.7357	10.7356	10.7398	10.7623
Triangular basis	12.4982	10.8987	10.7276	10.7135	10.7130	10.7922
Hard-limit	13.0039	10.7667	10.7727	10.8188	10.8289	10.7881
Radial basis	11.7583	10.8961	10.9217	10.8683	10.7957	10.8254

3.5. Forecasting Results of the New Hybrid Model

To comprehensively evaluate the prediction performance of the new hybrid model, four days in different quarters of 2004 are used to predict wind speed. The one-step and multi-step prediction (from two steps ahead to three steps ahead) results are shown in Figures 6–10.
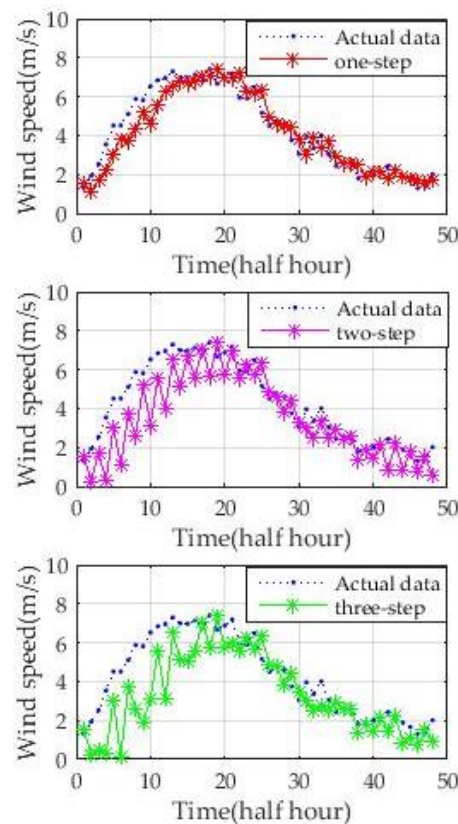

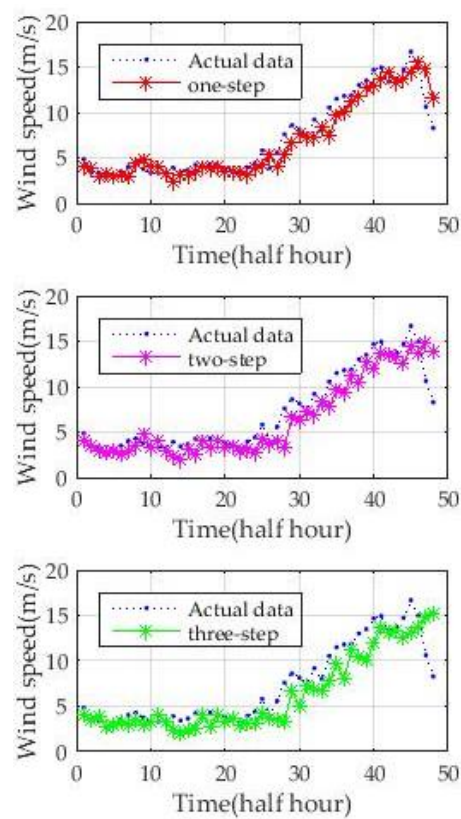**Figure 7.** Forecasting results on 20 February 2004.
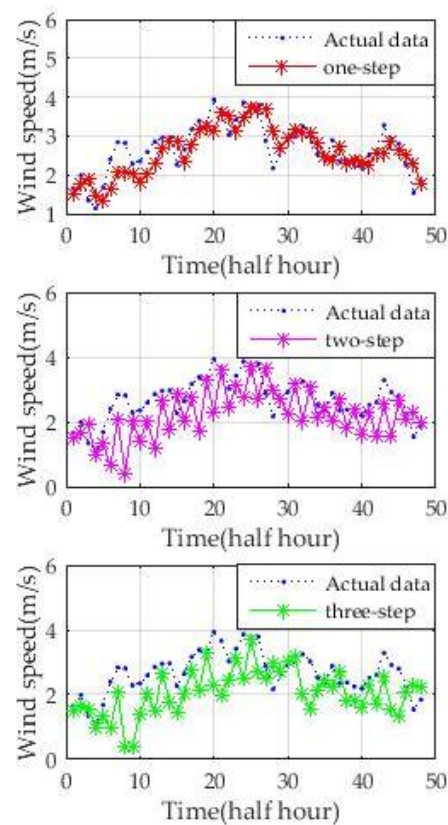
Figure 8. Forecasting results on 20 May 2004.

Figure 9. Forecasting results on 20 August 2004.
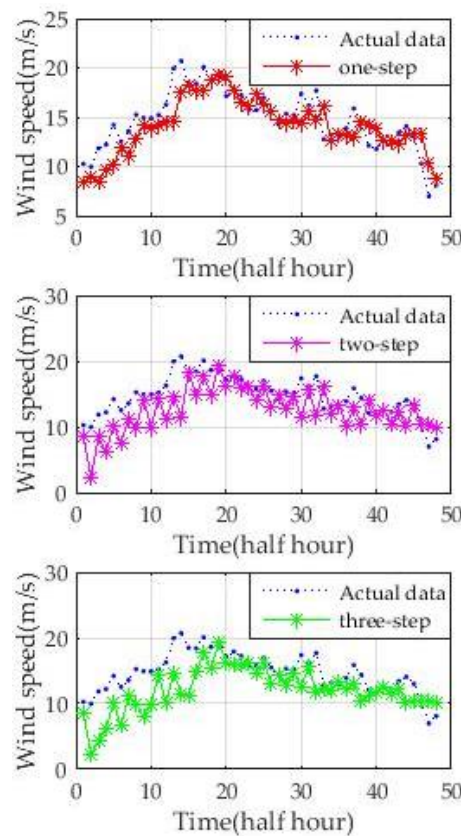
Figure 10. Forecasting results on 20 November 2004.

3.6. Comparison Experiments and Discussion

Figures 7–10 show that, although the wind speed curves on different dates have different fluctuations and ranges of speed, the new model could accurately predict wind speed. One-step-ahead prediction results are very close to the actual wind speed curve. The forecasting accuracy of two-step-ahead and three-step-ahead predictions declined to different degrees. However, the results of the multi-step-ahead forecasting of the new approach are still close to those of the real wind speed. The experiments illustrate the stability and adaptability of the new approach.

The experiment with five types of hybrid wind speed forecasting methods is proposed in this section in order to show the advancements of the new hybrid approach. The wind speed data from 20 February 2004, 20 May 2004, 20 August 2004, and 20 November 2004 are used to verify the accuracy of the different approaches. The one-step-ahead and multi-step-ahead forecasting results of the different approaches are shown in Figures 11–14. The mean forecasting errors with five types of wind speed approaches, in four days, are shown in Table 5.

The methods used for the experiments include VMD + ELM, VMD + WRELM, EMD + PACF + WRELM, VMD + PACF + ELM, and VMD + PACF + WRELM. The predictors without the feature selection step use the variables from lag 1 to lag 24 as the input vectors. The ELM characteristics of each method are optimized by cross-validation [21,22].

The contrast methods used to verify the advancement of the new approach do not consider methods with different intelligent algorithms because existing research has illustrated the advanced nature of ELM relative to traditional intelligent algorithms, such as Back Propagation Neural Networks (BPNN) and Support Vector Machine (SVM) [23,24]. The experiments mainly show the advancement of PACF and WRELM used in the new approach.

As shown in Figures 11–14, the new approach has the highest accuracy out of the five approaches. The forecasting curve of the new approach is closest to that of the true wind speed curve. Detailed analyses are performed based on Table 5.
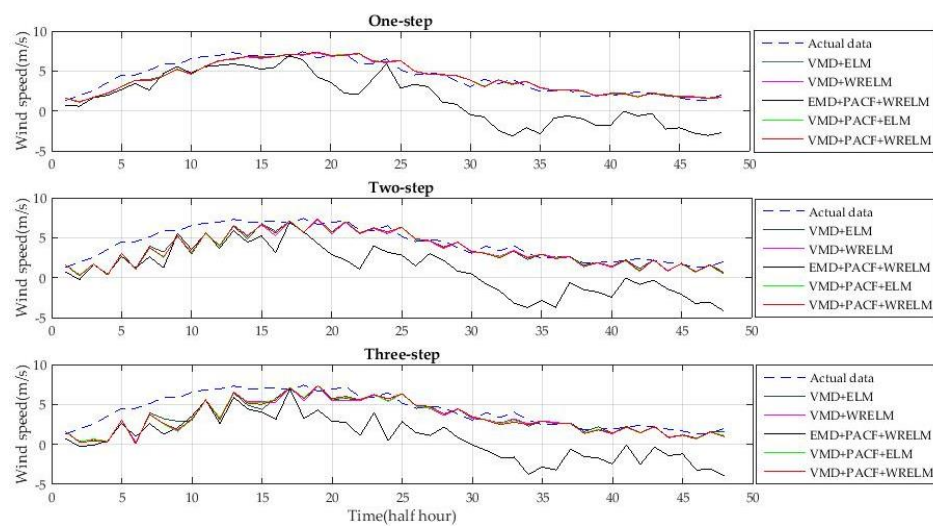
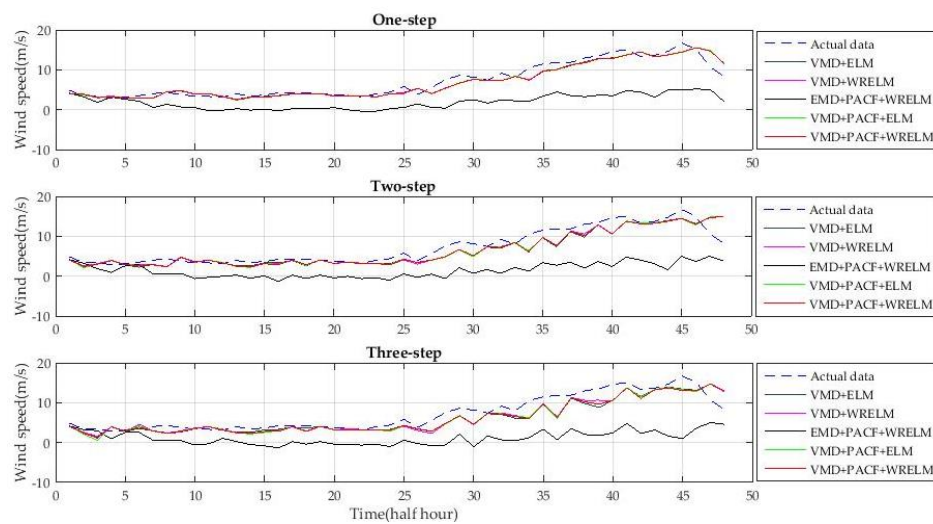**Figure 11.** Forecasting results on 20 February 2004, using different methods.

Figure 12. Forecasting results on 20 May 2004, using different methods.
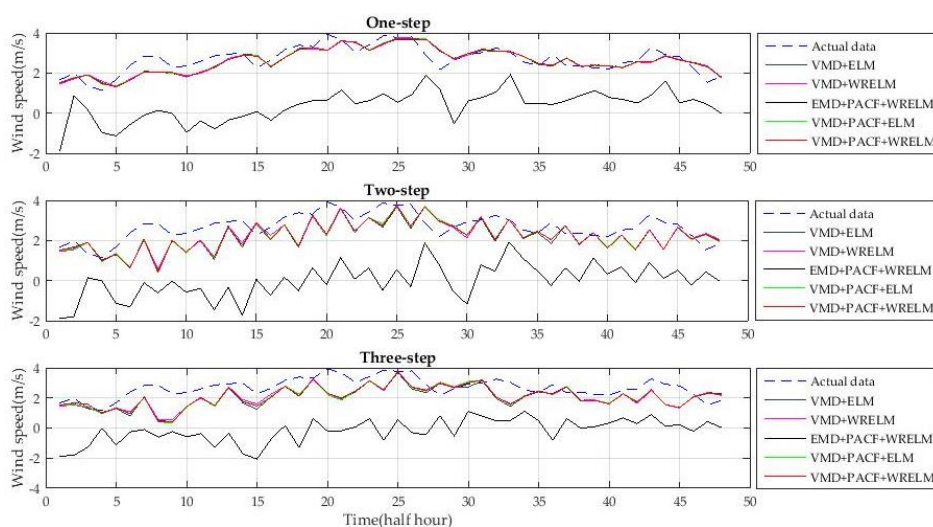

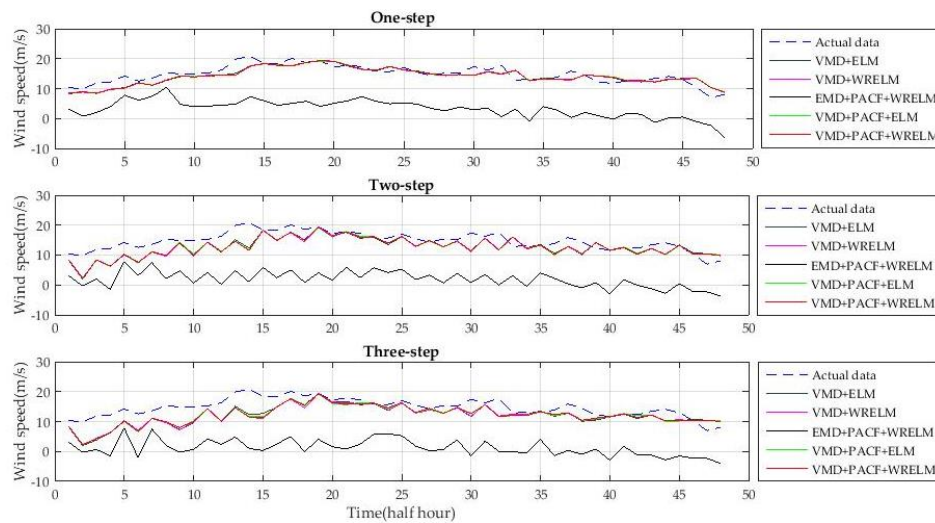**Figure 13.** Forecasting results on 20 August 2004, using different methods.

Figure 14. Forecasting results on 20 November 2004, using different methods.

Table 5. Mean forecasting error of different methods.

Methods	MAPE (%)			RMSE (m/s)		
	One-Step	Two-Step	Three-Step	One-Step	Two-Step	Three-Step
VMD + ELM	11.8011	16.4512	18.4901	0.2853	0.4845	0.5854
VMD + WRELM	11.4518	15.3523	17.6016	0.2634	0.4769	0.5536
EMD + PACF + WRELM	16.6350	18.7262	20.7317	0.5540	0.8835	1.2644
VMD + PACF + ELM	11.3644	15.8224	17.8260	0.2758	0.4691	0.5491
VMD + PACF + WRELM	11.0912	14.6710	16.7378	0.2590	0.4378	0.5382

Compared to VMD + ELM without PACF, VMD + PACF + ELM reduced MAPE from 11.8011, 16.4512, and 18.4901 to 11.3644, 15.8224, and 17.8260, respectively. Compared to the VMD + WRELM method, VMD + PACF + WRELM reduced MAPE from 11.4518, 15.3523, and 17.6016 to 11.0912, 14.6710, and 16.7378, respectively. The accuracy evaluated by the RMSE was also higher using the methods that included PACF. These verified the benefits of PACF relative to the improvement of the forecasting accuracy of the entire model.

Meanwhile, according to Table 5, the hybrid method with WRELM has a higher forecasting accuracy than methods with ELM in each type of multi-step forecasting. The method with VMD decomposition has a much higher forecasting accuracy than the EMD-based method. The experimental results verified the advancements of WRELM and VMD.

To further illustrate the effectiveness and advancement of the new approach, it was used to predict wind speeds in a longer time interval (five days). In this experiment, the new approach is compared with the Autoregressive Integrated Moving Average (ARIMA) and Support Vector Regression (SVR) based methods. SVR is used instead of WRELM to construct a contrast method with VMD, PACF, and SVR. The kernel function charaters of SVR are set according to References [10,39]. The optimal penalty factor and kernel function variance of SVR are determined using the cross-validation method. The characteristics of ARIMA are set according to References [5,12]. The experimental results of the long interval are shown in Figure 15 and Table 6.

Figure 15 shows that the new approach has the best prediction accuracy in the experiment. The prediction results of the new approach accurately track the change trends of the true wind speed, especially in the time interval of 100 to 150 sample points. When the real wind speed suddenly changed from a high speed to a low speed, the new approach predicted the change trend much more accurately than the other methods. Table 6 shows that the MAPE and RMSE of the new approach are much lower than those of the other methods. The experimental results fully verified the advancement of the new approach.
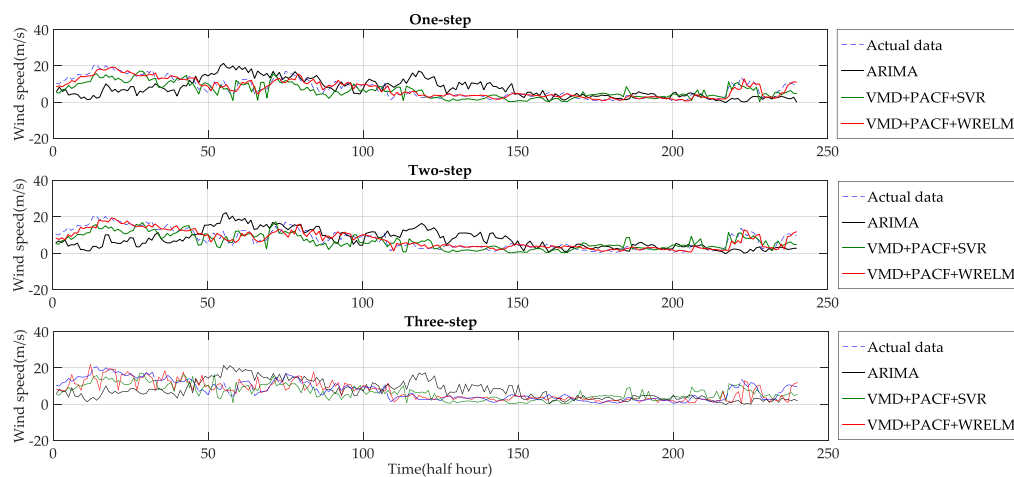
Figure 15. Forecasting results from 20–24 November 2004, using different methods.

Table 6. The mean forecasting errors of different methods.

Methods	MAPE (%)			RMSE (m/s)		
	One-Step	Two-Step	Three-Step	One-Step	Two-Step	Three-Step
ARIMA	20.3974	26.6640	30.6191	0.2773	0.5405	0.6149
VMD + PACF + SVR	14.9761	18.4143	20.1438	0.2662	0.4665	0.6093
VMD + PACF + WRELM	13.6665	16.6635	18.9372	0.2652	0.4512	0.5943

4. Conclusions

To improve the short-term wind speed forecasting accuracy, a new hybrid wind speed forecasting approach, based on VMD, PACF, and WRELM, is proposed. This new approach has the following advantages:

(1) The wind speed series is decomposed by VMD with an optimal decomposing layer selected through center frequency analysis in order to reduce the influences of randomness and volatility. The results of VMD are more precise than those of EMD, and show that it is suitable for hybrid forecasting model construction with fewer amplitude fluctuations in the IMFs of a low frequency range.
(2) PACF, which is used for the correlation analysis of IMFs, determines the optimal input vector of each predictor. It effectively reduces the feature dimension and complexity of the model and leads to the improvement of prediction efficiency and accuracy.
(3) WRELM is used to build the predictors of the hybrid model. The weights of WRELM are adjusted automatically according to different training sets. WRELM improves the prediction accuracy by avoiding the negative effects of the outlier samples in the data set.

The experimental results show that the new method is effective and advanced.

Author Contributions: Nantian Huang put forward to the main idea and designed the entire structure of this paper. Chong Yuan did the experiments and prepared the manuscript. Guowei Cai guided the experiments and paper writing. Enkai Xing completed data preprocessing.

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

References

1. Men, Z.X.; Yee, E.; Lien, F.S.; Wen, D.Y.; Chen, Y.S. Short-term wind speed and power forecasting using an ensemble of mixture density neural networks. *Renew. Energy* **2016**, *87*, 203–211. [CrossRef]
2. Potter, C.W.; Lew, D.; McCaa, J.; Cheng, S.; Eichelberger, S.; Grimit, E. Creating the dataset for the western wind and solar integration study (U.S.A.). *Wind Eng.* **2008**, *32*, 325–338. [CrossRef]
3. Heng, J.N.; Wang, C.; Zhao, X.J.; Xiao, L.Y. Research and application based on adaptive boosting strategy and modified CGFPA algorithm: A case study for wind speed forecasting. *Sustainability* **2016**, *8*, 235. [CrossRef]
4. Hu, Q.; Zhang, S.; Yu, M.; Xie, Z. Short-term wind speed or power forecasting with heteroscedastic support vector regression. *IEEE Trans. Sustain. Energy* **2016**, *7*, 241–249. [CrossRef]
5. Cadenas, E.; Rivera, W.; Campos-Amezcua, R.; Heard, C. Wind speed prediction using a univariate ARIMA model and a multivariate NARX model. *Energies* **2016**, *9*, 109. [CrossRef]
6. Noorollahi, Y.; Jokar, M.A.; Kalhor, A. Using artificial neural networks for temporal and spatial wind speed forecasting in Iran. *Energy Convers. Manag.* **2016**, *115*, 17–25. [CrossRef]
7. Doucoure, B.; Agbossou, K.; Cardenas, A. Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data. *Renew. Energy* **2016**, *92*, 202–211. [CrossRef]
8. Zhao, J.; Guo, Z.H.; Su, Z.Y.; Zhao, Z.Y.; Xiao, X.; Liu, F. An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed. *Appl. Energy* **2016**, *162*, 808–826. [CrossRef]
9. Sfetsos, A. A novel approach for the forecasting of mean hourly wind speed time series. *Renew. Energy* **2002**, *27*, 163–174. [CrossRef]
10. Santamaría-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. [CrossRef]
11. Wang, J.Z.; Zhang, F.Y.; Liu, F.; Ma, J.J. Hybrid forecasting model-based data mining and genetic algorithm-adaptive particle swarm optimisation: A case study of wind speed time series. *IET Renew. Power Gener.* **2016**, *10*, 287–298. [CrossRef]
12. Lydia, M.; Kumar, S.S.; Selvakumar, A.I.; Kumar, G.E.P. Linear and non-linear autoregressive models for short-term wind speed forecasting. *Energy Convers. Manag.* **2016**, *112*, 115–124. [CrossRef]
13. Hu, Q.H.; Zhang, R.J.; Zhou, Y.C. Transfer learning for short-term wind speed prediction with deep neural networks. *Renew. Energy* **2016**, *85*, 83–95. [CrossRef]
14. Hill, D.C.; McMillan, D.; Bell, K.R.W.; Infield, D. Application of auto-regressive models to U.K. wind speed data for power system impact studies. *IEEE Trans. Sustain. Energy* **2012**, *3*, 134–141. [CrossRef]
15. 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. [CrossRef]
16. 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. [CrossRef]
17. 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. [CrossRef]
18. Shrivastava, N.A.; Lohia, K.; Panigrahi, B.K. A multiobjective framework for wind speed prediction interval forecasts. *Renew. Energy* **2016**, *87*, 903–910. [CrossRef]
19. Huang, G.-B. What are extreme learning machines? Filling the gap between Frank Rosenblatt's dream and John von Neumann's puzzle. *Cogn. Comput.* **2015**, *7*, 263–278. [CrossRef]
20. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]
21. Wang, J.Z.; Hu, J.M.; Ma, K.L.; Zhang, Y.X. A self-adaptive hybrid approach for wind speed forecasting. *Renew. Energy* **2015**, *78*, 374–385. [CrossRef]
22. Liu, H.; Tian, H.Q.; Li, Y.F. Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms. *Energy Convers. Manag.* **2015**, *100*, 16–22. [CrossRef]
23. Zong, W.W.; Huang, G.B.; Chen, Y.Q. Weighted extreme learning machine for imbalance learning. *Neurocomputing* **2013**, *101*, 229–242. [CrossRef]
24. Zhang, K.; Luo, M. Outlier-robust extreme learning machine for regression problems. *Neurocomputing* **2015**, *151*, 1519–1527. [CrossRef]
25. Sun, W.; Liu, M.H. Wind speed forecasting using FEEMD echo state networks with RELM in Hebei, China. *Energy Convers. Manag.* **2016**, *114*, 197–208. [CrossRef]

26. Deng, W.; Zheng, Q.; Chen, L. Regularized extreme learning machine. In Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining, Nashville, TN, USA, 30 March–2 April 2009; pp. 389–395.

27. Martínez-Martínez, J.M.; Escandell-Montero, P.; Soria-Olivas, E.; Martín-Guerrero, J.D.; Magdalena-Benedito, R.; Gómez-Sanchis, J. Regularized extreme learning machine for regression problems. *Neurocomputing* **2011**, *74*, 3716–3721. [CrossRef]

28. Meng, A.B.; Ge, J.F.; Yin, H.; Chen, S.Z. Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm. *Energy Convers. Manag.* **2016**, *114*, 75–88. [CrossRef]

29. Fei, S.W. A hybrid model of EMD and multiple-kernel RVR algorithm for wind speed prediction. *Int. J. Electr. Power Energy Syst.* **2016**, *78*, 910–915. [CrossRef]

30. Wang, S.X.; Zhang, N.; Wu, L.; Wang, Y.M. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew. Energy* **2016**, *94*, 629–636. [CrossRef]

31. Hu, J.M.; Wang, J.Z. Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression. *Energy* **2015**, *93*, 1456–1466. [CrossRef]

32. Hu, J.M.; Wang, J.Z.; Ma, K.L. A hybrid technique for short-term wind speed prediction. *Energy* **2015**, *81*, 563–574. [CrossRef]

33. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]

34. Lahmiri, S. Intraday stock price forecasting based on variational mode decomposition. *J. Comput. Sci.* **2016**, *12*, 23–27. [CrossRef]

35. Sun, G.Q.; Chen, T.; Wei, Z.N.; Sun, Y.H.; Zang, H.X.; Chen, S. A carbon price forecasting model based on variational mode decomposition and spiking neural networks. *Energies* **2016**, *9*, 54. [CrossRef]

36. Wang, Y.X.; Markert, R. Filter bank property of variational mode decomposition and its applications. *Signal Process.* **2016**, *120*, 509–521. [CrossRef]

37. Zhang, Y.C.; Liu, K.P.; Qin, L.; An, X.L. Deterministic and probabilistic interval prediction for short-term wind power generation based on variational mode decomposition and machine learning methods. *Energy Convers. Manag.* **2016**, *112*, 208–219. [CrossRef]

38. Guo, Z.H.; Zhao, W.G.; Lu, H.Y.; Wang, J.Z. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. *Renew. Energy* **2012**, *37*, 241–249. [CrossRef]

39. Hu, Q.H.; Zhang, S.G.; Xie, Z.X.; Mi, J.S.; Wan, J. Noise model based v-support vector regression with its application to short-term wind speed forecasting. *Neural Netw.* **2014**, *57*, 1–11. [CrossRef] [PubMed]