

Article

Short-Term Wind Power Prediction Based on Data Decomposition and Combined Deep Neural Network

Xiaomei Wu ¹, Songjun Jiang ¹, Chun Sing Lai ^{1,2,*} , Zhuoli Zhao ¹ and Loi Lei Lai ^{1,*} 

¹ Department of Electrical Engineering, School of Automation, Guangdong University of Technology, Guangzhou 510006, China

² Brunel Interdisciplinary Power Systems Research Centre, Department of Electronic and Electrical Engineering, Brunel University London, London UB8 3PH, UK

* Correspondence: chunsing.lai@brunel.ac.uk (C.S.L.); l.l.lai@gdut.edu.cn (L.L.L.)

Abstract: A hybrid short-term wind power prediction model based on data decomposition and combined deep neural network is proposed with the inclusion of the characteristics of fluctuation and randomness of nonlinear signals, such as wind speed and wind power. Firstly, the variational mode decomposition (VMD) is used to decompose the wind speed and wind power sequences in the input data to reduce the noise in the original signal. Secondly, the decomposed wind speed and wind power sub-sequences are reconstructed into new data sets with other related features as the input of the combined deep neural network, and the input data are further studied for the implied features by convolutional neural network (CNN), which should be passed into the long and short-term memory neural network (LSTM) as input for prediction. At the same time, the improved particle swarm optimization algorithm (IPSO) is adopted to optimize the parameters of each prediction model. By superimposing each predicted sub-sequence, the predicting wind power could be obtained. Simulations based on a short-term power prediction in different months with huge weather differences is carried out for a wind farm in Guangdong, China. The simulated results validate that the proposed model has a high prediction accuracy and generalization ability.

Keywords: short-term wind power prediction; data decomposition; combined deep neural network; improved particle swarm optimization algorithm; optimal parameter



Citation: Wu, X.; Jiang, S.; Lai, C.S.; Zhao, Z.; Lai, L.L. Short-Term Wind Power Prediction Based on Data Decomposition and Combined Deep Neural Network. *Energies* **2022**, *15*, 6734. <https://doi.org/10.3390/en15186734>

Academic Editor: Adrian Ilinca

Received: 21 August 2022

Accepted: 12 September 2022

Published: 14 September 2022

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



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As a source of clean and renewable power generation, the installed capacity of wind power is increasing year by year. However, wind power is volatile and stochastic, improving the accuracy of wind power prediction contributes to control the power grid stability, and maximizes the wind energy consumption of the power grid operation. According to the following references, predicting model based on Artificial Intelligence (AI) has more potential in the short-term prediction. At the same time, to enhance the prediction model with better performance and to address the limitations of a single AI model in terms of prediction accuracy, combined prediction models have been used in wind power prediction in recent years. Combined prediction models have been improved from multiple perspectives, such as data decomposition [1], feature clustering [2], parameter optimization [3], and error correction [4], which have greatly improved the performance of wind power prediction.

As the first step in combined prediction models, mode decomposition of data is usually performed with the purpose to reduce the volatility of data. The commonly used decomposition methods for wind power prediction are empirical mode decomposition (EMD) [5], ensemble empirical mode decomposition (EEMD) [6], and improved methods are based on EMD [7–10]. At the same time, there are also Wavelet Decomposition (WPD) [11], singular spectrum analysis [12], and other methods for smoothing data. Variational mode decomposition (VMD) [13] was proposed by Konstantin Dragomiretskiy et al. in 2014, which can

effectively suppress the phenomenon of spectral aliasing during EMD by controlling its bandwidth [14]. The merits and shortcomings of the above-mentioned model decompositions are shown in Table 1. Compared with the original data, the decomposed sub-data are more likely to be mined by the prediction model for implied features, thus improving the training effectiveness. After one-step decomposition of the original wind speed and wind power, the sub-series may keep high complexity. To further reduce the noise of the original data, the framework of multi-step mode decomposition was firstly proposed in [15] to be applied to wind power prediction, and the experimental results verified the effectiveness of the method. In [16], after dividing the data into long and short time scales, the data of the two time scales are processed separately using EEMD and reinforcement learning, and the experimental results verify that the method is feasible. Parameter optimization, the use of intelligent algorithms to solve a problem, has the issues that the prediction model parameters are difficult to be determined. By setting the solution space of the prediction model parameters, the optimal solution of the parameters is continuously searched and optimized, which greatly maximizes the potential of the intelligent algorithm in parameter optimization [17]. Reference [18] used a combination of Adaboost and PSO algorithms to optimize the thresholds and weights of ELM to improve the prediction accuracy of the model and to avoid the impact on accuracy caused by manually setting parameters. Using the idea of using FAR and AR models for forecasting uncertain components in [19] to apply to short-term wind power forecasting, the decomposed components of the data are divided into deterministic and non-deterministic components, and adopting a combined forecasting model to deal with the deterministic component and applying an autoregressive model to deal with the uncertain component will be helpful for forecasting accuracy. Wind power prediction is not only related to the input at the current moment but also correlated with the past inputs and outputs since As LSTM is equipped with memory function and captures the temporal correlation between wind power and other meteorological features, such as wind speed and temperature, LSTM is more suitable for wind power prediction [20]. A detailed and in-depth study and analysis of the application of artificial intelligence techniques (deep learning models) to wind speed and wind power has been conducted in the literature [21] to verify the applicability of the combined depth model to wind power prediction from different perspectives, such as data pre-processing, data feature mining, and data relationship learning.

By summarizing the above results, the models used in wind power prediction mostly consider multiple perspectives and combine prediction models to ensure that prediction accuracy can be at a high level. Table 2 gives a detailed comparison of this paper with some other references. In this paper, a novel hybrid short-term wind power prediction model was proposed that is based on data decomposition (VMD) and combined deep neural network (CNN-LSTM). Firstly, the model uses VMD to decompose wind speed and wind power, with the aim of smoothing such time series as needed due to the volatility of wind speed and wind power. The decomposed wind speed and wind power sub-series are reconstructed with other correlated feature variables (temperature and humidity) into several new data sets as the multi-feature inputs to the prediction model. Secondly, to dig deeper into the implied relationships between the variables, the input data are subject to feature extraction using CNN. The prediction model is constructed independently for each new data set, and the network structure parameters are optimized using improved particle swarm algorithm (IPSO). Then, the optimized parameters are substituted back into the prediction model for wind power prediction. The final wind power is obtained by superimposing each prediction component. The simulated results verify the improvement of the proposed model in prediction accuracy and the effectiveness of combining IPSO to optimize the network structure parameters.

Table 1. Merit and demerit of different decomposition methods.

Decomposition Method	Merit	Demerit
EMD	1. Adaptive: automatic generation of basis functions based on the data itself, adaptive filtering characteristics, adaptive multi-resolution 2. Completeness: the properties of the original series can be reproduced by superimposing the decomposed components	1. Endpoint effect: there is no guarantee that the left and right endpoints of the signal are exactly the local poles 2. Mode mixing: the signals of different characteristic scales appear in one IMF component, and the other is that the signals of the same characteristic scales are dispersed into different IMF components
EEMD	White noise is added to the original signal before each decomposition to overcome the mode mixing in the EMD method	Due to the addition of noise, there is a possibility of inaccuracy in the reconstructed signal and noise cannot be ignored after reconstruction
CEEMD	The CEEMD effectively solves the problems of inaccurate reconstruction caused by noise pollution in the EEMD by adding positive and negative pairs of noise	Differences in the number of IMFs generated during decomposition lead to errors in decomposition
SSA	SSA can be applied to a wide range of time series without model restrictions	SSA does not have an explicit method to determine a threshold to distinguish the signal component from the noise component
VMD	VMD transforms signal decomposition into non-recursive, variational decomposition mode, which can effectively overcome the mode mixing phenomenon generated in EMD and has stronger noise robustness and weaker endpoint effect than EMD	Empirical knowledge is needed to adjust the choice of the K parameter in VMD

Table 2. Literature review of recent works.

	This Work	Anfeng Zhu, et al. [7]	Chao Zhang, et al. [9]	Hao Yin, et al. [15]	Rasoul Rahmani, et al. [22]
Country	China	China	China	Spain	Iran
Research context	Presenting a hybrid model combined with VMD and CNN-LSTM for short-term wind power prediction	A combined prediction model is proposed which employs CEEMD as a wind speed decomposition method as well as an Elman neural network to learn the relationship between wind speed and wind power	Presenting a novel hybrid wind power short-term prediction model for wind power prediction that is combined with CEEMD-LZC, ALO, and ELM network	Presenting a hybrid wind power prediction approach by applying a cascaded deep learning model and a two-layer of mode decomposition method	Presenting a new hybrid swarm technique applied in forecasting the wind energy that makes the best advantage of mixing search ability of ACO and PSO

Table 2. Cont.

	This Work	Anfeng Zhu, et al. [7]	Chao Zhang, et al. [9]	Hao Yin, et al. [15]	Rasoul Rahmani, et al. [22]
Prediction interval	15 min	15 min	1 h	1 h	1 h
Model considered	VMD CNN LSTM IPSO	CEEMD WOA Elman	CEEMD LZC LLE ALO ELM	EMD VMD CNN LSTM	ACO PSO HAP
Findings	Optimizing the network structure parameters of CNN-LSTM model by improved PSO algorithm to improve the prediction accuracy of wind power	The combined prediction model utilizes CEEMD as a wind speed decomposition method to effectively solve the problem of reconstruction error and low efficiency of wind speed decomposition by EEMD	The hybrid model uses LZC to evaluate the complexity of each IMF after applying CEEMD to decompose the wind power and merges the IMFs according to the value of complexity, thus reducing the overall complexity of the model and improving the training efficiency	The original time series is decomposed by EMD and the sub-series that need to be decomposed in two steps are filtered out and further decomposed by VMD, while the implied relationship between wind speed, wind direction, and wind power are extracted by CNN	With the hybrid algorithm of ACO and PSO for the search of parameter values in the mathematical model, it can converge more quickly than the single ACO and PSO algorithm, and the accuracy of the model is also optimal
Drawback	Further validation of the generalization capability of this model by obtaining data from wind farms in other regions would strengthen the generalization ability of this model	The high number of IMFs in the CEEMD decomposition may lead to a higher complexity of the combined model	The LLE method proposed in this paper, as a novel data dimensionality reduction method, effectively reduces the dimensionality of the data. However, traditional dimensionality reduction methods are missing for comparison, such as PCA and other methods	The two-step decomposition results in a complex model and an increase in the time used for model training	The mathematical model consists of the S-curve and parabola, but it would be useful for the simulation to provide reasons why the model is composed of both of two curves and how much the effect can be improved compared to a single curve case

This paper is organized as follows: Section 2 describes the methodology. Section 3 presents the construction of the wind power prediction model and the evaluation metrics. Section 4 provides case study. The conclusion is drawn in Section 5.

2. Methodology

2.1. Variational Mode Decomposition

VMD is an adaptive and completely non-recursive mode decomposition method, which essentially smooths the sequence signal [23]. VMD uses iterative search for the optimal solution of the variational model to select the center frequency and bandwidth of each component, so the decomposed modes can achieve the reproduction of the original sequence, and at the same time each mode demodulation to the baseband can be smooth. In this paper, VMD is chosen as the decomposition method because it can effectively reduce the non-smoothness of time series by overcoming the problem of mode component mixing in EMD. The specific decomposition process and the algorithmic steps are described in References [24–28]. VMD is used as a pre-processing tool for time series, and the decomposed mode quantities with different characteristics can be applied as input to subsequent analysis methods for further in-depth analysis.

2.2. Convolutional Neural Network

CNN has features, such as shared weights and local connectivity [29], which makes it an efficient processing tool for data feature extraction. CNN can be classified into 1D, 2D, and 3D, and each of the three types of CNN is applicable to different categories. The data sets that make up wind power prediction are essentially time series composed of multiple feature variables, and 1D CNN is mainly used for feature extraction of time series data. In this paper, the time series of temperature and humidity related to wind power are combined with the sub-series of wind speed and wind power decomposed by VMD to form the multi-feature input of the 1D-CNN, which aims to further explore the implied relationship between the feature variables and to improve the training efficiency of the model. Considering the small number of features in the data of this paper, no pooling layer is set in this CNN. The computation process of the 1D CNN is:

$$x_i^k = f\left(\sum_{j=1}^N x_i^{k-1} * \omega_{ji}^k + b_i^k\right) \quad (1)$$

Figure 1 shows the CNN-LSTM combined neural network model used in this paper.

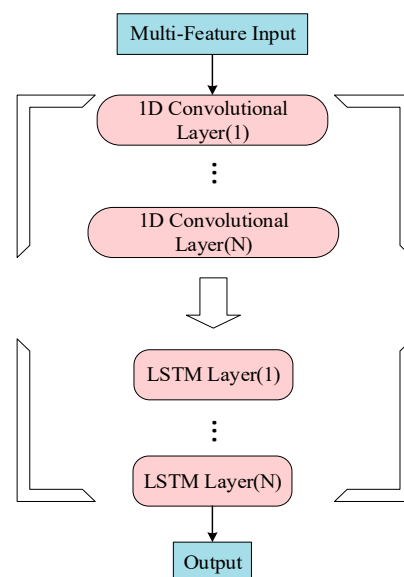


Figure 1. CNN-LSTM combined neural network model.

2.3. Long Short-Term Memory Neural Network

LSTM is the improved recurrent neural network (RNN), which aims to overcome the problem of gradient explosion and gradient disappearance in traditional RNN. The internal structure of LSTM is mainly composed of three gates, which are forget gate f_t , input gate i_t , and output gate o_t . LSTM has the function of memory and can capture the correlation between past and future information, so LSTM is very advantageous for its application on time series [30]. Considering that wind speed and wind power data are typically nonlinear and volatile time series, the prediction model uses LSTM as the neural network for modeling. The structure of LSTM is shown in Figure 2.

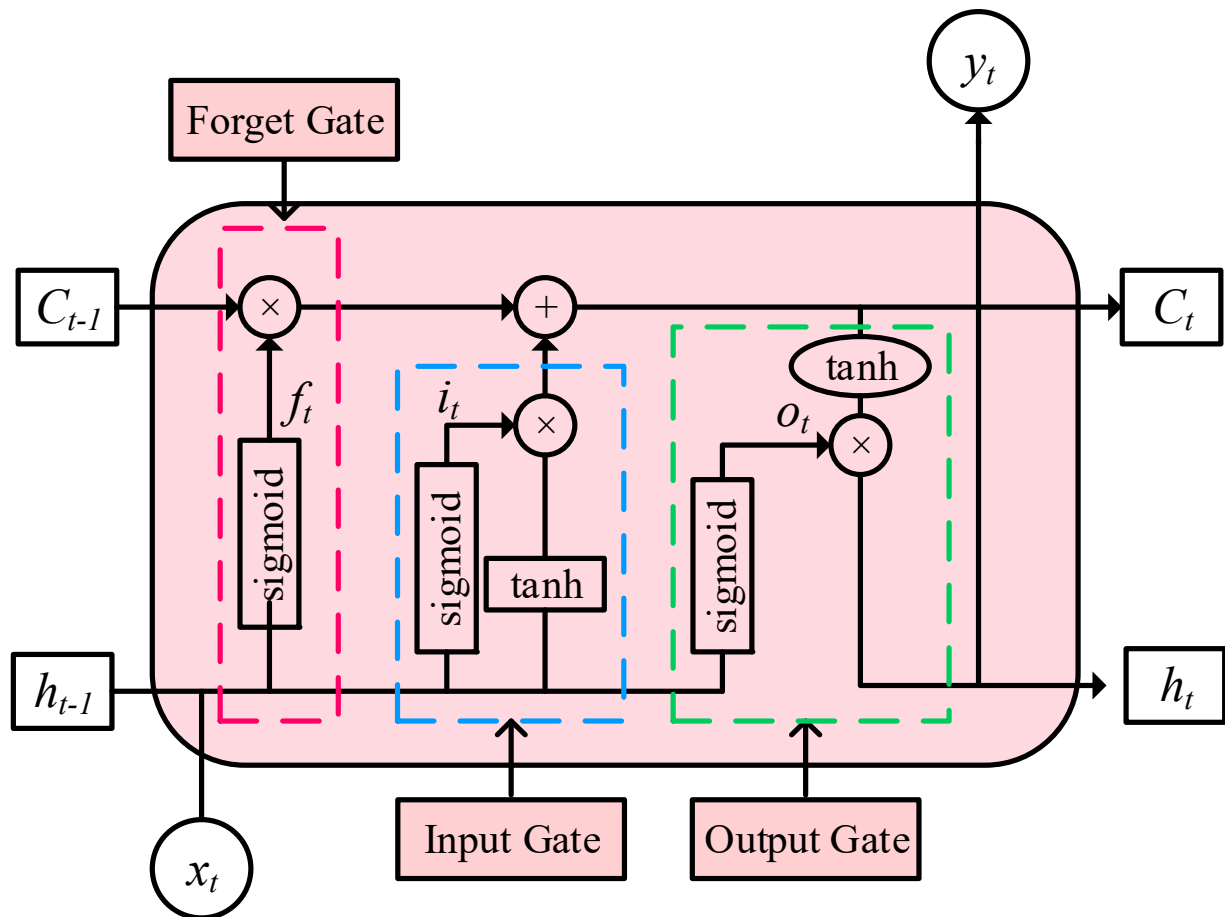


Figure 2. Internal structure of LSTM.

Firstly, the role of the forget gate f_t in LSTM is to decide whether to keep or delete the information in the previous cell state C_{t-1} according to the size of the output value. f_t input is the current moment input X_t and the previous moment output h_{t-1} . The input gate i_t is used to determine which part of X_t can be retained in C_t . The input of i_t is also used by the sigmoid function to generate an output value of size between 0 and 1 to control the tanh function to generate a new candidate cell state \tilde{C}_t . The cell state C_t at the current moment is updated by the combined action of the forget and input gates. The output of the LSTM is controlled by the output gate o_t . The output h_t of the LSTM at the current moment is equal to the product of the output weights and the pending output. The relationships among the variables in the LSTM are shown in Equations (2)–(7).

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \quad (2)$$

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_c \times [h_{t-1}, x_t] + b_c) \quad (4)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C} \quad (5)$$

$$\sigma_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t \times \tanh(C_t) \quad (7)$$

2.4. Adaptive Weighted Particle Swarm Algorithm Combined with Elimination Mechanism

The original data set is decomposed to form different data sub-sets, and manually setting the optimal network structure parameters corresponding to each data sub-set will increase the time cost, while the prediction error caused by manually setting parameters will cause the error to expand as the number of data sub-sets increases. Therefore, particle swarm optimization (PSO) algorithm, which has a strong search capability, is introduced to optimize the network structure parameters [31]. The particle swarm continuously updates its position during the search process to finally obtain the totally optimal solution. During each iteration, the velocity and position of the particles are updated as shown in Equations (8) and (9).

$$V_i^{k+1} = \omega \times V_i^k + c_1 r_1 (Pbest_i^k - X_i^k) + c_2 r_2 (Gbest_i^k - X_i^k) \quad (8)$$

$$X_i^{k+1} = X_i^k + V_i^{k+1} \quad (9)$$

When the standard PSO algorithm is used to optimize the parameters, multiple particles are usually involved in iterations to fully utilize the search capability of the particle swarm. In each round of iteration, some particles move to the position with a better fitness value, while others move to the position with a worse fitness value. If particles move to the position with a worse fitness value in the previous iteration, the inertial component of particles in the current iteration will lead to a worse fitness value for the particles, which is not good for the particles to converge to the global optimal solution. If the inertia direction of each particle is dynamically adjusted in each iteration according to the fitness value obtained in the previous iteration, it will be very beneficial to the particle convergence. Therefore, adaptive adjustment is used for the inertia weights, as shown in Equation (10).

$$\omega_i^k = \begin{cases} \omega_{\min} + (\omega_{\max} - \omega_{\min}) \times \frac{f(X_i^k) - f_{\min}^k}{f_{\text{average}}^k - f_{\min}^k}, & f(X_i^k) \leq f_{\text{average}}^k \\ \omega_{\max}, & f(X_i^k) \geq f_{\text{average}}^k \end{cases} \quad (10)$$

where f_{average}^k is the average fitness of all particles at the k th iteration; f_{\min}^k is the minimum fitness of all particles at the k th iteration.

To take further advantage of the total search capability of the particle swarm, this paper introduces the elimination mechanism in the genetic algorithm into the standard PSO algorithm [32]. After the adaptive adjustment of the inertia weights, particles are arranged in descending order according to the fitness value of each particle in this round, and the particles ranked in the bottom 50% of the fitness value are eliminated, while the particles ranked in the top 50% are used as the boundary. The new 50% of the particles are randomly generated within the boundary and combined with the top 50% of the retained particles to form a new population to start the next round of iterations, thus improving the total optimization capability of the algorithm. When the IPSO algorithm is used to optimize the network structure parameters, the positions of the particles are used to represent the number of LSTM layers, the number of DENSE layers, and the number of neurons in each layer of the prediction model. The flowchart of the IPSO algorithm for optimizing the network structure parameters is shown in Figure 3.

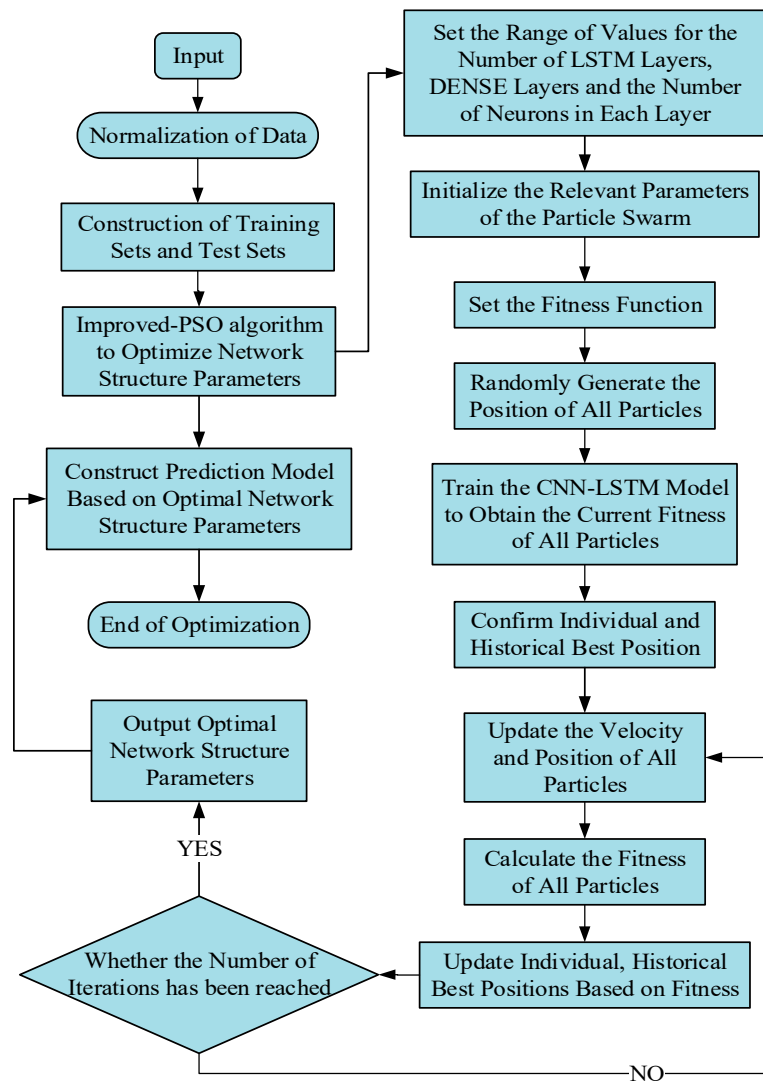


Figure 3. Flow chart of optimizing parameters for combined neural network structure by IPSO.

3. Construction of Wind Power Prediction Model and Evaluation Metrics

3.1. VMD-CNN-IPSO-LSTM Prediction Model

In this paper, VMD-CNN-IPSO-LSTM prediction model is used for short-term wind power prediction, which is mainly divided into six steps, and the prediction process is shown in Figure 4.

Step 1: Decomposition of the original wind speed and wind power using VMD after removing and repairing the outliers and missing values of the data set;

Step 2: Reconstruct the new dataset with other correlated meteorological features (temperature, humidity) after obtaining the sub-sequence and normalize the dataset at the same time;

Step 3: Constructing CNN- LSTM combined prediction models according to the different datasets;

Step 4: In the pending training model, the number of layers for LSTM layers and DENSE layers and the number of neurons in each layer are optimized by IPSO;

Step 5: The mean square error between the true and predicted values is used as the fitness function, and the optimal network structure parameters of each prediction model are obtained based on the merit of the fitness value and substituted back into the corresponding prediction model;

Step 6: Add all prediction components to obtain the final wind power.

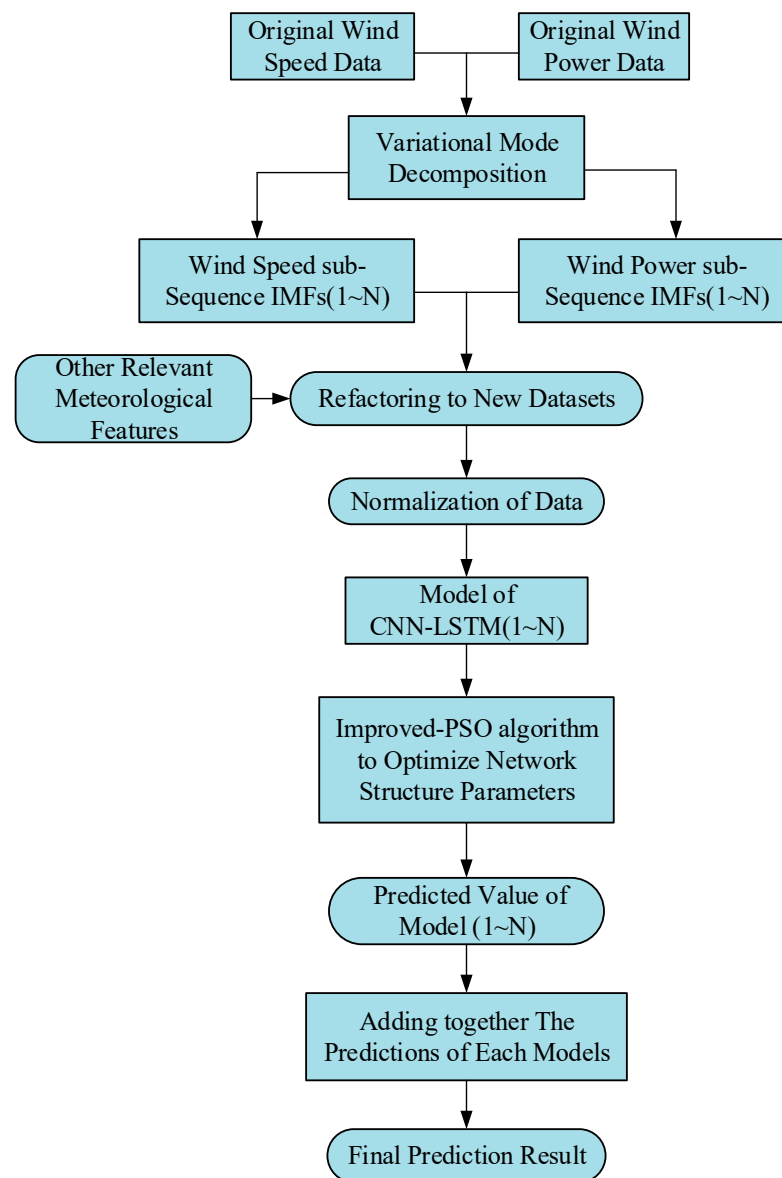


Figure 4. Flow chart of prediction model.

3.2. Evaluation Metrics

In this paper, four evaluation metrics are used to measure the prediction performance of this model. They are mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and Adjusted R-Square (adj-R^2) and are defined as follows:

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

$$MAPE = \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times \frac{100\%}{n} \quad (13)$$

$$\bar{y} = \frac{1}{n} \times \sum_{i=1}^n y_i \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (15)$$

$$adj-R^2 = 1 - \frac{(1 - R^2) \times (n - 1)}{n - p - 1} \quad (16)$$

The smaller the value of MAE, RMSE, and MAPE, the better the prediction performance of the model. The value of adj- R^2 is between [0, 1], and if the value is closer to 1, the model can perform better.

4. Case Study

In this paper, simulations are conducted with historical data from a wind farm in Guangdong, and the dataset include wind speed, temperature, humidity, and wind power. The dataset is selected from 1 January 2019 to 31 January 2019, with a sampling interval of 15 min. The training set is selected for the first 29 days, while the test set is from the last 2 days.

4.1. Decomposition of Wind Speed and Wind Power Series Using VMD

The decomposition of the wind speed and wind power series are performed by using VMD to reduce the self-contained volatility of the original series. The results of wind speed and wind power decomposition are shown in Figures 5 and 6, respectively.

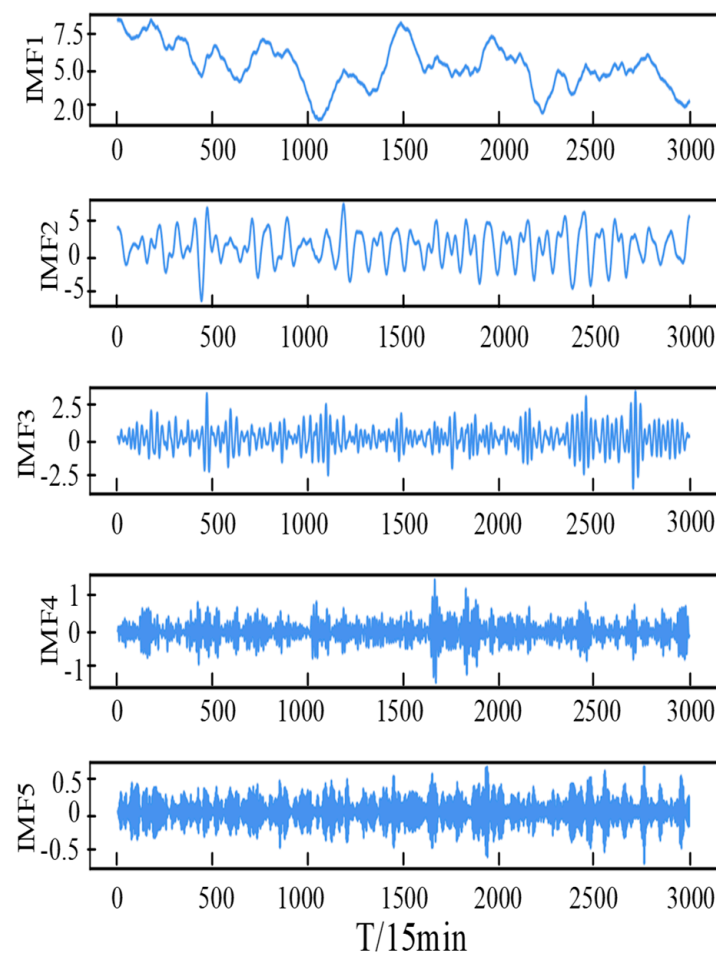


Figure 5. Decomposed sub-sequence of wind speed by VMD.

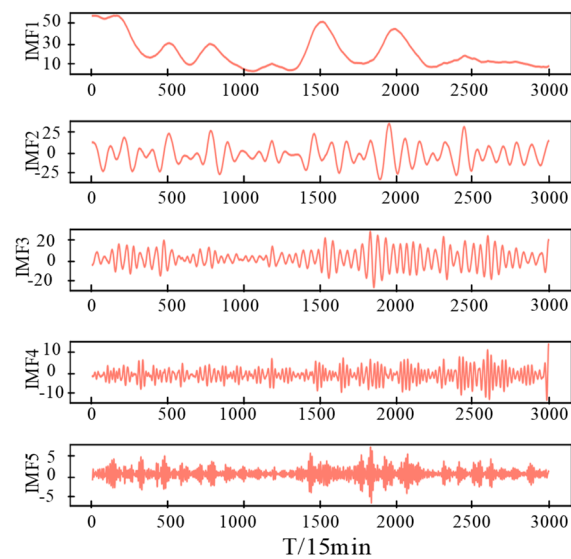


Figure 6. Decomposed sub-sequence of wind power by VMD.

After VMD processing, wind speed and wind power are decomposed into five modes, and the decomposed modes maintain the characteristics of the original series, while reducing their non-smoothness.

Before presenting the model, three groups of prediction models (Standard LSTM, EMD-LSTM, and VMD-LSTM) are introduced for comparison. Figure 7 shows the improvement of the prediction curves after EMD and VMD decomposition over the undecomposed prediction curves, which are closer to the true value in terms of trend. Table 3 gives the MAE, RMSE, MAPE, and adj- R^2 error analysis for the predictions of the three models. The LSTM model without data decomposition is the worst in MAE, RMSE and adj- R^2 for prediction, so it's reasonable to apply data decomposition as data pre-processing; meanwhile, the prediction model with VMD is higher than EMD in terms of accuracy, which proves that the decomposition method can learn deeper information on the data set used in this paper.

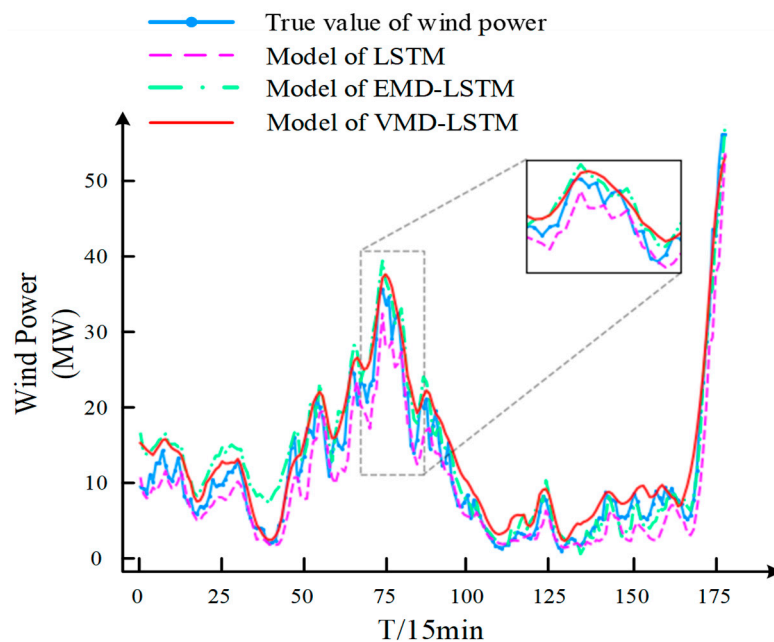


Figure 7. Predicted and real values for different decomposition models.

Table 3. Prediction performance analysis of single model and decomposed combined model.

Model	Prediction Performance Analysis			
	MAE	RMSE	MAPE	adj-R ²
LSTM	2.38732	3.32454	0.59477	0.87264
EMD-LSTM	2.31746	2.88094	1.29272	0.90436
VMD-LSTM	2.14829	2.63109	1.46155	0.92023

4.2. IPSO Optimized Prediction Model Network Structure Parameters

The decomposed wind speed and wind power sub-series were combined with other feature variables to reconstruct the data to obtain five datasets, and five CNN-LSTM combined prediction models were constructed simultaneously. The constructed combined neural network contains four 1D-CNN layers, the number of filters in each layer is 4, 4, 8, and 16, and the filter training step is 2. The number of training iterations of IPSO are set to 200; the number of particles is set to 10; the initial inertia weight ω is set to 0.9; c_1 and c_2 are set to 2.0. For the five combined neural networks, the LSTM layer and the DENSE layers are set at [1, 3] layers; the number of neurons in each layer is set at [8, 96], and the optimal number of layers and neurons are searched by IPSO. After 200 generations of training, the results of the optimal parameters of the network structure corresponding to the five prediction models are shown in Table 4 below.

Table 4. Optimal structural parameters of each prediction model.

Combined Prediction Model	LSTM Layers	DENSE Layers	The Number of Neurons
Model 1	3	2	85/59/13/33/58
Model 2	2	1	66/57/37
Model 3	2	2	38/94/63/35
Model 4	1	2	71/79/21
Model 5	3	1	64/26/86/26

To verify the effectiveness of this model, the VMD-LSTM model and VMD-CNN-LSTM model are introduced to compare the performance of the two models. The network structures of both models are two LSTM layers and two DENSE layers; the number of neurons in the two LSTM layers are 32 and 16, and the number of neurons in the two DENSE layers are 100 and 50, respectively.

Figure 8 shows the prediction values of each model compared to the true values. It is clear that the prediction values obtained by this model are the best fit and demonstrate the best prediction performance of the model.

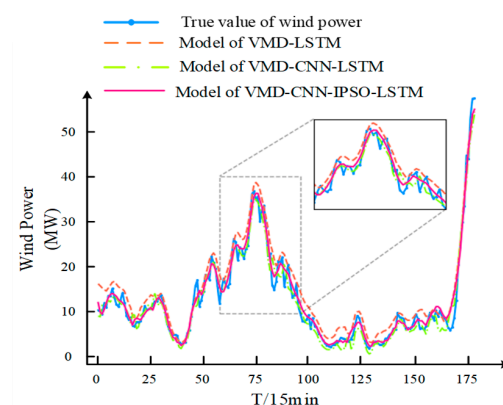
**Figure 8.** Wind power predicted value and real value of each model.

Table 5 shows the values of the prediction performance analysis of each model under the evaluation metrics. Firstly, the accuracy of the model has been improved after VMD processing. After VMD processing, using CNN to further capture the data features will lead to better prediction performance of the model. However, manually setting model structure parameters are not the optimal structure parameters, so IPSO is used to search for the optimal structure parameters of the prediction model corresponding to each VMD-CNN processed dataset. After substituting the optimal parameters in Table 4 back to each prediction model for prediction, it is proven that all the metrics of the VMD-CNN-IPSO-LSTM model are optimal. Compared with the VMD-CNN-LSTM model, the MAE is reduced by 25.7%, the RMSE is reduced by 20.87%, and the MAPE also showed a substantial decrease compared to the above comparative models, while $\text{adj-R}^2 = 0.97349$ also indicates that the proposed model is helpful for short-term wind power prediction.

Table 5. Analysis of the prediction performance of each model.

Model	Prediction Performance Analysis			
	MAE	RMSE	MAPE	adj-R ²
VMD-LSTM	2.14829	2.63109	1.46155	0.92023
VMD-CNN-LSTM	1.48819	1.91680	1.38401	0.95766
VMD-CNN-IPSO-LSTM	1.10571	1.51672	0.43060	0.97349

4.3. Verify the Generalization Ability of the Model under the New Data Set

To further evaluate the prediction performance and the generalization ability of this model, it was validated for different meteorological conditions of this wind farm. Considering the large differences of weather in Guangdong in January and June, 1 June 2019 to 30 June 2019 was chosen as the new data set for simulation.

The optimal parameters of the network structure after IPSO optimization for the five prediction models in the new dataset are shown in Table 6, while the wind power curves of the three models with the real values for 48 h are shown in Figure 9.

Table 6. Optimal structural parameters of each prediction model in the new data set.

Combined Prediction Model	LSTM Layers	DENSE Layers	The Number of Neurons
Model 1	3	3	68/74/46/15/78/82
Model 2	1	1	66/75
Model 3	3	1	64/54/78/52
Model 4	1	3	32/34/77/40
Model 5	2	3	58/77/14/24/39

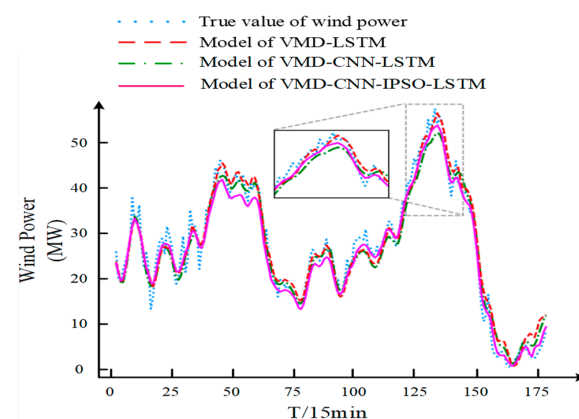


Figure 9. Wind power predicted value and real value of each model in the new data set.

Since this model is adopted with a deep neural network, it can sharply capture the temporal correlation between the feature in the new data set and thus quickly learn the coupling relationship implied by each feature. The prediction curves of the VMD-CNN-IPSO-LSTM model still fit the real wind power curve well in terms of both trend and turning points, as seen in Figure 9, thus demonstrating that the prediction performance of the model remains robust with the new data set. The prediction error analysis of each model in the new data set is given in Table 7.

Table 7. Prediction performance analysis of each model in new data set.

Model	Prediction Performance Analysis			
	MAE	RMSE	MAPE	adj-R ²
VMD-LSTM	3.30657	4.07535	0.35427	0.95683
VMD-CNN-LSTM	3.20033	4.01825	0.43912	0.95803
VMD-CNN-IPSO-LSTM	2.92668	3.59604	0.20147	0.96639

In the new data set, the MAE, RMSE, MAPE, and adj-R² of all three models have increased. However, the VMD-CNN-IPSO-LSTM model still outperforms the comparison models in all error metrics. Compared with the model processed by VMD-CNN only, the MAE of this model is reduced by 8.55%, the RMSE is reduced by 10.5%, and the MAPE is reduced by 43.1%. Additionally, the adj-R² of the proposed model is greater than 0.96 for both January and June data sets, which proves that the VMD-CNN-IPSO-LSTM model has strong generalization ability for short-term wind power prediction.

5. Conclusions

In this paper, a hybrid short-term wind power prediction model is proposed based on data decomposition (VMD) and combined deep neural network (CNN-LSTM) for wind speed and wind power series with the consideration of the characteristics of non-linearity and volatility. The effectiveness of the model is confirmed through comparative simulated analysis with a wind farm in Guangdong, China. Some of the observations can be summarized as follows:

- (1) The volatility and noise of wind speed and wind power series are effectively reduced after variational mode decomposition;
- (2) The combined deep neural network has a sharp learning ability for the data implicit feature. However, the overall consumption time is longer than that of the single model, so it is important to choose the number of decomposed modes reasonably. If the number of modes is too much or too little, it will have an impact on the hybrid model;
- (3) The network structure parameters of the prediction model can be optimized by the improved PSO algorithm to further enhance prediction accuracy;
- (4) To verify the generalized potential of the proposed model, the output of the same wind farm in two months with large climate difference is investigated. The analysis of the prediction results confirms that the model has a strong generalized ability;
- (5) For future research directions, the authors intend to use a multi-step decomposition model after which the sub-sequence obtained from the decomposition is passed through different neural networks for feature extraction, such as using graph convolutional neural networks for implicit mining, and to compare whether the model proposed in this paper makes a breakthrough in accuracy through simulation.

Author Contributions: L.L.L. and C.S.L. guided the framework of the paper and provided professional guidance. X.W. and S.J. did the simulation and wrote the paper. Z.Z. provided academic assistance and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Guangdong-Guangxi Joint Foundation of China: Research on Distributed Optimal Control of Renewable Multi-microgrids based on the Integration of Multi-Agent Dynamic Game and Prediction Mechanism [Project Number 2021A1515410009].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data in this paper are from a real wind farm and involve a confidentiality agreement. The dataset in this paper is not publicly available.

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

Abbreviations

The following abbreviations are used in this manuscript:

VMD	Variational Mode Decomposition
EMD	Empirical Mode Decomposition
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long and Short-term Memory Neural Network
IPSO	Improved Particle Swarm Optimization
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
MAPE	Mean Absolute Percentage Error
R^2	Coefficient of Determination
Adj- R^2	Adjusted R-Square
x_i^k, w_{ji}^k, b_i^k	Output, Weight, Threshold of the i th Convolution Kernel in the k th Layer
f	Activation Function
N	the Number of Input Convolution Kernels
$*$	Convolution Operation
f_t, i_t, o_t	Outputs of the 3 Gates of LSTM
W_f, W_i, W_c, W_o	Weight matrix of LSTM
b_f, b_i, b_c, b_o	Biased vector of LSTM
V_i^{k+1}, V_i^k	Updated, Current Velocity of the i th Particle
X_i^{k+1}, X_i^k	Updated, Current Position of the i th Particle
w	Inertia Weight
c_1, c_2	Learning Factor
r_1, r_2	Random Number between [0, 1]
n	the Total Amount of Data Set
p	Number of features in the dataset
y_i, \hat{y}_i, \bar{y}	True value, Predicted value, the Average of True Value

References

1. Zhou, J.; Xu, X.; Huo, X.; Li, Y. Forecasting models for wind power using extreme-point symmetric mode decomposition and artificial neural networks. *Sustainability* **2019**, *11*, 650. [\[CrossRef\]](#)
2. Wang, Y.; Wang, D.; Tang, Y. Clustered hybrid wind power prediction model based on ARMA, PSO-SVM, and clustering methods. *IEEE Access* **2020**, *8*, 17071–17079. [\[CrossRef\]](#)
3. Han, Y.; Tong, X. Multi-step short-term wind power prediction based on three-level decomposition and improved grey wolf optimization. *IEEE Access* **2020**, *8*, 67124–67136. [\[CrossRef\]](#)
4. Zhang, Y.; Han, J.; Pan, G.; Xu, Y.; Wang, F. A multi-stage predicting methodology based on data decomposition and error correction for ultra-short-term wind energy prediction. *J. Clean. Prod.* **2021**, *292*, 125981. [\[CrossRef\]](#)
5. Bokde, N.; Feijóo, A.; Villanueva, D.; Kulat, K. A review on hybrid empirical mode decomposition models for wind speed and wind power prediction. *Energies* **2019**, *12*, 254. [\[CrossRef\]](#)
6. Li, M.; Li, Y. Dispatch Planning of a Wide-Area Wind Power-Energy Storage Scheme Based on Ensemble Empirical Mode Decomposition Technique. *IEEE Trans. Sustain. Energy* **2020**, *12*, 1275–1288. [\[CrossRef\]](#)
7. Zhu, A.; Zhao, Q.; Wang, X.; Zhou, L. Ultra-Short-Term Wind Power Combined Prediction Based on Complementary Ensemble Empirical Mode Decomposition, Whale Optimisation Algorithm, and Elman Network. *Energies* **2022**, *15*, 3055. [\[CrossRef\]](#)
8. Liu, Z.; Hara, R.; Kita, H. 24 h-ahead wind speed forecasting using CEEMD-PE and ACO-GA-based deep learning neural network. *J. Renew. Sustain. Energy* **2021**, *13*, 046101. [\[CrossRef\]](#)

9. Zhang, C.; Ding, M.; Wang, W.; Bi, R.; Miao, L.; Yu, H.; Liu, L. An improved ELM model based on CEEMD-LZC and manifold learning for short-term wind power prediction. *IEEE Access* **2019**, *7*, 121472–121481. [\[CrossRef\]](#)
10. Chen, X.; Lai, C.S.; Ng, W.W.; Pan, K.; Lai, L.L.; Zhong, C. A stochastic sensitivity-based multi-objective optimization method for short-term wind speed interval prediction. *Int. J. Mach. Learn. Cybern.* **2021**, *12*, 2579–2590. [\[CrossRef\]](#)
11. Zhang, D.; Peng, X.; Pan, K.; Liu, Y. A novel wind speed forecasting based on hybrid decomposition and online sequential outlier robust extreme learning machine. *Energy Convers. Manag.* **2019**, *180*, 338–357. [\[CrossRef\]](#)
12. Safari, N.; Chung, C.; Price, G. Novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis. *IEEE Trans. Power Syst.* **2017**, *33*, 590–601. [\[CrossRef\]](#)
13. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2013**, *62*, 531–544. [\[CrossRef\]](#)
14. Ali, M.; Khan, A.; Rehman, N.U. Hybrid multiscale wind speed forecasting based on variational mode decomposition. *Int. Trans. Electr. Energy Syst.* **2018**, *28*, e2466. [\[CrossRef\]](#)
15. Yin, H.; Ou, Z.; Huang, S.; Meng, A. A cascaded deep learning wind power prediction approach based on a two-layer of mode decomposition. *Energy* **2019**, *189*, 116316. [\[CrossRef\]](#)
16. Zhang, H.; Yue, D.; Dou, C.; Li, K.; Hancke, G.P. Two-step wind power prediction approach with improved complementary ensemble empirical mode decomposition and reinforcement learning. *IEEE Syst. J.* **2021**, *16*, 2545–2555. [\[CrossRef\]](#)
17. Li, L.-L.; Chang, Y.-B.; Tseng, M.-L.; Liu, J.-Q.; Lim, M.K. Wind power prediction using a novel model on wavelet decomposition-support vector machines-improved atomic search algorithm. *J. Clean. Prod.* **2020**, *270*, 121817. [\[CrossRef\]](#)
18. An, G.; Jiang, Z.; Cao, X.; Liang, Y.; Zhao, Y.; Li, Z.; Dong, W.; Sun, H. Short-term wind power prediction based on particle swarm optimization-extreme learning machine model combined with AdaBoost algorithm. *IEEE Access* **2021**, *9*, 94040–94052. [\[CrossRef\]](#)
19. Jan, F.; Shah, I.; Ali, S. Short-Term Electricity Prices Forecasting Using Functional Time Series Analysis. *Energies* **2022**, *15*, 3423. [\[CrossRef\]](#)
20. Jaseena, K.; Kovoov, B.C. Decomposition-based hybrid wind speed forecasting model using deep bidirectional LSTM networks. *Energy Convers. Manag.* **2021**, *234*, 113944. [\[CrossRef\]](#)
21. Wang, Y.; Zou, R.; Liu, F.; Zhang, L.; Liu, Q. A review of wind speed and wind power forecasting with deep neural networks. *Appl. Energy* **2021**, *304*, 117766. [\[CrossRef\]](#)
22. Rahmani, R.; Yusof, R.; Seyedmahmoudian, M.; Mekhilef, S. Hybrid technique of ant colony and particle swarm optimization for short term wind energy forecasting. *J. Wind Eng. Ind. Aerodyn.* **2013**, *123*, 163–170. [\[CrossRef\]](#)
23. Wu, X.; Lai, C.S.; Bai, C.; Lai, L.L.; Zhang, Q.; Liu, B. Optimal kernel ELM and variational mode decomposition for probabilistic PV power prediction. *Energies* **2020**, *13*, 3592. [\[CrossRef\]](#)
24. Bo, H.; Niu, X.; Wang, J. Wind speed forecasting system based on the variational mode decomposition strategy and immune selection multi-objective dragonfly optimization algorithm. *IEEE Access* **2019**, *7*, 178063–178081. [\[CrossRef\]](#)
25. Ma, Z.; Chen, H.; Wang, J.; Yang, X.; Yan, R.; Jia, J.; Xu, W. Application of hybrid model based on double decomposition, error correction and deep learning in short-term wind speed prediction. *Energy Convers. Manag.* **2020**, *205*, 112345. [\[CrossRef\]](#)
26. Shi, X.; Lei, X.; Huang, Q.; Huang, S.; Ren, K.; Hu, Y. Hourly day-ahead wind power prediction using the hybrid model of variational mode decomposition and long short-term memory. *Energies* **2018**, *11*, 3227. [\[CrossRef\]](#)
27. Naik, J.; Bisoi, R.; Dash, P. Prediction interval forecasting of wind speed and wind power using modes decomposition based low rank multi-kernel ridge regression. *Renew. Energy* **2018**, *129*, 357–383. [\[CrossRef\]](#)
28. Li, C.; Tang, G.; Xue, X.; Saeed, A.; Hu, X. Short-term wind speed interval prediction based on ensemble GRU model. *IEEE Trans. Sustain. Energy* **2019**, *11*, 1370–1380. [\[CrossRef\]](#)
29. Ju, Y.; Sun, G.; Chen, Q.; Zhang, M.; Zhu, H.; Rehman, M.U. A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting. *IEEE Access* **2019**, *7*, 28309–28318. [\[CrossRef\]](#)
30. Liu, Y.; Guan, L.; Hou, C.; Han, H.; Liu, Z.; Sun, Y.; Zheng, M. Wind power short-term prediction based on LSTM and discrete wavelet transform. *Appl. Sci.* **2019**, *9*, 1108. [\[CrossRef\]](#)
31. Zhang, Y.; Sun, H.; Guo, Y. Wind power prediction based on PSO-SVR and grey combination model. *IEEE Access* **2019**, *7*, 136254–136267. [\[CrossRef\]](#)
32. Liu, J.; Shi, Q.; Han, R.; Yang, J. A Hybrid GA-PSO-CNN Model for Ultra-Short-Term Wind Power Forecasting. *Energies* **2021**, *14*, 6500. [\[CrossRef\]](#)