

Article



# Short-Term Wind Power Prediction for Wind Farm Clusters Based on SFFS Feature Selection and BLSTM Deep Learning

Xiaosheng Peng \*, Kai Cheng, Jianxun Lang, Zuowei Zhang, Tao Cai and Shanxu Duan

State Key Laboratory of Advanced Electromagnetic Engineering and Technology, School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; m201871391@hust.edu.cn (K.C.); langjianxun2012@126.com (J.L.); m201971642@hust.edu.cn (Z.Z.); caitao@hust.edu.cn (T.C.); duanshanxu@hust.edu.cn (S.D.)
\* Correspondence: XiaoshangPenc@hust.edu.cn

\* Correspondence: XiaoshengPeng@hust.edu.cn

**Abstract:** Wind power prediction (WPP) of wind farm clusters is important to the safe operation and economic dispatch of the power system, but it faces two challenges: (1) The dimensions of the input parameters for WPP of wind farm clusters are very high so that the input parameters contain irrelevant or redundant features; (2) it is difficult to build a holistic WPP model with high-dimensional input parameters for wind farm clusters. To overcome these challenges, a novel short-term WPP model for wind farm clusters, based on sequential floating forward selection (SFFS) feature selection and bidirectional long short-term memory (BLSTM) deep learning, is proposed in this paper. First, more than 300,000 input features of the wind farm cluster are constructed. Second, the SFFS method is applied to sort the high-dimensional features and analyze the rule that the forecasting accuracy changes with the number of feature to obtain the optimal number of features and feature sets. Finally, based on the results of feature selection, BLSTM is applied to build a WPP model for wind farm clusters with a combination of feature selection and deep learning. This case study shows that (1) SFFS is an effective method for selecting the core features for WPP of wind farm clusters; (2) BLSTM shows not only higher WPP accuracy than long short-term memory and backpropagation neural network but also outstanding performance in terms of reducing the phase errors of WPP.

Keywords: deep learning; feature selection; BLSTM; wind power prediction; wind farm clusters

## 1. Introduction

According to the statistical results from the Global Wind Energy Council (GWEC), at the end of 2019, the total installed capacity of global wind farms reached 651 GW [1]. With the rapid development of the global wind power industry, the distribution of wind farms is changing from the decentralized and small-scale distribution of the early stages to clustered and large-scale distribution [2]. A wind farm is a power plant composed of multiple wind turbine units. A wind farm cluster is a group of multiple wind farms in a specific area, in which the number of wind farms is generally several or dozens, depending on the size of the geographical area [3]. Wind power prediction (WPP) is the estimation of the expected production of wind power for a period of future time based on the meteorological data and historical operating data of the wind farm [3,4]. Short-term WPP is usually considered to forecast wind power output in the next few days, usually 2–3 days, which is mainly applied to short-term power generation scheduling and power trading [3]. In this paper, the prediction of the future 96 h is studied. WPP for wind farm clusters is the forecasting of the overall output of a wind farm cluster composed of multiple wind farms in a large spatial range [3]. Wind power output is random and uncertain. The integration of large-scale wind farms into the power system has a significant impact on the safe operation and economic dispatch of the power system, making the WPP of wind farm clusters increasingly important [4].



Citation: Peng, X.; Cheng, K.; Lang, J.; Zhang, Z.; Cai, T.; Duan, S. Short-Term Wind Power Prediction for Wind Farm Clusters Based on SFFS Feature Selection and BLSTM Deep Learning. *Energies* **2021**, *14*, 1894. https://doi.org/10.3390/ en14071894

Academic Editor: Mohamed Benbouzid

Received: 27 February 2021 Accepted: 26 March 2021 Published: 29 March 2021

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



**Copyright:** © 2021 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/). WPP methods for wind farm clusters mainly include the accumulation method and the statistical upscaling method [5–7]. The basic principle of these methods is to predict each or part of the wind farms in the cluster first, based on which the forecasting of output power for the wind farm cluster is obtained through accumulating or upscaling the forecasting results of individual farms [8,9]. With the improvement of computing techniques and the development of deep learning theories [10], holistic WPP modeling for wind farm clusters is becoming possible.

The holistic WPP modeling for wind farm clusters faces the following two challenges: (1) Many features affect the power output of wind farm clusters, including uncorrelated and redundant features, which need to be optimized; (2) for WPP of wind farm clusters, it is difficult to establish a holistic high-precision prediction model with high-dimensional input parameters. To overcome these two challenges, both feature selection and deep learning are studied in this paper.

Feature selection is removing irrelevant and redundant features and reducing the computational complexity of the algorithm by mining the intrinsic relationship between the features and the target sequences [11]. In past studies [12–15], random search methods, such as simulated annealing algorithm, tabu search algorithm, genetic algorithm, and random sampling with replacement algorithm, have been applied for feature selection, which could allow the quick securing of a locally optimal solution that meets the requirements. However, the disadvantages of these methods are the uncertainty of the results and a large amount of training data [12–15].

In reference [16], the heuristic search method was applied for feature selection of the siting and sizing of active and reactive power sources. Both the calculation time and the effectiveness of feature selection are considered within the heuristic search method, which is widely applied for feature selection of text analysis and image recognition. There are many common heuristic search methods, including the best individual feature (BIF) method, the sequence forward selection (SFS) method, the sequential floating forward selection (SFS) method, and so on [17–20]. The BIF method was adopted in reference [18], based on which the features are ranked by calculating the contribution value when each feature is applied individually. The calculation speed of BIF is fast, but the redundancy between feature selection of pattern recognition. Backtracking, a feature elimination mechanism, is applied in the feature selection process of the SFFS method, based on which the redundancy between features is reduced [20]. Therefore, the SFFS method is chosen as the feature selection method for WPP of wind farm clusters in this paper.

Feature selection results will be used as the input parameters for the WPP model. Deep learning is one of the foremost methods of WPP [21–27]. Compared with traditional shallow artificial neural networks, deep learning neural networks have excellent learning and generalization capabilities for massive data, including convolutional neural networks [22], deep belief networks [23], long short-term memory (LSTM) networks [24], stacked denoising autoencoders [25], and so on. Among these methods, LSTM networks show excellent characteristics for the forecasting of time series data [26]. In studies [24] and [27], LSTM was applied for WPP, which was proven to offer significant advantages compared to the traditional forecasting methods of backpropagation neural network (BPNN) and autoregressive integrated moving average.

LSTM has a memory function for historical data, so the trend of wind power changes (such as rise or fall) could be well served by the LSTM network. However, when the power sequence changes from one trend to another, such as from the rise to fall, certain phase delay errors might be caused by the LSTM network [28].

In recent years, the bidirectional long short-term memory (BLSTM) method has shown significant advantages in the areas of speech recognition [29,30], handwriting recognition [31], and protein structure prediction [32] compared with the LSTM method. For the BLSTM method, both historical and future time series are applied as the input of the model to offset the trend inertia appearing in a single time series direction and to effectively

reduce the phase error of the forecasting. Therefore, BLSTM is chosen as the WPP method for wind farm clusters in this paper.

A novel SFFS-BLSTM WPP model for wind farm clusters is proposed in this paper. Specifically, we make the following two contributions:

- (1) A high-latitude candidate feature set, with more than 300,000 features, for WPP of wind farm clusters is constructed based on feature transformation such as wavelet transformation (WT) and empirical mode decomposition (EMD) transformation, and a novel SFFS method is applied in feature selection for WPP of wind farm clusters;
- (2) Based on the results of feature selection, a short-term WPP model for wind farm clusters, named SFFS-BLSTM, combining SFFS feature selection and BLSTM deep learning, is proposed in the paper, which shows excellent characteristics of reducing prediction errors, especially phase errors.

## 2. The Combination Method of SFFS Feature Selection and BLSTM Deep Learning

2.1. The Overall Flowchart of SFFS-BLSTM

The overall flowchart of the WPP model for wind farm clusters based on SFFS-BLSTM is shown in Figure 1, which is divided into three phases: Phase 1: high-dimensional feature construction for wind farm clusters, Phase 2: feature selection of wind farm clusters based on SFFS, and Phase 3: short-term WPP for wind farm clusters based on BLSTM. These three stages are divided into nine work steps, which are described in detail as follows.

- Step 1: Feature extraction of wind farm clusters. Based on the wind power data and numerical weather prediction (NWP) data of the wind farm cluster, different parameters such as wind speed, wind direction, temperature, pressure, and humidity of each wind farm are extracted, and time-series features and statistical features of wind farm clusters are also constructed.
- Step 2: Feature transformation of wind farm clusters. Based on WT and EMD transformations, the time series features are decomposed into low-frequency and high-frequency components to obtain frequency-domain features. In total, more than 300,000 features are constructed in the paper.
- Step 3: Initial feature ranking based on BIF. The BIF method based on mutual information (MI) is applied to initially rank over 300,000 features [33].
- Step 4: Feature validity verification. Based on the results of the initial feature ranking, the number of input features of the LSTM WPP model is increased in increments of 500 to analyze the change in the WPP accuracy when the number of features increases with the feature ranking results and to initially determine the optimal number of features for WPP.
- Step 5: Feature ranking based on SFFS. Based on the initial feature selection results, the SFFS method is applied to further rank the features selected in step 4.
- Step 6: Feature validity verification. Based on the feature ranking results in step 5, the number of input features of the LSTM WPP model is increased in increments of 20, to analyze the change in the WPP accuracy when the number of features increases with the feature ranking results and to determine the optimal number of features and feature sets for WPP.
- Step 7: Statistical analysis of the selected features. Based on the results of optimal feature selection, statistical analysis is applied to obtain the most important factors affecting the WPP accuracy of wind farm clusters.
- Step 8: Deep learning-based WPP for wind farm clusters. Based on the results of feature selection, LSTM and BLSTM are comparatively applied to carry out WPP for wind farm clusters.
- Step 9: WPP results and error analysis. Based on the WPP results obtained in step 8, the root mean square error (RMSE) of the WPPs and wind power outputs of the WPPs for LSTM and BLSTM are comparatively analyzed to assess the two methods.



Figure 1. Overall flowchart of sequential floating forward selection-bidirectional long short-term memory (SFFS-BLSTM).

2.2. Stage 1: Feature Construction for Wind Farm Clusters

In this paper, three kinds of features are applied as candidate features for feature selection of WPP for wind farm clusters, including original NWP features, frequency domain features, and time-series features.

(1) Original NWP features and corresponding statistical features

The original NWP features of the wind farms applied in the paper are shown in Table 1. There are 11 NWP features for each wind farm, including wind speeds, and

wind directions at four different heights, atmospheric temperature, humidity, and sea-level pressure. Taking a wind farm cluster as an example, which contains 20 wind farms, the number of original NWP features is  $11 \times 20 = 220$ . For each original NWP feature, statistical features of 20 wind farms could be constructed, which reflect the overall output of the wind farm cluster.

Table 1. Original numerical weather prediction (NWP) features of wind farms.

Feature Types	Quantity	Description
Wind speed	4	Wind speed at 170 m, 100 m, 30 m, 10 m
Wind direction	4	Wind direction at 170 m, 100 m, 30 m, 10 m
Temperature	1	Atmospheric temperature
Humidity	1	Atmospheric humidity
Pressure	1	Sea-level pressure

As shown in Table 2, the mean, mode, upper quartile, median, lower quartile, and interquartile range of each original NWP feature of 20 wind farms were constructed.

Table 2. NWP s	statistical features	of wind farm	cluster
----------------	----------------------	--------------	---------

Feature Types	Quantity	Description
Mean	11	Mean of 11 original features for 20 wind farms
Mode	11	Mode of 11 original features for 20 wind farms
Upper quartile	11	Upper quartile of 11 original features for 20 wind farms
Median	11	Median of 11 original features for 20 wind farms
Lower quartile	11	Lower quartile of 11 original features for 20 wind farms
Interquartile range	11	Interquartile range of 11 original features for 20 wind farms

In total, there are  $11 \times 20 + 11 \times 6 = 286$  original NWP features and statistical features of the wind farm cluster containing 20 wind farms.

(2) Time series features

As shown in Figure 2, in addition to the NWP data at the time of WPP, the NWP data 12 h before and after the time to be predicted might also be applied as valid input parameters [34]. If the time interval of the NWP data is 15 min, there are 96 moments within 24 h: 12 h plus 12 h. Therefore, for each moment to be predicted, the valid number of input features is  $286 \times 96 = 27,456$ .



Figure 2. Input data for wind power prediction.

## (3) Frequency domain features

Frequency domain features are constructed based on 286 original NWP features and corresponding statistical features. WT and EMD transformations are applied to obtain 10 new features with different frequency components for each original NWP feature and corresponding statistical features. "db9" is chosen as the mother wavelet. After wavelet transform, the feature sequence is decomposed into four layers. The high-frequency components generated at each layer are named wavelet1, wavelet2, wavelet3, and wavelet4, and the low-frequency component generated at the fourth layer is named wavelet5. The features generated by the EMD transform are named emd1, emd2, emd3, emd4, and emd5 in order of frequency. The frequency ranges of the wavelet and EMD features are shown in Table 3.

Table 3. Frequency ranges o	f wavelet and e	mpirical mode dec	composition (EMD)	features.
			• • • • •	

Feature Type	Quantity	Main Frequency Components
wavelet1	27,456	$>4.55 \times 10^{-5} \text{ Hz}$
wavelet2	27,456	$3.35{\sim}4.55 imes10^{-5}~{ m Hz}$
wavelet3	27,456	$2.25{\sim}3.35 imes10^{-5}{ m Hz}$
wavelet4	27,456	$1.15{\sim}2.25 imes10^{-5}~{ m Hz}$
wavelet5	27,456	$< 1.15  imes 10^{-5}  { m Hz}$
Emd1	27,456	$>1.5 imes10^{-5}\mathrm{Hz}$
Emd2	27,456	$1.5{\sim}1.27 imes10^{-5}~{ m Hz}$
Emd3	27,456	$1.27{\sim}0.7 imes10^{-5}~{ m Hz}$
Emd4	27,456	$0.7{\sim}0.2 imes10^{-5}~{ m Hz}$
Emd5	27,456	$<0.2 imes10^{-5}~{ m Hz}$

After being decomposed by WT and EMD transformation, the 27,456 original features and corresponding features become:  $27,456 + 27,456 \times 10 = 302,016$  features.

## 2.3. Stage 2: Feature Selection Based on SFFS

If more than 300,000 features are applied as the input parameters for WPP, not only is the training of the WPP model difficult but also the computational efficiency and prediction accuracy make it difficult to achieve the ideal situation. Therefore, feature selection is applied to the high-dimensional input features. The heuristic feature selection method based on SFFS, which is developed from the SFS method, is adopted in this paper [35]. Unlike SFS, a feature elimination mechanism is added to the SFFS method. The feature selection and feature elimination are alternated within the SFFS process, which removes redundant features while selecting the effective features, essentially avoiding redundancy among features [20].

The flowchart of the SFFS method is shown in Figure 3, which is divided into five steps. The purpose of the SFFS method is to select a certain number of optimal features from the candidate features and add them to the target feature subset *S*. The number of target features is *d*, and the number of candidate features is *m*.

- Step 1: The optimal number of features, added to the target feature subsets, is determined, named as *L*. *L* is set to be the difference between the number of target features *d* and the number of selected features *n* multiplied by a coefficient, and the coefficient is recommended to be 10%, that is,  $L = (d n) \times 10\%$  [36].
- Step 2: According to the formulated criterion function, which is presented in the second part of Section 2.3 of this paper, *L* features that maximize the criterion function value are selected from the candidate features and added to the target feature subset *S*.
- Step 3: The number of target features and the threshold number of features are compared. If the number of target features reaches the threshold *d*, the loop is stopped, and the target feature subset that meets the requirements is obtained. Otherwise, step 4 is executed.

- Step 4: The optimal number of removing features is determined, named as *R*. *R* is set to be the number of selected features multiplied by a coefficient. The value of the coefficient is recommended to be 10%, that is,  $R = n \times 10\%$  [36].
- Step 5: *R* number of features that minimize the criterion function are selected and removed from the target feature subset *S*, and then step 1 is executed again, and the above steps are looped.



Figure 3. Flowchart of the SFFS algorithm.

The key points of the SFFS method are evaluation index and criterion function [37]. MI is applied as the evaluation index, and the minimum redundancy maximum relevance (mRMR) algorithm is applied as the criterion function, which is presented in detail as follows.

#### (1) Evaluation index

As a result of the strong nonlinear relationship between the output power of the wind farm clusters and the NWP, the MI has a stronger ability to represent the nonlinear correlation than other indicators such as Euclidean distance and consistency measure, so it is selected as the evaluation index for the feature selection method in the paper.

MI is a correlation parameter in information theory [33], which is the amount of information of one random variable contained in another random variable [38]. In other words, MI is a reduction in the uncertainty of a random variable as a result of knowing the laws of another random variable [39].

For example, if there are two random variables, *X* and *Y*, the joint probability distribution of *X* and *Y* is p(x, y). The edge probability distributions of *X* and *Y* are p(x) and p(y). The MI of *X* and *Y*, named as I(X; Y), is defined as the relative entropy of the joint

probability distribution p(x, y) and the edge probability distribution p(x)p(y), which is shown in Equation (1).

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

$$\tag{1}$$

## (2) Criterion function

To make the selected features meet the requirements of both effectiveness and low redundancy, the criterion function of the mRMR is selected as the criterion function of the feature selection method in the paper.

Based on the maximum relevance principle, the average value of the MI between the features, contained in the target feature subset S and wind power P of wind farm clusters, is maximized. The constraint condition based on the maximum relevance is shown in Equation (2).

$$\begin{cases} \max D(S, P) \\ D = \frac{1}{n} \sum_{i=1}^{n} I(v_i, P) \end{cases}$$
(2)

 $I(v_i, P)$  in Equation (2) is the MI between  $v_i$ , the ith feature contained in target feature subset *S*, and wind power *P* of wind farm clusters. There will be redundancy in the target feature subset *S* based on the maximum relevance principle, and there will be a large degree of correlation between features in the set *S*. Therefore, a constraint condition based on the minimum redundancy should be added to the criterion function. The constraint condition based on the minimum redundancy is shown in Equation (3).

$$\begin{cases} \min R(S) \\ R = \frac{1}{C_n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^n I(v_i, v_j) \end{cases}$$
(3)

 $I(v_i, v_j)$  in Equation (3) is the MI between features  $v_i$  and  $v_j$ , the *i*<sup>th</sup> and *j*<sup>th</sup> features contained in *S*. Based on the aforementioned two constraint conditions, the criterion function for feature selection based on the mRMR principle is obtained, as shown in Equation (4).

$$\begin{pmatrix}
\max \Phi(D, R) \\
\Phi(D, R) = D - R
\end{cases}$$
(4)

## 2.4. Stage 3: WPP for Wind Farm Clusters Based on BLSTM

The results of the SFFS feature selection will be applied as input parameters for the WPP model. The deep learning short-term WPP based on BLSTM for wind farm clusters is constructed in the paper. BLSTM networks not only have strong learning and generalization capabilities for massive data but also have strong mapping capabilities for time series data. BLSTM networks offer the advantage of eliminating phase errors [28]. The BLSTM network is developed from LSTM, and the two networks have similarities in structure.

(1) LSTM

LSTM is a deep learning network widely applied in time series data prediction, which is mainly composed of an input layer, hidden layer, and output layer. The structure of LSTM is shown in Figure 4.

In Figure 4, each LSTM unit is a cell with memory function, and the state of the cell at time *t* is recorded as  $c_t$ . The state of the cell at the last moment  $c_{t-1}$  will be inputted into each gate as internal information. The current input  $x_t$  and the output at the last time  $y_{t-1}$  are received by the LSTM unit.  $x_t$  and  $y_{t-1}$  are applied as control information, and  $c_{t-1}$  is modified and updated to get  $c_t$  based on the value of  $x_t$  and  $y_{t-1}$ . Finally, the state  $c_t$  of the



Figure 4. Structure of a single-layer long short-term memory (LSTM) neural network.

(2) BLSTM

One shortcoming of the LSTM model is that only the historical data transmitted from the forward sequence could be applied in the model. For WPP, the output at the time to be predicted is not only causal with historical data but also correlated with future data. Therefore, the predicted value of the future output from the reverse sequence is also critical to the accuracy of WPP. In BLSTM, complementary information from the past and the future are integrated for prediction, based on two independent hidden layers containing the data from both the forward and reverse directions. The structure of BLSTM is shown in Figure 5.



Figure 5. Single-layer BLSTM neural network structure.

(1) Training

The forward and reverse LSTM networks are trained in alternating order. BLSTM could be trained based on a similar algorithm as LSTM. The training process of BLSTM is as follows: For forward networks, the forward and reverse states are processed first, and then the output is calculated. For reverse networks, the output is processed first, and then the forward and backward states are processed. After the forward and backward networks are processed, the weights will be updated [40].

(2) Forecasting

The forward and reverse LSTM networks are applied for WPP in parallel. The prediction flowchart of the BLSTM neural network is shown in Figure 6. In Figure 6, the input data of the testing data set are inputted to the forward and reverse LSTM networks in forward and reverse order respectively. The results of the two sequences are averaged to obtain the hidden layer results of the BLSTM. The results of the BLSTM are inputted to the second hidden layer of the forward and reverse network in the forward and reverse order, respectively. Finally, the output results of each hidden layer are obtained, the results of the last hidden layer are inputted to the output layer, and the results of the BLSTM network are obtained [40].



Figure 6. Prediction flowchart of BLSTM neural network.

## 3. Case Study

In this paper, based on the overall flowchart shown in Figure 1, feature construction, feature selection, and the prediction model based on deep learning are carried out and evaluated with data from industrial applications. The data are from a wind farm cluster in Ning Xia province of China, which contains 20 wind farms. The time range of the data is 660 days, ranging from 1 January 2017 to 1 November 2018. The data for the first 440 days are selected as training data, and the data for the last 220 days are selected as testing data. NWP data are forecast for the next 4 days, with a time-lapse of 15 min. The geographical distribution of the wind farms within the wind farm cluster is shown in Figure 7. The NWP numbers and the installed capacities of the 20 wind farms are shown in Table 4.



Figure 7. Geographical distribution of the 20 wind farms in the Ning Xia province of China.

NWP_NUM	CAP(MW)	NWP_NUM	CAP(MW)
CN0014	49.5	CN0263	99.0
CN0016	99.0	CN0286	99.0
CN0018	94.5	CN0029	48.0
CN0017	102.0	CN0351	172.5
CN0015	90.0	CN0437	99.0
CN0029	79.5	CN0505	198.18
CN0136	99.0	CN0351	150.0
CN0136	69.5	CN0287	100.0
CN0287	148.5	CN0449	96.0
CN0199	297.0	CN0449	97.5

Table 4. Installed capacities of the 20 wind farms.

#### 3.1. Results of Feature Selection

A total of 302,016 features are preliminarily ranked based on BIF. According to the ranking results, the number of input features of the LSTM prediction model is successively increased in increments of 500, and the change rule of RMSE of the WPP model for the next 4 days with the increase in the number of features is analyzed; the results are shown in Figure 8a. As shown in Figure 8a, the WPP error of the wind farm cluster first drops sharply and then rises slowly with the increase in the number of features, and the number of optimal features is about 1000.



**Figure 8.** The curve graph of the wind power prediction (WPP) accuracy as a function of the number of input features.

To determine a more accurate number of optimal features, the number of input features of the LSTM prediction model is successively increased in increments of 20 for the first 2000 features based on the order ranked by BIF, and the change rule of RMSE of the WPP model for the next 4 days with the increase in the number of features is shown in Figure 8b. As seen in Figure 8b, the number of optimal features is 980. The MI values of the first 2000 features are shown in Figure 8c. The MI of the 980th feature is 0.6891, so in the data of the case, the features with MI higher than 0.6891 are effective features that could promote

the WPP accuracy. The dotted green lines in Figure 8a,b show the RMSE of WPP of the wind farm clusters without feature construction and selection. No feature construction and selection means that the raw NWP data of the 20 wind farms are inputted into the WPP model, comprising 220 features of the wind speed and wind direction of 4 different altitudes, sea-level pressure, atmospheric moisture, and temperature corresponding to each wind farm.

The three methods of feature selection—BIF, mRMR, and SFFS—are applied to rank the top 1000 features ranked by BIF. According to the ranking results, the change rules of RMSE of the WPP model for the next 4 days with the increase in the number of features are analyzed, and the results of the change rule are shown in Figure 9.



**Figure 9.** Comparison of the three feature selection methods (**a**) root mean square error (RMSE) of WPP changing with the number of input features; (**b**) mean absolute error (MAE) of WPP changing with the number of input features.

From Figure 9, conclusions could be drawn as follows: (1) At the primary stage of feature selection, the number of selected features is less than 100; the WPP error of the SFFS method declines more rapidly than that of the other two methods, so more effective features were selected by the SFFS method at the primary stage than the other two methods. (2) When the number of features selected by the SFFS method is about 130, the WPP accuracy is higher than that without feature construction and selection. When the number of selected features is about 660, the optimal accuracy is achieved by SFFS, which is 0.37% lower than that with no feature construction and selection. The number of selected features is about 780 by the mRMR method and about 980 by the BIF method when the optimal WPP accuracies are achieved. (3) When the optimal accuracy is achieved, the RMSE of the SFFS method is the lowest, which is 11.99%, followed by the use of the mRMR method, which is 12.05%, and the RMSE of the BIF method is the highest, which is 12.07%.

The RMSEs of WPP with an increase in the number of features corresponding to three feature selection methods by daily statistics are shown in Figure 10.

From Figure 10, conclusions could be drawn as follows: (1) Among the three methods, WPP errors of the SFFS method decline most rapidly, with an increase in the number of features, and the optimal WPP accuracy of the SFFS method is first achieved with minor fluctuations of the curve, so very few redundant features are selected by this method. (2) When the timescales of WPP are different, the numbers of optimal features are different. Based on the SFFS method, the number of optimal features is 720 for the first day, while it is 700 for the second day, 560 for the third day, and 560 for the fourth day. (3) Compared with no feature construction and selection, after feature construction and selection based on the SFFS method, the errors of WPP for different days have different degrees of decline.

0.116 0.132 BIF BIF mRMR mRMR 0.112 0.128 SFFS SFFS RMSE 0.108 BSW8 0.124 No feature selection No feature selection 0.104 0.120.10.116 200 400 600 800 1000200 400 600 800 10000 0 Number of features Number of features (a) RMSE of the WPP for the first day (b) RMSE of the WPP for the second day BIF BIF 0.14 0.148mRMR mRMR SFFS SFFS BSW2 0.136 No feature selection No feature selection 0.140.128 0.136 400 600 200600 800 200 800 1000400 1000 0 Number of features Number of features (c) RMSE of the WPP for the third day (d) RMSE of the WPP for the fourth day

RMSE decreases by 0.33% for the first day, 0.52% for the second day, 0.35% for the third day, and 0.5% for the fourth day.

Figure 10. Comparison of three feature selection methods by daily statistics.

Statistical analysis is carried out for the top 660 features selected by the SFFS method. Some aspects of the features, in terms of their frequency level and the wind farm that they belong to, are enumerated in Table 5.

1170 m wind speedwavelet1Wind farm 10230 m wind speedemd4Wind farm 11310 m wind speedemd2Wind farm 74170 m wind speedemd3Wind farm 19	NUM	NAME	FREQ	WIND FARM	
230 m wind speedemd4Wind farm 11310 m wind speedemd2Wind farm 74170 m wind speedemd3Wind farm 19	1	170 m wind speed	wavelet1	Wind farm 10	
310 m wind speedemd2Wind farm 74170 m wind speedemd3Wind farm 19	2	30 m wind speed	emd4	Wind farm 11	
4 170 m wind speed emd3 Wind farm 19	3	10 m wind speed	emd2	Wind farm 7	
	4	170 m wind speed	emd3	Wind farm 19	
	•••	•••			
660 10 m wind speed emd2 Wind farm 18	660	10 m wind speed	emd2	Wind farm 18	

Table 5. First 660 features selected by the SFFS method.

The statistical results of the valid features, which are grouped into four types—original features, statistical features, WT decomposition features, and EMD decomposition features—are shown in Figure 11.

From Figure 11, the conclusion could be drawn that the original NWP features represent the largest quantity of the selected features, with 283 selected, accounting for 43%, while the other features—WT and EMD decomposition features and statistical features—account for 57%.

The statistical results of the subdivided valid features are shown in Figure 12.



Figure 11. Statistical results of the valid features.



Figure 12. Statistical results of subdivided valid features.

From Figure 12, conclusions could be drawn as follows: (1) In the selected 283 original features, shown in Figure 12a, wind speed features account for the most quantity of valid original features, and 90, 69, 65, and 26 features of wind speed at the height of 170 m, 100 m, 30 m, and 10 m are selected, respectively. A total of 33 other NWP features are selected, much less than the features of wind speed. (2) Among the selected wind speed features, shown in Figure 12a, 90 are wind speeds at the height of 170 m, which account for the largest part, 69 are wind speeds at 100 m, and 65 are wind speeds at 30 m. The

results match the situation that the height of most wind turbine hubs is about 100 m. (3) Among the 100 statistical features shown in Figure 12b, 37 are mean value and 27 are mode value, accounting for the highest number, which reflects the overall situation of the wind farm cluster. Thus, constructing statistical features reflecting the overall situation of the wind farm cluster will contribute to the promotion of the WPP accuracy. (4) Among the five different bands of WT and EMD decomposition features, shown in Figure 12c,d, when the frequency is lower, more features are selected, and when the frequency is higher, fewer features are selected.

## 3.2. Comparison of the WPP Results Based on BPNN, LSTM, and BLSTM

With the same input features, the WPP results of BPNN, LSTM, and BLSTM are compared. The change rule of the RMSEs of the three WPP models for 4 days with the increase in the number of features is shown in Figure 13.



**Figure 13.** Comparison of the WPP RMSEs of backpropagation neural network (BPNN), LSTM, and BLSTM with the increase in the number of features (**a**) RMSE of WPP changing with the number of input features; (**b**) MAE of WPP changing with the number of input features.

From Figure 13, conclusions could be drawn as follows: (1) The optimal number of features of the three WPP models is about 660 for each one. (2) During the increase of the input number of features from 20 to 1000, except for a few dots, BLSTM shows higher WPP accuracy than LSTM. (3) When 660 optimal features are selected, the WPP RMSE of BLSTM is 11.80%, while those of LSTM and BPNN are 11.99% and 12.34%, respectively. There is a 0.19 percentage improvement by BLSTM compared with LSTM, and a 0.54 percentage improvement by BLSTM compared with BPNN.

The one-day-ahead WPP results of LSTM and BLSTM for 750 h, in terms of the wind power curve, are shown in Figure 14. From the details of the WPP results shown in Figure 14b,d, conclusions could be drawn as follows: For the prediction of a wave process of wind power, LSTM is ideal for the prediction of the uphill stage. However, after the crest, the actual wind power trend changes from uphill to downhill, while the uphill inertial of the prediction curve is maintained because of LSTM's memory, leading to a predictive value higher than the actual value at the downhill stage. As predicted from both historical and future directions, BLSTM does not have the problem of advance or lag of wind power crest, so the WPP accuracy is higher than LSTM. In other words, BLSTM shows better performance in terms of reducing the phase errors of WPP than LSTM.



Figure 14. Comparison of WPP results of LSTM and BLSTM (one day in advance, 750 h WPP).

## 4. Conclusions

A short-term WPP method for wind farm clusters based on SFFS feature selection and BLSTM deep learning is presented in this paper and validated with data from 20 wind farms. The conclusions are summarized as follows:

- (1) Based on the data of the wind farm cluster and the 302,016 features in the paper, the feature selection and validation results show that the WPP errors of the wind farm cluster first drop sharply and then rise slowly with the increase in the number of features (Figure 8a). When the timescale of WPP is different, the number of optimal features and the optimal feature sets are different (Figure 10).
- (2) The comparison of BIF-, mRMR-, and SFFS-based feature selection shows that the SFFS method selects more effective features than the other two methods.
- (3) When the number of the features selected by the SFFS method is about 130, the WPP accuracy is higher than that without feature construction and selection. When the number of selected features is about 660, the optimal accuracy is achieved, which is 0.37% lower than that without feature construction and selection (Figure 9). Compared with no feature construction and selection, after feature construction and selection, the errors of different prediction models have different degrees of decline (Figures 8–10).
- (4) The results of statistical analysis of the optimal feature set show that the following features are effective for the overall WPP modeling of wind farm clusters: wind speed of the height of the wind turbine hub, statistical features reflecting the overall situation of the wind farm cluster, low-frequency features in the frequency decomposition features, and so on (Figure 12).
- (5) Based on SFFS feature selection, a short-term WPP model for wind farm clusters based on BLSTM is presented in this paper. The case study demonstrates that BLSTM shows higher WPP accuracy than LSTM (Figure 13). Compared with LSTM, BLSTM can predict from both historical and future directions, which contributes to the outstanding performance of reducing the phase errors (Figure 14).

**Author Contributions:** Conceptualization, X.P. and K.C.; methodology, X.P.; software, K.C.; validation, J.L.; formal analysis, Z.Z.; investigation, T.C.; resources, X.P.; data curation, K.C.; writing—original draft preparation, K.C.; writing—review and editing, X.P.; visualization, K.C.; supervision, S.D.; project administration, X.P.; funding acquisition, X.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key R&D Program of China (Technology and application of wind pow-er/photovoltaic power prediction for promoting renewable energy consumption, 2018YFB0904200) and the Complement S&T Program (Research on short term wind power prediction technology of wind farm clusters based on deep learning methods, DUKZZZ-YBHT-2019-JSC0405-0085) of Inner Mongolia Power (Group) Co., Ltd., Inner Mongolia 015000, China.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

## References

- 1. Wiser, R.; Bolinger, M. 2019 Wind Technologies Market Report; Lawrence Berkeley National Laboratory: Foshan, China, 2020.
- Cutler, N.J.; Outhred, H.R.; MacGill, I.F.; Kepert, J.D. Predicting and presenting plausible future scenarios of wind power production from numerical weather prediction systems: A qualitative ex ante evaluation for decision making. *Wind Energy* 2011, 15, 473–488. [CrossRef]
- 3. Lobo, M.G.; Sanchez, I. Regional wind power forecasting based on smoothing techniques, with application to the Spanish peninsular system. *IEEE Trans. Power Syst.* **2012**, *27*, 1990–1997. [CrossRef]
- 4. Zhang, Y.; Xu, Y.; Zhou, X.; Guo, H.; Zhang, X.; Chen, H. Compressed air energy storage system with variable configuration for accommodating large-amplitude wind power fluctuation. *Appl. Energy* **2019**, *239*, 957–968. [CrossRef]
- 5. Burgas, L.; Colomer, J.; Melendez, J.; Gamero, F.I.; Herraiz, S. Integrated Unfold-PCA Monitoring Application for Smart Buildings: An AHU Application Example. *Energies* **2021**, *14*, 235. [CrossRef]
- 6. Drew, R.; Cannon, J.; Barlow, F.; Coker, J.; Frame, H.A. The importance of forecasting regional wind power ramping: A case study for the UK. *Renew. Energy* 2017, *114*, 1201–1208. [CrossRef]
- 7. Li, Z.; Han, X.; Yang, M.; Ma, Y. Multi-stage power source and grid coordination planning method considering grid uniformity. *Glob. Energy Interconnect.* **2020**, *3*, 303–312. [CrossRef]
- 8. Chen, N.; Wang, Q.; Yao, L.; Zhu, L.; Tang, Y.; Wu, F.; Chen, M.; Wang, N. Wind power forecasting error-based dispatch method for wind farm cluster. *J. Mod. Power Syst. Clean Energy* **2013**, *1*, 65–72. [CrossRef]
- Nedaei, M.; Assareh, E.; Walsh, P.R. A comprehensive evaluation of the wind resource characteristics to investigate the short term penetration of regional wind power based on different probability statistical methods. *Renew. Energy* 2018, 128, 362–374.
   [CrossRef]
- 10. Hong, Y.-Y.; Rioflorido, C.L.P.P. A hybrid deep learning-based neural network for 24-h ahead wind power forecasting. *Appl. Energy* **2019**, *250*, 530–539. [CrossRef]
- 11. Feng, C.; Cui, M.; Hodge, B.-M.; Zhang, J. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. *Appl. Energy* **2017**, *190*, 1245–1257. [CrossRef]
- 12. Ekren, O.; Ekren, B.Y. Size optimization of a PV/wind hybrid energy conversion system with battery storage using simulated annealing. *Appl. Energy* **2010**, *87*, 592–598. [CrossRef]
- 13. Lumbreras, S.; Ramos, A. Offshore wind farm electrical design: A review. Wind Energy 2012, 16, 459–473. [CrossRef]
- 14. Şişbot, S.; Turgut, Ö.; Tunç, M.; Çamdalı, Ü. Optimal positioning of wind turbines on Gökçeada using multi-objective genetic algorithm. *Wind Energy* **2009**, *13*, 297–306. [CrossRef]
- 15. Antal, E.; Tillé, Y. Simple random sampling with over-replacement. J. Stat. Plan. Inference 2011, 141, 597-601. [CrossRef]
- 16. Bayat, A.; Bagheri, A. Optimal active and reactive power allocation in distribution networks using a novel heuristic approach. *Appl. Energy* **2019**, 233–234, 71–85. [CrossRef]
- 17. Zhang, C.; Zhou, J.; Li, C.; Fu, W.; Peng, T. A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting. *Energy Convers. Manag.* **2017**, *143*, 360–376. [CrossRef]
- 18. Ma, Y.; Han, X.; Yang, M.; Lee, W. Multi-timescale robust dispatching for coordinated automatic generation control and energy storage. *Glob. Energy Interconnect.* **2020**, *3*, 355–364. [CrossRef]
- 19. Aydemir, O.; Ergün, E. A robust and subject-specific sequential forward search method for effective channel selection in brain computer interfaces. *J. Neurosci. Methods* **2019**, *313*, 60–67. [CrossRef] [PubMed]
- Choi, K.-S.; Zeng, Y.; Qin, J. Using sequential floating forward selection algorithm to detect epileptic seizure in EEG signals. In Proceedings of the 2012 IEEE 11th International Conference on Signal Processing, Beijing, China, 21–25 October 2012; Volume 3, pp. 1637–1640.
- 21. Ni, K.; Wang, J.; Tang, G.; Wei, D. Research and Application of a Novel Hybrid Model Based on a Deep Neural Network for Electricity Load Forecasting: A Case Study in Australia. *Energies* **2019**, *12*, 2467. [CrossRef]

- 22. Li, D.; Mei, F.; Zhang, C.; Sha, H.; Zheng, J. Self-Supervised Voltage Sag Source Identification Method Based on CNN. *Energies* **2019**, *12*, 1059. [CrossRef]
- 23. Li, C.; Ding, Z.; Yi, J.; Lv, Y.; Zhang, G. Deep Belief Network Based Hybrid Model for Building Energy Consumption Prediction. *Energies* 2018, 11, 242. [CrossRef]
- 24. Delgado, I.; Fahim, M. Wind Turbine Data Analysis and LSTM-Based Prediction in SCADA System. *Energies* 2020, 14, 125. [CrossRef]
- 25. Liu, P.; Zheng, P.; Chen, Z. Deep Learning with Stacked Denoising Auto-Encoder for Short-Term Electric Load Forecasting. *Energies* **2019**, *12*, 2445. [CrossRef]
- Zheng, H.; Yuan, J.; Chen, L. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies* 2017, 10, 1168. [CrossRef]
- He, F.; Zhou, J.; Feng, Z.-K.; Liu, G.; Yang, Y. A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. *Appl. Energy* 2019, 237, 103–116. [CrossRef]
- 28. Huang, C.; Huang, H.; Li, Y. A Bi-Directional LSTM prognostics method under multiple operational conditions. *IEEE Trans. Ind. Electron.* **2019**, *66*, 8792–8802. [CrossRef]
- Graves, A.; Fernández, S.; Schmidhuber, J. Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2005; pp. 799–804.
- Graves, A.; Jaitly, N.; Mohamed, A.-R. Hybrid speech recognition with Deep Bidirectional LSTM. In Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, Olomouc, Czech Republic, 8–12 December 2013; pp. 273–278.
- 31. Frinken, V.; Fischer, A.; Baumgartner, M.; Bunke, H. Keyword spotting for self-training of BLSTM NN based handwriting recognition systems. *Pattern Recognit.* 2014, 47, 1073–1082. [CrossRef]
- 32. Baldi, P.; Brunak, S.; Frasconi, P.; Soda, G.; Pollastri, G. Exploiting the past and the future in protein secondary structure prediction. *Bioinformatics* **1999**, *15*, 937–946. [CrossRef]
- Liu, S.; Shi, R.; Huang, Y.; Li, X.; Li, Z.; Wang, L.; Mao, D.; Liu, L.; Liao, S.; Zhang, M.; et al. A Data-Driven and Data-Based Framework for Online Voltage Stability Assessment Using Partial Mutual Information and Iterated Random Forest. *Energies* 2021, 14, 715. [CrossRef]
- 34. Xu, Q.; He, D.; Zhang, N.; Kang, C.; Xia, Q.; Bai, J.; Huang, J. A Short-Term Wind Power Forecasting Approach with Adjustment of Numerical Weather Prediction Input by Data Mining. *IEEE Trans. Sustain. Energy* **2015**, *6*, 1283–1291. [CrossRef]
- Gacav, C.; Benligiray, B.; Topal, C. Sequential forward feature selection for facial expression recognition. In Proceedings of the 2016 24th Signal Processing and Communication Application Conference (SIU), Zonguldak, Turkey, 16–19 May 2016; pp. 1481–1484.
- Setiawan, D.; Kusuma, W.A.; Wigena, A.H. Sequential forward floating selection with two selection criteria. In Proceedings of the 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Bali, Indonesia, 28–29 October 2017; pp. 395–400.
- 37. Gan, J.Q.; Hasan, B.A.S.; Tsui, C.S.L. A filter-dominating hybrid sequential forward floating search method for feature subset selection in high-dimensional space. *Int. J. Mach. Learn. Cybern.* **2012**, *5*, 413–423. [CrossRef]
- Wu, Z.; Du, X.; Gu, W.; Ling, P.; Liu, J.; Fang, C. Optimal Micro-PMU Placement Using Mutual Information Theory in Distribution Networks. *Energies* 2018, 11, 1917. [CrossRef]
- 39. Iranmanesh, H.; Abdollahzade, M.; Miranian, A. Mid-Term Energy Demand Forecasting by Hybrid Neuro-Fuzzy Models. *Energies* 2011, *5*, 1–21. [CrossRef]
- Ning, Y.; Wu, Z.; Li, R.; Jia, J.; Xu, M.; Meng, H.; Cai, L. Learning cross-lingual knowledge with multilingual BLSTM for emphasis detection with limited training data. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 5615–5619.