Probabilistic prediction of short term wind power considering temporal and spatial dependence of prediction error

Yong Sun[1], *Zhenyuan* Li[1], *Chaoyu* Shi[2, *], *Baoju* Li[1], and *Mao* Yang[2]

[1]State Grid Jilin Electric Power Supply Company, Changchun 130012, China
[2]Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education (Northeast Electric Power University), Jilin 132012, China

Abstract. The short-term probabilistic prediction of wind power has the characteristics of spatial dependence and time-series dependence. Considering the two characteristics at the same time can improve the prediction level. In this paper, a probabilistic short-term wind power prediction model considering the temporal and spatial dependence of prediction error is proposed. Considering the coupling relationship between the two properties, the NWP(Numerical Weather Prediction) wind speed point prediction error in the historical period is hierarchical clustering, and the empirical distribution model is used to fit the probability distribution of the error under different wind conditions; the cumulative empirical distribution probability value corresponding to the NWP wind speed at the time to be predicted is bootstrap sampling; under the given confidence level, the possible wind speed at each time point to be predicted in the short term is calculated The power fluctuation range of the generator. The test results show that this method can ensure the statistical significance and fitting stability of the sub sample set at the same time, and improve the classification accuracy of the days to be predicted. Compared with considering a single property, the result of probability prediction performs better on multiple evaluation indexes.

1 Introduction

High precision wind power prediction has become a necessary operation technology for high proportion new energy to access power system[1]. Point prediction is the most widely used form of wind power forecasting. Because of the limited information contained in the single numerical prediction results, it can not fully reflect all the information of wind power uncertainty. It is easy to cause conservative or risk in the optimal operation and standby decision-making arrangement of power grid, thus restricting the wind power consumption or threatening the operation safety. Probability prediction is a necessary supplement to point prediction, which can quantitatively reflect the fluctuation range of wind power, optimize system backup, improve wind power consumption and reduce operation risk. The

---

* Corresponding author: 2201900156@neepu.edu.cn

short-term probability prediction in this paper refers to the cumulative distribution function (CDF) or interval for predicting the wind power of the next day.

The uncertainty of wind power has many characteristics. According to whether the probability distribution of power / error is conditional dependent, it can be classified into conditional modeling and non conditional modeling. The conditional dependence property refers to the correlation between the distribution function and the wind speed, wind direction, point prediction results, etc. In the early literature, the non conditional modeling was mostly used[2]. The wind turbine power curve was proposed in document[3], which describes the conversion relationship between the wind turbine speed and the power output. Pinson first proposed the view of conditional modeling in 2010[4]. Then, it is gradually improved, such as the segmentation model of prediction error, the method of error stratification analysis, the method of multi condition combined prediction, and the consideration of spatial correlation. The time dependence of wind power uncertainty means that its distribution function will change with time, which is due to the nonstationarity of wind power stochastic process[5]. At present, the existence of this property has been proved by comparing probability density function and hypothesis test. The paper [6] proposes Markov transfer autoregression model, etc., but the wind power sequence has only low order Markov property. Considering the condition or / or timing characteristics, the probability prediction level can be improved, but the error density distribution established by this method may be asymmetric and has thick tail effect. Some literatures use the distribution fitting of double or multi parameters (β distribution, α stable distribution[7], generalized error distribution) and nonparametric distribution. The nonparametric fitting does not presuppose the expression of error distribution, but directly obtains the distribution or the quantile through data driven methods such as quantile regression and kernel density estimation[8], which is more adaptable to irregular distribution function, but a lot of data is needed. Considering the time-space dependence, the sample separation must be carried out, which contradicts the requirement of non parametric modeling for large sample size.

In order to solve the above problems, this paper first analyzes the characteristics of wind power prediction error, proposes a sample separation method considering spatial and temporal characteristics, and establishes the joint probability density distribution function of wind speed power prediction error; Then, the bootstrap method is used to sample the classified error samples; Finally, the probability prediction method considering the coupling of the two properties is given. The test data of a wind farm in Jilin Province of China show that the proposed method can guarantee the statistical significance and fitting stability of the sub sample set at the same time, and improve the classification accuracy of the time period to be predicted. Compared with considering the single property, the probabilistic prediction result performs better on multiple evaluation indexes.

2 Wind power prediction error characteristics

2.1 Hierarchical clustering

Spatial dependence and time dependence are two typical characteristics of point prediction error. It is necessary to separate multiple error sample subsets and establish probability density function on each subset. Considering the spatial dependence, it is required to separate samples according to the size of wind speed, while considering the temporal correlation, it is required to extract samples continuously according to the time axis. Due to the continuity of weather system, wind power is non-stationary, which is also reflected in the short-term prediction error distribution of wind power changing with time.

Hierarchical clustering algorithm can be divided into the aggregation type from the bottom to the top and the split type from the top to the bottom. There is no better or worse of the two methods. Only in the practical application, the decomposition order should be considered according to the data characteristics. The advantage of hierarchical clustering algorithm is that it can get a complete cluster tree directly, and the number of classification can be increased or reduced without recalculation of the distance between different categories.

This paper uses hierarchical clustering method to cluster sample data, the specific steps are as follows:

1) The feature matrix Q of the sample is constructed by using 100m high wind speed and wind direction in the data with time and space characteristics.

2) The DM(Mahalanobis Distance) between the initial samples is calculated and the distance matrix D is obtained. The formula is as follows:

$$D_M(x, y) = \sqrt{(\boldsymbol{x} - \boldsymbol{y})^T \sum{}^{-1}(\boldsymbol{x} - \boldsymbol{y})} \tag{1}$$

X and Y represent two different data categories, and σ is the covariance matrix of multidimensional random variables. If covariance matrix is unit vector, that is, each dimension is distributed independently, the Markov distance becomes Euclidean distance.

3) The minimum DM classes are condensed into a new category by calculating the spacing.

4) Repeat steps 1) to 3) until the complete cluster tree is obtained, that is, the number of categories is 1.

2.2 Joint probability density distribution model

Traditional wind power interval prediction only considers the distribution of prediction error under different power levels, ignoring the influence of meteorological factors on prediction error. In this paper, combined with the characteristics of wind power prediction error time series distribution under different spatial wind conditions, the prediction performance of the model is improved, and the empirical distribution estimation method is used to establish the joint probability density distribution model of wind speed and wind power prediction error under different spatial wind conditions. The wind speed is discretized into four categories ([0,4), [4,8), [8,12), [12,16]). Taking the wind speed value as the sample classification condition of statistical interval, four sub sample sets are formed correspondingly. The cross section of the joint probability density distribution map is the probability distribution of the error under the NWP wind speed at the time to be predicted. The joint probability density distribution of wind speed power prediction error obtained by classifying NWP wind speed is shown in Figure 1. It can be seen that there are significant differences in the error distribution corresponding to different prediction power intervals.
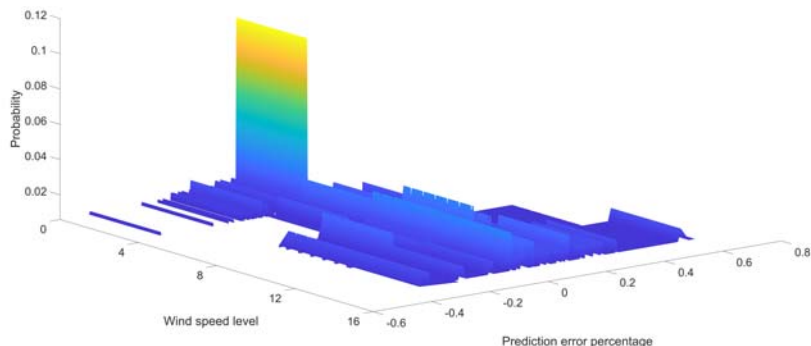
Fig. 1. Joint probability density distribution of wind speed power prediction error

3 Build prediction model

3.1 Probabilistic prediction of short term wind power considering temporal and spatial dependence of prediction error

Because the empirical distribution is a discontinuous probability density distribution function, it is difficult to use the traditional quantile regression method based on continuous distribution function. However, for the uniform distribution of cumulative density function on (0,1), bootstrap sampling method can be used to sample, and then the power uncertainty fluctuation interval can be obtained according to the given confidence level. The specific steps are as follows:

1) According to the historical deterministic prediction error, the probability density distribution model of wind power considering the temporal and spatial dependence of the error is established.

2) According to the NWP wind speed information of the time to be predicted, the error cumulative probability density distribution function under the wind condition is matched.

3) Bootstrap sampling is used to sample the cumulative probability density, and then the corresponding error sequence is obtained according to the cumulative probability density. Finally, the sampling error sequence is reordered according to the size of the cumulative probability density.

4) Given the confidence interval, such as 90%, the power prediction interval at this time can be obtained according to the error sequence distribution.

The short-term interval prediction process of wind power based on bootstrap sampling is shown in Figure 2.
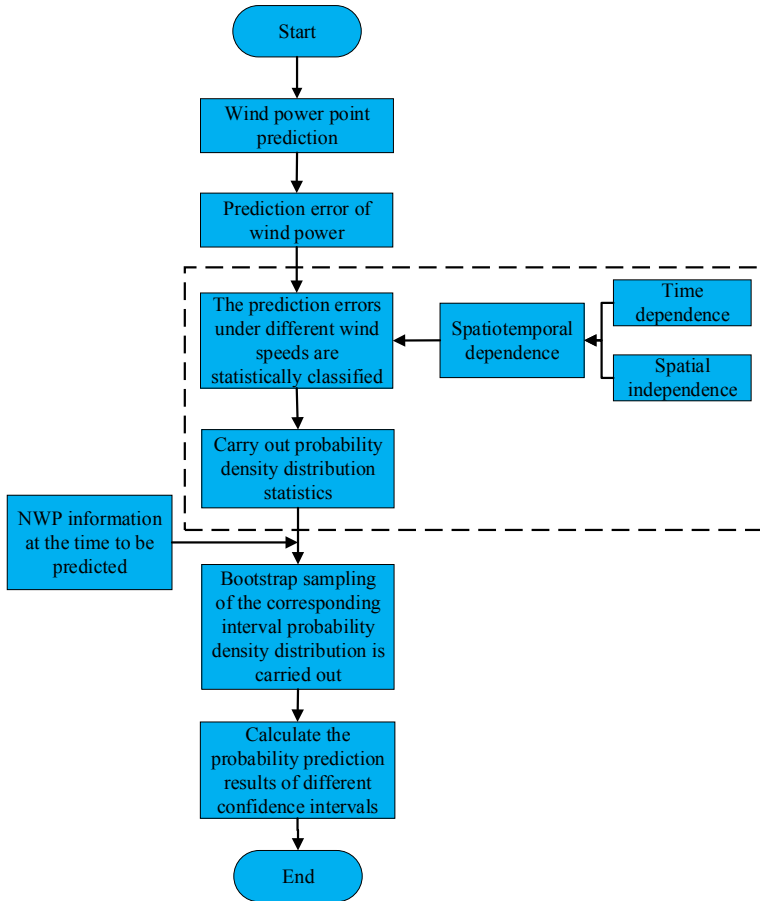
Fig. 2. Probabilistic short term wind power forecasting process considering temporal and spatial dependence of forecasting errors

3.2 Evaluating indicator

In this paper, the interval coverage, average bandwidth and reliability index are selected to analyze and evaluate the short-term wind power interval prediction results.

The index of PI(Predictive Index) coverage probability (PICP) reflects the probability that the actual power falls within the prediction range, which can evaluate the reliability of the prediction model.

$$PICP = \frac{1}{N} \sum_{i=1}^{N} k_i^{(a)} \qquad (2)$$

Where: n is the number of points to be predicted, this paper takes 96; α is the given confidence; K is a Boolean quantity, when k = 1, it means that the actual power value falls into the prediction interval; when k = 0, it means that the actual power value falls out of the prediction interval. In practice, the predicted interval coverage should exceed the given confidence level as much as possible. If the PICP value is less than the confidence α, the prediction is invalid; on the contrary, the prediction is effective, and the larger the PICP

value is, the greater the probability of the actual power falling between the upper and lower limits of the prediction is, and the better the prediction effect is.

The index of PI normalized average width (PINAW) can evaluate the clarity of prediction model. It is used to reflect the average value of the width between the upper and lower limits of the prediction.

$$PINAW = \frac{1}{NR} \sum_{i=1}^{N} \left[U_t(x_i) - L_t(x_i) \right] \tag{3}$$

Where R is the change range of predicted power, which is the normalized reference value. When the PICP values of the prediction results are the same, the smaller PINAW value corresponds to the better prediction effect.

The reliability index R is the difference between the interval coverage rate and the preset confidence level. If the actual wind power curve falls within the prediction range, the prediction result is effective and reliable.

$$r^{(\alpha)} = PICP - \alpha = \frac{1}{N} \sum_{i=1}^{N} k_i^{(\alpha)} - \alpha \tag{4}$$

The reliability level of the model is judged by the positive and negative of the reliability index. When the reliability index R is greater than zero, it means that the model has a favorable deviation and the reliability is higher than the given confidence level; when the reliability index R is less than zero, it means that the model has a harmful deviation and the reliability is lower than the given confidence level.

4 Example analysis

In this paper, the data of a wind farm with 400.5MW installed capacity in Jilin Province of China is selected for example analysis. The prediction error data of the last month of each quarter in 2017 is used for the probability density statistics of error, and the data of the last month of each quarter in 2018 is used for the short-term interval prediction of wind power.
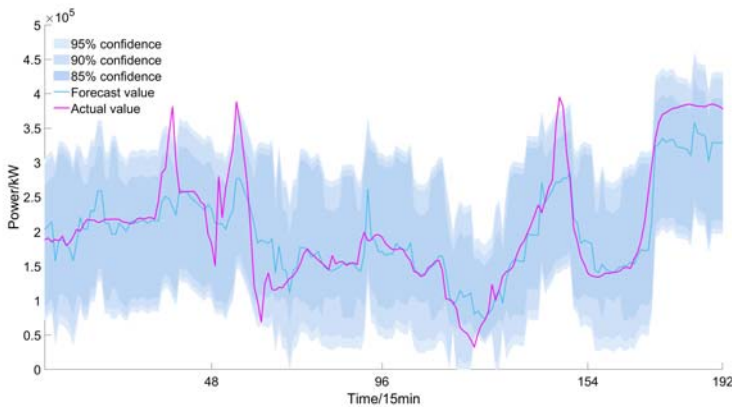

Fig. 3. Forecast effect picture of two days in spring of 2018

Figure 3 shows the forecast effect of two days in the last month of spring in 2018.It can be seen that under the same confidence level, the probability prediction model based on NWP wind speed and bootstrap sampling can closely follow the trend of wind power

sequence, and get a narrower average bandwidth and higher interval coverage, which can provide more accurate prediction information for decision makers.

In order to further study the performance of this model, the prediction and evaluation indexes of quantile regression method, kernel density estimation method and bootstrap resampling method are calculated respectively. The results are shown in Table 1.

Table 1. Comparison of prediction results of different methods under wind speed segmentation

	Confidence interval	PICP	PINAW
Quantile regression	85%	84.83	27.13
	90%	90.34	29.52
	95%	95.21	34.68
Kernel density estimation	85%	85.17	26.84
	90%	90.20	29.85
	95%	95.06	34.93
Bootstrap resampling	85%	85.52	26.58
	90%	90.77	29.34
	95%	95.83	33.28

It can be seen from Table 1 that the short-term probability prediction method based on NWP wind speed classification and bootstrap sampling proposed in this paper is the best, and the interval coverage rate greater than the preset confidence level can be obtained under different confidence levels. Compared with other nonparametric methods, the average interval coverage is increased by 1.62%, and the average bandwidth is reduced by 2.76%. At the same time, it shows that the accuracy of probability prediction can be effectively improved by considering the temporal and spatial dependence of errors.

5 Conclusion

In this paper, a new empirical distribution model of combined probability density of errors based on historical error data and NWP wind speed classification is established. Based on this model, a short-term wind power probability prediction method is proposed considering the dependence of prediction error time and space. The conclusion is as follows:

1) Considering the spatiotemporal dependence, the statistical analysis of errors can better reflect the adaptability of the model in different wind conditions, and improve the prediction accuracy of probability prediction.

2) Compared with the quantile regression method and the kernel density estimation method, bootstrap method has higher reliability under different confidence levels, and the interval coverage rate is increased by 1.62% on average.

In this paper, the statistical characteristics of meteorological factors and errors are fully considered in the construction of the prediction model, but the environmental factors such as topography and geomorphology will indirectly change the wind power output by changing the wind conditions, which is not involved in this paper. In the future, the uncertainty analysis of wind power prediction will focus on the comprehensive consideration of the statistical characteristics of prediction error, the dynamic changes of physical weather process, topography and other environmental characteristics.

Acknowledgements

References

1. A. Costa, A. Crespo, J. Navarro, et al. Renew. Sust. Energ. Rev., *A review on the young history of the wind power short-term prediction*, **12**(6):1725-1744(2008)

2. A. Khosravi, S. Nahavandi, D. Creighton. IEEE Trans Sustain Energy, *Prediction intervals for short term wind farm power generation forecasts*, **4**(3): 602-610(2013)

3. M. Yang, C.Y. Shi, H.Y. Liu., Energy, *Day-ahead wind power forecasting based on the clustering of equivalent power curves*, **218**(2021)

4. P. Pinson, G. Kariniotakis., IEEE Trans Power Systems, *Conditional prediction intervals of wind power generation*, **25**(4): 1845-1856(2010)

5. A. Yamaguchi, K. Enoki, T. Ishihara, et al., J Wind Engineer, *Wind power forecasting with physical model and multi time scale model*, **2007**: 251-264(2010)

6. P. Pinson, H. Madsen. J. Forecast, *Adaptive modeling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models*, **31**(4): 281-313(2012)

7. K. Bruninx, E. Delarue. IEEE Trans Sustain Energy, *A statistical description of the error on wind power forecasts for probabilistic reserve sizing*, **5**(3): 995-1002(2014)

8. J. Juban, L. Fugon, G. Kariniotakis. Eur. Wind Energy C. E., *Probabilistic short-term wind power forecasting based on kernel density estimators*, **11**(2007)