Medium and long-term wind energy forecasting method considering multi-scale periodic pattern

Yisha Lin[1,*], *Zongxiang* Lu[1], *Ying* Qiao[1], *Mingjie* Li[2] and *Zhifeng* Liang[2]

[1] State Key Lab of Control and Simulation of Power Systems and Generation Equipment (Department of Electrical Engineering, Tsinghua University), Haidian District, Beijing 100084, China
[2] State Grid Corporation of China, Xicheng District, Beijing 100031, China

Abstract. Medium and long-term weather sequence forecast becomes unreliable beyond two weeks since the weather is a chaotic system. Using values of same months for electricity prediction of wind power is the usual method. This approach defaults wind power output with annual cycle law. However, the periodic pattern can be very complicated in fact with multiple time scales. This paper proposes an approach with multi-scale periodic pattern considered. The application of parametric estimation on cumulative distribution function avoids the difficulty of predicting the power curve. Meteorological condition is considered to some extent via multi-scale periodic pattern explored basing on historical energy data. This work is an exploration for medium and long-term wind power forecasting that can well adapt to existing conditions. It has better prediction accuracy than the method without multi-scale periodicity considered.

1 Introduction

Reliable forecast on mid and long-term generation of wind power can be significant for energy balance, generation scheduling and maintenance scheduling [1] with the penetration of renewable power raising.

However, medium and long-term power timing curves are difficult to obtain under current technical conditions. For one thing, Numerical models for simulation and prediction of atmospheric flows are subject to deterministic chaos and are likely to give unrealistic solutions [2]. The theoretical upper limit of deterministic forecast for weather sequence is about two weeks. After reaching the upper limit of two weeks, the error of deterministic weather forecast is almost equal to the level of natural variability [3]. For another, statistical methods based on historical sequence extrapolation are poorly adaptable [4].

Above problems considered, the main form of wind power forecasting at present for medium and long-term scale is to predict the amount of electricity monthly or yearly. Probabilistic prediction based on principal component analysis and quantile regression is carried out in [5]. However, the prediction quality is strongly correlated with the accuracy of the weather forecast for the next 30 days and so it can be ineffective for areas that are lack of weather forecast meeting the requirements. Gray model is often used in situations over which historical data is insufficient [6], but it cannot handle situation with new installed capacity. Another common practice is to find the representative wind farm of targeted area [7], and then extrapolate the output of the representative wind farm. Forecast on representative wind farm is used as the result for the region. This method is not universally applicable for places with complex terrain.

This paper proposes a forecasting method for monthly electricity generation of wind power. The approach is based on cumulative distribution function (CDF), which avoids the difficulty of predicting the power curve. Multi-scale periodic pattern of wind resource is taken into account, and so the meteorological condition is considered to some extent.

2 Wavelet analysis for multi-scale periodicity

Wavelet analysis is a powerful tool that can reveal the dominant modes of variability. One important feature that makes it distinguished is it decomposes a time series into time-frequency space, and so it can offer both time-domain and frequency-domain information. Because of that, wavelet analysis has better performance in analysing nonstationary time series compared with other signal representation approaches, such as Fourier transform [8]. Wind power series is typically non-stationary with different frequencies at various scales [9], and thus it is ideal to be analysed by Wavelet transform.

Continuous wavelet transform is applied in the following evaluation in view of its good performance on feature extracting [10]. A function that satisfies the admissible condition can be the mother wavelet as the base of transformation. Admissible condition is given as formula (1).

* Corresponding author: lys_ncepu@126.com

$$\int_{-\infty}^{\infty} |\omega|^{-1} |\psi(t)|^2 \, d\omega < \infty \tag{1}$$

The continuous wavelet transform of a signal $f(t) \in L^2(R)$ is defined as the convolution of $f(t)$ with the daughter wavelets $\psi_{a,b}(t)$ which are the translated and scaled versions of the mother wavelet $\psi(t)$.

$$(W_\psi f)(a,b) = \int_{-\infty}^{\infty} f(t) \overline{\psi_{a,b}(t)} dt \tag{2}$$

$$\psi_{a,b}(t) = |a|^{-1/2} \psi((t-b)/a) \quad a,b \in R, a \neq 0 \tag{3}$$

Here, a is scale factor, and b is shift factor. $(W_\psi f)(a,b)$ is also named wavelet coefficient, which indicates the correlation between the wavelet and the input array. Overline indicates complex conjugate.

In this paper, Morlet complex wavelet is chosen to be the mother wavelet. The real part of the wavelet coefficient is convenient for us to study the scale variability over the entire time domain [11].

The wavelet variance Var(a) is the integral of the squared norm of the wavelet coefficients in the time domain, which indicates the periodic wave energy distribution under the different scales. When the scale is fixed, the greater the wavelet variance is, the more obvious the periodic characteristics are, which means that the time series signal has the main period. The wavelet variance is as shown in equation (4).

$$\text{Var}(a) = \int_{-\infty}^{\infty} |(W_\psi f)(a,b)|^2 \, db \tag{4}$$

3 Electricity generation forecasting method based on cumulative distribution function

Restricted by the accuracy of weather forecasting for long term, we gave up the common idea to calculate generation by the integral of power. We rather choose the CDF as the intermediary, and make the computation concise and easy to process. The transformation from CDF to generation can be expressed as equation (5)

$$W = \tau \int_0^{P_M} (1 - F(x)) dx \tag{5}$$

Here, $F(x)$ represents CDF, and P_M is the upper bound of power output, that is capacity in this work. τ represents the time sampling interval.

Since quantities of wind farms were constructed in the recent years, and the old ones are often lack of good data storage conditions. This leads to the situation at present that parametric estimation is chosen to be the method in this paper. Since complex non-parametric estimation is hard to be achieved, which is likely to have

better precision. The scheme for forecast of wind power generation is proposed as displayed in Fig.1.
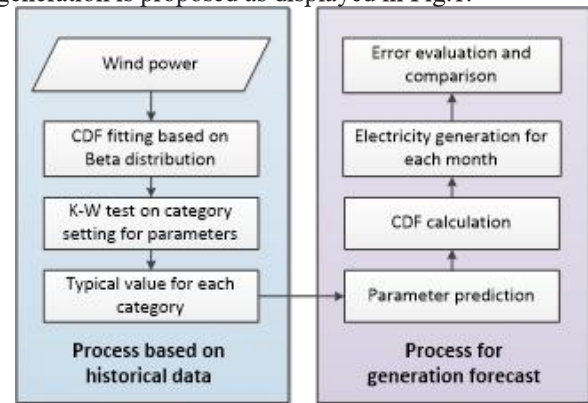

Fig. 1. Overall steps for wind electricity generation.

3.1 CDF fitting based on Beta distribution

Beta distribution is applied for our analysis as it leads to lower error for CDF fitting in most circumstances compared to Weibull distribution and Lognormal distribution, both of which are commonly used [12-14]. This will be reflected in section 4.2.

Probability density function (PDF) of beta distribution is

$$f(x;\alpha,\beta) = x^{\alpha-1}(1-x)^{\beta-1} B(\alpha,\beta)^{-1} \tag{6}$$

$$B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt \tag{7}$$

Correspondingly, CDF of beta distribution is

$$F(x;\alpha,\beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1} dt \, B(\alpha,\beta)^{-1} \tag{8}$$

Maximum Likelihood Estimation is used to obtain the value of two parameters of Beta distribution: α and β.

3.2 Parameter prediction based on periodicity

Parameter estimation is expected to be done according to the multi-scale periodic pattern from existing data. That is, several significantly different categories will be distinguished based on periodicity, and then typical values will then be figured out for predicting.

Assume that there are m major periods found, and then at least $C_m^1 + C_m^2 + ... + C_m^m$ classification schemes are optional. Kruskal-Wallis nonparametric test (K-W test) is used to find the best division, for which the differences between categories are most significant. If null hypothesis that distribution of different category is the same cannot be rejected, it means that the class definition method is not reasonable.

Once categories are settled, the median of parameters for each class is defined as the typical value. If no effective classification based on periodicity is obtained,

then other information, for example, the values of previous months will be adopted directly.

3.3 Evaluation indices

The difference of empirical CDF obtained from historical data and estimated CDF reflects the error of forecast. Root mean square error (RMSE) is used to quantify the deviation.

$$RMSE = (\frac{1}{N}\sum_{i=1}^{N}(F(x_i;\alpha,\beta) - \hat{F}(x_i))^2)^{1/2} \tag{9}$$

Here, $\hat{F}(x)$ refers to empirical CDF.

Relative error (*RE*) is applied to measure the offset of forecasting results for electricity generation.

$$RE = (W - \hat{W})\hat{W}^{-1} \tag{10}$$

Therein, $\hat{W}$ represents the actual electricity generation.

4 Case Study

Case study is designed based on the data of wind farms in north Hebei province. Time ranges from 2015 to 2017, and the time step is 1 minute.

4.1 Multi-scale periodic pattern

Since our goal is to analyze the medium and long-term pattern of wind power, which corresponds to monthly or even yearly scales, we choose series of monthly wind power generation as input. Electricity generation also reflects the abundance of wind resources. Contour map over the real part of wavelet coefficients is shown in Fig.2.
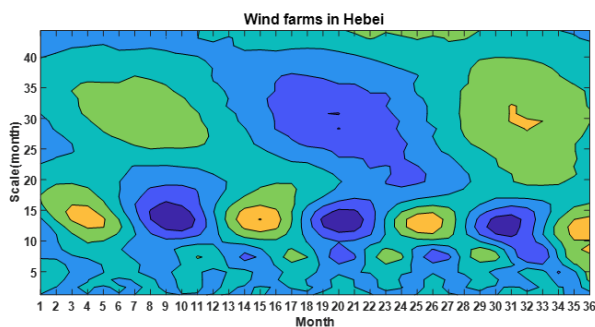

Fig. 2. Contour map based on wind power generation of Hebei.

In the figure, the horizontal axis represents time in the unit of month and the vertical axis represents the wavelet scale. If the region of the contour map is filled with the warm color, the real part is positive which means that the wind resource is abundant, and the wind power generation is vast. Opposite conclusion will be made if the filling color is cold.

The wavelet variance of wind farms in Hebei is exhibited in Fig.3 to figure out the main periodicity. The leading periodic component is 13.54 months, the second-domain component is 29.54 months, and the third is 7.385 months. They are reflectively defined as medium, long and short period for the following discussion.
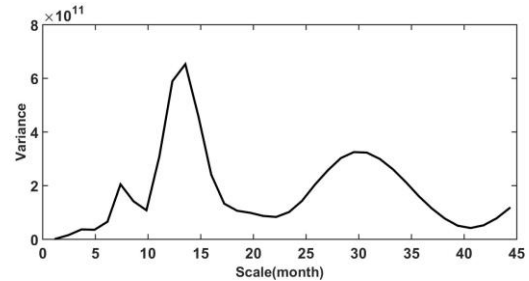

Fig. 3. Wavelet variance diagram.

4.2 CDF fitting

Distributions that are widely used for wind power are tried for each month, including Beta distribution, Weibull distribution and Lognormal distribution. RMSE for each situation is shown in Fig.4, and illustrates that applying Beta distribution has best performance for this case.
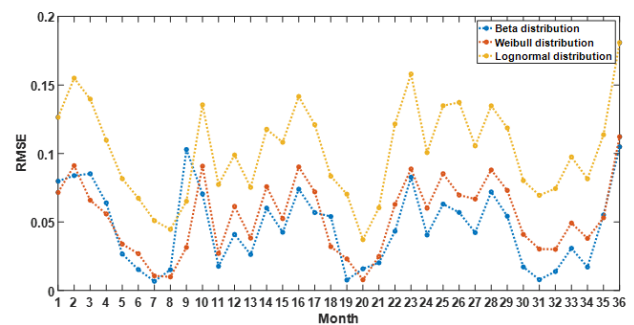

Fig. 4. RMSE for CDF fitting based on different distributions.

Table 1. Parameter estimation and RMSE of CDF fitting for 2015.

Month	1	2	3	4	5	6
α	1.981	0.725	2.915	2.068	1.477	0.859
β	10.672	5.079	13.267	11.101	6.985	5.005
RMSE	0.067	0.071	0.071	0.054	0.027	0.015
Month	7	8	9	10	11	12
α	0.877	0.734	1.002	1.063	0.915	0.968
β	7.582	7.478	3.194	3.356	6.736	3.992
RMSE	0.004	0.009	0.083	0.069	0.018	0.043

Estimation of parameters for Beta distribution and RMSE for CDF fitting are given as Table 1(results for 2015) and Table 2(results for 2016). The maximus RMSE is 0.085. It indicates that the method to forecast electricity generation using CDF based on Beta distribution is feasible.

Table 2. Parameter estimation and RMSE of CDF fitting for 2016.

Month	1	2	3	4	5	6
α	1.609	1.368	1.819	1.452	1.676	0.617
β	6.056	4.481	5.455	3.880	4.443	3.095
RMSE	0.029	0.059	0.046	0.074	0.059	0.053
Month	7	8	9	10	11	12
α	0.883	0.705	0.772	1.289	2.467	1.284
β	5.806	8.941	4.762	3.833	4.696	3.946
RMSE	0.013	0.010	0.015	0.051	0.085	0.045

4.3 Parameter prediction

K-W test is applied to different modes, and statistical significances are listed in table 3 for parameter α and β. Significance level is set as 0.05. Long, medium and short period was defined in section 4.1.

For every period considered, there are two classes, one is that the wind resource is abundant, and another is not. For example, if "Long-Medium" periods are concerned, there are four categories for parameters. The first is that wind resource presents abundant for both periods, and the second shows scarce characteristic for both periods. Other two correspond to the situations that wind resource is ample for one period, and necessitous for another period.

Table 3. Significance for different classifications.

Periods considered for classification	Statistical significance for α	Statistical significance for β
Long	0.835	0.728
Medium	0.002	0.321
Short	0.08	0.448
Long-Medium	0.01	0.638
Medium-Short	0.006	0.196
Long-Short	0.354	0.084
Long-Medium-Short	0.048	0.116

Whether takes medium period as the base for categories affects the conclusion a lot. When analyzing on the Medium-Short mode, we found situation that wind resources are in shortage for both periods (defined as category1 in Fig.5) has significantly different data distribution with others (defined as category2 in Fig.5). We then made two categories according to the result, and the statistical significance became 0.001, lower than any as shown in Table 3. It is what we use in the following analysis. The distribution of α for different group is displayed in Fig.5.

α for category 1 is obviously small than category 2. We select medians as the typical value for each class, which is 0.7527 and 1.4101 separately.

As we can refer from Table 3, difference is not significant between different categories no matter the way of defining groups. This indicates that the parameter prediction method according to periodic pattern is not suitable for β. Hence, we use the average of the last two months to be the predicted value for β as the lack of quantitative data.
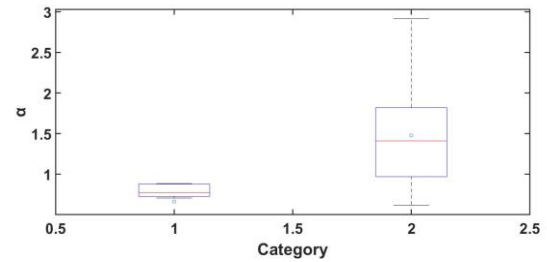

Fig. 5. Distribution of for different category.

4.4 Wind electricity generation forecast

Since the category belonging is very important for selection of α, forecast on abundance situation for medium and short periods should be done firstly. Fig.6 shows the abundance condition for months of 2015 and 2016, and it is consistent with Fig.2. The red square corresponds to the warm part in Fig.2, and the blue one to the cool tone. The duration of stages of abundant and dry wind resources shows a strong regularity. Therefore, we do random sampling to forecast the mode for the future months according to the distribution reflected in the historical data. Predicted results are displayed as hollow squares.
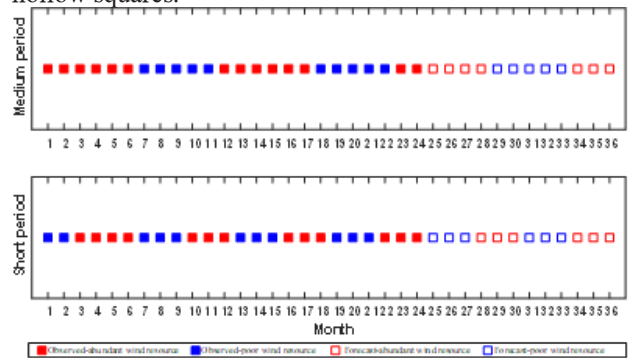

Fig. 6. Forecast on abundance of wind resources.

Parameters can be set then for months that generation is expected to be predicted. To reflect the preponderance with multi-scale periodic pattern concerned for forecast, we make the benchmark that parameters are set only based on the data of same months in the previous years. The absolute values of RE for electricity generation forecast for 2017 is demonstrated in Fig.7.
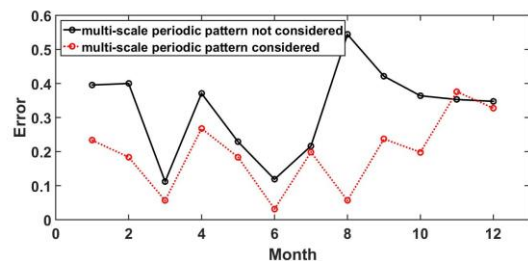

Fig. 7. Forecast error of electricity generation.

The mean error with multi-scale periodic pattern not considered is 0.3226, and the value drops to 0.1958 with the method proposed in this paper. However, the error is higher than the benchmark in November, the probable cause is that power dispatching institution increase the

accommodation of wind power to fulfill the promise of reducing wind abandon made in the early of the year.

Error of electricity generation forecast and CDF fitting is compared as Fig.8 shows. The larger the error of CDF fitting, the higher the absolute value of RE for electricity generation. Moreover, the prediction error of electricity generation amplifies the error of CDF fitting during integral operation.
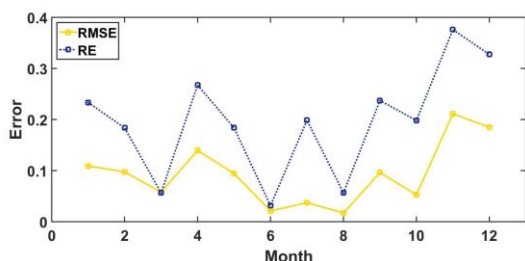

Fig. 8. Error of electricity generation forecast and CDF fitting.

5 Discussion and conclusion

This work provides a forecast method for monthly electricity generation of wind power. The method is put forward under the condition that no credible weather sequence forecast for medium and long term could be referred. Multi-scale periodic pattern is in consideration to predict the parameters of Beta distribution, which is chosen to fit the probability density distribution of wind power. The method has better prediction accuracy than the method without periodic pattern considered, since the mean relative error of electricity generation decrease from 0.3226 to 0.1958.

However, this paper is just a preliminary work for medium and long-term wind power forecasting. The absolute value of relative error reaches 0.3901 for the worst situation, and it is not a satisfying result for practice. Further efforts could be made for the following points:

- Use complicated distribution, for example, mixture distribution, with accumulated data increasing.
- More information is deserved to be explored from historical data, not only periodicity for wind resources.
- Try to integrate weather forecasting information to make prediction for wind electricity generation.

References

1. S. Han, Y. Qiao, J. Yan, Y. Liu, L. Li, Z.Wang. Appl. Energy. **239**,181-91 (2019).
2. A.Mary Selvam, S. Fadnavis. Chaos Solitons & Fractals. **10**,1577-82 (1999).
3. F. Zhang, Y. Sun, L. Magnusson, R. Buizza, S. Lin, J. Chen, et al. Journal of the Atmospheric Sciences. **76**, 1077-91 (2019).
4. D. Alencar, C.Affonso, R.Oliveira, J. Moya Rodríguez, J. Leite, J. Reston Filho. Energies. **10**, 1976 (2017).
5. D. Li, Z.Ren, W. Yan, J.Zhu, X. Zhao, J.Yu. Zhongguo Dianji Gongcheng Xuebao. **37**, 5238-47 (2017).
6. H. Citakoglu, A. Aydemir. 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT); 2019 9-11 April 2019.
7. Z. Guo, Y.Dong, J.Wang, H.Lu. Math. Probl. Eng. **2010**, 17 (2010).
8. P. Chowdhury, B.N. Dwivedi. Planet. Space Sci., **67**, 92-100 (2012).
9. X. Bao, C. Li. Journal of Wind Engineering and Industrial Aerodynamics, **193**, 103982 (2019).
10. A.C. Furon, C. Wagner-Riddle, C.R. Smith, J.S. Warland. Agricultural and Forest Meteorology, **148**, 1305-17 (2008).
11. X. Chen, L. Yin, L. Fan, L.Song, T.Ji, Y. Liu, et al. Sci. Total Environ., 134244 (2019).
12. X. Yuan, C. Chen, M. Jiang, Y. Yuan. Applied Soft Computing, **82**, 105550 (2019).
13. P. Wais. Renewable Energy, **103**, 15-29 (2017).
14. I. Pobočíková, Z. Sedliačková, M. Michalková. Procedia Eng, **192**, 713-8 (2017).