

Article

Deterministic and Interval Wind Speed Prediction Method in Offshore Wind Farm Considering the Randomness of Wind

Qin Chen ¹, Yan Chen ^{1,2,*} and Xingzhi Bai ¹

¹ Key Laboratory of Intelligent Manufacturing Technology, Ministry of Education, Shantou University, Shantou 515063, China; 18qchen5@stu.edu.cn (Q.C.); 17xzbai@alumni.stu.edu.cn (X.B.)

² Institute of Energy Science, Shantou University, Shantou 515063, China

* Correspondence: ychen@stu.edu.cn; Tel.: +86-1382-968-2901

Received: 9 September 2020; Accepted: 22 October 2020; Published: 26 October 2020



Abstract: In order to improve the prediction accuracy of wind speed, this paper proposes a hybrid wind speed prediction (WSP) method considering the fluctuation, randomness and nonlinear of wind, which can be applied to short-term deterministic and interval prediction. Variational mode decomposition (VMD) decomposes wind speed time series into nonlinear series Intrinsic mode function 1 (IMF1), stationary time series IMF2 and error series (ER). Principal component analysis-Radial basis function (PCA-RBF) model is used to model the nonlinear series IMF1, where PCA is applied to reduce the redundant information. Long short-term memory (LSTM) is used to establish a stationary time series model for IMF2, which can better describe the fluctuation trend of wind speed; mixture Gaussian process regression (MGPR) is used to predict ER to obtain deterministic and interval prediction results simultaneously. Finally, above methods are reconstructed to form VMD-PRBF-LSTM-MGPR which is the abbreviation of hybrid model to obtain the final results of WSP, which can better reflect the volatility of wind speed. Nine comparison models are built to verify the availability of the hybrid model. The mean absolute percentage error (MAE) and mean square error (MSE) of deterministic WSP of the proposed model are only 0.0713 and 0.3158 respectively, which are significantly smaller than the prediction results of comparison models. In addition, confidence intervals (CIs) and prediction interval (PIs) are compared in this paper. The experimental results show that both of them can quantify and represent forecast uncertainty and the PIs is wider than the corresponding CIs.

Keywords: wind characteristic; VMD; hybrid model; deterministic prediction; interval prediction

1. Introduction

Renewable sources of energy are gradually replacing tradition sources in many countries, not only because of the limited of fossil and nuclear fuel energy sources but unprecedented and irreversible damage to the environment caused by them, as well as the continuing reduction in the cost of renewable technologies. Wind energy has been growing rapidly in recent years. In 2018, the global installed capacity of offshore wind power increased by 4.3 GW with a total installed capacity of 23 GW. It has grown by nearly 30% a year since 2010 [1]. According to a new report by Ember, a European think tank on climate and energy, wind and solar power supply 10% of the global electricity demand in the first half of 2020, double the level of 2015.

Compared with other major power generation, wind has nonlinear characteristics [2], and reliable wind speed prediction (WSP) can reduce the impact of these characteristics on the power system [3,4]. Reference [5] study the existence of low-dimensional deterministic chaos in wind times series through

characteristic extracting, estimation of invariant measures and noise removal. The results of reference [6] indicate wind has diverse characteristics such as stochastic, deterministic and chaotic behavior and a cluster of deterministic models built upon separate frequency components of a wind speed time series can enhance the prediction accuracy. Reference [7] build an adaptive WSP model based on the chaotic characteristic combined with the spatial reconstruction theory.

At present, WSP is mainly divided into deterministic prediction and interval prediction. Deterministic prediction can obtain the prediction value of wind speed, while the interval prediction can obtain the uncertainty of the measured wind speed occurring at a certain time in the future. When large-scale wind power is connected to the power grid, wind power fluctuation will affect the balance of supply and demand of the power grid, and may endanger the security of power system in serious cases [3]. Interval prediction of wind speed can reduce the adverse impact of wind randomness and fluctuation on power system reliability, and interval prediction is crucial in the planning of economical dispatch, aiming to an efficient and economical wind power integration and operation [8]. In addition, wind speed is not only the main factor affecting wind turbine operation, but also is an important indicator of wind field location, which mainly involves wind speed deterministic prediction. Since the wind power has a cubic relationship with wind speed, any error in the wind speed forecast leads to a larger error in wind power production. So WSP is a key to wind power forecasting.

WSP is directly related to the characteristics of original data, data processing methods and data prediction models. In recent years, research reports on WSP is divided into four main categories [9–17].

- Physical modes, which mainly use numerical weather prediction (NWP) data to complete wind speed prediction by establishing variable ratio expressions of wind speed and air pressure, air density, air humidity, etc. [13].
- Statistical models, which mainly use time series modeling, including autoregressive integrated moving average model (ARIMA) and Kalman filtering, etc. [14].
- Temporal-spatial correlation models, which mainly use data from different measuring points to predict wind speed [15].
- Artificial intelligence models, which are hot spots for WSP, such as recurrent neural network (RNN), support vector machine (SVM), and fuzzy logic method, etc. [16].

Most scholars choose one or more of above methods for prediction, but studies show that the single statistical method is not accuracy enough in wind speed prediction. Therefore, the academia proposes a hybrid model based on statistical methods of which prediction accuracy is better than the single model [17]. In reference [18], the effective information is extracted by CNN sequence feature extraction ability, and the data is input to LSTM network after other information is removed. Reference [19] proposes using momentum item to speed up the convergence rate of BP, and taking the advantage of the genetic algorithm (GA) to optimize the structure, weights and bias of BP. Reference [20] performs grey correlation analysis on other meteorological factors such as temperature to achieve dimension reduction, then build LSTM model based on that. Reference [21] uses EMD to decompose wind speed series into different IMFs, and build LSTM model, which has high prediction accuracy. In reference [22], EMD and ensemble empirical mode distribution (EEMD) are used for decomposition of wind speed data into IMFs, and an ANN is proposed to predict the wind speed with IMFs given as the inputs. Reference [23] proposes a WSP method based on deep learning combined by CNN and RNN using the big data of a certain wind form from 2014 to 2015.

Considering the characteristics of randomness, volatility, low energy density and non-linearity and above analysis, wind speed should be decomposed into several components and studied separately. At present, wavelet transform (WT), EMD methods and their modified versions such as complementary EEMD, empirical wavelet transform (EWT), EEMD, and other signal decomposition methods have been widely used [24]. Variational mode decomposition (VMD) was developed by Dragomiretskiy and Zosso, which is a novel type of complex signal decomposition on the basis of EMD [25].

VMD is a completely intrinsic, self-adaptive and non-recursive decomposition technique, by which a signal can be decomposed into a finite broadband mode set and show the partial feature of the signal. It has significant advantages in solving the problem of signal noise and avoiding mode overlap [25]. Wind speed series can be decomposed by VMD algorithm into intrinsic mode functions (IMFs), which can effectively reflect the characteristics of randomness and fluctuation. In addition, wind speed is affected by meteorological factors, such as wind direction, temperature, humidity, air pressure and other parameters. Principal component analysis algorithm (PCA) can effectively remove redundant data and realize data dimension reduction. In order to ensure the reliability of the data, the number of data samples involved in this paper is 350, which belongs to small sample data. It is more suitable to solve the problem by means of machine learning, such as LSTM. On the contrary, deep learning is more suitable for the case of large quantity, because more parameters and data quantity are needed to fully train the deep learning model. If the data quantity is small, the parameters cannot be fully trained, and the accuracy of the deep learning model is not as good as machine learning. This paper mainly includes two parts, deterministic prediction and interval prediction. For deterministic prediction part, RBF neural network model is used to model IMF with obvious nonlinear characteristic in this paper, because RBF has the advantages of infinite approximation to nonlinear functions and overcoming local optimal solution [26]. LSTM is used to model and fit the mode with small fluctuation and non-obvious nonlinear characteristics. Long short term memory network (LSTM) extended from RNN network is a special algorithm for processing time-series problem, which solves the problem of gradient explosion or disappearance of RNN and can learn long-term dependent information and is suitable for time series classification and prediction [17]. The noise sequence reflects the fluctuation and randomness of wind speed. In this paper, mixture Gaussian process regression (MGPR) model is used to model the noise sequence and the prediction results of error sequence are obtained. Finally, the final wind speed prediction results are obtained by summing and reconstructing the initial prediction results of each mode and the error sequence. As for interval prediction, confidence interval (CIs) and prediction interval (PIs) are two well-known methods for quantifying and representing the uncertainty of predictions [27]. A quantitative comparison is made between CIs and PIs of wind speed under a certain confidence degree, where CIs are obtained by a single model MGPR and the hybrid model proposed in this paper respectively and PIs are obtained by the hybrid model combining the lower and upper bound estimation method (LUBE).

In this paper, a new hybrid model is constructed by combining a variety of methods to predict wind speed, and the proposed model in this paper can be applied for short-term deterministic and interval WSP. Deterministic prediction can help optimize power system scheduling and improve the reliability of wind power grid connection. Interval prediction can quantify the volatility and intermittent of wind power, and improve the security and economy of wind grid. The primary innovations are as follows.

Firstly, VMD is used to decompose the original wind speed sequence into different IMFs, so as to achieve de-noising while retaining the original information and enhance the predictability of wind speed. Additionally, the comparison between EMD and VMD shows VMD is more suitable for this paper. Secondly, as for prediction model with multiple input variables, PCA is applied to data dimension reduction to remove redundant information. The experimental results show that the prediction results with PCA is superior to those without PCA. Thirdly, corresponding reasonable model is established for each IMF obtained from VMD algorithm to improve prediction accuracy. Fourthly, the hybrid model VMD-PRBF-LSTM-MGPR which is obtained by reconstructing models for each mode can be applied to deterministic prediction and interval prediction simultaneously, where deterministic prediction has high accuracy and can represent the fluctuation of wind speed. The mean absolute percentage error (MAE), relative mean square error (RMSE) and mean square error (MSE) of deterministic prediction of the hybrid model proposed in this paper are only 0.0713 and 0.3158 respectively, which are more accurate than the prediction results of comparison models. In addition, CIs and PIs are compared in this paper. The results show that the prediction interval is wider than the corresponding confidence interval.

This paper is divided into the five parts. The second section describes the basic introduction of methods used in this paper. The third section mainly introduces the process of WSP including deterministic and interval prediction, which consists of three parts, data processing, establishment of comparison models and hybrid model, and validation of the proposed model. The fourth section shows the WSP results and discussion through a numeral case, followed by conclusion.

2. Introduction of Basic Algorithm for Building the Model

2.1. Variational Mode Decomposition

Variational modal decomposition (VMD) is proposed by Desomimickiy and Zosso on the basis of EMD, and is an adaptive, quasi-orthogonal and completely non-recursive decomposition method based on Hilbert transform, classical Weiner filtering and mixed frequency variational problem. EMD is a classical mode decomposition method proposed by Huang et al., which is lack of mathematical theory, ability to properly cope with noise and the hard band-limits of wavelet approaches [25]. Compared with EMD, VMD decomposes the signal into finite bandwidth with different center frequency according to the preset model number K . VMD uses alternating direction multiplier method (ADMM) to constantly update each mode and its center frequency, and gradually adjusts each mode to the corresponding base band, then extracts each mode and its corresponding center frequency, and eventually obtains each component with no central frequency.

The essence of the VMD is the variational problem, which is divided into the construction and solution of the variational problems, and the process is as follows.

Hilbert transform was used to calculate the correlation analysis signal of each mode u_k , and the unilateral frequency spectrum was obtained.

For each mode u_k , the frequency spectrum of the mode is shifted to base-band by tuning to the exponential mixing of the respective estimated center frequency.

The bandwidth is estimated by the H^1 Gaussian smoothness of demodulation signal. The resulting constrained variational problem is the following:

$$\begin{aligned} \min_{\{u_k\}, \{w_k\}} & \left\{ \sum_k \|\partial_t [(\delta(t) + \frac{j}{\pi t}) * u_k(t)] e^{-jw_k t}\|_2^2 \right\} \\ \text{s.t.} & \sum_k u_k = f \end{aligned} \quad (1)$$

where, $*$ denotes convolution, u_k represents the k -th mode, w_k represents the center frequency, K represents the total number of modes, ∂_t represents the Dirac distribution, $f(t)$ is the original signal.

In order to eliminate the constraints of the above problems, quadratic penalty term and Lagrangian multiplier are introduced to transform the constrained optimization into the unconstrained optimization problem:

$$\begin{aligned} L(\{u_k\}, \{w_k\}, \lambda) = & \alpha \left\{ \sum_{k=1}^K \|\partial_t [(\delta(t) + \frac{j}{\pi t}) \otimes u_k(t)] e^{-jw_k t}\|_2^2 \right\} + \|f(t) - \sum_{k=1}^K u_k\|_2^2 \\ & + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle \end{aligned} \quad (2)$$

where, α is the penalty parameter. λ is the Lagrange multiplier and \otimes represents the convolution operator.

The solution of the original minimization problem (1) is transformed into an optional model, which can be solved by multipliers alternating direction method. Update u_k and w_k according to ADMM. The expression is as follows:

Minimization u_k and w_k :

$$\hat{u}_k^{N+1}(w) = \frac{\hat{f}(w) - \sum_{i \neq k} \hat{u}_i(w) + \frac{\hat{\lambda}(w)}{2}}{1 + 2\alpha(w - w_k)^2} \quad (3)$$

$$w_k^{N+1}(w) = \frac{\int_0^\infty w |\hat{u}_k(w)|^2 dw}{\int_0^\infty |\hat{u}_k(w)|^2 dw} \quad (4)$$

where, N represents the number of iterations. $\hat{u}_k^{n+1}, \hat{f}(w), \hat{u}_i(w), \hat{\lambda}(w)$ is the Fourier transform of $u_k^{n+1}, f(t), u_i(t), \lambda(t)$.

2.2. PCA-RBF Model

Principal component analysis is to reflect data characteristics and rules with as little information as possible. It not only retains the information of the original data, but also achieves data dimensional-reduction and reduces the complexity of the data itself [28,29]. The input variables of RBF include six variables, namely wind speed (V), wind direction (WD), temperature (T), humidity (H), air pressure (P) and battery energy (B) respectively in this paper. In order to achieve dimensional-reduction and reduce the redundant data, PCA is selected in this paper.

Firstly, data are collected to determine the original data input matrix X and output variables $Y = (y_1, y_2, \dots, y_n)'$ of RBF neural network.

$$X = (X_1, X_2, \dots, X_n) = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1h} \\ x_{21} & x_{21} & \cdots & x_{2h} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nh} \end{bmatrix} \quad (5)$$

where, n represents the number of samples, h represents the number of input data.

Calculate the mean value and standard deviation of each variable, which can be expressed as follows:

$$\bar{x}_j = \left(\sum_{i=1}^n x_{ij} \right) / n, (j = 1, 2, \dots, h) \quad (6)$$

$$s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}, (j = 1, 2, \dots, h) \quad (7)$$

The mean value and standard deviation are used to standardize the original data to eliminate the dimensional influence, which can be expressed as:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j}, (i = 1, 2, \dots, n; j = 1, 2, \dots, h) \quad (8)$$

The co-variance matrix R is obtained from the data after standardized processing, which can be expressed as:

$$R = \left[\frac{1}{n-1} \sum_{i=1}^n x_{ks}^* x_{kt}^* \right], (s, t = 1, 2, \dots, p) \quad (9)$$

Secondly, calculate the cumulative variance contribution rate.

$$a(m) = \sum_{j=1}^m \lambda_j / \sum_{j=1}^h \lambda_j \quad (10)$$

where, λ_j represents the j -th eigenvalue. $a(m)$ represents the cumulative contribution rate of the first m principal components, and h represents the number of original variables.

Extract principal components information, which can be expressed as:

$$Y = X \cdot l \quad (11)$$

where, Y is the principal component part of the original variable matrix X . l represents the eigenvectors corresponding to the eigenvalues. The first m principal components whose eigenvalue is greater than 1 and whose cumulative contribution rate is greater than 90% are selected, while the rest components are discarded.

Finally, the PCA-RBF model is established, and the principal components matrix Y is used as the input matrix to build a new Radial Basis Function (RBF) model. Because RBF Network has strong nonlinear fitting ability and can approximate arbitrary nonlinear functions. It has good generalization ability and can map any complex nonlinear relationship. The structure of RBF neural network is shown in Figure 1. RBF neural network is not repeated in this paper, detailed references [22].

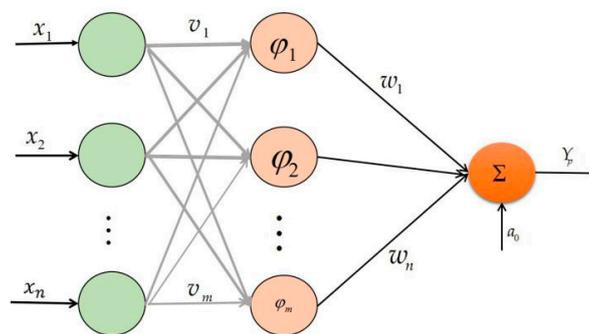


Figure 1. The structure of RBF neural network.

2.3. LSTM Model

Long-short term memory (LSTM) network is a special recurrent structure which can improve the ability of RNN network to handle long-term dependent tasks and resists the problem of vanishing gradient disappearance [17]. Compared with the traditional neural network, a feedback connection is added in the hidden layer of RNN. The interconnection of nodes between hidden layers in RNN makes the output of the hidden layer enter into the hidden layer of output and the next time step simultaneously, as shown in Figure 2. The network can generate the memory state of the past data, and establish the dependency relationship between the data of different time periods through this structure, so it can better deal with the time series problems [30]. LSTM solves the problem that the back propagation across time steps may lead to vanishing gradient problem of RNN when handle long-term dependent tasks by introducing a memory unit instead of the implicit node of the traditional RNN. The memory units of LSTM in hidden layer are shown in Figure 3. The repetition module of LSTM consists of three sigmoid functions and one tanh function, and introduces threshold structure, inputting gate, forgetting gate and outputting gate, so that the LSTM can selectively remember new information or delete information.

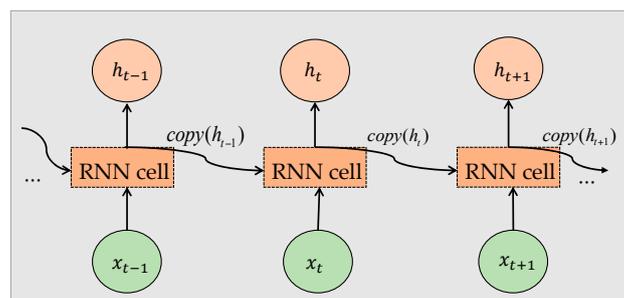


Figure 2. The expanded structure of recurrent neural network (RNN).

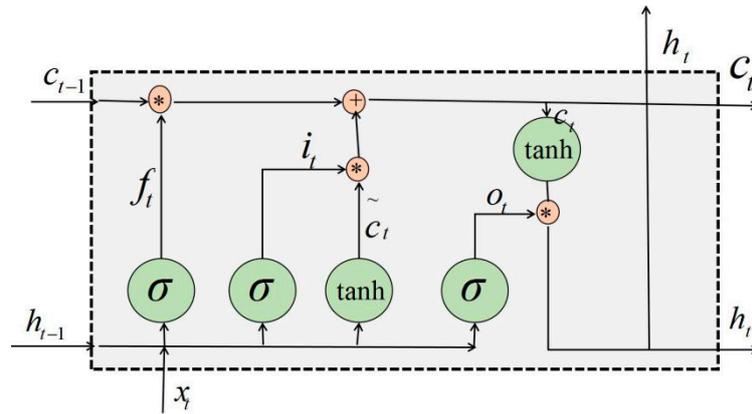


Figure 3. The hidden layer of long short-term-memory (LSTM) cell structure.

The process of information processing about LSTM model can be expressed by the following equation respectively:

$$i_t = \sigma(W_i x_t + H_i h_{t-1} + b_i) \quad (12)$$

$$f_t = \sigma(W_f x_t + H_f h_{t-1} + b_f) \quad (13)$$

$$o_t = \sigma(W_o x_t + H_o h_{t-1} + b_o) \quad (14)$$

$$\tilde{c}_t = \tanh(W_c x_t + H_c h_{t-1} + b_c) \quad (15)$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (16)$$

$$h_t = o_t * \tanh(c_t) \quad (17)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (18)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (19)$$

where, i_t is the input gate, which determines the updated information within the unit. f_t is the forget gate, that determines the deleted information in the unit. o_t is the output gate, that determines how much information is output. \tilde{c}_t is the candidate value of the element state at time t . c_t is the state of the unit at time t , calculated by the combination of parameters $i_t, \tilde{c}_t, f_t, c_{t-1}$ through Element-wise multiplication (*). h_t is the output value filtered by the output gate. σ represents the Sigmoid function in the range of 0 to 1, and uses the tanh function to trust the value between -1 and 1 . x_t represents the input of memory unit at time t . $W_i, W_f, W_o, W_c, H_i, H_f, H_o, H_c$ are the weight matrix. b_i, b_f, b_o, b_c is the deviation vectors.

The flow chart of LSTM model for wind speed prediction is shown in Figure 4.

2.4. Mixture Gaussian Process Regression (MGPR)

Gaussian process regression model (GPR) is a non-parametric probability model based on nuclear [31], and MGPR is based on the mixed Gaussian process expert support GP approaching the temptation of formalism, as shown in Figure 5. Each expert uses a set of guidance to supplement, and according to the proximity of data points to the experts, and the data point are defined and assigned to the experts probability.

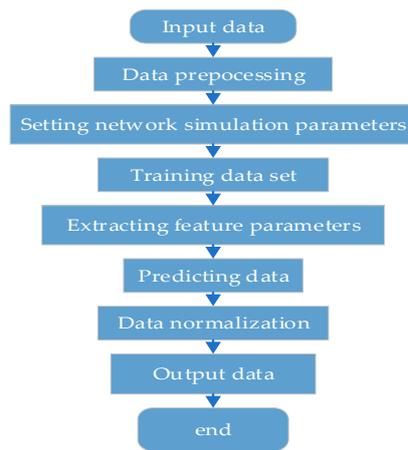


Figure 4. Wind speed prediction flow chart based on LSTM model.

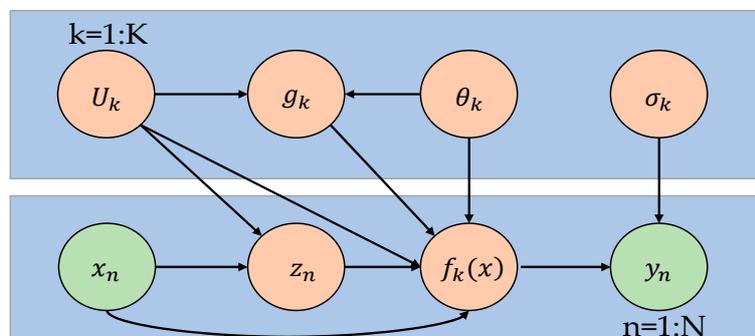


Figure 5. Mixture Gaussian process regression (MGPR) model representation.

Set the input eigenvector be $X = \{x_n\}_{n=1}^N$ and the actual output signal be $Y = \{y_n\}_{n=1}^N$.

Defined function space $f_k(x)$ conforms to Gaussian distribution, and each independent expert model contains a potential function, which is expressed as:

$$f_k(x) \sim GP(0, k(x, x'; \theta_k)) \tag{20}$$

where, $GP(\cdot)$ is gaussian distribution function, $k(x, x'; \theta_k)$ is the kernel function of expert K and θ_k is hyper parameter.

Each expert is supplemented by a set of M induction inputs $U_k = \{u_{k1}, \dots, u_{kM}\}$, whose inducement points are expressed as:

$$g_k = (f_k(u_{k1}), \dots, (f_k(u_{kM}))^T \tag{21}$$

Suppose that each observation x_n, y_n are associated with only one expert identified by z_n , denoted by $z, g, f, U, \theta, \sigma$ for $z_k, g_k, f_k, U_k, \theta_k, \sigma_k$. The total joint distribution of the model is expressed as:

$$p(y, z, f, g|U, \theta, \sigma, X) = p(y|f, \sigma) \cdot p(f|z, g, U, \theta, X) \cdot p(g|U, \theta) \cdot p(z|U, X) \tag{22}$$

where,

$$p(y|f, \sigma) = \prod_{k=1}^K N(y_k; f_k, \sigma_k^2 I) \tag{23}$$

$$p(g|U, \theta) = \prod_{k=1}^K N(g_k; 0, K(U_k, U_k)) \tag{24}$$

$$p(z|U, X) = \prod_{k=1}^K p(z_n | x_n, U) \quad (25)$$

$$p(f \left| z, g, U, \theta, X \right.) = \prod_{k=1}^K N(f_k; \bar{f}_k(X_k), \Lambda_k(X_k)) \quad (26)$$

where,

$$\bar{f}_k(X_k) = K(X_k, U_k)K(X_k, U_k)^{-1}g_k \quad (27)$$

$$\Lambda_k(X_k) = \text{diag}(K(X_k, X_k) - K(X_k, U_k)K(X_k, U_k)^{-1}K(X_k, U_k)) \quad (28)$$

$$p(z_n = k | x_n, U) = \frac{N(X_n; m_k, V)}{\sum_{j=1}^K N(X_n; m_j, V)} \quad (29)$$

$$m_k = \frac{1}{M} \sum_{m=1}^M u_{km} \quad (30)$$

$$v_j = \frac{1}{K(M-1)} \sum_{k=1}^K \sum_{m=1}^M (u_{kmj} - m_{kj})^2 \quad (31)$$

where, $N(\cdot, \cdot)$ represents Gaussian distribution. $K(X_k, U_k)$ is the co-variance matrix evaluated at all points in X_k and U_k . And the results of interval WSP in the sense of probability distribution can be obtained. In addition, the subscript k of the co-variance matrix is specified by the super-parameter of expert K . Each mean m_k is called the center of expert K corresponding to the output of point prediction. $V = \text{diag}(v_1, \dots, v_d)$ represents the co-variance.

3. Wind Speed Prediction Process

3.1. Data Analysis

3.1.1. Data Preprocessing

The research data of this paper are from the offshore small wind turbine test station of Shantou University located on Nan'ao Island, Guangdong Province. Nan'ao Island is the largest island wind farm traversed by the Tropic of Cancer with obvious marine climate characteristics in Asia and abundant in wind resources with average wind speed of 8.54 m/s. We selected samples collected for three consecutive days in September 2017 (1 September 2017–3 September 2017), and the number of sample is 350, and the sampling time interval is 10 min. The first 300 samples of all samples is taken as the training set, and the remaining 50 samples is taken as the test set. The method of this paper is applicable to short term WSP. The sampled variables include six variables, namely wind speed (V), wind direction (WD), temperature (T), humidity (H), air pressure (P) and battery energy (B) respectively. Figure 6. shows the wind speed series distribution. As we can see from Figure 6, the strong stochastic characteristic of wind speed is observed, with daily average wind speeds values changing from 1 to 15 m/s, where Y axis represents the wind speed and X axis represents the sample with an interval of 10 min.

There are correlations between sampled variables, but multidimensional data may lead to information redundancy and increase the computational complexity. The main objective is to find how the additional sampled variables can help improving wind speed prediction. Therefore, PCA is used to eliminate redundant information between input parameters, and replace the original input parameters with new uncorrelated input parameters, so as to reduce the dimension of the original data on the premise of retaining the original information. PCA algorithm is used to obtain the principal

component scores of each variable, and the standard orthogonal principal component score coefficient of each variable is visualized, as shown in the Figure 7.

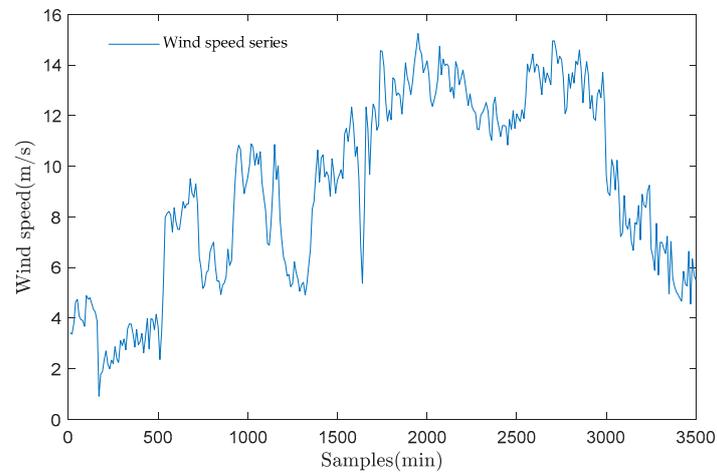


Figure 6. The original wind speed series.

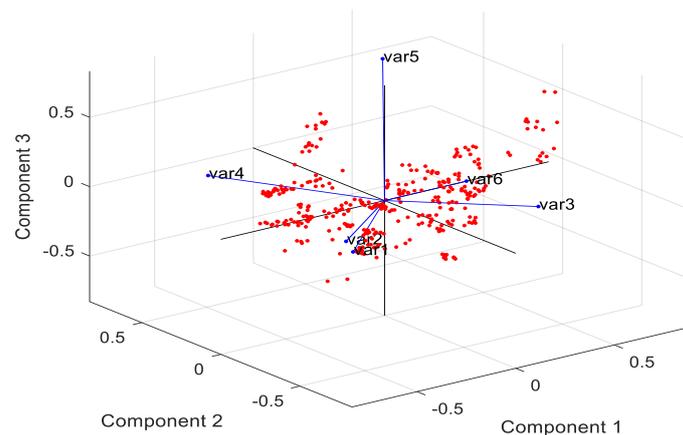


Figure 7. The visualization of principal component scores of each variable.

The eigenvalue of each principal component and the cumulative contribution rate of the sample data are shown in Table 1. As we can see from Table 1, the eigenvalues of the first two principal components of the sample data are both greater than 1, but the cumulative contribution rate is only 68%. The eigenvalues of the third and fourth principal components are close to 1, and the cumulative contribution rate of the first four principal components reaches 93.3%, therefore the first four principal components are selected as the model input data.

Table 1. Principle component analysis.

Principal Component	Principal Component Eigenvalues	Eigenvalue Contribution Rate	Cumulative Contribution Rate
First	3.053	0.508	0.508
Second	1.031	0.172	0.680
Third	0.929	0.155	0.835
Fourth	0.587	0.098	0.933
Fifth	0.352	0.059	0.992
Sixth	0.049	0.008	1

3.1.2. Decomposition of Wind Speed Data

Variational mode decomposition (VMD) is used to extract the wind speed series into different intrinsic mode functions. As shown in Figure 8, it is the original wind speed series results decomposed by Empirical mode decomposition (EMD) and VMD respectively, which include decomposition modes and corresponding spectrum. If the number of modes K is too small, it will lead to undersegmentation of the sample data, and some components are contained in other modes or discarded as “noise”. On the contrary, if K is too big, it will lead to capture extra noise or mode to mode duplication. As shown in Figure 8b, EMD can not preset K , which is not convenient to observe the fluctuation of wind speed. Optimization of VMD is on the basis of the EMD algorithm, which can preset K , as well as can observe the fluctuation and randomness of wind speed. K is 3 in this paper. Therefore, the mode is extracted by denoising VMD algorithm in this paper.

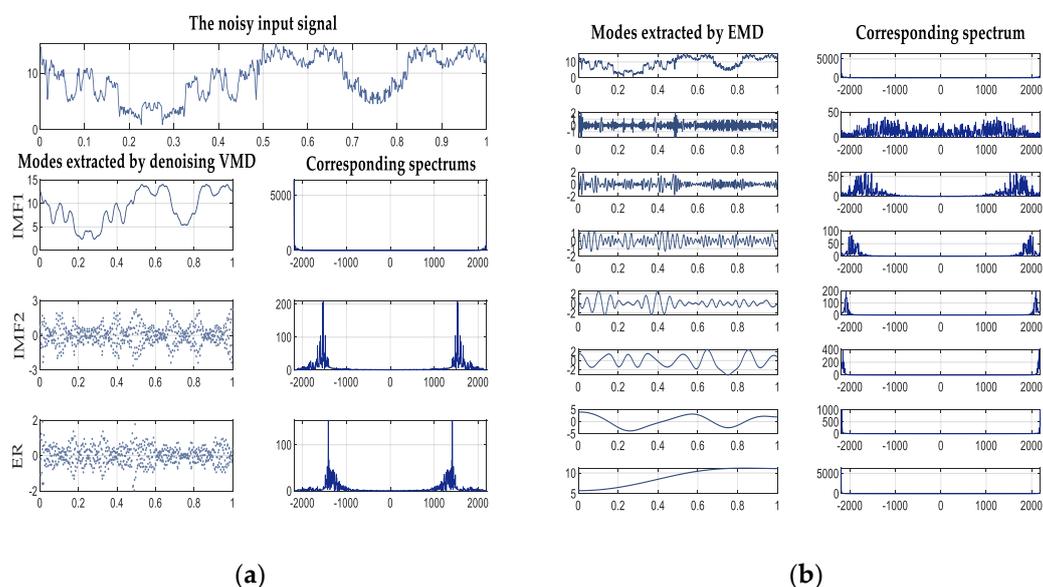


Figure 8. The decomposed results of wind speed. (a) The decomposed results of wind speed by VMD; (b) the decomposed results of wind speed by EMD.

Figure 9 shows the 3D visualization of VMD decomposition modes. As we can see from Figure 9, IMF1 has strong volatility and obvious nonlinear characteristics. IMF2 has a weak fluctuation range from -3 to 3 , and is randomly distributed around 0 without obvious nonlinear characteristic and the fluctuation range is small. IMF3 is randomly distributed around $[-1.5, 1.5]$ and scatter points fluctuate around 0 randomly, and they do not have distinct trend and fluctuation. Therefore, IMF3 can be considered as a noise series. The ADF in Matlab is used to test whether the time series is stable. The return value of IMF1 is 0 but IMF2 and ER series return a value of 1 . Thus, according to the ADF test and the characteristic analysis of the decomposed series, as can be seen that, IMF1 is non-stationary series, IMF2 and ER are all stationary series. VMD algorithm decomposes the original wind speed series into nonlinear part, stationary time series part and noise part, and corresponding reasonable models are built for each part, which can reduce the impact of randomness and fluctuation.

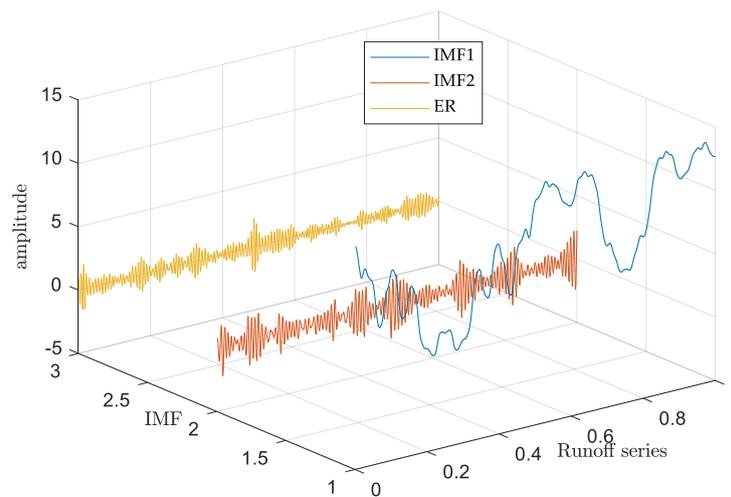


Figure 9. The three modes decomposed by VMD 3D visualization.

3.2. The Process of Modeling

The WSP modeling process mainly consists of three parts: data pre-processing, establishment of hybrid models and comparison models including deterministic prediction and interval prediction, and validation of the proposed model. Finally, a numeral case is given to introduce these three parts in details respectively. The specific modeling process is shown in Figure 10, where the green part shows the entire process of the hybrid model proposed in this paper and the dotted lines are connected to the comparison models flow.

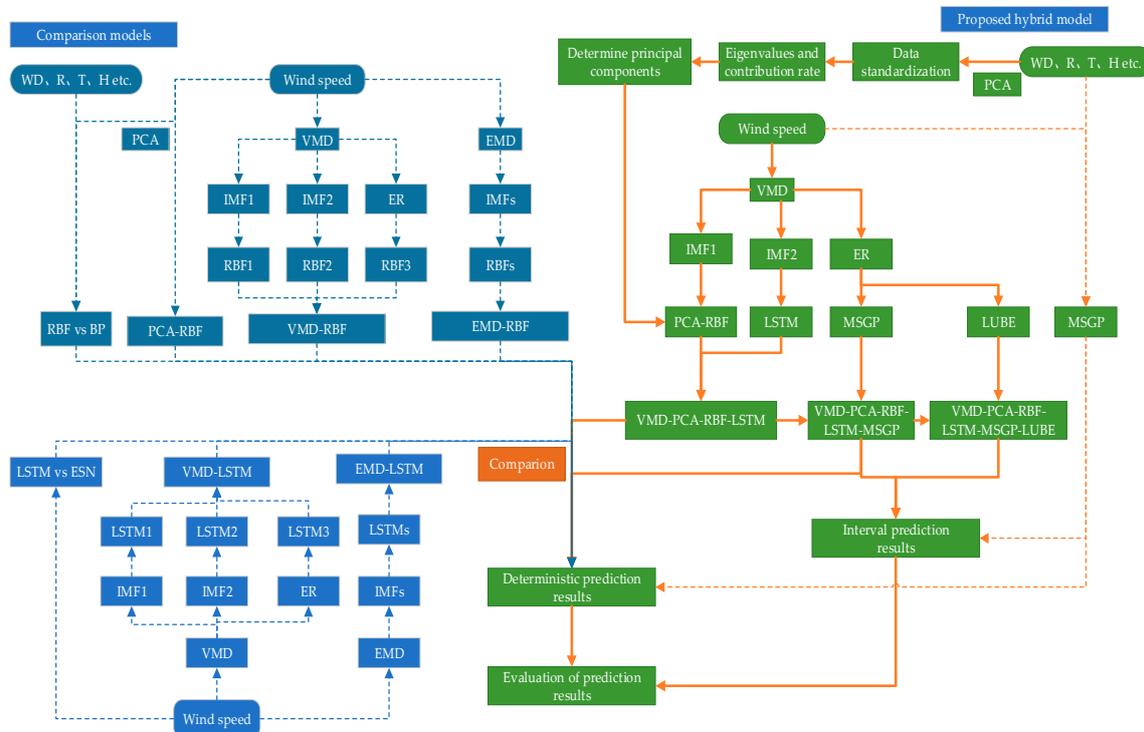


Figure 10. The flow chart of the proposed wind speed prediction (WSP) model and comparison models.

1. Data processing section.

First, PCA is used to extract the principal components of multidimensional variables (like wind speed (V), wind direction (WD), temperature (T), humidity (H) and air pressure (P)), which the cumulative

contribution rate is greater than 90%. The partial auto-correlation function is used in time series to determine the state variables and construct the appropriate training set. Figure 11 shows the auto-correlation function and partial correlation function of the original time series. As we can see from Figure 11, the auto-correlation function has the trailing feature, while the partial auto-correlation function truncates the tail, so that the wind speed series satisfies the autoregression (AR) model.

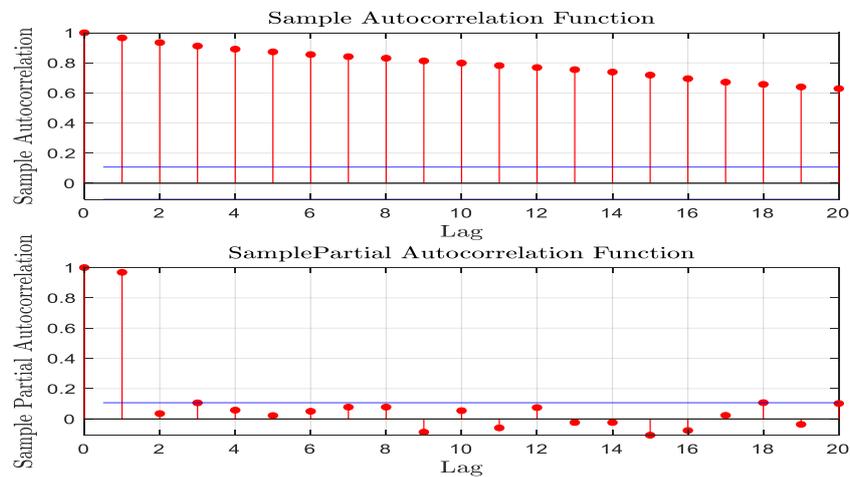


Figure 11. The auto-correlation function and partial correlation function of wind speed series.

Secondly, VMD decomposition is used to extract the original wind speed series into IMFs and error series. A nonlinear model is established for IMF1 with strong volatility and obvious nonlinear characteristics. IMF2 is a stationary time series, and a linear time series prediction model is established for IMF2. MGPR is used to predict ER series.

2. Modeling part.

First of all, according to the previous analysis, the RBF nonlinear model is established for IMF1 obtained by VMD with obvious nonlinear characteristic, where the input of RBF model is the principal component matrix obtained by PCA. Combined with PCA, this model is referred to as VMD-PCA-RBF, and the prediction result is v_{pi1} .

As for IMF2, a stationary time series model is established by LSTM, which is named as VMD-LSTM and the prediction result is v_{pi2} . So, we can obtain the deterministic prediction result of hybrid model VMD-PRBF-LSTM v_{pi} , which is expressed as follows:

$$v_{pi} = v_{pi1} + v_{pi2} \tag{32}$$

The ER series satisfies Gaussian distribution according to the JARque-Bera test in Matlab, and the error distribution is shown in Figure 12. The MGPR model is built to predict ER series, and the deterministic prediction results \hat{E}_i and interval prediction results \hat{v}_{pi} of ER are obtained simultaneously, where the upper limits and lower limits of interval prediction of ER are denoted as v_{upi} and v_{lowi} respectively. The final deterministic prediction result of hybrid model v_{pi} is obtained by sum reconstruction. The hybrid model combined above methods is named as the VMD-PCA-RBF-LSTM-MGPR.

$$v_{pi} = v_{pi1} + v_{pi2} + \hat{E}_i \tag{33}$$

$$\hat{v}_{pi} = [v_{pi1} + v_{pi2} + v_{upi}, v_{pi1} + v_{pi2} + v_{lowi}] \tag{34}$$

where, \hat{v}_{pi} is the CIs under the confidence degree $100(1 - \alpha)\%$.

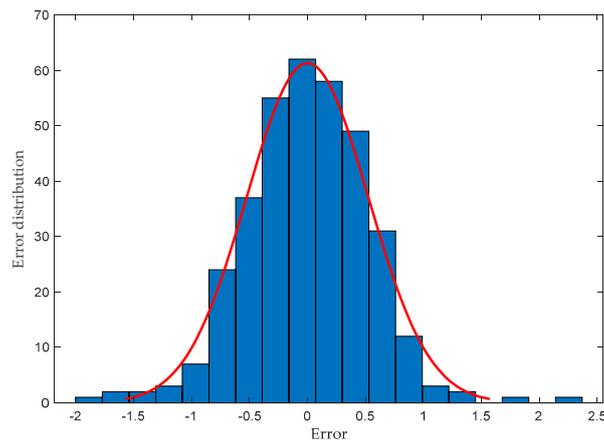


Figure 12. Error distribution of ER series.

Based on the deterministic prediction of VMD-PCA-RBF-LSTM-MGPR model, an interval prediction model is made by using the lower upper estimation method [16]. The wind speed PIs is obtained under the confidence degree $100(1 - \alpha)\%$, where α stands for significance level. Which can be expressed as:

$$\hat{v}_{pi} = [L_t^\alpha(v_{pi}), U_t^\alpha(v_{pi})] \quad (35)$$

where $L_t^\alpha(v_{pi})$ means the lower limit of PIs and CIs, $U_t^\alpha(v_{pi})$ means the upper limit.

Secondly, establish comparison models. In this paper, a total of nine comparison models are established. RBF and BP neural network models are built respectively with taking wind speed (V) and other variables as input data. Only use PCA in data processing to obtain the input matrix of RBF, so as to establish PCA-RBF model. As for each mode obtained by VMD, establish RBF model and then build VMD-RBF model by reconstituting three RBF. Similarly, LSTM, ESN and VMD-LSTM comparison models are established respectively, where all three models have only one input, wind speed. Meanwhile we build two other comparison models, EMD-RBF and EMD-LSTM. In addition, a single model MGPR is used for deterministic WSP and interval WSP, so the MGPR comparison model is obtained.

3. The validity of the model is illustrated in detail in Section 4.

3.3. Evaluation of Prediction Results

For deterministic prediction, this paper takes mean absolute error (MAE), mean square error (MSE), relative mean square error (RMSE) and mean absolute percent error (MAPE) as evaluation indexes of the prediction model. The larger the evaluation indexes value are, the larger the prediction error is. They are expressed as follows respectively.

$$MAE = \frac{1}{n} \sum_{i=1}^n |v_{pi} - v_{ri}| \quad (36)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (v_{pi} - v_{ri})^2 \quad (37)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (v_{pi} - v_{ri})^2} \quad (38)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{v_{pi} - v_{ri}}{v_{ri}} \right| \quad (39)$$

where, n represents the number of samples. v_{ri} represents the actual wind speed. v_{pi} represents the prediction wind speed.

The PIs and CIs indicators include mean absolute percentage error I_{MAPE} , prediction interval coverage percentage I_{PICP} , mean prediction interval width I_{MPIW} . The smaller the I_{MAPE} and I_{MPIW} values, the higher the forecasting accuracy. The larger I_{PICP} value, the more actual value falling into the interval, that represents the reliability of interval prediction is higher.

Mean absolute percentage error I_{MAPE} is used to evaluate the error between true and predicted value. Its expression is as follows.

$$I_{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{v_{ri} - v_{pi}}{v_{ri}} \right| \times 100\% \quad (40)$$

Mean prediction interval width I_{MPIW} is used to test the suitability of the width of CIs and PIs, which used to prevent excessive interval width. Its expression is as follows.

$$I_{MPIW} = \frac{1}{n} \sum_{i=1}^n (U(v_{pi}) - L(v_{pi})) \quad (41)$$

where, $U(v_{pi})$ represents the upper limit of the prediction interval. $L(v_{pi})$ denotes the lower limit of the prediction interval.

Prediction interval coverage percentage I_{PICP} is used to express the probability of the actual value in the prediction interval. Its expression is as follows:

$$I_{PICP} = \frac{1}{n} \sum_{i=1}^n c_i \quad (42)$$

where, If $v_{pi} \in [L(v_{pi}), U(v_{pi})]$, $c_i = 1$, or $c_i = 0$.

4. Analysis of Prediction Results

4.1. Analysis of Deterministic Prediction Results

In order to verify the availability of the model, a total of nine comparison models are proposed in this paper. Figures 13–16 respectively show the deterministic prediction results of different WSP models, where Y axis represents the wind speed and X axis represents prediction samples with an interval of 10 min. The error evaluation indexes are listed in Table 2. Figure 13 shows the deterministic prediction results of Back Propagation model (BP), Radial Basis Function model (RBF), Principal component analysis-Radial Basis Function model (PCA-RBF) and Variational mode decomposition-Radial Basis Function model (VMD-RBF) respectively. According to Table 2 and Figure 13, the prediction performance of RBF model is better compared with the single BP model, where all of the evaluation index of RBF is lower than BP. The MAE of VMD-RBF is 2.6233, which is 10.7243 lower than BP model. The prediction results of VMD-RBF and PCA-RBF have some improvement effect but compared with the single RBF model. It proves PCA and VMD can help improve the accuracy of WSP. The prediction results of RBF, VMD-RBF and PCA-RBF can only reflect the volatility trend of some actual values, however, there is a big difference between the prediction values and the actual values. In addition, VMD-RBF and PCA-RBF both have stronger volatility. The VMD-RBF model improves the predicted value of RBF model to some extent, but it does not improve the volatility tendency of the predicted values significantly. MAE, MSE and RMSE of VMD-RBF model are the lowest among all the models in Figure 13, and MAPE of VMD-RBF model is only 0.0263 different from that of PCA-RBF model but the prediction effect is not evident.

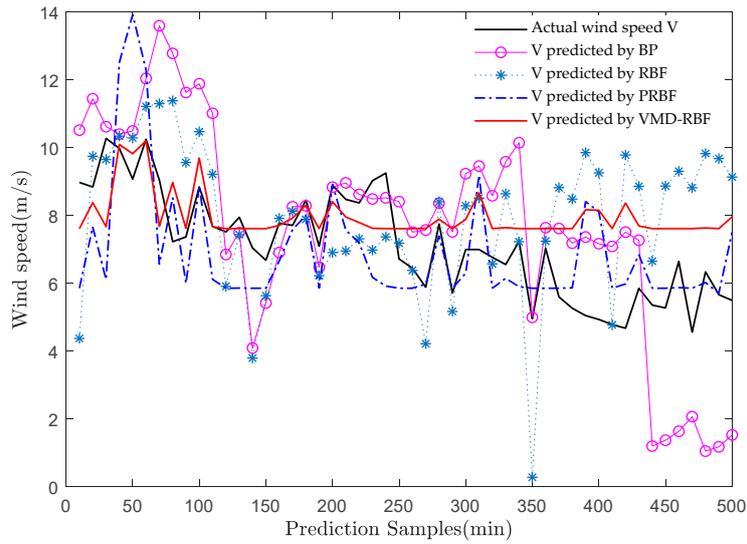


Figure 13. The results of comparison models (BP and RBF and PRBF and VRBF).

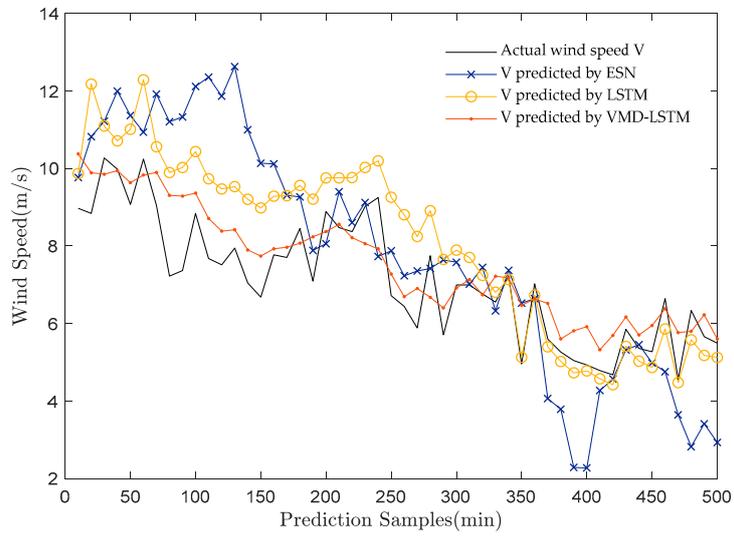


Figure 14. The results of comparison models (ESN and LSTM and VLSTM).

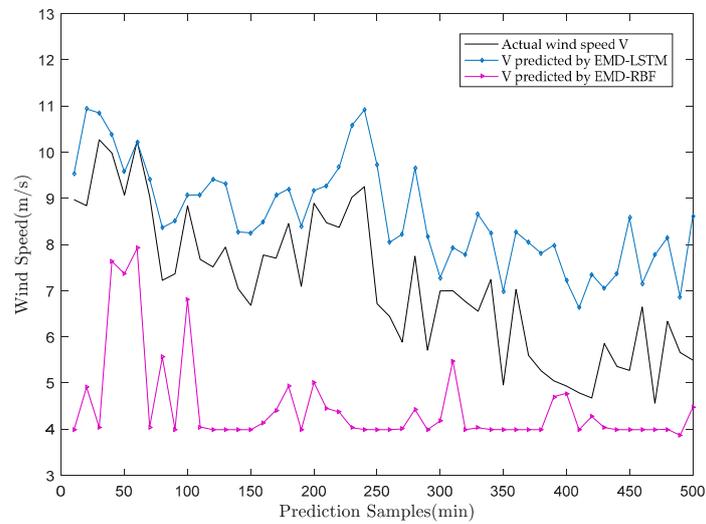


Figure 15. The results of comparison models (EMD-LSTM and EMD-RBF).

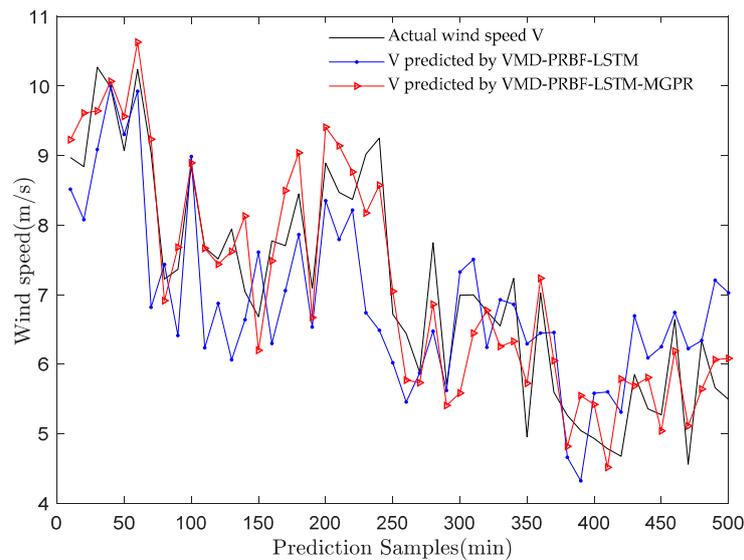


Figure 16. The results of hybrid models (VMD-PRBF-LSTM-VMD-PRBF-LSTM-MGPR).

Table 2. Analysis table of point prediction evaluation index.

	MAE	MSE	RMSE	MAPE
BP	2.9010	13.3476	3.6534	0.4426
RBF	3.1535	6.0185	2.4533	0.3128
PCA-RBF	1.3507	3.0501	1.7465	0.1886
VMD-RBF	1.2884	2.6233	1.6197	0.2149
EMD-RBF	2.6530	8.9093	2.9848	0.3539
ESN	83.5093	4.5839	2.141	0.2413
LSTM	58.4243	2.0863	1.4444	0.1586
VMD-LSTM	33.594	0.6765	0.8225	0.1008
EMD-LSTM	75.3289	3.0046	1.7334	0.2431
MGPR	2.2739	7.0604	2.6571	0.3714
VMD-PCA-RBF-LSTM	0.8049	1.0229	1.0114	0.1183
VMD-PCA-RBF-LSTM-MSGP	0.4796	0.3158	0.5619	0.0713

Figures 14 and 17 describe the Echo State Network model (ESN), Long-Short Term Memory model (LSTM) and Variational mode decomposition-Long-Short Term Memory model (VMD-LSTM) and MGPR four comparison models for deterministic prediction. Compared with the single ESN model, LSTM can better describe the fluctuation of wind speed series. Compared with the single LSTM model, from Table 2 combined with Figures 14 and 17, VMD-LSTM improve the prediction value with lower RMSE 0.6765, which reduces 3.9074 than ESN model. MAE of LSTM is 33.594, which shows the prediction value is far from actual wind speed even though it is lower than ESN model with 49.9153. On the contrary, MAE of MGPR is just 2.2739, but the MSE 7.0604 is second only to BP neural network model with 13.3476, and there is a large gap between the prediction value and the actual value as we can see from Figure 16.

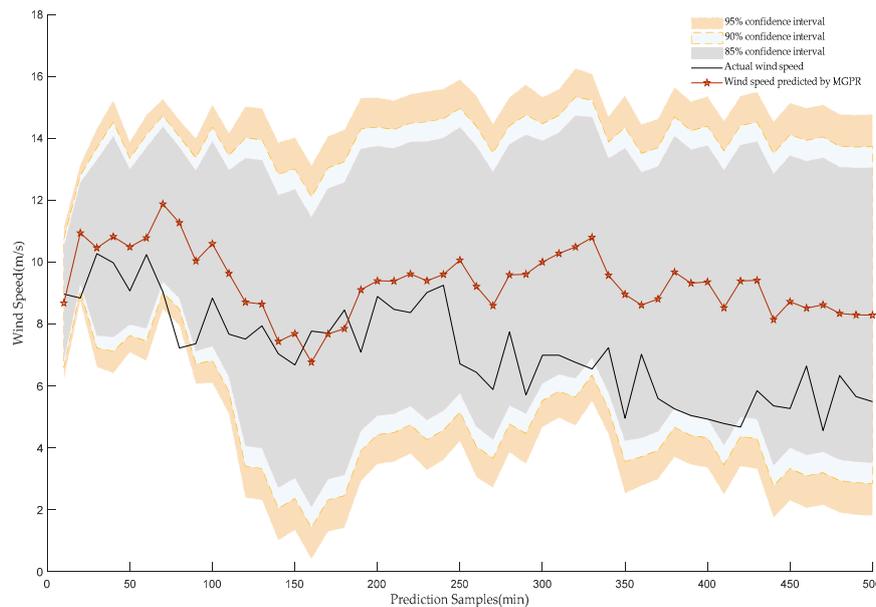


Figure 17. Interval and point prediction results of mixture Gaussian process regression (MGPR).

Figure 15 shows the deterministic prediction results of Empirical mode decomposition-Radial Basis Function model (EMD-RBF) and Empirical mode decomposition-Long-Short Term Memory model (EMD-LSTM). As we can see from Figure 15 and Table 2, the prediction result of EMD-RBF is far from the actual wind speed and MSE of EMD-RBF is 8.9093, which is the second among all models. Even though the prediction results of EMD-LSTM is closer to the actual wind speed, but MAE of EMD-LSTM is 75.3289, and there is a difference between prediction value and actual value. The prediction accuracy of both models is not ideal. In addition, the RMSE of VMD-LSTM is only 0.8225 which is lower than all model mentioned in Figures 14 and 15. The modes obtained by EMD is 7 which cannot be preset, it is complex to build prediction model for each modes and too modes will lead to capture extra noise or mode to mode duplication, so as to cause poor prediction. Thus, VMD is more suitable to build prediction model in this paper.

The deterministic prediction results of VMD-PCA-RBF-LSTM and the hybrid model VMD-PCA-RBF-LSTM-MGPR proposed in this paper are shown in Figure 16. As we can see from Figure 16, the prediction results of both models are observably closer to the actual values than other comparison models. From Table 2 combined with Figure 16, MSE, RMSE and MAE of VMD-PRBF-LSTM-MGPR is 0.3158, 0.5619 and 0.4796 respectively, and both of them are the lowest among all the models, of which MAE of hybrid model is lower than ESN with 83.0297. Therefore, the prediction accuracy of VMD-PRBF-LSTM-MGPR is greater than other comparison models. The prediction results show that the proposed hybrid model has higher accuracy, and the distribution characteristics of prediction results are closer to the actual series through combining the advantages of nonlinear model and stationary time series model. Thus, the proposed model is more suitable for WSP than other nine comparison models.

4.2. Analysis of Interval Prediction Results

According to the previous analysis, CIs of wind speed are obtained by MGPR and VMD-PRBF-LSTM-MGPR-MGPR showed in Figures 17 and 18 respectively, and PIs is obtained by VMD-PRBF-LSTM-MGPR-LUBE showed in Figure 19, where Y axis represents the wind speed and X axis represents prediction samples with an interval of 10 min. As it can be seen from those figures, wind speed CIs and PIs of 95% are greater than that of 90% which are greater than that of 85%, which accords with the general statistical law, so the prediction results are correct and has statistical significance. CIs are obtained based on deterministic prediction results, thus the results of MGPR cannot

reflect the fluctuation and randomness clearly, while CIs of the proposed method can. The interval width of the hybrid model is obviously narrower than that of single model MGPR from above pictures. Figure 17 shows the deterministic prediction and interval prediction results of wind speed obtained by the single model MGPR, where the interval prediction contains the results with confidence of 95%, 90% and 85% respectively. As can be seen from Table 3 and Figure 18, when the confidence is 85%, the I_{PICP} of MGPR is 86%, and the I_{MPIW} is 8.1148. Compared with the 90% confidence level, the I_{PICP} increased by 10%, but the I_{MPIW} also increased by 1.1575. When the confidence is 95%, the interval coverage rate is 98%, but the I_{MPIW} is 11.0486, therefore it can be seen that the prediction results of the single model MGPR are poor.

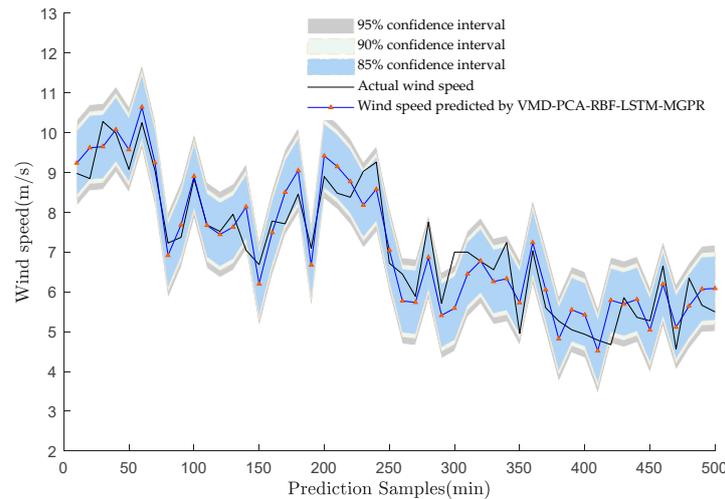


Figure 18. Interval prediction results of VMD-PRBF-LSTM-MGPR.

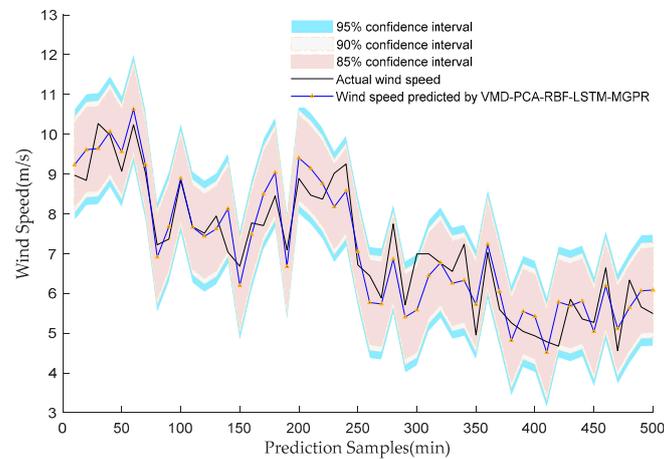


Figure 19. Interval prediction results of VMD-PRBF-LSTM-MGPR-LUBE.

Table 3. Analysis table of interval forecast evaluation index.

	$100(1-\alpha)\%$	I_{MAPE}	I_{PICP}	I_{MPIW}
MGPR	85%	0.3714	0.86	8.1148
	90%		0.96	9.2723
	95%		0.98	11.0486
VMD-PCA-RBF-LSTM-MSGP	85%	0.0713	0.86	1.5865
	90%		0.94	1.8127
	95%		0.94	2.16
VMD-PCA-RBF-LSTM-MSGP-LUBE	85%	0.0713	0.94	2.1240
	90%		0.98	2.3996
	95%		0.98	2.8079

Figure 18 shows the interval prediction results obtained by the hybrid model VMD-PRBF-LSTM-MGPR proposed in this paper, which includes the deterministic prediction results and the confidence intervals under different confidence levels. Obviously, CIs of this method can significantly reflect the stochastic volatility of wind speed and the interval width is narrower than CIs of MGPR. Combined with Table 3, when the confidence level is 85%, the mean width of the confidence interval I_{MPIW} is only 1.5865, but the interval coverage rate is only 86%, which is the lowest I_{PICP} among the three confidence levels. Compared with 90% confidence, the interval coverage rate is 94% with an increase of 8%, and the I_{MPIW} is 1.8127. However, when the confidence is 95%, the interval coverage rate is not improved and the I_{MPIW} increases by 0.3473.

Figure 19 shows the PIs obtained on the basis of the deterministic prediction results of VMD-PRBF-LSTM-MGPR combining LUBE. Combined with Table 3, when the confidence level is 85%, the I_{PICP} reaches 94% and the I_{MPIW} is 2.1240. When the confidence is 90%, the I_{PICP} only increases by 4% and the I_{MPIW} increases to 2.3996. When the confidence is 95%, the interval coverage rate remains the same. The I_{MPIW} increases to 2.8079. The interval prediction results show that PIs of 95% and 90% contain most of the actual wind speed values. In addition, CIs and PIs can quantify and represent forecast uncertainty. CIs predict uncertainty of unknown but fixed value, PIs express the uncertainty in the prediction of a future realization of a random variable, and prediction width is greater than corresponding confidence interval.

5. Conclusions

A new WSP model is proposed considering volatility and randomness of wind speed in this paper, which can be applied to short-term deterministic and interval prediction with high accuracy and strong stability. Based on the deterministic WSP, the randomness, volatility and low energy density characteristics of wind speed can be expressed clearly and accurately, which is beneficial to the grid scheduling plan optimization and can increase the reliability of wind power grid. Based on the interval WSP, the volatility range of wind speed can be significantly predicted, which is beneficial to quantify the intermittent risk of wind turbine. The prediction results show that this method is valid and feasible.

In this paper, the wind speed series is exacted into different IMFs by VMD decomposition, and the corresponding models are built for each part. PCA analysis is used to realize multidimensional reduction of the original input parameters, and the PCA-RBF model is established to reduce the computational complexity of the model. LSTM model is established for IMF2 with low fluctuation and a stationary time series model is carried out. The MGPR model is used to predict the error series, and the results of both deterministic prediction and interval prediction are obtained meanwhile. The prediction results obtained from the above models are reconstructed to obtain the final results of WSP and complete the construction of VMD-PRBF-LSTM-MGPR model.

The experimental results show that the proposed hybrid model can effectively reduce the prediction error caused by randomness and fluctuation of wind speed. MAPE, MSE and RMSE of deterministic prediction of the hybrid model proposed in this paper are only 0.0713, 0.3158 and 0.5619 respectively, which are more accurate than the prediction results of comparison models. The average absolute error is 0.4796, which is 83.0297 lower than the single model ESN, proving the validity of the model. In addition, the interval prediction results obtained by the model presented in this paper show that, compared with CTs obtained by the single model MGPR, when the confidence degree is 85%, the average interval width decreases by 80.45%, and the interval coverage is about 95% with the confidence of 90% and 95% respectively. However, the average interval width of the model proposed in this paper is narrower, indicating that the prediction accuracy is higher. In this paper, we make a comparison between CIs and PIs, and it shows that both of them can quantify and represent forecast uncertainty and the width of PIs is wider than CIs under the same confidence degree.

Author Contributions: Methodology, Q.C. and X.B.; software, Q.C.; writing—review and editing, X.B. and Q.C.; project administration, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Science and Technology Planning Project of Guangdong Province, grant number 2015B020240003, the National Natural Science Foundation of China, grant number 51976113, the National High-tech R&D Program of China (863 Program), grant number 2012AA051301.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Nomenclature

WSP	Wind speed prediction	NWP	Numerical weather prediction
VMD	Variational mode decomposition	MSE	Mean square error
PCA	Principal component analysis algorithm	MAE	Mean absolute error
RBF	Radial Basis Function	RMSE	Relative mean bias error
RNN	Recurrent neural network	MAPE	Mean absolute percent error
ARIMA	Autoregressive integrated moving average model	ER	Error series
SVM	Support vector machine	IMF	Intrinsic mode function
ESN	Echo State Network	CI	Confidence interval
EMD	Empirical mode decomposition	PI	Prediction interval
WT	Wavelet transform	ADF	Augmented Dickey-Fuller
EWT	Empirical wavelet transform	MPIW	Mean prediction interval width
GPR	Gaussian process regression	BP	Back propagation
PRBF	The method combined with PCA and RBF	LUBE	Lower and upper bound estimation method
PICP	Prediction interval coverage percentage	VMD-RBF	The method combined with VMD and RBF
LSTM	Long-Short Term Memory network	VMD-LSTM	The method combined with VMD and LSTM
MGPR	Mixture Gaussian Process Regression	VMD-PRBF-LSTM	Data is pre-processing by VMD and PCA, and RBF is used in IMF1, LSTM is used in IMF2. The model combined IMF1 and IMF2 is named as VMD-PRBF-LSTM.
VMD-PCA-RBF-LSTM-MGPR			MGPR is built to predicted ER, and the model of recombining IMF1, IMF2 and ER is named as VMD-PCA-RBF-LSTM-MGPR
VMD-PCA-RBF-LSTM-MGPR-LUBE			Based on the deterministic result of VMD-PCA-RBF-LSTM-MGPR, LUBE is used to predictive the PIs of wind speed. The model combined with VMD-PCA-RBF-LSTM-MGPR and LUBE is named as VMD-PCA-RBF-LSTM-MGPR-LUBE

References

1. QianYao Technology. Polaris Wind Power Grid. Available online: <http://news.bjx.com.cn/html/20200720/1090384.shtml> (accessed on 20 July 2020).
2. Calif, R.; Schmitt, F. Multiscaling and Joint Multiscaling of the Atmospheric Wind Speed and The Aggregate Power Output From A Wind Farm. *Nonlinear Process. Geophys.* **2014**, *21*, 379–392. [[CrossRef](#)]
3. Yang, X.Y.; Liang, G.S. Development of wind power generation and its market prospect. *Pow. Syst. Technol.* **2003**, *27*, 78–79.
4. Yang, J.; Zhao, X.; Wei, H.; Zhang, K. Sample Selection Based on Active Learning for Short-Term Wind Speed Prediction. *Energies* **2019**, *12*, 337. [[CrossRef](#)]
5. Karakasidis, T.E.; Charakopoulos, A. Detection of Low-dimensional Chaos in Wind Time Series. *Chaos Solitons Fractals* **2009**, *41*, 1723–1732. [[CrossRef](#)]
6. Drisya, G.V.; Asokan, K.; Kumar, K.S. Diverse dynamical characteristics across the frequency spectrum of wind speed fluctuations. *Renew. Energy* **2018**, *119*, 540–550. [[CrossRef](#)]
7. Luo, S.; Ding, Q.W.; Li, C.; Huang, Z.Q. Chaotic Characteristic Analysis and Nonlinear Short-Term Prediction of Wind Speed Time Series. *Energy Eng.* **2019**, 50–56.
8. Shivam, K.; Tzou, J.-C.; Wu, S.-C. Multi-Step Short-Term Wind Speed Prediction Using a Residual Dilated Causal Convolutional Network with Nonlinear Attention. *Energies* **2020**, *13*, 1772. [[CrossRef](#)]
9. Huang, Y.; Yang, L.; Liu, S.; Wang, G. Multi-Step Wind Speed Forecasting Based on Ensemble Empirical Mode Decomposition, Long Short Term Memory Network and Error Correction Strategy. *Energies* **2019**, *12*, 1822. [[CrossRef](#)]
10. Adnan, R.M.; Liang, Z.; Yuan, X.; Kisi, O.; Akhlaq, M.; Li, B. Comparison of LSSVR, M5RT, NF-GP, and NF-SC Models for Predictions of Hourly Wind Speed and Wind Power Based on Cross-Validation. *Energies* **2019**, *12*, 329. [[CrossRef](#)]
11. Sfetos, A. A Comparison of Various Forecasting Techniques Applied to Mean Hourly Wind Speed Time Series. *Changzhou Univers.* **2016**, *21*, 23–35. [[CrossRef](#)]
12. Huang, X.L.; Liu, Z.R.; Zhu, R.J.; Yang, Z.; Zhou, G.; Wang, C.; Ji, Z. Impact of Power System Integrated with Large Capacity of Variable Speed Constant Frequency Wind Turbines. *Transact. China Electrotech. Soc.* **2010**, *25*, 142–149.
13. Administration of Quality Supervision, Inspection and Quarantine of the People’s Republic of China. State Electricity Regulatory Commission of People’s Republic of China. In *GB/T 19963-2011 Technical Rule for Connecting Wind Farm to Power System*; General: Beijing, China, 2011.
14. Yang, M.; Chen, X.X.; Huang, B.Y. Ultra-short-term Multi-step Wind Power Prediction Based on Fractal Scaling Factor Transformatio. *Renew. Sustain. Energy* **2018**, *10*, 053310. [[CrossRef](#)]
15. Tang, B.; Chen, Y.; Chen, Q.; Su, M. Research on Short-Term Wind Power Forecasting by Data Mining on Historical Wind Resource. *Appl. Sci.* **2020**, *10*, 1295. [[CrossRef](#)]
16. Erick, L.; Carlos, V.; Héctor, A.; Esteban, G.; Henrik, M. Wind Power Forecasting Based on Echo State Networks and Long Short-Term Memory. *Energies* **2018**, *11*, 526.
17. Wang, J.; Li, X.; Zhou, X.D.; Liu, B.; Zhao, S. Ultra-short-term Wind Speed Prediction Based on VMD-LSTM. *Power Syst. Prot. Control.* **2020**, *48*, 45–52.
18. Li, Y.; Peng, C.H.; Fu, Y.; Sun, H.J. Short Term Wind Power Prediction Based on CNN-LSTM Network Model. *East China Jiaotong Univers.* **2020**, *37*, 109–115.
19. Li, X.; Liu, Y.; Xin, W. Wind Speed Prediction Based on Genetic Neural Network. In Proceedings of the 2009 4th IEEE Conference on Industrial Electronics and Applications, Xi’an, China, 25–27 May 2009; pp. 2448–2451. [[CrossRef](#)]
20. Li, J.Y.; Wang, D.F.; Zhang, Y. Short-term Wind Speed Prediction Method Based on Grey Relation Analysis. *Shandong Electr. Power.* **2020**, *47*, 15–19.
21. Lu, B.J.; Zhou, P.; Wang, X.; Zhou, K. Research on Short-term Wind Speed Forecasting Model Based on EEMD and LSTM. *Softw. Eng.* **2020**, *23*, 43–48.
22. Sinha, S.; Tiwari, S. An Improvement in Performance and Computational Cost of ANN Based Wind Speed Prediction System. In Proceedings of the 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 17–19 July 2019; pp. 542–546.

23. Li, D.Z.; Li, Y.Y.; Wang, C. Short-term Wind Speed Prediction Based on Convolutional Recurrent Neural Network and Deep Learning. *Electr. Power. Sci. Eng.* **2019**, *35*, 1–6.
24. 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. A. Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [[CrossRef](#)]
25. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [[CrossRef](#)]
26. Ye, G.Q.; Li, W.G.; Wang, H. Study of RBF Neural Network Based on PSO Algorithm in Nonlinear System Identification. In Proceedings of the 2015 8th International Conference on Intelligent Computation Technology and Automation (ICICTA), Nanchang, China, 14–15 June 2015; pp. 852–855. [[CrossRef](#)]
27. Khosravi, A.; Nahavandi, S.; Creighton, D.; Atiya, A.F. Lower Upper Bound Estimation Method for Construction of Neural Network-Based Prediction Intervals. *IEEE Trans. Neural Netw.* **2010**, *22*, 337–346. [[CrossRef](#)] [[PubMed](#)]
28. Zhou, S.L.; Mao, M.Q.; Su, J.W. Prediction of Wind Power Based on Principal Component Analysis and Artificial Neural Network. *Power Grid Technol.* **2011**, 128–132.
29. He, D.; Liu, R.Y. Ultra-Short-Term Wind Power Prediction Using Ann Ensemble Based on The Principal Components Analysis. *Power syst. Protect. Control* **2013**, *41*, 50–54.
30. Xiao, J.H. *Research on Hybrid Model of Short-term Wind Speed Forecast for Wind Farm Based on Machine Learning*; Dong Hua University: Shanghai, China, 2020.
31. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 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/>).