

Article



# **Short-Term Wind Speed Forecasting Based on Low Redundancy Feature Selection**

Nantian Huang<sup>1,\*</sup>, Enkai Xing<sup>1</sup>, Guowei Cai<sup>1</sup>, Zhiyong Yu<sup>2</sup>, Bin Qi<sup>1</sup> and Lin Lin<sup>3</sup>

- <sup>1</sup> School of Electrical Engineering, Northeast Electric Power University, Jilin 132012, China; xingenkai@126.com (E.X.); caigw@mail.nedu.edu.cn (G.C.); qibinlehehe@126.com (B.Q.)
- <sup>2</sup> Economic Research Institute, State Grid Xinjiang Electric Power Limited Company, Urumchi 830000, China; yuzhiyongnedu@126.com
- <sup>3</sup> College of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin 132022, China; jllinlin@126.com
- \* Correspondence: huangnantian@126.com; Tel.: +86-135-7855-5676

Received: 27 May 2018; Accepted: 19 June 2018; Published: 22 June 2018



**Abstract:** Wind speed forecasting is an indispensable part of wind energy assessment and power system scheduling. In the modeling of wind speed forecasting, there are problems of insufficiency of the high input feature dimension, weak pertinence of the model and a lack of consideration about the redundancy between features. To address these problems, a short-term wind speed forecast method based on low redundancy feature selection is proposed. Firstly, complementary ensemble empirical mode decomposition (CEEMD) is used to pretreat the wind speed data to reduce the randomness and fluctuation of wind speed data. Secondly, conditional mutual information (CMI) is used to analyze the correlation between the input features on different predicted days and wind speed series. The feature order based on conditional mutual information is used to reduce the redundancy between candidate features and establish subsets with candidate features. After that, according to different candidate feature subsets of different predicted days, the outlier-robust extreme learning machine (ORELM) is used to carry out the forward feature selection and obtain optimal feature subsets for different predicted days. Finally, the optimal prediction model is constructed by using the optimal feature subset and the short-term wind speed forecasting is carried out. The validity and advance of the new method are verified by measured data through comparison experiments.

**Keywords:** wind speed forecasting; low redundancy; feature selection; complementary ensemble empirical mode de-composition

## 1. Introduction

Wind energy is one of the renewable energy sources that could replace fossil fuels. It has grown rapidly in the past decades. By the end of 2015, the cumulative installed capacity of wind energy around the world reached 43.29 GW [1,2]. Large-scale wind energy integration into power grid has brought operational problems because of the randomness and volatility of wind power generation [3,4]. High precision wind power prediction is one of the solutions for optimizing the power reserves, which is used to balance the fluctuations of wind power. With the assistance of accurate wind power forecasting, the stability of power system operation and the adopt capacity for wind power could be improved [5,6].

Wind speed forecasting is one of the basic components of wind power forecasting. The wind speed forecasting for the next 30 min to 6 h can be classified as short-term wind speed forecasting [7], which can meet the need of power producer on managing the grid operations, reducing the negative impact of wind power volatility in power grid operation [8]. The existing wind speed forecasting

model construction can be generally divided into three steps: data preprocessing, feature selection and optimal model construction.

The collected original wind speed data has strong randomness and volatility, the outlier data contained can influence the training effect of the prediction model. The existing research uses signal processing to preprocess original wind speed data, which can reduce the effect of outlier wind speed data on the prediction model [9]. At present, among the signal processing method applied to wind speed data preprocessing. Empirical Mode Decomposition (EMD) has remarkable self-adaptability and suitable for processing nonlinear data, but it is prone to modal-mixing issues [10]. To solve the problem of modal-mixing, Ensemble Empirical Mode Decomposition (EEMD) introduces white noise signals into the original signals, but the decomposition result is contaminated by noise components [11]. Complementary Ensemble Empirical Mode Decomposition (CEEMD) is an improvement based on EEMD, which can address the defects of noise components by counteracting the noise components in the decomposition result with two groups of noise signals which have opposite phase [12]. Therefore, CEEMD can effectively reduce the randomness and volatility of wind speed data under the premise that the wind speed signal is not contaminated by noise components.

Feature selection is an effective approach to reduce the feature dimensions of wind speed forecasting models that can directly improve the prediction accuracy. By introducing as many meteorological factors as possible, the prediction model can reflect the effect of complex external conditions on wind speed to a certain extent. However, too many input features also greatly complicated the model. High redundancy between input features reduced the prediction accuracy and efficiency of the prediction model. To reduce the prediction complexity, the existing researches choose the optimal feature subset through feature selection process [13]. Among the existing feature selection methods, filter methods use some predefined evaluation criteria to evaluate the importance of features for prediction or classification. On this basis, forward feature selection or backward feature selection is carried out through a special feature order obtained by correlation analyzing methods. Filter methods have the characteristics of fast speed and small calculation, which offers strong advantages in engineering applications [14].

The relevance and redundancy between features can directly affect the result of filter feature selection methods. Among the correlation analyzing methods, Mutual Information (MI) can analyze the correlation between feature subset and prediction target, but the result of MI lacks consideration of the redundancy features inside the feature subset [15]. When analyzing the relevance between a feature and the predicted object, Conditional Mutual Information (CMI) also considered the redundancy between the feature and other selected features, so CMI can maintain low redundancy among the feature subset under the premise of ensuring the feature subset is strongly relevant to the predicted object [16]. With the assistance of the feature order obtained by CMI, filter feature selection can effectively improve the prediction accuracy. Meanwhile, existing studies usually analyze the relationship between wind speed and related features based on annual historical data, not considering the correlation between wind speed and complex meteorological factors in different periods of a year, which can't fully meet the needs of the wind speed forecasting in the different periods [17].

The methods of wind speed predictor construction can be divided into physical methods, statistical methods, intelligent methods and hybrid methods [18]. Statistical methods only need historical data to establish the mapping relationship between input features and time series of wind speed, and then carry out the prediction through this mapping relationship [19]. Representative statistical methods include the Kalman filter [20] and autoregressive moving average (ARMA) methods [21]. The intelligent methods can dig into the potential relationship between the input feature and the time series of wind speed through historical data, which is more suitable to deal with complicated relationships than the traditional statistical methods [22]. The intelligent methods include artificial neural networks [23] and Recurrent Neural Networks (RNNs) [24]. The artificial neural network methods [23–25] have advantage in constructing nonlinear prediction models, which have

excellent generalization performance and are suitable for very short-term and short-term wind speed forecasting [26].

Among the artificial neural network methods, Extreme Learning Machine (ELM) has the advantages of extremely fast learning speed and generalization performance compared with traditional neural network, but it easily runs into partial optimization problems [27]. Outlier-Robust Extreme Learning Machine (ORELM) improves the ELM generalization ability by introducing standard parameters, and is more suitable for forecasting wind speeds which have characteristics of high randomness [28].

According to the deficiency of existing methods, a multi-step prediction method based on low redundancy feature selection is proposed. Firstly, the CEEMD method was used to pretreat the training set wind speed data. Then, a low redundancy forward feature selection was conducted based on ORELM and the order of feature importance gained by CMI. Finally, the optimal short-term wind speed forecasting model is constructed with the optimal feature subset to predict the specific period wind speed. The feasibility and effectiveness of the new method are proved through the measurement data of the American wind energy technology center.

#### 2. Structure and Methodology of the New Hybrid Model

The data of training set was pretreated by CEEMD to reduce the effect of outlier data on prediction model. CMI is used to reduce the redundancy of feature selection results. Construct ORELM predictor with low redundancy optimal feature subset to improve the generalization ability of the predictor.

## 2.1. Complementary Ensemble Empirical Mode Decomposition (CEEMD)

CEEMD is an improved method on the basis of EMD and EEMD, the specific iteration of EMD is generated as follows [29]:

- (a) The local maximum and minimum points of the original signal *S* are connected by a cubic spline to obtain the upper envelope  $e_{max}$  and the lower envelope  $e_{min}$ .
- (b) The sequence  $m_1 = [e_{\text{max}} + e_{\text{min}}]/2$  is obtained by averaging two envelopes.
- (c) The first component  $h_1$  is obtained by removing  $m_1$  from *S*:

$$h_1 = S - m_1 \tag{1}$$

- (d) Repeat the above steps with  $h_1$  until the number of extrema and zero crossings is equal or differ at most by one, and the mean value of  $e_{max}$  and  $e_{min}$  is zero. The remaining signal is the first Intrinsic Mode Functions (IMF).
- (e) Remove IMF<sub>1</sub> from the original *S* and repeat the iterations above until the signal cannot be decomposed, the remaining signal is the remainder function.

With the assistance of noise signals, EEMD makes up for the defects of EMD's prone to model-mixing by using the characteristics of the uniform distribution of the noise spectrum [11]. However, the residue noise during the signal reconstruction is difficult to be tolerated, which affects the efficiency of EEMD decomposition. The fallowing improvement has been carried out based on EEMD:

- (a) Two groups of white noise signals with the same amplitude *N* and opposite phase are introduced into the original signal *S* respectively, get two generated signals  $M_1 = S + N$  and  $M_2 = S N$ ;
- (b) Decomposition two groups of sequences with EMD method, obtain two groups of IMFs  $imf_{x1}$  and  $imf_{x2}$ . Then obtain the IMFs of CEEMD by averaging the components of the two groups of IMFs:

$$imf = (imf_{x1} + imf_{x2})/2$$
 (2)

(c) Repeat the above steps with the data which removed the IMFs until the signal cannot be decomposed. The remainder function  $r_n(t)$  is the remainder of the signal, and the final decomposition result of CEEMD is:

$$x(t) = \sum_{i=1}^{n} imf_i(t) + r_n(t)$$
(3)

CEEMD solves the defect of EMD being prone to mode-mixing, and eliminates the effect of white noise induced by EEMD by frequency domain complementation [30].

#### 2.2. Conditional Mutual Information (CMI)

The CMI method can calculate the correlation between the target feature and the predicted target with condition of the lowest redundancy between the target feature and the selected features. The MI method uses the probability density function to define the correlation between variables X and Y, the formula is as follows [31]:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} P(x,y) \log\left(\frac{P(x,y)}{P(x)P(y)}\right)$$
(4)

where P(x) and P(y) represent the marginal probability distribution functions of sample *X* and *Y* respectively. P(x, y) represents the joint probability density function for sample *x* and sample *y*. The larger the MI value of the feature is, the higher the correlation between it and the predicted target [16].

In the condition of a given discrete random variable *Z*, the CMI between *X* and *Y* can be expressed as I(X;Y|Z). In the process of wind speed forecasting, assuming that the original feature set is *V*, the given condition is the selected feature set  $V_j$ , CMI between the target variable *C* and the selected feature  $V_i$  is:

$$I(C; V_i | V_j) = I(C; V_i) - I(C; V_i; V_j)$$
(5)

where  $I(C; V_i)$  represents MI between the target variable *C* and the selected feature  $V_i$  (the correlation between features),  $I(C; V_i; V_j)$  represents that the information overlaps between feature  $V_i$  and feature  $V_i$  meanwhile as the target variable (the redundancy between features).

In conclusion, based on the MI between features and target variables, the CMI method reduces the redundancy between features as another indicator of feature evaluation. It has not only evaluated the contribution of features to the accuracy of prediction model, but also ensured the low redundancy of the corresponding feature arrangement modes. Therefore, CMI can reduce the effect of redundant information between features on feature selection results.

### 2.3. Outlier-Robust Extreme Learning Machine (ORELM)

ELM performs the prediction by minimizing the training error, but it is prone to reduce the generalization performance of the model. ORELM introduces specification parameters *C* to improve ELM generalization ability. ORELM uses  $\ell_1$  – norm to replace  $\ell_2$  – norm, and converts the target function from:

$$\min \|\mathbf{y} - \mathbf{H}\boldsymbol{\beta}\|_2^2 \tag{6}$$

to:

$$\min_{\beta} \|\mathbf{e}\|_1 + \frac{1}{C} \|\boldsymbol{\beta}\|_2^2 \tag{7}$$

where  $\mathbf{e} = \mathbf{y} - \mathbf{HB}$ ,  $\mathbf{y}$  represents the output matrix,  $\beta$  represents the weight matrix between the hidden layer and the output layer.  $\mathbf{H}$  represents the hidden layer output matrix.

When establishing wind speed forecasting model, assuming a data set with *N* training samples  $(\mathbf{x}_i, \mathbf{y}_i), i \in [1, N]$ , where  $\mathbf{x}_i$  represents the input matrix,  $\mathbf{y}_i$  represents the output matrix. Assume that g(x) represents the activation function, and the number of hidden layer nodes is *L*. The iterations of the ORELM algorithm are as follows:

- (a) The implicit layer node parameters, namely the weight matrix  $\mathbf{w}_i$  and the threshold  $b_i$  are randomly generated, where  $i \in [1, L]$ ;
- (b) Calculate the output matrix of the hidden layer [32]:

$$\mathbf{H}(\mathbf{w}_{1},\ldots,\mathbf{w}_{L},\mathbf{x}_{1},\ldots,\mathbf{x}_{N},b_{1},\ldots,b_{L}),$$

$$=\begin{bmatrix}g(\mathbf{w}_{1}\cdot\mathbf{x}_{1}+b_{1})\cdots g(\mathbf{w}_{L}\cdot\mathbf{x}_{1}+b_{L})\\\vdots\\g(\mathbf{w}_{1}\cdot\mathbf{x}_{N}+b_{1})\cdots g(\mathbf{w}_{L}\cdot\mathbf{x}_{N}+b_{L})\end{bmatrix}$$
(8)

- (c) Parameter initialization:  $\mu = 2N/||\mathbf{y}||_1$ , where  $\mu$  represents penalty coefficient,  $\mathbf{e}_1 = 0$ , the Lagrange multiplier  $\lambda_1 = 0$ ;
- (d) This constraint optimization problem is solved by using the augmented Lagrange multiplier (ALM) method, execute the following iteration process until the loop parameter *k* reaches the maximum number of iterations:

$$\begin{cases} \boldsymbol{\beta}_{k+1} = (\mathbf{H}^T \mathbf{H} + 2/C \mu \mathbf{I})^{-1} \mathbf{H}^T (\mathbf{y} - \mathbf{e}_k + \lambda_k / \mu) \\ \mathbf{e}_{k+1} = shrink (\mathbf{y} - \mathbf{H} \boldsymbol{\beta}_{k+1} + \lambda_k / \mu, 1/\mu) \\ \lambda_{k+1} = \lambda_k + \mu (\mathbf{y} - \mathbf{H} \boldsymbol{\beta}_{k+1} - \mathbf{e}_{k+1}) \end{cases}$$
(9)

ORELM avoided solving the sparse matrix by converting the ELM's target function into a manageable convex relaxation problem. In addition, the ALM method is adopted to deal with the convex relaxation problem, and the ability of prediction model to deal with discrete data is strengthened, the generalization performance of ELM is improved [28].

#### 2.4. Forecasting Accuracy Evaluations

For performance evaluation model, using Root-Mean-Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) as the index to evaluate model performance, which are widely used in the wind speed forecasting field. The RMSE and MAPE are calculated as follows:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (x_t - \hat{x}_t)^2}$$
(10)

$$MAPE = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{x_t - \hat{x}_t}{x_t} \right| \times 100\%$$
(11)

where  $\hat{x}_t$  represents the predicted value corresponding to the real wind speed  $x_t$ , T represents the number of wind speed samples.

#### 3. The Composition of Data Sets

#### 3.1. The Features of The Input Set

The wind speed at different time periods of a year has different characteristics and is affected by the surface roughness and air density. Wind shear (1/s) can reflect the surface roughness around the wind tower and the rate of wind velocity at different heights [33]. Temperature (°C), pressure (mBar) and humidity are variables that able to influence the air density and can affect wind speed [34]. Therefore, when building the original feature set, the basic features include wind speed  $x_t$  (m/s), temperature  $T_t$  (°C), relative humidity  $R_t$  (%), absolute humidity  $S_t$  (g/Kg), atmospheric pressure  $P_t$  (mBar) and wind shear  $C_t$  (1/s) of the 16 historical time points (15 min of sampling interval), including the prediction time point t. Meanwhile, the extreme values (max, min), mean, Standard Deviation (std) and Variance (var) of each feature were counted as supplementary features, the final feature dimension is 126. The original historical features and their feature numbers are shown in Table 1. The original statistical features and their feature numbers are shown in Table 2.

<b>Feature Types</b>	<b>Historical Features</b>	Numbers
Wind speed $(x_t)$	$x_{t-1}, x_{t-2}, x_{t-3}, \ldots, x_{t-16}$	1–16
Temperature $(T_t)$	$T_{t-1}, T_{t-2}, T_{t-3}, \ldots, T_{t-16}$	17–32
Relative humidity $(R_t)$	$R_{t-1}, R_{t-2}, R_{t-3}, \ldots, R_{t-16}$	33–48
Absolute humidity $(S_t)$	$S_{t-1}, S_{t-2}, S_{t-3}, \ldots, S_{t-16}$	49-64
atmospheric pressure $(P_t)$	$P_{t-1}, P_{t-2}, P_{t-3}, \dots, P_{t-16}$	65-80
Wind shear $(C_t)$	$C_{t-1}, C_{t-2}, C_{t-3}, \dots, C_{t-16}$	81–96

**Table 1.** The original historical feature set and the features' serial number.

Feature Types	Statistical Features	Numbers
Wind speed $(x_t)$	$\max(x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-16}); \min(x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-16}); \\ \max(x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-16}); \\ \operatorname{std}(x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-16}); \operatorname{var}(x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-16})$	97–101
Temperature ( $T_t$ )	$\max(T_{t-1}, T_{t-2}, T_{t-3}, \dots, T_{t-16}); \min(T_{t-1}, T_{t-2}, T_{t-3}, \dots, T_{t-16}); \\ \max(T_{t-1}, T_{t-2}, T_{t-3}, \dots, T_{t-16}); \\ \operatorname{std}(T_{t-1}, T_{t-2}, T_{t-3}, \dots, T_{t-16}); \\ \operatorname{var}(T_{t-1}, T_{t-2}, T_{t-3}, \dots, T_{t-16}); \\ \operatorname{var}(T_{t$	102–106
Relative humidity $(R_t)$	$\max(R_{t-1}, R_{t-2}, R_{t-3}, \dots, R_{t-16}); \min(R_{t-1}, R_{t-2}, R_{t-3}, \dots, R_{t-16}); \\ \max(R_{t-1}, R_{t-2}, R_{t-3}, \dots, R_{t-16}); \\ \operatorname{std}(R_{t-1}, R_{t-2}, R_{t-3}, \dots, R_{t-16}); \\ \operatorname{var}(R_{t-1}, R_{t-16}, \dots, R_{t-16}); \\ \operatorname{var}(R_{t-1}, R_{t-16}, \dots, R_{t-16}); \\ \operatorname{var}(R_{t-1}, R_{t-16}, \dots, R_{t-16}); \\ \operatorname{var}(R_{t-1},$	107–111
Absolute humidity $(S_t)$	$\max(S_{t-1}, S_{t-2}, S_{t-3}, \dots, S_{t-16}); \min(S_{t-1}, S_{t-2}, S_{t-3}, \dots, S_{t-16}); \\ \max(S_{t-1}, S_{t-2}, S_{t-3}, \dots, S_{t-16}); \\ \operatorname{std}(S_{t-1}, S_{t-2}, S_{t-3}, \dots, S_{t-16}); \\ \operatorname{var}(S_{t-1}, S_{t-2}, S_{t-3}, \dots, S_{t-16}); \\ \operatorname{var}(S_{t$	112–116
atmospheric pressure ( $P_t$ )	$\max(P_{t-1}, P_{t-2}, P_{t-3}, \dots, P_{t-16}); \min(P_{t-1}, P_{t-2}, P_{t-3}, \dots, P_{t-16}); \\ \max(P_{t-1}, P_{t-2}, P_{t-3}, \dots, P_{t-16}); \\ \operatorname{std}(P_{t-1}, P_{t-2}, P_{t-3}, \dots, P_{t-16}); \\ \operatorname{var}(P_{t-1}, P_{t-3}, P_{t-16}, \dots, P_{t-16}); \\ \operatorname{var}(P_{t-1}, P_{t-3}, P_{t-16}, \dots, P_{t-16}); \\ \operatorname{var}(P_{t-1}, P_{t-3}, \dots, P_{t-16}, \dots, P_{t-16}, \dots, P_{t-16}, \dots, P_{t-16}); \\ \operatorname{var}(P_{t-1}, P_{t-1}, P_{t-1}, P_{t-16}, \dots, P_{t-16}, \dots, P_{t-16}, \dots, P_{t-16}, \dots, P_{t-16}, \dots, P_{t-16}); $	117–121
Wind shear ( $C_t$ )	$\max(C_{t-1}, C_{t-2}, C_{t-3}, \dots, C_{t-16}); \min(C_{t-1}, C_{t-2}, C_{t-3}, \dots, C_{t-16}); \\ \max(C_{t-1}, C_{t-2}, C_{t-3}, \dots, C_{t-16}); \\ \operatorname{std}(C_{t-1}, C_{t-2}, C_{t-3}, \dots, C_{t-16}); \\ \operatorname{var}(C_{t-1}, C_{t-3}, \dots, C_{t-16}); \\ \operatorname{var}(C_{t-1}, C_{t-3}, C_{t-3}, \dots, C_{t-16}); \\ \operatorname{var}(C_{t-1}, C_{t$	122–126

Table 2. The original statistical feature set and the features' serial number.

Data were collected using measured data from the National Wind Energy Technology Center (NWTC) M2 wind tower. The geographical location of NWTC and M2 wind tower is shown in Figure 1 [35]. The region is located in north latitude 39.91° and west longitude 105.29°, the measured height of the wind speed and temperature is 80 m. The test environment is a personal computer with 16 GB of memory and the Intel(R) Core(TM) i7-6820hk processor which has an operating frequency of 2.7 GHz. The experimental platform is Matlab (R2016a, MathWorks, Natick, MA, USA). The multi-input and multi-output Strategy (MIMO) is more accurate than the rolling prediction method [7]. Therefore, the new method adopts the MIMO prediction structure, which is constructed with the optimal feature subset obtained by low-redundancy forward feature selection, and builds a four-step wind speed forecasting model that meets the requirements of wind speed forecasting in different time periods (the step length is 15 min).



Figure 1. The geographical location of date measurement in NWTC.

#### 3.2. Data Preprocessing

CEEMD is used to pretreat the signals of wind speed time series. The signal is decomposed into a series of IMFs:  $IMF_1, ..., IMF_n$ . CEEMD parameters are determined by reference to existing literature and statistical experimental conclusions [7], adding a 5 db signal-to-noise ratio white noise to the original signal and the number of iteration is 500. If the high-frequency IMF of wind speed signal fluctuates repeatedly in a very short time, it is indicated that the mode contains more non-cyclical components. Although the volatile component contains a small amount of wind speed information, the large number of outliers in the mode can greatly affect the accuracy and stability of the prediction model. Therefore, CEEMD was used to decompose the wind speed samples to 8 to 10 IMFs, and filter out the highly volatile IMFs, the remaining IMFs reconstructs the new wind speed time series. Thus, the effect of outliers on prediction model is reduced, and the prediction accuracy of model is improved.

To compare the pretreatment impact of CEEMD, EEMD and EMD, the wind speed time series of 31 March 2009 is taken as an example. Figure 2 shows the pretreatment results of three decomposition methods. As shown in Figure 2, the wind speed curve obtained after CEEMD pretreatment reduces the volatility of wind speed, and is more accurately follow the trend of wind speed than other methods in the period from 15:45 to 23:45. Wind speed series after EMD and EEMD pretreatment is difficult to reflect the detailed trend of wind speed.



Figure 2. Wind speed curves reconstructed by different decomposition methods.

Meanwhile, in order to test the effect of pretreatment to wind speed forecasting accuracy, a wind speed forecasting experiment is carried out with predictor based on ORELM. The original feature set is used to construct the input feature set. The annual wind speed sample of 2008 is used as the training set, and the data from 30 March to 5 April 2009 is used as the test set.

It can be seen from Table 3 that the wind speed sequence obtained after pretreatment with CEEMD is forecasted and got the highest the accuracy. Compared with raw data, RMSE decreased 0.8437 m/s and MAPE decreased 13.93%, which proved that CEEMD effectively reduced the effect of outliers on wind speed forecasting and improves the prediction accuracy.

Table 3. Result of multistep forecasting for wind speed with data before and after pretreatment.

Pretreatment Method	EMD	EEMD	CEEMD	Raw Data
MAPE	31.96%	29.79%	24.23%	38.16%
RMSE (m/s)	1.5211	1.4013	1.1465	1.9902

## 3.3. Data Set Construction

The correlation between wind speed and meteorological factors was different in different time of year. In order to meet the requirements of wind speed forecasting model in different time periods, a method of daily low redundancy wind speed features selection is proposed. Expect the data of the predicted day as target date, the data of k days before and after the target date in the former four years is selected as training set, then analyze the correlation between features with this data set. Meanwhile, data of the week before target date is selected as the validation set, which can ensure feature selection process get feature subsets with strong pertinence based on the meteorological characteristics of different forecast periods. The validation set can also the optimal feature set guaranteed to the meet of target date requirements. Figure 3 shows the data set on 6 April 2009 as an example.



Figure 3. Prediction of data set construction on 6 April 2009.

To ensure the new method meets the needs of wind speed forecast at different periods of a year, one day is selected randomly from each quarter of 2009 as the target for prediction (the test set). To test the generalization performance of the new method, new method is used to build the optimal prediction model for every predicted day.

## 3.4. Data Set Determination

During the training set construction, the parameter of k effects the prediction accuracy and efficiency of models directly. To obtain the appropriate value of parameter k, multi-step prediction models was constructed with different parameters k for 364 days (total 52 weeks) in 2009, and series of statistical experiments was carried out. Figure 4 shows the weekly average error curves with different values of parameter k.



Figure 4. Average weekly error curves with different parameter *k*.

It can be seen from Figure 4 that the training time of models increases with the increase of parameter k, and the MAPE elevation of the parameter k in the interval from 30 to 45 is only 0.03%. In order to ensure the efficiency and accuracy of the new method, the data set with the value of parameter k is set as 30.

## 4. Time-Sharing Low Redundancy Feature Selection

#### 4.1. Analysis of Feature Correlation

To analyze the necessity of the feature selection with historical neighborhood data and the analysis advantage of redundancy feature set gained by CMI, the feature correlation on 21 January, 17 April, 18 July and 7 November 2009 was analyzed according to the Data of Adjacent (AD) and the Data of the whole Year (YD). Figure 5 shows the after the normalization analysis results of Pearson Correlation coefficient [36] (PCC), MI and CMI. Table 4 shows numbers of the most important 30-dimensional features.

As can be seen from Figure 5, the importance of the same type of features varies greatly in different dates according to the analysis result with AD dataset. For example, in the analysis results of PCC, feature  $P_t$  is higher than the rest periods on 17 April significantly. Table 4 shows that the same correlation analysis method has different importance order for the same feature in different time dates. The analysis results with YD data sets cannot reflect the differences.



Figure 5. Cont.



**Figure 5.** Results of the features' correlation analysis. (**a**) The correlation analysis results on 21 January; (**b**) The correlation analysis results on 17 April; (**c**) The correlation analysis results on 18 July; (**d**) The correlation analysis results on 7 November; (**e**) The correlation analysis results of the whole year.

Table 4 can be used to compare the redundancy analytical capability between features with different correlation analysis methods. The overlap between features is the redundancy of information, and the redundancy of the same type of features is higher. As can be seen from Table 4, the historical wind speed category features  $x_t$  has the highest frequency.

In the first 30 dimensional feature set on 17 April, the PCC method selected 15 dimensions wind speed category features, and the MI method selected the 14 dimensions wind speed category features, while the CMI method only selected the 11 dimensions wind speed category features. Meanwhile, the CMI method took the standard deviation of absolute humidity (115 dimensional) and the variance of air pressure (121 dimensional) into the first 30 dimensional feature set to replenish the information on humidity and air pressure. Similar phenomena occurred in other periods. It can be seen above that CMI method can improve the information integrity of feature subset.

Periods of Time	Analysis Methods	1–10 Dimensional Features	11–20 Dimensional Features	21–30 Dimensional Features
	PCC	1,2,3,4,81,5,82,97,109,6	83,98,114,7,84,8,107,85,9,10	86,108,11,87,12,88,13,14,89,15
21 January	MI	1,2,3,81,4,97,82,5,109,98	83,6,114,84,7,107,8,85,9,10	86,108,11,87,12,13,88,14,15,89
	CMI	1,2,3,81,4,82,97,5,83,109	98,6,84,114,107,7,121,115,85,8	108,9,86,10,11,87,12,88,13,126
	PCC	1,2,3,4,81,5,97,82,109,6	98,83,7,114,84,8,85,9,107,10	86,108,11,12,87,13,88,14,89,15
17 April	MI	1,2,3,4,81,97,5,82,98,109	83,6,7,84,114,8,107,85,9,108	10,86,11,87,12,88,13,121,115,14
	CMI	1,2,3,81,4,97,82,5,98,109	83,6,84,114,7,115,121,107,85,108	8,9,86,10,87,11,33,34,126,120
	PCC	1,2,3,4,81,82,5,97,109,83	6,98,114,84,7,107,85,8,9,86	108,10,87,11,115,12,88,13,89,14
18 July	MI	1,2,3,81,4,82,97,5,98,109	83,6,84,114,107,7,85,8,86,9	108,10,87,11,121,115,88,12,13,89
	CMI	1,2,3,81,4,82,97,5,98,83	109,6,84,107,114,115,121,7,85,108	8,86,9,10,120,126,87,11,122,116
	PCC	1,2,3,4,97,81,5,109,82,6	98,7,83,114,8,84,9,107,10,85	11,12,86,108,13,14,87,15,16,88
7 November	MI	1,2,3,4,81,97,5,109,82,98	6,83,114,7,107,8,84,9,85,10	11,86,12,108,13,14,87,15,16,88
	CMI	1,2,3,81,4,97,82,5,109,98	83,6,114,84,7,107,8,115,121,108	85,9,10,86,11,12,13,14,87,15
	PCC	1,2,3,4,5,97,81,109,6,82	98,7,83,8,114,84,9,10,85,11	107,12,108,86,13,14,87,15,16,88
YD	MI	1,2,3,4,81,97,5,82,109,98	6,83,7,114,84,8,107,9,85,10	11,108,86,12,13,87,14,88,15,16
	CMI	1,2,3,81,4,97,82,5,98,109	83,6,84,7,114,107,121,115,8,85	9,108,86,10,11,87,12,13,88,14

<b>Table 4.</b> Feature numbers of the most important 30 dimensions
---

#### 4.2. Forward Feature Selection

The reorder feature gained by PCC, MI and CMI are respectively combined with ORELM, ELM and Back-Propagation Neural Network [25] (BPNN) for forward feature selection. Parameters of ELM and BPNN are set according to relevant references [26,27], the specification parameter *C* of ORELM is set to  $2^{-10}$  [28].

To compare the effect of the correlation with AD and YD data sets on predictive models, the feature subset gained with different data sets are combined with the predictors respectively. The forward feature selection is carried out for different forecast days in the condition of different training set and same validation set. MAPE is used to evaluate the prediction accuracy of different method with different feature subset.

Figure 6 shows the error curve of test set in the four predicted days include 21 January, 17 April, 18 July and 7 November 2009, respectively. As shown in Figure 6, the optimal feature subset is determined by the MAPE. Comparing Figure 6a,b, error curves of feature selection with the AD data set converged rapidly and achieved the minimum MAPE, while the error curves of feature selection based on the YD data set converged slowly and the minimum of MAPE is larger.



**Figure 6.** Optimal feature selection curves. (**a**) Analysis curves with AD data; (**b**) Analysis curves with YD data.

Tables 5 and 6 show the prediction effect of predictors based on different optimal feature subsets. As Tables 5 and 6 show, the optimal MAPE of AD-ORELM decreased by an average of 4.8% compared

with YD-ORELM, the optimal MAPE of AD-ORELM decreased by an average of 3.5% compared with YD-ELM, the optimal MAPE of AD-BPNN decreased by an average of 3.7% compared with YD-BPNN, which proved that analyzing feature correlation with AD data can improve the performance of feature selection.

	An alarsta Mathada	AD-O	RELM	AD-I	ELM	AD-BPNN	
Periods of Time	Analysis Methods	MAPE	DIM	MAPE	DIM	MAPE	DIM
	PCC	10.45%	17	11.56%	14	11.39%	103
21 January	MI	10.44%	17	11.33%	21	11.40%	118
	CMI	10.34%	18	11.00%	19	11.03%	116
	PCC	10.75	22	12.19	10	11.93	34
17 April	MI	10.76	23	11.99	18	11.98	90
_	CMI	10.37	23	11.56	21	11.87	30
	PCC	10.20	17	11.73	16	11.88	24
18 July	MI	10.23	19	11.82	15	11.97	106
	CMI	10.08	19	11.37	17	11.44	44
	PCC	11.25	15	12.19	15	12.50	88
7 November	MI	11.20	14	12.68	52	12.76	55
	CMI	11.02	15	12.51	15	12.64	59

Table 5. The optimal feature subsets of different predict days obtained with AD data.

Table 6. The optimal feature subsets of different predict days obtained with YD data.

	Analysis Mathada	YD-OI	RELM	YD-H	ELM	YD-BPNN	
Periods of Time	Analysis Methods	MAPE	DIM	MAPE	DIM	MAPE	DIM
	PCC	11.11%	106	11.62%	125	11.90%	120
21 January	MI	10.93%	113	11.63%	88	11.83%	121
	CMI	10.71%	83	11.51%	109	11.62%	68
17 April	PCC	11.34%	83	12.77%	53	12.46%	67
	MI	11.23%	97	12.72%	20	12.43%	111
	CMI	11.23%	38	12.32%	56	12.38%	61
	PCC	10.99%	22	12.29%	22	12.34%	21
18 July	MI	10.99%	20	12.22%	21	12.29%	29
-	CMI	10.70%	29	12.16%	16	12.24%	27
	PCC	11.52%	81	12.50%	63	12.97%	69
7 November	MI	11.42%	62	12.72%	95	12.90%	57
	CMI	11.37%	94	12.65%	103	12.81%	75

Meanwhile, AD-CMI-ORELM has the highest prediction accuracy in every predicted days, which can be preliminarily proved that the redundancy among the features of the optimal feature subset is reduced by CMI, which improved the prediction accuracy of models.

Table 7 shows types of the optimal features subset obtained by the AD-ORELM method. In combination with Tables 5–7, it can be seen that the redundancy of feature subsets can be reduced after appropriate reduction of similar features. This reduction also improved the prediction accuracy. Meanwhile, adding new types of features with high correlation can enhance the information integrity of the feature subset and further reduce MAPE of the optimal feature subset.

The CMI method can control the redundant between features in four predicted days accurately. In the first three predicted days, the feature subsets obtained by CMI introduces the standard deviation of absolute humidity (115 dimensional) and pressure variance (121 dimensional) in the premise of low redundancy, makes the optimal subset of features obtained by CMI is more advantageous than those from MI and PCC in information integrity, and the smaller MAPE is obtained. On 7 November, CMI ensured that the information integrity in the optimal features subset was achieved, resulting in lower

MAPE than MI and PCC. In this review, it can be proved that using CMI to carry out low redundancy forward feature selection can effectively improve the prediction accuracy and reduce the dimension of feature subsets.

Periods of Time	Analysis Method	Historica	l Features	Statistical Features			
Tenous of Time		$x_t$	C <sub>t</sub>	$x_t$	$R_t$	$S_t$	$P_t$
	PCC	8	4	2	2	1	0
21 January	MI	8	4	2	2	1	0
-	CMI	7	4	2	2	2	1
	PCC	10	6	2	3	1	0
17 April	MI	11	6	2	3	1	0
-	CMI	9	6	2	3	2	1
	PCC	7	5	2	2	1	0
18 July	MI	9	5	2	2	1	0
- 5	CMI	7	5	2	2	2	1
7 November	PCC	8	3	2	1	1	0
	MI	7	3	2	1	1	0
	CMI	7	4	2	1	1	0

Table 7. The feature variety of the optimal feature subset obtained by AD-ORELM method.

#### 5. Predictive Effect and Model Comparison

In order to verify the effectiveness and advancement of the new method, the optimal feature subset selected by the AD data feature selection is combined with different predictors. Meanwhile, to ensure the comparison results have wider ramifications, Classification and Regression Tree [37] (CART) which can automatically complete the feature selection process according to the training set sample and obtain the optimal feature set is used to predict the wind speed samples of different period of time. Figure 7 shows the prediction curves of the AD-CMI-ORELM method by taking the four predicted days as examples.

On 21 January, Figure 6 shows that the range of wind speed is very wide (minimum 1.55 m/s to 26.07 m/s), and wind speed increased from 2.43 m/s to 18.52 m/s rapidly, which brought extremely high requirement to prediction models. In the other three predicted days, the wind speed was less than 10 m/s and there was a plummet in wind speed. In particular, wind speed on 18 July was up to 17 h in a continuous random fluctuation period below 5 m/s. However, it can be seen from Figure 7 that, although situations of the four predicted days brought challenges to the prediction model, the new method can accurately fit the trend of wind speed, which proves the effectiveness and advancement of the new method.

Table 8 shows the prediction error of predictors construct with different optimal feature subsets. It can be seen from Table 8 that, no matter in which predicted day, the MAPE and RMSE generated with the optimal feature subset obtained by CMI are significantly less than the ones based on PCC and MI. This indicates that low redundancy feature subset can effectively improve the prediction accuracy of the model.

Models established with ORELM generated lower MAPE and RMSE, which proves that ORELM introduced the specification parameters *C* and adjusted the target function of ELM effectively reduces the effect of outlier data on the prediction precision and improves the generalization performance of the model.



**Figure 7.** Predicts curves of the proposed method on different predict days (**a**) Forecast curve of 21 January; (**b**) Forecast curve of 17 April; (**c**) Forecast curve of 18 July; (**d**) Forecast curve of 7 November.

The comparison experiment proves that the new method can obtain better forecast results than AD-CART method, which once again proved that the performance of feature selection can be improved by using adjacent samples as validation set and the effectiveness of ORELM as a predictor. AD-CMI-ORELM model also obtained the minimum MAPE and RMSE in each prediction day. Take the experimental result of 21 January as an example. The AD-CMI-ORELM model decreased by 15.81% compared with MAPE of the worst model AD-PCC-ELM, while RMSE decreased by 20.6%. AD-CMI-ORELM was reduced by 4.6% and RMSE by 11.4%, compared with the suboptimal model AD-MI-ORELM. There is a similar increase in the other three predicted days, which proves the effectiveness and advancement of the new method. Meanwhile, due to the frequent and larger fluctuation of wind speed on 21 January, the RMSE was slightly worse than that of the other three predicted days.

However, compared with the other methods, the new method still has the same proportion of improvement, which further proves the effectiveness and advancement of the new method.

In order to further illustrate the effectiveness of the new method, the 7-day data was randomly selected from each season in 2009 to constitute a test set and verify the prediction effect of different models. The average error of each model in different seasons is shown in Table 9.

It can be seen from Table 9 that the error indexes of AD-AMI-ORELM model adopted by the new method still have obvious advantages in the prediction results obtained in all models, and the validity of the new method is proved again.

Bariada of Time Analysis Methods		AD-ORELM			AD-ELM		AD-BPNN		AD-CART				
Periods of Time	Analysis Wethous	RMSE (m/s)	MAPE	DIM	RMSE (m/s)	MAPE	DIM	RMSE (m/s)	MAPE	DIM	RMSE (m/s)	MAPE	DIM
	PCC	1.7284	11.42%	17	1.8966	12.84%	14	1.7655	12.51%	103			
21 January	MI	1.6987	11.33%	17	1.7187	12.52%	21	1.7295	12.44%	118	1.6725	11.03%	32
	CMI	1.5056	10.81%	18	1.6244	12.03%	19	1.7083	12.37%	116			
	PCC	0.5545	9.68%	22	0.5655	11.96%	10	0.5097	11.49%	34			
17 April	MI	0.5118	9.77%	23	0.5401	11.81%	18	0.6424	11.9%	90	0.5261	9.75%	25
	CMI	0.4545	9.17%	23	0.5005	11.43%	21	0.5024	11.41%	30			
	PCC	0.5054	11.54%	17	0.5481	13.22%	16	0.5468	13.43%	24			
18 July	MI	0.5021	11.15%	19	0.5476	13.12%	15	0.5015	12.94%	106	0.4855	11.09%	27
	CMI	0.4141	11.06%	19	0.4603	12.74%	17	0.4439	12.66%	44			
	PCC	0.4561	10.37%	15	0.4211	11.49%	15	0.4972	12.28%	88			
7 November	MI	0.3988	10.45%	14	0.453	11.61%	52	0.4484	11.89%	55	0.372	10.41%	20
	CMI	0.3269	9.72%	15	0.3704	10.95%	15	0.4154	11.52%	59			

 Table 8. Prediction accuracy on different predict days.

**Table 9.** Prediction results on all selected predict days.

Second Analysis Methods		AD-ORELM		AD-E	AD-ELM		AD-BPNN		AD-CART	
Seasons	Analysis Wethous	RMSE (m/s)	MAPE	RMSE (m/s)	MAPE	RMSE (m/s)	MAPE	RMSE (m/s)	MAPE	
	PCC	1.3397	11.23%	1.3878	11.82%	1.397	12.06%			
Spring	MI	1.3474	11.15%	1.3927	11.79%	1.3856	11.92%	1.3482	11.16%	
	CMI	1.2881	10.72%	1.3241	11.36%	1.3328	11.17%			
	PCC	0.5894	11.32%	0.6633	12.45%	0.6648	12.45%			
Summer	MI	0.5911	11.27%	0.6519	12.40%	0.6467	12.49%	0.5872	11.3%	
	CMI	0.5321	10.84%	0.5864	11.86%	0.5942	11.89%			
	PCC	0.5290	10.88%	0.5843	11.97%	0.5798	11.98%			
Autumn	MI	0.5204	10.95%	0.5774	11.96%	0.5760	12.02%	0.4957	10.83%	
	CMI	0.4684	10.25%	0.5175	11.28%	0.4995	11.46%			
	PCC	0.9491	11.63%	0.9825	12.54%	0.9858	12.89%			
Winter	MI	0.9521	11.71%	0.9908	12.75%	0.9832	12.89%	0.943	11.25%	
	CMI	0.9026	10.93%	0.9493	12.02%	0.9375	12.27%			

In order to explain the difference in the prediction performances of different seasons, Table 10 presents the statistical indicators of wind speed in different seasons. As shown in Table 10, the range of wind speed is similar in the four seasons, but according to variance value of the wind speed, wind speed in spring and winter is more volatile and violent. Therefore, the RMSE of the new method in spring and winter is higher than that in summer and autumn.

Feature	Seasons	Range of Variation	Variance Value
	Spring	[0.29, 32.48]	22.34
Wind speed	Summer	[0.29, 31.57]	10.30
	Autumn	[0.25, 30.52]	7.12
	Winter	[0.25, 32.65]	21.76

Table 10. The statistical indicators of wind speed in different seasons.

## 6. Conclusions

To overcome the lack consider of the redundancy between features in the modeling process and the variation of wind speed characteristics at different time periods. This paper makes the following innovations to improve the pertinence and prediction accuracy of wind speed forecasting model:

- (1) A low redundancy feature selection is carried out with the CMI method, which reduced the effect of redundancy between features on prediction accuracy and complexity of a model.
- (2) Building prediction model with ORELM, which improves the generalization performance of the prediction model based on the high prediction efficiency by adding the standard parameter adjustment training error and output layer weight into the target function of ELM.
- (3) The targeted feature selection process and modeling are carried out on different forecast days, overcomes the shortcoming of carrying out feature selection with annual data, which is difficult to show the correlation between wind speed and complex meteorological factors in different time periods, improved the pertinence and prediction of wind speed forecasting model.

The experimental results verified the effectiveness and advancements presented of the new method.

**Author Contributions:** N.H. put forward to the main idea and designed the entire structure of this paper. E.X. and B.Q. did the experiments and prepared the manuscript. G.C. guided the experiments and paper writing. Z.Y. and L.L. completed data preprocessing.

**Funding:** This research was funded by [the National Key Research and Development Program of China] grant number [2016YFB0900104], [the Key Scientific and Technological Project of Jilin Province] grant number [20160204004GX], [Science and Technology Project of Jilin Province Education Department] grant number [JJKH20170219KJ], [Major science and technology projects of Jilin Institute of Chemical Technology] grant number [2018021], [Science and Technology Innovation Development Plan Project of Jilin City] grant number [201750239].

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

## References

- 1. Men, Z.X.; Yee, E.; Lien, F.S.; Wen, D.Y.; Chen, Y.S. Short-term wind speed and power forecasting using an ensemble of mixture density neural networks. *Renew. Energy* **2016**, *87*, 203–211. [CrossRef]
- Ye, L.; Zhao, Y.; Zeng, C.; Zhang, C. Short-term wind power prediction based on spatial model. *Renew. Energy* 2017, 101, 1067–1074. [CrossRef]
- 3. Wang, H.Z. Deep learning based ensemble approach for probabilistic wind power forecasting. *Appl. Energy* **2017**, *118*, 56–70. [CrossRef]
- 4. Jadidoleslam, M.; Ebrahimi, A.; Latify, M.A. Probabilistic Transmission Expansion Planning to Maximize the Integration of Wind Power. *Renew. Energy* **2017**, *114*, 866–878. [CrossRef]

- Yin, H.; Dong, Z.; Chen, Y.; Ge, J.; Lai, L.L.; Vaccaro, A.; Meng, A. An effective secondary decomposition approach for wind power forecasting using extreme learning machine trained by crisscross optimization. *Energy Convers. Manag.* 2017, 150, 108–121. [CrossRef]
- 6. Yuan, X.; Tan, Q.; Lei, X.; Yuan, Y.; Wu, X. Wind Power Prediction Using Hybrid Autoregressive Fractionally Integrated Moving Average and Least Square Support Vector Machine. *Energy* **2017**, *129*, *122–137*. [CrossRef]
- Wang, J.; Song, Y.; Liu, F.; Hou, R. Analysis and application of forecasting models in wind power integration: A review of multi-step-ahead wind speed forecasting models. *Renew. Sustain. Energy Rev.* 2016, 60, 960–981. [CrossRef]
- Ssekulima, E.B.; Anwar, M.B.; Al Hinai, A.; El Moursi, M.S. Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: A review. *IET Renew. Power Gener.* 2016, *10*, 885–989. [CrossRef]
- Zhao, J.; Wang, J.; Liu, F. Multistep Forecasting for Short-Term Wind Speed Using an Optimized Extreme Learning Machine Network with Decomposition-Based Signal Filtering. J. Energy Eng. 2016, 142, 04015036. [CrossRef]
- 10. Guo, Z.; Zhao, W.; Lu, H.; Wang, J. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. *Renew. Energy* **2012**, *37*, 241–249. [CrossRef]
- Chen, D.; Lin, J.; Li, Y. Modified complementary ensemble empirical mode decomposition and intrinsic mode functions evaluation index for high-speed train gearbox fault diagnosis. *J. Sound Vib.* 2018, 424, 192–207. [CrossRef]
- 12. Niu, M.; Wang, Y.; Sun, S.; Li, Y. A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM 2.5, concentration forecasting. *Atmos. Environ.* **2016**, *134*, 168–180. [CrossRef]
- 13. Wang, Y.W.; Feng, L.Z. A new feature selection method for handling redundant information in text classification. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 221–234. [CrossRef]
- 14. Wang, Y.; Feng, L. Hybrid feature selection using component co-occurrence based feature relevance measurement. *Expert Syst. Appl.* **2018**, *102*, 83–99. [CrossRef]
- 15. Lin, Y.; Hu, Q.; Liu, J.; Li, J.; Wu, X. Streaming Feature Selection for Multilabel Learning Based on Fuzzy Mutual Information. *IEEE Trans. Fuzzy Syst.* **2017**, *25*, 1491–1507. [CrossRef]
- 16. Li, S.; Wang, P.; Goel, L. A Novel Wavelet-Based Ensemble Method for Short-Term Load Forecasting with Hybrid Neural Networks and Feature Selection. *IEEE Trans. Power Syst.* **2016**, *31*, 1788–1798. [CrossRef]
- 17. Liu, D.; Wang, J.; Wang, H. Short-term wind speed forecasting based on spectral clustering and optimised echo state networks. *Renew. Energy* **2015**, *78*, 599–608. [CrossRef]
- 18. 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]
- 19. Hu, Q.; Su, P.; Yu, D.; Liu, J. Pattern-Based Wind Speed Prediction Based on Generalized Principal Component Analysis. *IEEE Trans. Sustain. Energy* **2014**, *5*, 866–874. [CrossRef]
- 20. Poncela, M.; Poncela, P.; Perán, J.R. Automatic tuning of Kalman filters by maximum likelihood methods for wind energy forecasting. *Appl. Energy* **2013**, *108*, 349–362. [CrossRef]
- 21. Erdem, E.; Shi, J. ARMA based approaches for forecasting the tuple of wind speed and direction. *Appl. Energy* **2011**, *88*, 1405–1414. [CrossRef]
- 22. Ma, X.; Jin, Y.; Dong, Q. A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting. *Appl. Soft Comput.* **2017**, *54*, 296–312. [CrossRef]
- 23. Gouveia, H.T.; de Aquino, R.R.; Ferreira, A.A. Enhancing Short-Term Wind Power Forecasting through Multiresolution Analysis and Echo State Networks. *Energies* **2018**, *11*, 824. [CrossRef]
- 24. Bonanno, F.; Capizzi, G.; Sciuto, G.L. A neuro wavelet-based approach for short-term load forecasting in integrated generation systems. In Proceedings of the International Conference on Clean Electrical Power, Alghero, Italy, 11–13 June 2013; pp. 772–776.
- Chen, Y.; Cai, K.; Tu, Z.; Nie, W.; Ji, T.; Hu, B.; Chen, C.; Jiang, S. Prediction of benzo[a]pyrene content of smoked sausage using back-propagation artificial neural network. *J. Sci. Food Agric.* 2018, *98*, 3022–3030. [CrossRef] [PubMed]
- 26. Noorollahi, Y.; Jokar, M.A.; Kalhor, A. Using artificial neural networks for temporal and spatial wind speed forecasting in Iran. *Energy Convers. Manag.* **2016**, *115*, 17–25. [CrossRef]

- 27. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]
- 28. Zhang, K.; Luo, M. Outlier-robust extreme learning machine for regression problems. *Neurocomputing* **2015**, 151, 1519–1527. [CrossRef]
- 29. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Math. Phys. Eng. Sci.* **1998**, 454, 903–995. [CrossRef]
- 30. Yeh, J.R.; Shieh, J.S.; Huang, N.E. Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. *Adv. Adapt. Data Anal.* **2010**, *2*, 135–156. [CrossRef]
- 31. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: New York, NY, USA, 1991; pp. 155–183.
- Peng, X.; Zheng, W.; Zhang, D.; Liu, Y.; Lu, D.; Lin, L. A novel probabilistic wind speed forecasting based on combination of the adaptive ensemble of on-line sequential ORELM (Outlier Robust Extreme Learning Machine) and TVMCF (time-varying mixture copula function). *Energy Convers. Manag.* 2017, 138, 587–602. [CrossRef]
- 33. Li, H.; Zhou, M.; Guo, Q.; Wu, R.; Xi, J. Compressive sensing-based wind speed estimation for low-altitude wind-shear with airborne phased array radar. *Multidimens. Syst. Signal Process.* **2016**, *7*, 1–14. [CrossRef]
- 34. Groves, G.V. Determination of Air Density Temperature & Winds at High Altitudes; University Coll: London, UK, 1966.
- 35. NWTC Information Portal. Available online: http://wind.nrel.gov/MetData/ (accessed on 15 June 2018).
- 36. Mu, Y.; Liu, X.; Wang, L. A Pearson's correlation coefficient based decision tree and its parallel implementation. *Inf. Sci.* 2018, 435, 40–58. [CrossRef]
- 37. Huang, N.; Peng, H.; Cai, G.; Chen, J. Power Quality Disturbances Feature Selection and Recognition Using Optimal Multi-Resolution Fast S-Transform and CART Algorithm. *Energies* **2016**, *9*, 927. [CrossRef]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).