An Ultra-short-term Wind Power Forecasting Method with Special Error Distribution

Shuheng Wei¹, Deping Ke¹, *, Jian Yang², Shangguang Jiang², Yu Liu² and Yanqing Wu¹

1 School of Electrical Engineering and Automation, Wuhan University, Wuhan, China
2 North China Branch of State Grid Corporation of China, Beijing, China
*E-mail: kedeping@whu.edu.cn.

Abstract. The randomness and uncontrollability of wind energy have brought many difficulties to the development of wind power generation. How to obtain accurate forecasting results of wind power has become an increasingly important topic these years. This paper proposes an ultra-short-term wind power forecasting method using Elman neural network with an enhanced gradient descent training algorithm, which is able to compute the probabilistic distribution function (PDF) of the forecasting error by utilizing two kinds of series expansion for the purpose of regarding it as the loss function of the forecasting model. To validate the accuracy and efficiency of the purposed method, the conventional least mean square error (LMSE) based model is served as a benchmark in the comparison of simulation results. At last, historical wind power statistics of one month collected from a wind farm in the northern Hebei Province of China are used to perform single-point and probabilistic predictions in order to verify the effectiveness of the purposed method.

1. Introduction
As the increasing depletion of traditional fossil energy, new power generation technologies represented by wind power generation has developed rapidly in the past few years [1-3]. However, the integration of large-scale wind turbines imposes a great impact on the safety and stable operation of the power system. Thus with the assistance of accurate predictions of wind power, system dispatchers can make significant system control decisions to ensure the stability of the power grid [4]. Generally, the forecasting methods of wind power can be classified into two groups, single-point prediction, and probabilistic prediction. The single-point forecasting methods can be divided into numeric weather prediction (NWP) methods and statistical methods [5-7]. NWP methods are based on the numeric weather prediction model which outputs the power using a large amount of meteorological data and geographic information. Consequently, NWP methods tend to have higher forecasting accuracy when it comes to long-term prediction. But the modelling process can be quite difficult considering the unfamiliarity with meteorology and geography. Statistical methods are mainly utilized for short-term and ultra-short-term wind power prediction, such as support vector machines (SVM) and artificial neural networks (ANN) [8-11]. These methods are based on a large amount of historical statistical data, and the key lies in establishing the relationship between the model input variables and the output prediction values. When the specific input dataset is given, forecasting models based on single-point methods usually give the forecasting curve of wind power in the form of discontinuous points. However, these single-point methods cannot provide probabilistic forecasting results. Thus probabilistic prediction methods are becoming more and more favourable in practical uses [12].
Traditional forecasting models based on neural networks mainly use root mean-square error (RMSE) or mean absolute error (MAE) as the loss function [13]. However, this kind of forecasting method still cannot provide probabilistic forecasting results. This paper focuses on the problem that how to control the form of the probabilistic density curve of the prediction error. Therefore, the probabilistic feature of the forecasting error is applied to build the model.

This paper proposes a wind power forecasting model based on Elman neural network with an enhanced gradient descent training algorithm to effectively control the probabilistic density curve of the forecasting error for ultra-short-term wind power prediction. The forecasting time step is 15 minutes and the model is able to forecast the wind power output of the next 0-4 hours in the future. In the first step, the theoretical background related to the forecasting process is introduced. Then, an ameliorated gradient descent training algorithm is proposed. The forecasting process of the proposed training model can be summarized as follows: Step 1: Build an Elman neural network, then the single-point prediction results can be obtained. Step 2: According to the single-point prediction results, the forecasting error can be calculated. Then use the Gram-Charlier series and Edgeworth series to compute the probability density function (PDF) of forecasting error. Step 3: Using probabilistic features of the forecasting error to improve the gradient descent algorithm based on the calculated PDF. Step 4: Finally, apply the training algorithm to the Elman neural network to get both the single-point and probabilistic forecasting results of wind power.

Eventually, real observed wind power data of one month from an operating wind farm in the northern Hebei Province of China are selected to test the forecasting accuracy of the proposed model. The simulation result shows that the proposed model is able to effectively control the probabilistic density curve of the forecasting error. Matlab (Version R2020a) is utilized to build the Elman neural network without using the Matlab Toolbox in this paper.

The structure of this paper can be illustrated as follows: Section II introduces the theoretical background of this paper, including the Elman neural network and two kinds of series expansion. Section III shows the modeling process and proposes a forecasting model using Elman neural network with the improved gradient descent algorithm. Section IV shows the simulation results to evaluate the forecasting accuracy of the proposed model, then the model’s effectiveness of controlling the PDF is also examined. Section V summarizes the conclusion of this paper.

2. Related theories

According to the direction of information transmission in the network, ANN can be divided into feed-forward and feedback neural networks [14]. The output of the feed-forward neural network is only determined by the current input and the weight matrix, which means the feed-forward neural network is static. When dealing with dynamic time-series forecasting problems, this kind of neural network tends to overfit and fall into local optimal solutions [13].

2.1. Elman Neural Network

The feedback neural network is a dynamic neural network. Its current output is not only related to the current input and weight matrix but also related to the previous output of the network. Elman neural network was proposed by J.L. Elman in 1990 [15]. It is a typical local-feedback neural network with dynamic short-term memory, which is very advantageous for time-series forecasting problems.

The structure of Elman neural network is shown in figure 1. With the addition of a set of context units, the network has a local feedback section. After calculated by the input-hidden weights, the data in the hidden layer are not directly propagated to the output layer. Instead, it is transmitted to the context units to go through the local-feedback process and re-enter the hidden layer. In this way, the context units can act as a delay operator, so that the network’s capability to process time-series forecasting problem is significantly improved.

From figure 1 the nonlinear state-space representation of Elman neural network can be obtained as:
\[
\begin{align*}
    y(k) &= g\left( w^2 x(k) + b_y \right) \\
    h(k) &= f\left( w^1 u(k-1) + w^3 x_i(k) + b_h \right) \\
    x_i(k) &= h(k-1)
\end{align*}
\] (1)

Where \( u(k-1) \) is the input vector; \( y(k) \) is the output vector; \( h(k) \) is the hidden layer vector; \( x_i(k) \) is the vector of context units; \( w^1, w^2, \) and \( w^3 \) are the interlayer weights of this neural network; \( b_y \) and \( g(*) \) are the threshold and the transfer function of the neurons in the output layer, respectively; \( b_h \) and \( f(*) \) are the threshold and the transfer function of the neurons in the hidden layer, respectively; \( k \) is the number of iteration.

Generally, the gradient descent algorithm is used to train the parameters of neural networks. Once a specific parameter matrix enables the value of the loss function to satisfy the training target, the training process is completed. Training algorithms regard mean square error (MSE) as the loss function traditionally:

\[
E(w) = \sum_{i=k}^{\infty} \left( \tilde{y}_i(w) - y_i(w) \right)^2
\] (2)

Where \( \tilde{y}_i(w) \) is the predicted value; \( y_i(w) \) is the actual value; \( k \) is the number of iteration.

2.2. Gram-Charlier and Edgeworth Series

Gram-Charlier expansion is one of the methods to obtain the PDF of a random variable [16]. It can calculate the PDF of a random variable according to its moments and cumulants [17]. The expansion concludes Hermite polynomials of different orders and the PDF of standard normal distribution. When dealing with practical problems, Gram-Charlier series of infinite order is able to compute the PDF of random variables accurately. However, when limited order of series is used for calculating, the high-order one’s fitting performance is not necessarily better than the low-order series [18]. Therefore, lower-order Gram-Charlier expansion is often used for PDF fitting when solving practical problems. The specific process of using Gram-Charlier series to fit the PDF of wind power forecasting error can refer to literature [17] and the derived PDF can be mathematically written as follows:

\[
f(x) = \varphi(x) \left[ 1 + \frac{c_1}{3!} H_{c_1}(x) + \frac{c_2}{4!} H_{c_2}(x) + \ldots \right]
\] (3)

Where \( H_{c_i}(x) \) represents the Hermite polynomial of order \( i \); \( \varphi(x) \) represents the probabilistic density function of the standard normal distribution; \( c_i \) is the constant coefficient of order \( i \) for the expansion. The first four-order Gram-Charlier expansion is often used for PDF fitting in engineering problems. When using the Gram-Charlier series for PDF fitting, the moment and cumulant of every order need to be calculated. The application of Edgeworth series can overcome the time-consuming process because it is realized through the iterative process of Hermite polynomial [18].

**Figure 1.** Structure of Elman neural network.
The form of Edgeworth expansion is similar to Gram-Charlier expansion. Correspondingly, the first four-order Edgeworth series is used for PDF fitting in many cases. Thus the density function of the standardized variable \( \hat{e} \) is written as:

\[
\phi(x) = \varphi(x) \left[ 1 + \frac{c_1}{3!} H_1(x) + \frac{c_2}{4!} H_2(x) + \frac{10c_3}{6!} H_3(x) \right]
\]  

Comparing equation (3) and equation (4), it can be seen that the Edgeworth expansion contains a Hermite polynomial more than the Gram-Charlier expansion. However, the number of coefficients in the Edgeworth expansion does not change.

3. Modelling process

According to the related theories in Section II, the single-point method based on neural networks can only forecast a series of discontinuous wind power values in a certain period in the future, but cannot give the probabilistic information of the wind power output in this period. Therefore, a forecasting method that combines the expansion of Gram-Charlier series and Edgeworth series with Elman neural network is proposed in this section to obtain the probabilistic forecasting results.

3.1. Input and output variables

Iterative forecasting method can be used for time series prediction based on the historical statistics and the predicted value obtained in the current period to forecast the data in the next period. The predicted value is regarded as the model input in the iterative process. However, the forecast accuracy drops dramatically as the number of iterations increases. Aiming to overcome this problem, a direct forecasting method is used to make ultra-short-term predictions of wind power by 1 to 16 steps in advance. The forecasting time step is 15 minutes. An Elman neural network is established for every step of prediction. And the input of all the networks is the same, which is the actual value of wind power at \( t-(n-1), t-(n-2), \ldots, t \). The output is the predicted power at every 15 minutes in the next 0-4 hours. In this way, the input of each forecasting model is the true value that already obtained, thus there are 16 Elman neural networks that need to be established altogether. The input and output of the networks are as follows:

\[
X = \begin{bmatrix}
1 & 2 & \ldots & p-1 & p \\
2 & 3 & \ldots & p & p+1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
 n-1 & n & \ldots & p+n-3 & p+n-2 \\
n & n+1 & \ldots & p+n-2 & p+n-1
\end{bmatrix}
\]  

\[
Y_i = [n+i \ n+1+i \ \ldots \ p+n-2+i \ p+n-1+i]
\]

Where \( p \) is the number of samples of the training set; \( i = 1,2,\ldots,16 \).

3.2. Enhanced training algorithm

Iterative forecasting method can be used Neural network training algorithms are generally based on the least mean square error (LMSE) criterion. When MSE is selected as the loss function, it is usually necessary to set a relatively nominal target value to obtain a globally satisfying forecasting performance. When there is a large deviation for some points in the training statistics, the global forecasting accuracy may not be guaranteed because the square error of those individual points tends to be very large.

In order to overcome the limitations of the LMSE based training algorithm, this paper first computes moments and cumulants of the forecasting error to obtain its PDF fitted by Gram-Charlier or Edgeworth series. According to the obtained PDF, some of the probabilistic features of the PDF are applied to the training algorithm to form a custom loss function. By setting different training targets, the probabilistic distribution function of the forecasting error can be controlled, and the proposed model can output both the single-point and probabilistic forecasting results of wind power.

Based on the gradient descent algorithm, 4th order of Gram-Charlier expansion is taken as an example here.
Step 1: Calculate the moments and cumulants of the forecasting error, then the PDF fitted by Gram-Charlier Series can be expressed as:

\[
f(x) = \varphi(x) \left[ 1 + \frac{c_3(w)}{3!} H_{e_3}(x) + \frac{c_4(w)}{4!} H_{e_4}(x) \right]
\] (7)

Where \( f(x) \) represents the PDF of the standardized error; \( w \) represents the weight matrix of the neural network.

According to equation (2), the forecasting error is a function of the weight matrix \( w \). The coefficients of Gram-Charlier expansion are also the computational result of moments and cumulants of the error. Consequently, the coefficients \( c_3(w), c_4(w) \) are also a function of the weights of the neural network in equation (7).

Step 2: Determine an interval \( [e_-, e_+] \), then calculate the integration of the obtained PDF over this interval:

\[
p(x) = \int_{e_-}^{e_+} f(x) \, dx = \int_{e_-}^{e_+} \varphi(x) \left[ 1 + \frac{c_3(w)}{3!} H_{e_3}(x) + \frac{c_4(w)}{4!} H_{e_4}(x) \right] \, dx
\] (8)

Where \( e_- < 0, e_+ > 0 \); \( p(w) \) represents the probability that the forecasting error lies in the interval \( [e_-, e_+] \).

According to equation (7), the coefficients of the PDF are also the function of the parameters in the neural network. Then it can be inferred that \( p(w) \) is also the function of \( w \).

Step 3: Regard \( p(w) \) as the loss function of the neural network, the updating formula of the weight matrix is as follows:

\[
w = w + \Delta w
\]

\[
\Delta w = \eta \frac{\partial p(w)}{\partial w}
\] (9) (10)

Where \( \eta \) is the learning rate; \( p(w) \) is the custom loss function.

In the enhanced training algorithm, the updating direction of the weights is changed to the positive gradient direction of the loss function \( p(w) \) to the weight matrix \( w \).

Step 4: The training process terminates when the value of the loss function reaches the target threshold or the number of iterations reaches the upper limit.

In general, the proposed training algorithm excludes a few points with particularly large deviations from the target interval. The main training goal is to let the majority of forecasting error lies in the target interval.

According to probability theory, the probability that the error obtained is within the target interval after the training process should be as close to the target confidence level as possible. When making predictions, firstly input the test dataset, and then calculate the loss function according to equation (8). When the value is greater than the ideal confidence level, it proves that the accuracy of the model meets the requirements, and the probability distribution of the error is effectively controlled by the model. Otherwise, the model needs further improvement.

3.3. Forecasting model

The ultra-short-term prediction model of wind power based on Elman neural network with the ameliorated training algorithm now can be established, and its structure is shown in figure 2.

Firstly, perform autocorrelation analysis on the input series to determine the number of backtracking points used in equation (5). Secondly, there are 16 Elman neural networks established to predict the wind power values corresponding to a total of 16 forecasting moments in the next 0-4 hours at \( t+1 \), \( t+2 \), ..., \( t+16 \). Then the Gram-Charlier and Edgeworth series are utilized to compute the PDF of the forecasting error, and the integration of the obtained PDF over the target interval is regarded as the loss function of each Elman neural network. Finally, the model outputs both the single-point and probabilistic forecasting results of wind power. Where GC stands for the Gram-Charlier expansion; EW represents the Edgeworth expansion.
4. Case study and result comparison

4.1. Data and settings
Historical wind power data of one month (from March 25th, 2018 to April 25th, 2018) from an operating wind farm in the northern Hebei Province of China are selected to test the forecasting accuracy of the proposed model in this paper. The wind farm is equipped with 33 wind turbines with a total installed capacity of 49.5 MW. The total number of the data used for calculation is 2500, and the temporal resolution is 15 minutes.

To determine the number of the backtracking points in equation (5), an autocorrelation analysis is performed on the historical wind power series, and the result is shown in figure 3.

Figure 2. Structure of purposed forecasting model.

Figure 3. Autocorrelation analysis of wind power series.

Figure 3 shows that there is a strong correlation between wind power at the current moment and the power values at the previous 12 moments. To simplify the calculation of the model, the number of the backtracking points used in equation (5) is set to be 8.

The training samples and test samples of the model can be obtained at a ratio of 4:1, where the number of training samples is about 2000 (from March 25th, 2018 to April 19th, 2018) and the number of test samples is about 500 (from April 20th, 2018 to April 25th, 2018). After repeated trials, the structure of Elman neural network is determined to be 8-10-10-1, and the learning rate is set to be 0.05. The target interval of the loss function is taken as the ±10% absolute error (±10% installed capacity), the ideal training confidence level is set to 90%, and the upper limit of iteration is 200 times. As for the LMSE based benchmark, the parameter settings remain the same.

To evaluate the efficiency of the proposed model and LMSE based model, several indices should be taken into consideration. In this paper, normalized mean absolute error (NMAE) and normalized root mean square error (NRMSE) are used to appraise the forecasting accuracy [19].

\[ e_{\text{NMAE}} = \frac{1}{x_{\text{cap}} \cdot n} \sum_{i=1}^{n} |x_i' - x_i| \]  \hspace{1cm} (11)

\[ e_{\text{NRMSE}} = \frac{1}{x_{\text{cap}}} \sqrt{\frac{\sum_{i=1}^{n} (x_i' - x_i)^2}{n}} \]  \hspace{1cm} (12)

Where \( x_i' \) is the predicted result; \( x_i \) is the real wind power value; \( n \) is the number of samples; \( x_{\text{cap}} \) is the installed capacity of the wind farm.

Apart from the single-point prediction, two kinds of series expansion are used to calculate the PDF of the error, and then the training algorithm of the network is further improved to obtain the probabilistic forecasting results. Since the target interval of the loss function of each neural network is taken as the ±10% absolute error, the training confidence level is set to 90%. Therefore, the probability that the error output by the model lies in the target interval should be greater than the ideal confidence level of 90%.

The prediction interval coverage probability (PICP) is utilized to evaluate the effectiveness of probabilistic prediction in this paper [20], which can be written as:
Where $e_i$ is the coverage factor, if the value of the $i^{th}$ error is in the confidence interval, then $e_i = 1$, otherwise $e_i = 0$; $n$ is the number of samples.

PICP is often used to characterize the coverage of the target confidence interval [19]. The higher the PICP, the better the coverage of the target interval. When PICP is greater than or equal to the ideal training level in this paper (90%), it means that the probabilistic forecasting results meet the expected requirements, otherwise the model needs to be further improved.

4.2. Single-point forecasting results

Use the 4th order of Gram-Charlier expansion and Edgeworth expansion to compute the PDF, then a model based on the LMSE criterion is served as a benchmark. These three models are denoted by GC, EW, and LMSE, respectively.

![Figure 4. Comparison of single-point forecasting curves.](image)

Figure 4 presents the comparison of single-point forecasting curves of these three models at $t+1$, $t+8$, and $t+16$ (15 minutes ahead, 2 hours ahead, and 4 hours ahead). From figure 4 we can intuitively observe that the proposed models (blue and green lines) have much better forecasting performance compared with the LMSE model (red line) at $t+8$ and $t+16$. When the look-ahead time steps increase, the LMSE model tends to gradually lose its capability to track the real power curve. Due to the improved training algorithm based on probabilistic features of the forecasting error, the proposed model can still maintain its accuracy as the time interval between the forecasting moment and the current moment increases.

Table 1 demonstrates that several evaluating metrics of the proposed models (GC and FW) are less than those of the LMSE model. The average NMAE index and NRMSE index of the GC & FW models are around 7% and 10%, respectively. And the model using Edgeworth expansion to compute the PDF has the highest forecasting accuracy for single-point prediction, whose NMAE index is 7.23%, and NRMSE index is 10.65%.

| Moment | LMSE   | GC     | FW     |
|--------|--------|--------|--------|
|        | $e_{NMAE}$ | $e_{NMAE}$ | $e_{NRMSE}$ | $e_{NRMSE}$ | $e_{NMAE}$ | $e_{NMAE}$ | $e_{NRMSE}$ | $e_{NRMSE}$ |
| $t+1$  | 5.30   | 7.11   | 3.66   | 5.72   | 3.22   | 5.04   |
| $t+2$  | 7.63   | 10.61  | 4.54   | 6.73   | 5.16   | 7.16   |
| $t+3$  | 9.54   | 12.47  | 4.76   | 7.52   | 5.00   | 7.61   |
8

t+4  10.64  13.57  5.86  9.08  4.48  7.74
t+5  13.02  15.90  5.40  8.76  5.34  8.70
t+6  16.60  20.27  5.79  9.22  5.84  9.39
t+7  14.96  18.05  6.40 10.01  6.40 10.09
t+8  13.02  15.90  5.40  8.76  5.34  8.70
t+9  16.60  20.27  5.79  9.22  5.84  9.39
t+10 14.96  18.05  6.40 10.01  6.40 10.09

Table 2. PICP for Evaluating the Probabilistic Forecasting Results

| Moment | GC  | FW  | Moment | GC  | FW  |
|--------|-----|-----|--------|-----|-----|
|        | $\delta_{GC}$ | $\delta_{FW}$ |        | $\delta_{GC}$ | $\delta_{FW}$ |
| t+1    | 91.71 | 92.89 | t+9    | 81.50 | 77.99 |
| t+2    | 89.63 | 93.12 | t+10   | 75.89 | 74.70 |
| t+3    | 90.53 | 91.10 | t+11   | 73.48 | 71.62 |
| t+4    | 86.61 | 87.57 | t+12   | 69.25 | 67.95 |
| t+5    | 91.43 | 86.30 | t+13   | 64.33 | 64.60 |
| t+6    | 89.89 | 84.15 | t+14   | 61.81 | 61.55 |
| t+7    | 86.71 | 81.57 | t+15   | 58.11 | 58.50 |
| t+8    | 84.55 | 79.36 | t+16   | 56.21 | 56.03 |
| Average| 78.23 | 76.81 |

Table 2 shows that the average PICP index of the FW model has a better performance compared with the GC model, which reaches 78.23% under the confidence level of 90%. Both models can maintain a relatively satisfying forecasting accuracy when the number of look-ahead time steps is not too large (1 to 8 steps), indicating that the probabilistic prediction accuracy is high for two look-ahead hours forecast, and the minimum PICP indexes are about 85% and 80% for the GC model and FW model, respectively. When the number of advance steps continues to increase, the PICP indexes of both models drop to a certain extent.

Figure 5 shows the PDF fitting performance of the GC model. The original error in the first column of figure 5 is the absolute error obtained from each forecasting moment. Using the ksdensity function in Matlab to fit the absolute error series, then its PDF curves at t+1, t+8, and t+16 can be calculated. It can be observed from the graphs in the first column of figure 5 that the forecasting accuracy deteriorates as the number of advance steps in prediction increases. The second column in figure 5 is the fitting performance of Gram-Charlier series. From the graphs in the second column, we can observe that the PDF curves computed by Gram-Charlier expansion at t+1 and t+16 can effectively fit the standardized error (the standardized error series is obtained by subtracting the mean value from the original error series and dividing it by its standard deviation). However, the PDF of standardized error calculated using the ksdensity function at t+8 has a bigger kurtosis, resulting in a slightly poorer fitting effect.

The grey area within the target confidence interval in figure 5 represents the probability that the error is within ±10%. As the number of look-ahead forecasting steps increases, the area gradually decreases,
which means the probability that the error is within ±10% also drops. The area in the confidence interval at t+1, t+8, and t+16 is 0.9171, 0.8455, and 0.5621, respectively. Figure 6 shows the confidence intervals of the wind power output by the GC model. It can be clearly seen from figure 6 that the forecast curve at t+1 has a good effect in tracking the actual power curve. The wind power fluctuating interval is relatively narrow at this time, indicating that the error is limited to a small range under the 90% confidence level. When the look-ahead forecasting time steps are 8 and 16, the deviation between the forecast curve and the actual curve and the width of the 90% confidence interval becomes larger.

5. Conclusion
This paper proposes an ultra-short-term wind power forecasting method using Elman neural network with an ameliorated gradient descent training algorithm to meet special requirements for the PDF of forecasting error. Comparing with the traditional LMSE based model, the proposed model shows higher accuracy when utilized to perform single-point wind power prediction. Moreover, probabilistic forecasting results show that the proposed method enables the error’s PDF to meet the requirements when the advance forecasting steps is not too large. The probability of error within ±10% is approximately 85% when the forecast moment lies in the upcoming two hours. When the number of look-ahead steps gradually increases, the model can still effectively provide the probabilistic information of the forecasting error and the fluctuation interval of wind power. In summary, the proposed method uses the probabilistic distribution function of error as the loss function to train the model. The corresponding benefit is that it can directly target the PDF of forecasting error and fully consider the requirements of subsequent stochastic economic dispatching for the probabilistic distribution of the error, thereby improving the reliability of the dispatching solution. Future research will consider combining this method with NWP statistics to achieve longer-term wind power prediction.

Figure 5. Fitting performance of the GC model.

Figure 6. The 90% confidence intervals of the wind power output by the GC model.

References
[1] Kumar Y, Ringenberg J, and Depuru S S, et al. “Wind energy: Trends and enabling technologies,” Renewable and Sustainable Energy Reviews, vol. 53, pp. 209-224, Jan. 2016.
[2] Vargas S A, Esteves G R T, and Maçaira P M et al. “Wind power generation: A review and a research agenda,” Journal of Cleaner Production, vol. 218, pp. 850-870, May. 2019.

[3] Blaabjerg F, and Ma K, “Wind energy systems,” Proceedings of the IEEE, vol. 105, no. 11, pp. 2116-2131, May. 2017.

[4] Piwko R., Osborn D, and Gramlich R et al. “Wind energy delivery issues: transmission planning and competitive electricity market operation,” IEEE Power and Energy Magazine, vol. 3, no. 6, pp. 47-56, Oct. 2005.

[5] Foley A M, Leahy P G, and Marvuglia A et al. “Current methods and advances in forecasting of wind power generation,” Renewable Energy, vol. 37, no. 1, pp. 1-8, Jan. 2012.

[6] Landberg L, “A mathematical look at a physical power prediction model,” Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology, vol. 1, no. 1, pp. 23-28, Sep. 1998.

[7] Kariniotakis G, Waldl I H, and Marti I et al. “Next generation forecasting tools for the optimal management of wind generation,” in 2006 International Conference on Probabilistic Methods Applied to Power Systems, Stockholm, Sweden: IEEE, Jun. 2006, pp.1-6.

[8] Wen J, Wang X, and Li L, et al. “Short-term wind power forecasting based on lifting wavelet, SVM and error forecasting,” in Unifying Electrical Engineering and Electronics Engineering, New York: Springer New York, Jun. 2013, pp.1037-1045.

[9] Methaprayoon K, Yingvivatanapong C, and Lee W et al. “An integration of ANN wind power estimation into unit commitment considering the forecasting uncertainty,” IEEE Transactions on Industry Applications, vol. 43, no. 6, pp. 1441-1448, Nov. 2007.

[10] Khosravi A, Machado L, and Nunes R O, “Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil,” Applied Energy, vol. 224, pp. 550-566, Aug. 2018.

[11] Sideratos G, and Hatzigryriou N D, “Probabilistic wind power forecasting using radial basis function neural networks,” IEEE Transactions on Power Systems, vol. 27, no. 4, pp. 1788-1796, Mar. 2012.

[12] Bludszuweit H, Domínguez-Navarro J A, and Llombart A, “Statistical analysis of wind power forecast error,” IEEE Transactions on Power Systems, vol. 23, no. 3, pp. 983-991, May. 2008.

[13] Chen L, Hao H, and Zheng G, “An audio watermarking of wavelet domain based on BP neural network,” in 2011 International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, China: IEEE Nov. 2011, pp.1-5.

[14] Golden R M, “Exploring the Diversity of Artificial Neural Network Architectures. Review of Neural Networks: A Comprehensive Foundation, by Simon Haykin,” Journal of Mathematical Psychology, vol. 41, no. 3, pp. 287-292, Sep. 1997.

[15] Elman J L, “Finding structure in time,” Cognitive science, vol. 14, no. 2, pp. 179-211, Mar. 1990.

[16] Chateau J P, and Dufresne D, “Gram-Charlier Processes and Applications to Option Pricing,” Journal of Probability and Statistics, vol. 2017, pp. 1-19, Feb. 2017.

[17] Zhang P, and Lee S T, “Probabilistic load flow computation using the method of combined cumulants and Gram-Charlier expansion,” IEEE transactions on power systems, vol. 19, no. 1, pp. 676-682, Feb. 2004.

[18] DasGupta A, “Edgeworth expansions and cumulants,” in Asymptotic theory of statistics and probability. New York, NY: Springer Science & Business Media, 2008, pp. 187-202.

[19] Souza L V D, Pozo A T R, and Rosa J M C D et al. “The boosting technique using correlation coefficient to improve time series forecasting accuracy,” in 2007 IEEE Congress on Evolutionary Computation, Singapore, Singapore, Sep. 2007, pp.1288-1295.

[20] Chu Y, and Coimbra C F, “Short-term probabilistic forecasts for direct normal irradiance,” Renewable Energy, vol. 101, pp. 526-536, Feb. 2017.

Acknowledgments
This work is supported by the Science Project of State Grid Corporation of China (No.520101180052).