Article

Day-Ahead Wind Power Forecasting Based on Wind Load Data Using Hybrid Optimization Algorithm

Guangyu Qin [1,2], **Qingyou Yan** [1,2,*], **Jingyao Zhu** [1,2], **Chuanbo Xu** [1,2,*] **and Daniel M. Kammen** [3]

[1] School of Economic & Management, North China Electric Power University, Beijing 102206, China; qinguangyu@ncepu.edu.cn (G.Q.); zhujingyao@ncepu.edu.cn (J.Z.)
[2] Beijing Key Laboratory of New Energy & Low Carbon Development, North China Electric Power University, Beijing 102206, China
[3] Renewable and Appropriate Energy Laboratory, University of California, Berkeley, CA 94720, USA; kammen@berkeley.edu
[*] Correspondence: yanqingyou@ncepu.edu.cn (Q.Y.); 90102479@ncepu.edu.cn (C.X.)

Abstract: Accurate wind power forecasting is essential to reduce the negative impact of wind power on the operation of the grid and the operation cost of the power system. Day-ahead wind power forecasting plays an important role in the day-ahead electricity spot trading market. However, the instability of the wind power series makes the forecast difficult. To improve forecast accuracy, a hybrid optimization algorithm is established in this study, which combines variational mode decomposition (VMD), maximum relevance & minimum redundancy algorithm (mRMR), long short-term memory neural network (LSTM), and firefly algorithm (FA) together. Firstly, the original historical wind power sequence is decomposed into several characteristic model functions with VMD. Then, mRMR is applied to obtain the best feature set by analyzing the correlation between each component. Finally, the FA is used to optimize the various parameters LSTM. Adding the forecasting results of all sub-sequences acquires the forecasting result. It turns out that the proposed hybrid algorithm is superior to the other six comparison algorithms. At the same time, an additional case is provided to further verify the adaptability and stability of the proposed hybrid model.

Keywords: wind power forecast; variational mode decomposition; maximum relevance & minimum redundancy algorithm; long short-term memory neural network; firefly algorithm; optimization

1. Introduction

Nowadays, the global environment is deteriorating and energy resources are experiencing persistent shortages. The development of renewable energy has become an issue of increasing concern to the international community. The proportion of wind power capacity in the power grid continues to increase. China's wind power has developed rapidly in recent years. According to official data from the National Energy Administration of China; China's grid-connected installed capacity of wind power gradually increased from 2013 to 2019, with a compound annual growth rate of 20.94% [1]. By the end of 2019, the cumulative grid-connected installed capacity was 210 million kWh and the wind power generation was 405.7 billion kWh, accounting for 5.5% of total power generation [2]. In the process of continuous development of wind power, the relationship between wind power and the power grid is getting closer and closer. However; because of the characteristics of intermittent, uncontrollable and volatile of wind energy, there are many difficulties in the grid connection of large-scale wind power. Therefore, it is necessary to adopt an accurate and effective wind power forecasting technology to ensure the security and stability of the power grid. Through short-term wind power forecasting, the electric power department can make timely scheduling plans for wind farm output power

changes in advance, thereby reducing the system's reserve capacity, the operating cost of the power system and the adverse impact on the grid.

According to the literature, the forecasting methods of wind power are mainly divided into two categories. One is the physical method; the Numerical Weather Prediction (NWP) data, as well as related information affecting the surroundings and wind speed are used to establish the correlation model, and the forecasting result is obtained after tedious calculation [3]. However, this method requires high-precision and complete data. The correlation model is relatively rough, and the forecasting accuracy is poor. In addition, it is more suitable for long-term wind power forecasting. The other one is statistical methods. The statistical methods mainly include time series, regression analysis, Kalman filtering, and so on [4–6]. The advantage of the statistical methods is that the forecasting spontaneously adapts to the position of the wind farm to automatically reduce the system error. However, they require long-term measurement data and additional training that refers to testing under various weather conditions and correcting forecasts. Due to the instability and non-linearity of wind energy itself, the predicted results are unsatisfactory.

In recent years, with the continuous deepening of researches, a new branch of statistical methods, artificial intelligence methods, has been developed, which has been successfully used in the forecasting of wind power, and the forecasting results have been recognized [7–9]. These kinds of methods include support vector machine (SVM) [5], long-term and short-term memory (LSTM) model [10], artificial neural network (ANN) [11], and so on. However, as a branch of statistical method, it also has the same disadvantages as the statistical methods. The single artificial intelligence model is hard to master the law of wind power change, so it cannot meet the needs of forecasting accuracy. As the wind power data are random and volatile without stationary sequences, it is necessary to pre-process the data through other ways. Otherwise, it will seriously affect the accuracy of the forecast results.

There are various data pre-processed methods, which can effectively improve the accuracy of forecasting [12]. The decomposition methods, as one of the data pre-processing methods, mainly include the wavelet decomposition [13], empirical mode decomposition (EMD) [14], variational mode decomposition (VMD) [15,16], Fourier decomposition **Error! Reference source not found.**, ensemble empirical mode decomposition (EEMD) **Error! Reference source not found.**, and so on. The decomposition methods can pre-process the complex and changeable original wind power sequence to obtain more regular model characteristics. However, the results obtained by the decomposition have certain shortcomings; for example, decomposition results obtained by the wavelet decomposition method have residual noise. The EMD will produce modal aliasing during the decomposition process, which affects the decomposition performance and reduces the forecasting accuracy **Error! Reference source not found.**. This problem can be solved by EEMD, but its construction components remain noisy. As for the Fourier decomposition method, it has poor adaptability, while the wind power sequence is a non-stationary signal and will be affected by many factors **Error! Reference source not found.**. In contrast, the VMD is a good way to decompose and process the original data, which reduces the non-stationarity of the wind power sequence and improves the anti-interference ability and robustness of the model. Compared with other decomposition models such as EMD, ensemble empirical mode decomposition (EEMD), etc., VMD can decompose the original data into fewer sub-data as long as the appropriate convergence function is selected, thereby reducing the difficulty of modeling **Error! Reference source not found.**. Besides, minimal redundancy and maximal relevance (mRMR) can be used to recognize patterns and select features after decomposition **Error! Reference source not found.**. mRMR has the advantage that it not only considers the correlation between characteristics and target variables, but obtains redundant information among features. In reference **Error! Reference source not found.**, mRMR was applied in wind speed forecasting. In reference **Error! Reference source not found.**, mRMR was utilized to forecast global solar radiation.

Nowadays, more and more researchers tend to use hybrid algorithms to reduce forecasting error by combining the advantages of multiple methods [23–26]. Sun et al. [27] proposed a combination model of EEMDCAN, ARFIMA, and PSOSVM to forecast wind power. The results showed that the EEMDCAN-ARFIMA-PSOSVM hybrid model effectively improved forecasting accuracy. Zhao and Huang [23] used the ultra-short-term power forecasting model of EMD and the support vector regression (SVR) optimized by the simulated annealing (SA) algorithm (EMD-SA-SVR). This combined method had higher forecasting accuracy and stronger forecasting ability, and the optimization time was significantly shorter than other algorithms. Zhang et al. [28,29] employed a singular spectrum analysis (SSA) algorithm to decompose the original wind power sequence data and then optimized the support vector machine (SVM) to predict the wind power through the least square support algorithm. Wang et al. [30] proposed forecasting models integrating the back propagation (BP) algorithm, wavelet decomposition, and SVM, and used Gaussian cloud models to reflect the uncertainty in the forecasting process. The simulation results showed that each forecasting method had its limitations, each which can cause large errors, while the combined forecasting model was significantly better than the single forecasting models. Wang et al. [31] proposed a wind power forecasting method based on SSA, the opposition transition state transition algorithm (OTSTA), Laguerre polynomial, and neural network. By comparing with other popular methods, the results showed that the combination of SSA, OTSTA, and other methods indirectly or directly improved the forecasting accuracy of the model. Lang et al. [32] developed an improved long-term and short-term memory (LSTM) model based on VMD. This combination method had higher forecasting accuracy than other forecasting methods, but still had a relatively large error. Liang [33] used a multi-variable stacking LSTM model to predict uncertain short-term wind speeds, which allowed the ingestion of multiple weather parameters for real-time weather forecasting. Lopez et al. [34] established a LSTM-ESN (Echo State Network) hybrid algorithm, which was an improved LSTM training process to predict wind power and wind generator power.

The LSTM model is the development of Recurrent Neural Networks (RNN). It solves the inherent 'gradient dispersion' problem of RNN in the process of long-term sequences, which greatly improves the time series forecasting capabilities, and realizes a single point worth predicting. This compares favorably with traditional machine learning techniques, such as BP neural network, SVM, and so on, which only regard the wind speed forecasting problem as a static modeling problem. The LSTM model adds a cyclic structure increasing the connection between hidden layers so that it has a strong nonlinear mapping ability and memory function. However, the LSTM model has the drawback that the shadow layer is overloaded, causing it to be low in efficiency [35]. To solve this problem, the forecast models can be optimized by algorithm, which can directly optimize the parameters and improve the search ability of the forecasting models. The forecast model can be optimized by a genetic algorithm **Error! Reference source not found.**, cuckoo search algorithm **Error! Reference source not found.**, firefly algorithm (FA) **Error! Reference source not found.**, and so on. The firefly algorithm (FA) is used in this study to improve efficiency, which is relatively simple without strict continuous, micro or other conditions, and has high calculation efficiency compared with the genetic and cuckoo search algorithms. This paper not only considers the inherent randomness and the uncertainty of wind energy but aims to improve the structure of LSTM.

On this basis, in order to obtain more accurate wind power results, a multi-step hybrid wind power forecasting model is proposed. Firstly, for the purpose of obtaining more regular model features, the VMD is used to decompose the wind power data, which is a new method of signal decomposition. Then the algorithm of mRMR is used to select features. The algorithm considers the correlation and redundancy between features **Error! Reference source not found.**. Finally, the improved LSTM model combined with FA is applied to forecast wind power, which can be used directly to improve the neural structure of the network without any interaction.

The innovations of this article are as follows:

(1) In this paper, we firstly propose an improved LSTM model optimized by the FA to predict wind power. The FA is simple to implement with few parameters. Thus, it has high efficiency that can compensate for the efficiency problem of the LSTM model, which is comparatively inefficient due to the excessive hidden layer load. What's more, the FA is feasible and effective in continuous space and discrete space optimization that can improve forecast accuracy.

(2) A hybrid model VMD-mRMR-LSTM-FA is constructed for the first time: data decomposition preparation, data feature selection, algorithm optimization modeling optimization, and forecasting. This reasonable multi-step method is clear to people at a glance and lays a solid foundation for the following research.

The rest of the paper is organized as followed: Section 2 introduces the theories of methods that are used in the hybrid wind power forecasting model, which includes VMD, mRMR, FA, and LSTM. Section 3 establishes the hybrid forecasting model and there is a case study to test the model. Section 4 compares the model proposed in this paper with SVM, LSTM, FA-LSTM, and mRMR-FA-LSRM methods to prove its effectiveness. In Section 5 there is a further case study to validate the adaptability of the proposed model. The conclusion is presented in Section 6.

2. Theoretical Framework

The second part mainly introduces the methodology used in this paper, mainly introduces the principles and processes of each method.

2.1. Variational Mode Decomposition

VMD is a newly proposed non-recursive technology to decompose a multi-component signal model in several bandwidth-limited modes completely. Unlike EMD, VMD overcomes the shortcomings such as modal aliasing phenomenon and end effect. The specific process is as follows [40–42]:

(1) For each observed signal, assume that it is an independent Gaussian noise superimposed by the original signal. Firstly, perform noise reduction and reconstruction on the sampled signal f_0. The objective function is expressed as:

$$f_0 = f + \eta \tag{1}$$

where f is obtained through the regularization method: $\min_f \left\| f - f_0 \right\|_2^2 + \alpha \left\| \partial_t f \right\|_2^2$.

(2) Calculating the analytical signal of each mode u_k by Hilbert transform. It aims to obtain the unilateral frequency spectrum of the mode components.

(3) Estimating the center frequency by mixing the modal function u_k and the exponent; then the spectrum of each mode is transferred to the respectively estimated center frequency.

(4) The Gaussian smoothing method of the demodulated signal is applied to estimate the bandwidth of each mode function. The variational constraints are followed:

$$\min_{\{u_k\},\{\omega_k\}} \{\sum_{k=1}^{K} \partial_t \, \mathrm{P}[(\delta(t) + \frac{j}{\pi t})u_k(t)]e^{-j\omega_k t} \mathrm{P}_2^2\} \tag{2}$$

$$s.t. \sum_{k=1}^{K} u_k = f(t) \tag{3}$$

where u_k is the subsequence; ω_k is the center frequency; K is the total number of subsequences; ∂_t is Dirac distribution; and $f(t)$ is the original signal.

(5) In order to solve the above constraint problem, the quadratic penalty term $\hat{u}_k^{n+1}$、$\hat{f}(w)$、$\hat{u}(w)$、$\hat{\lambda}(w)$ and Lagrange multiplier λ are introduced, which can be

expressed as:

$$L(\{u_k\},\{\omega_k\},\lambda = \alpha \sum_{k=1}^{K} P\partial_t [(\delta(t) + \frac{j}{\pi t}) \otimes u_k(t)] e^{-jw_k t} P_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle \tag{4}$$

where $\otimes$ is the convolution operator.

(6) Through the alternating direction multiplier method (ADMM), the above functions can be solved by following expressions:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i^n(\omega) + \dfrac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \tag{5}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 \, d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 \, d\omega} \tag{6}$$

where $\hat{u}_k^{n+1}$、$\hat{f}(w)$、$\hat{u}(w)$、$\hat{\lambda}(w)$ are Fourier transformation of u_k^{n+1}、$f(t)$、$u_t(t)$、$\lambda(t)$ respectively.

2.2. Max-Relevance, and Min-Redundancy

The mRMR is a feature selection method based on mutual information, which selects features according to the maximum statistical dependence criterion **Error! Reference source not found.**. Therefore, the mRMR is used to evaluate the features through the mutual information and then find features from the feature space that have the greatest correlation with the target category and the least redundancy among them. The details are as follows:

(1) Defining the maximum correlation and minimum redundancy [44,45]:

$$\max D(S,c), D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) \tag{7}$$

$$\min R(S), R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j) \tag{8}$$

where S is the feature set, c is the target category, and function I represents the mutual information between the two variables: $I(x; y) = \iint \rho(x, y) \log \dfrac{\rho(x, y)}{\rho(x)\rho(y)} dxdy$.

(2) The feature selection criteria of mRMR is as follows:

$$\max \phi(D, R), \phi = D - R \tag{9}$$

$$\max \phi_1(D, R), \phi_1 = \frac{D}{R} \tag{10}$$

(3) Based on the principle of maximum correlation and minimum redundancy, the optimal feature set S_m is selected. Assuming that the feature set S_{m-1} composed of m-1 features has been obtained, then the next feature m can be searched by the operator in the following formula:

$$\max \Delta_{mid}, \Delta_{mid} = \max \{ I(x_j, c) - \frac{1}{m-1} \sum_{x_i \in S_{m-1}} I(x_i, x_j) \} \tag{11}$$

$$\max \Delta_{mid}, \Delta_{mid} = \max\{I(x_j, c) / \{\frac{1}{m-1}\sum_{x_i \in S_{m-1}} I(x_i, x_j)\}\} \tag{12}$$

where x_j is the other features in the original feature set that do not contain the feature quantities in S_{m-1}.

2.3. Firefly Algorithm

The firefly algorithm (FA) was proposed by Yang in 2008. Light intensity and attractive function are important matters in FA [46–48]. The mathematical description and analysis of the algorithm is as follows:

Hypothesis: For any two fireflies, the fireflies with higher brightness will attract the other, but the brightness will become weaker with the distance increase, which also means if the given firefly is the brightest one, it will move randomly. This connects the brightness with the objective attractive function.

(1) Defining brightness function:

$$I(r) = \frac{I_0}{1 + \gamma r^2} \tag{13}$$

where I is the maximum fluorescence emitted by the fireflies; γ is the coefficient of light intensity absorption; r is the distance between the two fireflies.

(2) Defining the attractive function:

$$\beta(r) = \beta_0 e^{-\gamma r^2} \tag{14}$$

where β is the maximum attraction of fireflies.

(3) During the attraction of two fireflies, the position update is expressed by Cartesian distance.

$$r_{ij} = \mathrm{P} x_i - x_j \ \mathrm{P} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{15}$$

Every time the location is updated, follow the formula:

$$x_j = x_i + \beta_0 e^{-\gamma r^2}(x_i - x_j) + \alpha[rand - 0.5] \tag{16}$$

where α is the step factor, and *rand* is a random function of number generated between [0,1].

(4) Designing of fitness function: In most intelligent algorithms, the design of the fitness function directly affects the convergence speed of the algorithm and the choose of optimal solution. Therefore, it is of great significance to design a reasonable fitness function in the intelligent algorithm. The design of its fitness function includes the following key formulas:

(1) The formula of the distance between the individual fireflies and the target point, which can be obtained by Formula (15);
(2) The average absolute error (MAE) between each point;
(3) The root mean square error (RMSE) between each point;
(4) Final construction: $Fitness = \mathrm{MAE} + \mathrm{RMSE}$.

2.4. Long Short-Term Memory

LSTM is proposed by Hochreiter and Schmidhuber [49] to learn long-term dependence information. It is a time recurrent neural network that based on recurrent neural network (RNN). RNN as the traditional neural network is different from the general neural network especially in the way of neuron connection: the information of general neural network flows unidirectional, while the information transmission of RNN has a directional loop **Error! Reference source not found.**. After the improvement, the LSTM replaces the hidden layer neurons of the RNN with memory units, then it introduces "gates" to select and control discarding or adding information. Through the gate structure on the unit state, the neural network can choose to remember or forget information **Error! Reference source not found.**. The control theory of the gate unit is what determines whether

the data is updated or discarded, so that it can solve the problems of gradient disappearance and gradient explosion in the later stage of RNN network training. The sequence principle of LSTM is shown in Figure 1. The core calculation formula of the LSTM model is as follows [52]:

(1)　Defining the function of the input gate, forget gate, output gate respectively:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{17}$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{18}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{19}$$

where x_t is the input data for the current time step; $[h_{t-1}, x_t]$ is a vector composed of two vectors; and σ is the sigmoid activation function.

(2)　In the input gate, the new information will selectively recorded into the cell state. During the process, the target is the memory cell:

$$\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{20}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \tag{21}$$

where the $W_{f,i,c,o}$ and $b_{f,i,c,o}$ are the weight matrices and bias vectors respectively; c_t is the memory cell; and $\tilde{c}_t$ is the element-by-element multiplication symbol between vectors.

(3)　In hidden layer output, the required output value can be determined, which target on h_t.

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \tag{22}$$

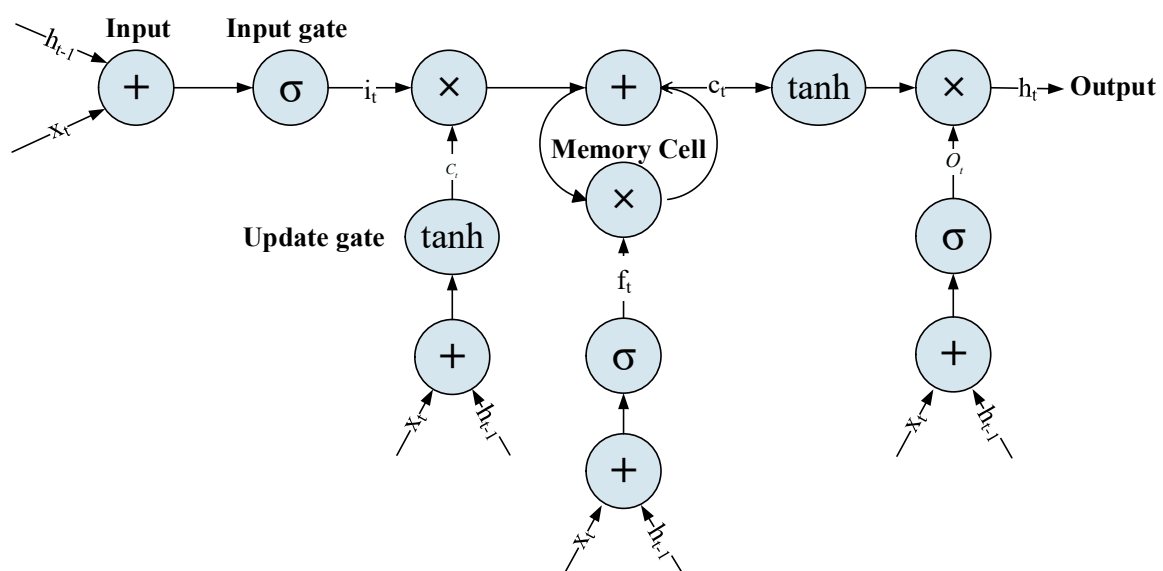where h_t is the hyperbolic tangent nonlinear function.

Figure 1. The sequence principle of long short-term memory (LSTM).

3. Case Study

The third part is data processing and empirical analysis. It uses actual data to verify the hybrid model and compares it with other single models to further verify the accuracy and stability of the hybrid model.

3.1. Wind Power Sequence Decomposition

This paper uses the measured wind power of Beijing Lumingshan Wind Power Plant from 10 May to 28 May 2016 as the research object. The data sampling interval was 5min, and the original wind power sequence was decomposed by VMD. The original sequence is shown in Figure 2. The decomposition results are shown in Figure 3.
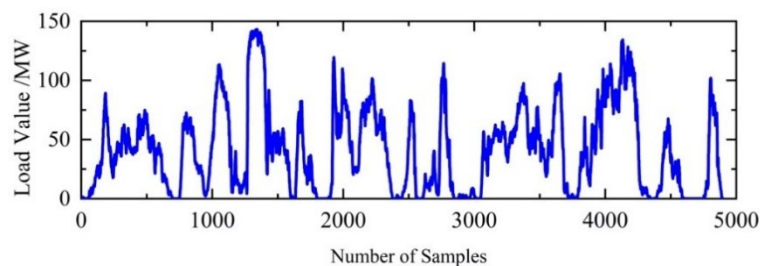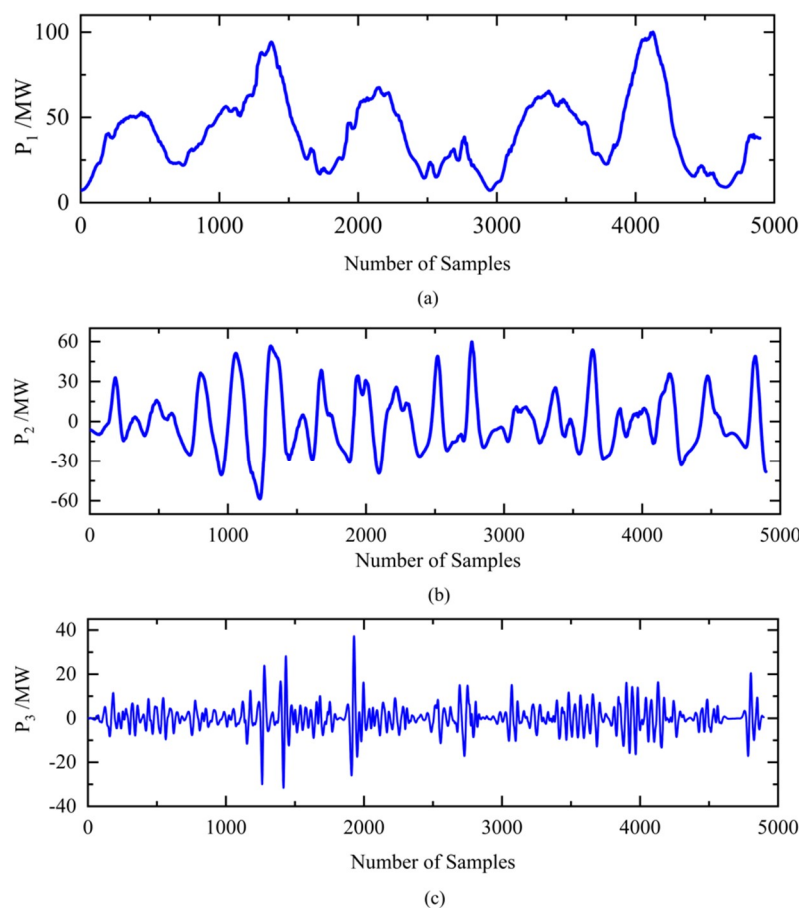

Figure 2. The original load sequence.

(a)

(b)

(c)

Figure 3. (**a**) Modal functions P_1 and (**b**) Modal functions P_2 and (**c**) Modal functions P_3.

It can be seen from Figure 2 that the modal functions P_1 and P_2 have good regularity and obvious periodic correction. Among them, P_1 can represent the long-term change of wind power and P_2 can represent the short-term change. The modal function P_3 has the smallest average amplitude, large fluctuations, and poor regularity, which can indicate the randomness of wind power.

3.2. Finding the Best Feature Set Using mRMR

Based on the results of wind power scenario analysis, screen out the key influencing factors of wind energy to improve the efficiency of load forecasting, which is shown in Table 1.

Table 1. Key influencing factors of wind energy.

Feature	Description
AL_{t-n}	Load at the time period t-n
TP_t	Temperature at the predicted period t
AP_t	Air pressure at the predicted period t
HP_t	Humidity at the predicted period t
WS_t, WD_t	Wheel height wind speed and wind direction at the predicted period t
TS_t, TD_t	10m wind speed and wind direction at the predicted period t
THS_t, THD_t	30m wind speed and wind direction at the predicted period t
FS_t, FD_t	50m wind speed and wind direction at the predicted period t
SS_t, SD_t	70m wind speed and wind direction at the predicted period t

After calculation, the best feature numbers of the three components P1, P2, and P3 are 12, 17, and 18. The corresponding optimal input feature sets are shown in Table 2.

Table 2. Best input features of each component.

Load Component	The Best Input Feature Sets
P1	AL_{t-136}, AL_{t-194}, AL_{t-152}, SS_t, AL_{t-56}, AL_{t-89}, AL_{t-134}, AL_{t-217}, TP_t, AL_{t-205}, AL_{t-26}, WS_t,
P2	AL_{t-112}, AL_{t-148}, THD_t, AL_{t-240}, TS_t, AL_{t-280}, FS_t, AL_{t-260}, AL_{t-103}, SD_t, AL_{t-265}, AL_{t-26}, AL_{t-27}, AL_{t-101}, THS_t, FD_t
P3	AL_{t-239}, AL_{t-50}, FD_t, AL_{t-3}, AL_{t-52}, TS_t, AL_{t-184}, THS_t, AL_{t-178}, AL_{t-257}, FS_t, AL_{t-26}, SS_t, AL_{t-171}, AL_{t-158}, WD_t, TD_t, THD_t

3.3. Load Forecasting Based on FA-LSTM

This paper selects the data from 10 May 2016 to 27 May 2016 as the training set, and uses the remaining data from 28 May 2016 as the test set. On the basis of determining the best input feature set of each component, the firefly algorithm was used to optimize the weights and thresholds of LSTM. The test simulation environment for this paper was Python 3.7. The parameter settings of the firefly algorithm are shown in Table 3.

Table 3. Parameter settings of the firefly algorithm.

Parameter Name	Description	Value
n	Number of individuals	50
k	Number of iterations	100
l_0	Fluorescein initial value	5
η	Fluorescein update rate	0.5
ρ	Volatility coefficient	0.3
β	Update speed of decision domain	0.07
r_s	Individual perception radius	5
L	Moving distance	0.05
P_k	Threshold of individual aggregation	5

During the optimization process of the firefly algorithm, the optimal individual fitness value changes as shown in Figure 4. It can be seen from Figure 4 that the firefly algorithm converges to the optimal fitness value of 0.06 after 60 evolutions in the case of

a population of 50. This shows that the firefly algorithm can find the optimal parameters of the LSTM neural network at a small cost.
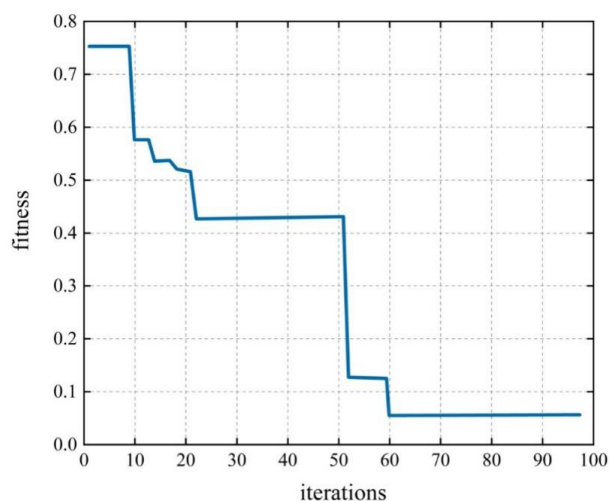

Figure 4. The fitness curve of firefly algorithm.

After the optimization of the firefly algorithm, it was determined that the parameter combination of the LSTM neural network is as follows: the number of hidden layers is 120, the time window step is 6, the number of training times is 160, and the learning rate is 0.015.

The forecasting results of the respective components are summed up to obtain the final forecasting result, as shown in Figure 5.
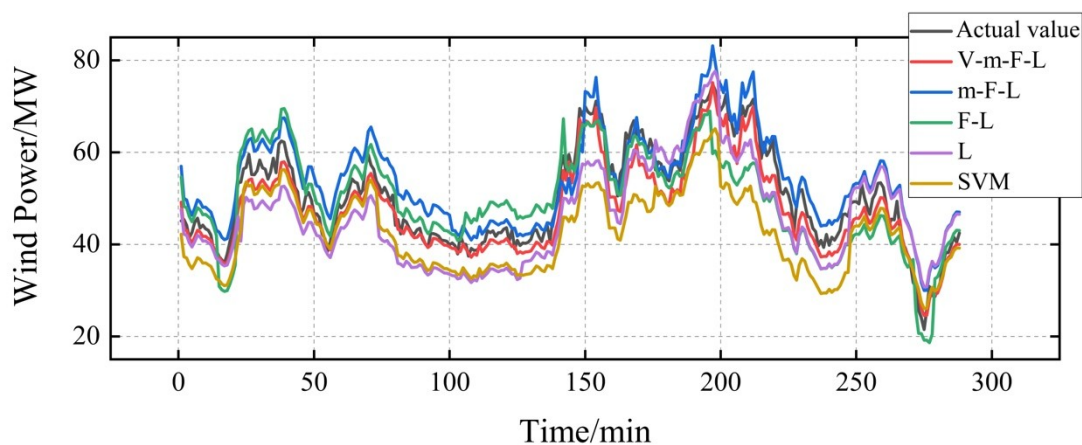
Figure 5. The final forecasting results.

4. Comparison Analysis

In this section, SVM, LSTM, FA-LSTM, mRMR-FA-LSRM will be used to compare with the proposed VMD-mRMR-FA-LSTM in order to verify the forecasting performance of the model proposed in this paper. Table 4 lists the parameter settings and input options of the comparison model.

Table 4. Parameter settings and input option of comparison models.

Comparison Model	Parameter Settings	Input Option
mRMR-FA-LSTM	Same as Section 3.3	After mRMR feature extraction: AL_{t-144}, AL_{t-185}, AL_{t-140}, AL_{t-90}, AL_{t-181}, HP_t, AL_{t-215}, AL_{t-214}, TP_t, AL_{t-217}, AL_{t-150}, THD_t, AL_{t-70}, AL_{t-182}, AP_t, THS_t
FA-LSTM	Same as Section 3.3	AL_{t-1}, AL_{t-48}, AL_{t-96}, AL_{t-192}, AL_{t-240}, TP_t, AP_t, HP_t, WS_t, WD_t, TS_t, TD_t, THS_t, THD_t, FS_t, FD_t, SS_t, SD_t
LSTM	Same as Section 3.3	AL_{t-1}, AL_{t-48}, AL_{t-96}, AL_{t-192}, AL_{t-240}, TP_t, AP_t, HP_t, WS_t, WD_t, TS_t, TD_t, THS_t, THD_t, FS_t, FD_t, SS_t, SD_t
SVM	Penalty parameters:30 Width parameter of kernel function:60	AL_{t-1}, AL_{t-48}, AL_{t-96}, AL_{t-192}, AL_{t-240}, TP_t, AP_t, HP_t, WS_t, WD_t, TS_t, TD_t, THS_t, THD_t, FS_t, FD_t, SS_t, SD_t

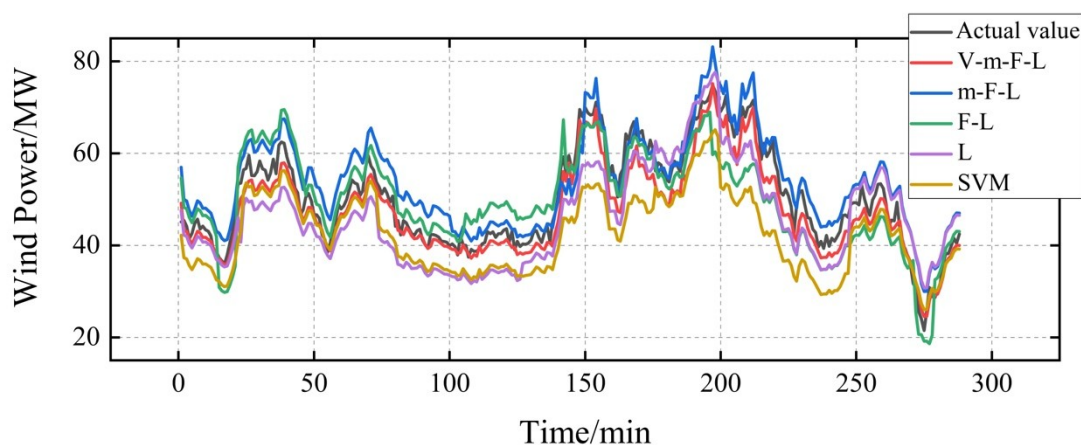The forecasting results of each model are shown in Figure 6.

Figure 6. Load forecasting curves of each model.

Where V-m-F-L, m-F-L, F-L, L refer to VMD-mRMR-FA-LSTM, mRMR-FA-LSTM, FA-LSTM and LSTM, respectively.

It can be seen from Figure 6 that the VMD-mRMR-FA-LSTM short-term load forecasting model can better approximate the true value and have better forecasting accuracy.

In order to quantitatively analyze the forecasting accuracy of each model, five evaluation indicators were introduced for forecasting results, that is the coefficient of determination (R^2), mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE). This can be defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(\hat{X}(i) - X(i))^2}{\sum_{i=1}^{m}(X(i) - \bar{X}(i))^2} \tag{23}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|\hat{X}(i) - X(i)| \tag{24}$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{X}(i) - X(i))^2} \tag{25}$$

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}|\frac{\hat{X}(i) - X(i)}{X(i)}| \times 100\% \tag{26}$$

$$TIC = \frac{\sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{X}(i)-X(i))^2}}{\sqrt{\frac{1}{m}\sum_{i=1}^{m}X(i)^2}+\sqrt{\frac{1}{m}\sum_{i=1}^{m}\hat{X}(i)^2}} \tag{27}$$

where m is the number of data points; $\overline{X}(i)=\frac{1}{m}\sum_{i=1}^{m}X(i)$. The smaller the values of RMSE, MAE, and MAPE are, the more accurate the forecasting result is. R^2 can measure the regression fitting effect of the model; the larger R^2 is, the better the fitting effect of the model will be. TIC can measure the predictive ability of the model.

Table 5 lists the comparison of R^2, MAE, RMSE, MAPE, and TIC of the training results and the calculation time of each model.

Table 5. The comparisons of the related forecasting results for power load.

	Time(s)	R^2	MAE	RMSE	MAPE	TIC
VMD-mRMR-FA-LSTM	263.4638	0.9578	2.9596	3.0435	0.0569	0.0305
mRMR-FA-LSTM	193.3369	0.8958	3.2188	3.2950	0.0673	0.0311
FA-LSTM	241.3697	0.9014	3.0929	3.1483	0.0635	0.0306
LSTM	30.8964	0.7618	5.2297	4.2191	0.0979	0.0431
SVM	28.1637	0.7023	7.0238	7.4105	0.1345	0.0695

Obviously, from the perspective of R^2, MAE, RMSE, MAPE and TIC, the forecasting results of the model established in this paper are optimal, which are equal to 0.9578, 2.9596, 3.6435, 0.0569, and 0.0365, respectively.

Compared with mRMR-FA-LSTM, after using VMD for data decomposition, the combined model established increased by 6.9%, 8.8%, 8.3%, 18.3%, and 19.7% on the five indicators mentioned above, respectively. Compared with the FA-LSTM model, the mRMR-FA-LSTM model decreased 0.6%, 4.1%, 4.7%, 5.9%, and 1.6% on the five indicators, respectively. Comparing the FA-LSTM model and the LSTM model, it can be found that the firefly algorithm improved the forecasting results on the five indicators by 18.3%, 69.1%, 34.0%, 54.1%, and 40.8%, respectively. Comparing the LSTM model with the classic SVM model, we found that the forecasting results of LSTM improved by 8.5%, 34.3%, 75.6%, 37.4%, and 61.2% on the five indicators, respectively.

The running time of FA-LSTM and mRMR-FA-LSTM was 241.3697 s and 193.3369 s. It can be seen that in this training, the application of mRMR reduced the training time by 24.8% compared to FA-LSTM.

Finally, based on the above, we can summarize the following points:

(1) In terms of forecasting accuracy, both the VMD decomposition and the firefly algorithm effectively improve the forecasting accuracy of the model. At the same time, LSTM has a greater advantage in wind power forecasting than the traditional SVM model. It can be found that after using mRMR for feature selection, the forecasting accuracy of FA-LSTM will slightly decrease because the indicators selected in the cases of this paper were chose based on experience.

(2) In terms of computing efficiency, the application of mRMR, which can perform feature extraction with maximum correlation and minimum redundancy, significantly accelerates the calculation speed of the model and reduces the calculation scale, thereby improving forecasting efficiency.

(3) After the FA algorithm is used to optimize the LSTM model, the influence of the initial value selection is reduced. The initial parameters of the model can be set more flexibly, which avoids the shortage of manual selection of model parameters.

5. Further Case Study

For the purpose of verifying the adaptability of the hybrid model, another case is included which selects additional load data (from 10 June 2016 to 16 June 2016) from Lumingshan Wind Power Plant. The training error of each model is shown in Figures 7–11.
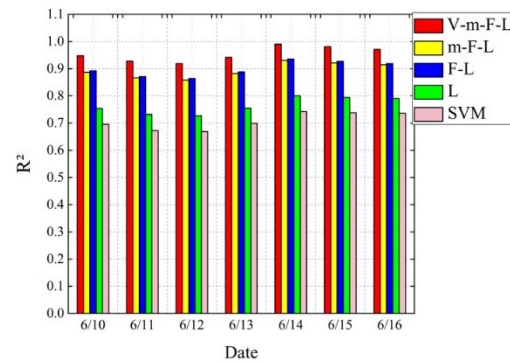

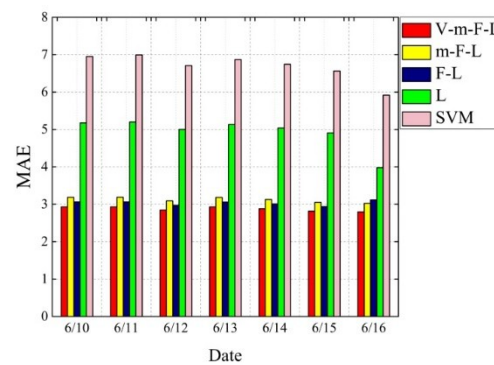**Figure 7.** R^2 of forecasting models.
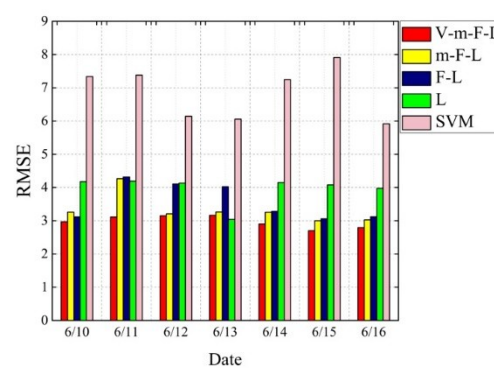

Figure 8. MAE of forecasting models.

Figure 9. RMSE of forecasting models.
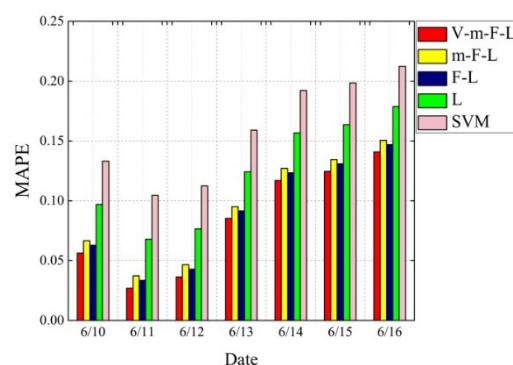
Figure 10. MAPE of forecasting models.
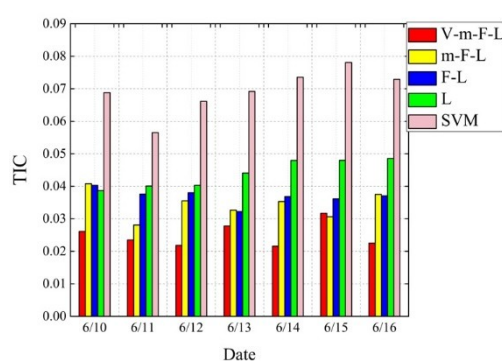
Figure 11. TIC of forecasting models.

Noted: V-m-F-L, m-F-L, F-L, L refer to VMD-mRMR-FA-LSTM, mRMR-FA-LSTM, FA-LSTM, and LSTM, respectively

As shown in the Figure 11, the forecasting results obtained by the model VMD-mRMR-FA-LSTM established in this paper are the best in the five parity indicators compared to the other four comparative models. It can be seen that the VMD-mRMR-FA-LSTM model can predict wind power with high accuracy.

6. Conclusions

Aiming at the instability of day-ahead wind power forecasting, a hybrid short-term wind power forecasting model, namely the VMD-mRMR-FA-LSTM model, is proposed. Firstly, the VMD decomposes the original load, and then mRMR is applied to obtain the best feature set by analyzing the correlation between each component and the features, including temperature, wind speed, wind direction, and so on. Secondly, different LSTM forecasting models for each new sequence based on the mRMR selection result are constructed. Finally, the FA is used to optimize the parameters of LSTM. The case study of the proposed hybrid model shows that:

(1) The hybrid model has higher forecasting accuracy than those of benchmarking models, and has broad application prospects in day-ahead wind power forecasting.

(2) Compared with the single LSTM model, FA can optimize the parameters and function of LSTM to obtain higher forecasting accuracy, which indicates that FA-LSTM model has stronger global search ability and more stable forecasting performance.

(3) Compared with other data preprocessing strategy, VMD-mRMR has better performance and effective improves the forecasting accuracy. The hybrid model proposed in this paper can be well applied to day-ahead wind power forecasting.

Abbreviations

ANN	Artificial neural network
ARFIMA	Autoregressive fractionally integrated moving average
BP	Back propagation
EMD	Empirical mode decomposition
EEMD	Ensemble empirical mode decomposition
EEMDCAN	EEMD with complementary adaptive noise
ESN	Eco state network
FA	Firefly algorithm
LSTM	Long short-term memory neural network
MAPE	Mean absolute percentage error
MAE	Mean absolute error
mRMR	Maximum relevance & minimum redundancy
NWP	Numerical weather prediction
OTSTA	Opposition transition state transition algorithm
PSO	Particle swarm optimization
R^2	Determining factor
RNN	Recurrent neural network
RMSE	Root mean square error
SVM	Support vector machine
SA	Simulated annealing
SSA	Singular spectrum analysis
TIC	Predictive ability
VMD	Variational mode decomposition

References

1. National Energy Administration. China Power Industry Annual Development Report: 2019 Wind Power Grid Operation. Available online: http://www.nea.gov.cn/2020-02/28/c_138827910.htm (accessed on 28 February 2020).
2. Yang, J.B.; Liu, Q.Y.; Li, X.; Cui, X.D. Overview of Wind Power in China: Status and Future. *Sustainability* **2017**, *9*, 1454.
3. Lange, M.; Focken, U. *Physical Approach to Short-Term Wind Power Prediction*; Springer: Berlin, Germany, 2006.
4. Lahouar, A.; Slama, J.B.H. Hour-ahead wind power forecast based on random forests. *Renew. Energy* **2017**, *109*, 529–541.
5. Jiang, P.; Yang, H.; Heng, J. A hybrid forecasting system based on fuzzy time series and multi-objective optimization for wind speed forecasting. *Appl. Energy* **2019**, *235*, 786–801.
6. Heydari, A.; Nezhad, M.M.; Pirshayan, E.; Garcia, D.A.; Keynia, F.; De Santoli, L. Short-term electricity price and load forecasting in isolated power grids based on composite neural network and gravitational search optimization algorithm. *Appl. Energy* **2020**, *277*, 115503.
7. Zhen, H.; Niu, D.X.; Yu, M.; Wang, K.K.; Liang, Y.; Xu, X.M. A Hybrid Deep Learning Model and Comparison for Wind Power Forecasting Considering Temporal-Spatial Feature Extraction. *Sustainability* **2020**, *12*, 9490.

8. Chen, K.S.; Lin, K.P.; Yan, J.X.; Hsieh, W.L. Renewable Power Output Forecasting Using Least-Squares Sopport Vector Regression and Google Data. *Sustainability* **2019**, *11*, 3009.

9. Zhou, J.G.; Yu, X.C.; Jin, B.L. Short-Term Wind Power Forecasting: A New Hybrid Model Combined Extreme-Point Symmetric Mode Decomposition, Extreme Learning Machine and Particle Swarm Optimization. *Sustainability* **2018**, *10*, 3202.

10. Zhou, J.; Shi, J.; Li, G. Fine tuning support vector machines for short-term wind speed forecasting. *Energy Convers. Manag.* **2011**, *52*, 1990–1998.

11. Shi, X.Y.; Lei, X.W.; Huang, Q.; Huang, S.Z.; Ren, K.; Hu, Y.Y. Hourly day-ahead wind power prediction using the hybrid model of variational model decomposition and long short term memory. *Energies* **2018**, *11*, 3227.

12. Amjady, N.; Abedinia, O. Short Term Wind Power Prediction Based on Improved Kriging Interpolation, Empirical Mode Decomposition, and Closed-Loop Forecasting Engine. *Sustainability* **2017**, *9*, 2104.

13. Li, G.; Shi, J. On comparing three artificial neural networks for wind speed forecasting. *Appl. Energy* **2010**, *87*, 2313–2320.

14. De Giorgi, M.G.; Campilongo, S.; Ficarella, A.; Congedo, P.M. Comparison between wind power prediction models based on wavelet decomposition with least-squares support vector machine (LS-SVM) and artificial neural network (ANN). *Energies* **2014**, *7*, 5251–5272.

15. Liu, H.; Tian, H.Q.; Liang, X.F.; Li, Y.F. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. *Appl. Energy* **2015**, *157*, 183–194.

16. Han, L.; Jing, H.T.; Zhang, R.C.; Gao, Z.Y. Wind power forecast based on improved Long Short Term Memory network. *Energy* **2019**, *189*, 116300.

17. Sang, B.Y.; Wang, D.S.; Yang, B.; Ye, J.L.; Tao, Y.B. Energy storage optimal allocation method for smoothing new energy output fluctuation. *Proc. CSEE* **2014**, *34*, 3700–3706.

18. Wang, C.; Zhang, H.L.; Fan, W.H.; Ma, P. A new chaotic time series hybrid prediction method of wind power based on EEMD-SE and full-parameters continued fraction. *Energy* **2017**, *138*, 977–990.

19. Ma, X.B. Short-term wind power prediction based on wavelet analysis and BP neutral network. *J. Electr. Power Sci. Technol.* **2015**, *30*, 92–97.

20. Niu, M.F.; Hu, Y.Y.; Sun, S.L.; Liu, Y. A novel hybrid decomposition-ensemble model based on VMD and HGWO for container throughput forecasting. *Appl. Math. Model.* **2018**, *57*, 163–178.

21. Liang, Y.; Niu, D.X.; Hong, W.C. Short term load forecasting based on feature extraction and improved general regression neural network model. *Energy* **2019**, *166*, 653–663.

22. Wang, Q.; Guan, T.S.; Qin, B.S. Short-term wind speed forecasting of ORELM based on MRMR. *Renew. Energy Resour.* **2018**, *36*, 85–90.

23. Bouzgou, H.; Gueymard, C.A. Minimum redundancy—Maximum relevance with extreme learning machines for global solar radiation forecasting: Toward an optimized dimensionality reduction for solar time series. *Solar Energy* **2017**, *158*, 595–609.

24. Zhao, H.R.; Zhao, H.R.; Guo, S. Short-Term Wind Electric Power Forecasting Using a Novel Multi-Stage Intelligent Algorithm. *Sustainability* **2018**, *10*, 881.

25. Wu, Q.L.; Lin, H.X. Short-Term Wind Speed Forecasting Based on Hybrid Variational Mode Decomposition and Least Squares Support Vector Machine Optimized by Bat Algorithm Model. *Sustainability* **2019**, *11*, 652.

26. Yang, W.; Wang, J.; Niu, T.; Du, P. A hybrid forecasting system based on a dual decomposition strategy and multi-objective optimization for electricity price forecasting. *Appl. Energy* **2019**, *235*, 1205–1225.

27. Sun, H.R.; Zhang, G.; Wang, R.J. Short-term wind power prediction based on combinatorial optimization algorithm. *J. North China Electr. Power Univ.* **2020**, *47*, 33–42.

28. Zhao, Q.; Huang, J.T. On ultra-short-term wind power prediction based on EMD-SA-SVR. *Power Syst. Prot. Control* **2020**, *48*, 89–96.

29. Zhang, Y.C.; Le, J.; Liao, X.B.; Zheng, F.; Li, Y.H. A novel combination forecasting model for wind power integrating least square support vector machine, deep belief network, singular spectrum and analysis and locality-sensitive Hashing. *Energy* **2019**, *168*, 558–630.

30. Wang, H.Z.; Li, G.Q.; Wang, G.B.; Peng, J.C.; Jiang, H.; Liu, Y.T. Deep learning based ensemble approach for probabilistic wind power forecasting Physical modeling methods. *Appl. Energy* **2016**, *188*, 56–70.

31. Wang, C.; Zhang, H.L.; Ma, P. Wind power forecasting based on singular spectrum analysis and a new hybrid Laguerre neural network. *Appl. Energy* **2019**, *259*, 114–139.

32. Lang, W.M.; Ma, X.J.; Zhou. B.W.; Yang, D.S.; Luo, Y.H.; Liu, L.Q. Wind power probabilistic intervals prediction based on LSTM and nonparametric kernel density Estimation. *New Energy* **2020**, *48*, 31–38.

33. Liang, S.; Nguyen, L.; Jin, F. A multi-variable stacked long-short term memory network for wind speed forecasting. IEEE international conference on big data (big data). In Proceedings of the IEEE International Conference on Big Data (Big Data) 2018, Seattle, WA, USA, 10–13 December 2018; pp. 4561–4564.

34. Lopez, E.; Valle, C.; Allende, H.; Gil, E.; Madsen, H. Wind power forecasting based on echo state networks and long short-term memory. *Energies* **2018**, *10657*, 347–355.

35. Bao, Z.Q.; Zhao, Y.; Hu, X.T.; Zhao, Y.Y.; Huang, Q.D. Minimal peephole long short-term memory. *Comput. Eng. Des.* **2020**, *41*, 134–138.

36. Flores, J.J.; Graff, M.; Rodriguez, H. Evolutive design of ARMA and ANN models for time series forecasting. *Renew. Energy* **2012**, *44*, 225–230.

37. Zhao, J.; Guo, Z.H.; Su, Z.Y.; Zhao, Z.Y.; Xiao, X.; Liu, F. An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed. *Appl. Energy* **2016**, *162*, 808–826.
38. Wang, J.; Liu, G.Y. Firefly algorithm based on dynamic step change. *Comput. Eng. Des.* **2019**, *40*, 1001–1007.
39. Ju, Z.; He, J.J. Prediction of lysine glutarylation sites by maximum relevance minimum redundancy feature selection. *Anal. Biochem.* **2018**, *550*, 1–7.
40. Li, H.; Zhang, R.C.; Wang, X.S.; Bao, A.C.; Jing, H.T. Multi-step wind power forecast based on VMD-LSTM. *IET Renew. Power Gener.* **2019**, *13*, 1690–1700.
41. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Proc.* **2014**, *62*, 531–544.
42. Wang, S.; Jiang, X.; Zeng, L.; Chang, Y.F. Ultra-short-term Photovoltaic Power Prediction Based on VMD-DESN-MSGP Model. *Power Syst. Technol.* **2020**, *44*, 917–926.
43. Sun, Z.X.; Zhao, S.S.; Zhang, Z.X. Short-Term Wind Power Forecasting on Multiple Scales Using VMD Decomposition, K-Means Clustering and LSTM Principal Computing. *IEEE Access* **2019**, *7*, 166917–166929.
44. Peng, H.C.; Long, F.H.; Ding, C. Feature selection based on mutual information criteria of max-dependency and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238.
45. Hui, Z.; Chen, P. Oil Spills Identification in SAR Image Using mRMR and SVM Model. In Proceedings of the 2018 5th International Conference on Information Science and Control Engineering (ICISCE), Zhengzhou, China, 20–22 July 2018.
46. Adil, B.-H.; Youssef, G.; Abderrahim, E.Q. HVS-MRMR wrapper method for variables selection. In Proceedings of the Second International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 17–19 April 2017.
47. Chen, H.D.; Zhuang, P.; Xia, J.K.; Dai, W.Z.; Lu, Y.; Gao, Q.; Chen, T. Optimal power flow of distribution network with distributed generation based on modified firefly algorithm. *Power Syst. Prot. Control* **2016**, *44*, 149–154.
48. Wang, X.; Huang. K.; Zheng, Y.H.; Li, L.X.; Shao, F.P.; Jia, L.K.; Xu, Q.S. Combined PV Power Forecast Based on Firefly Algorithm Generalized Regression Neural Network. *Power Syst. Technol.* **2017**, *42*, 455–461.
49. Rukslin; Haddin, M.; Suprajitno, A. Pitch angle controller design on the wind turbine with Permanent Magnet Synchronous Generator (PMSG) base on Firefly Algorithms (FA). In Proceedings of the 2016 International Seminar on Application for Technology of Information and Communication (ISemantic), Semarang, Indonesia, 5–6 August 2016.
50. Liu, H.; Mi, X.W.; Li, Y.F. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. *Energy Convers. Manag.* **2018**, *156*, 498–514.
51. Cheng, R.; Jin, Y. A social learning particle swarm optimization algorithm for scalable optimization. *Inf. Sci.* **2015**, *291*, 43–60.
52. Li, F.L.; Chen, S.; Fan, X.J.; Liu, Y. Path Planning Based on Firefly Algorithm in Dynamic Unknown Environment. *Autom. Instrum.* **2019**, *34*, 53–58.