Uncertainty analysis of wind power prediction based on quantile regression

Yongqian Liu, Jie Yan*, Shuang Han
State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (North China Electric Power University)
Beijing, China
jiejie3000@sina.com

Yuhui Peng
Delft University of Technology
Delft, the Netherlands

***Abstract*—Short-term wind power prediction is an effective way to mitigate the impact of large-scale wind power variability incurring to the electric power system. Given the fluctuation of wind energy is random; uncertainty analysis of wind power prediction is very important for engineering application. A risk assessment index of wind power prediction named PaR (Predict at Risk) was proposed based on quantile regression. And an uncertainty analysis model for wind power prediction was established to provide a possible fluctuation range of predicted wind power at any confidence level. Operation data and predicted power from a wind farm in north China are used as a test case to validate proposed model. The results show that the model can tolerate a wide range of different conditions without hypothesis of the error distribution of wind power prediction and is an effective, practical way to provide uncertainty information.**

Keywords-uncertainty analysis; PaR; quantile regression; wind power prediction; risk index.

I. INTRODUCTION

With the rapid expansion of the scale of wind power development, the fluctuation of wind power output has posed serious challenges to the secure, stable and economic operation of the electric power system. Short-term wind power prediction is an effective way to solve this problem. However, wind power fluctuation is random, so it is particularly important to provide uncertainty information of wind power prediction to facilitate decision-making for electric power system operators [1-2].

Over the past decade, many scholars' focus has gradually shifted to the uncertainty analysis of wind power prediction, and achieved fruitful results. Lange and Heinemann [3] identified relations between typical weather situations and the magnitude of the forecast error. Based on the above conclusion, they classified weather into several types utilizing principal component analysis method and cluster analysis method to obtain the prediction error and confidence interval. Han Shuang [4] established an uncertainty assessment model of conditional probability based on Independent Component Analysis Method. This model can provide the probability of each predicted wind power under a certain confidence interval. P.Pinson and G.N.Kariniotakis [5] used consecutive forecasts, and base on these, to define a quantity called the Meteo-Risk index'. This quantity measures the agreement between the consecutive forecasts and is used to predict the uncertainty of the wind power forecast [6]. Above methods suffers from poor practicability and accuracy[7], and some are on the basis of the hypothesis that prediction errors follow Gaussian distribution. However, it is not the case for wind power prediction errors, because wind power prediction errors always exhibit some skewness, and confidence intervals of prediction point are asymmetric[8].

To overcome above shortcomings, quantile regression without hypothesing the distribution of prediction errors has been introduced in uncertainty analysis of wind power prediction. It probably reflects the actual prediction errors distribution more accurately. Moreover, its less calculation cost ensures practicability.

In section II, quantile regression theory and its advantages are illustrated. In section III, inputs and structure of uncertainty analysis model based on quantile regression is established. Then PaR which is the risk assessment index of wind power prediction is proposed. In section IV, the actual data and predicted power from a wind farm in North China is used to verify the proposed method and assessment index. Finally, the paper carries out the conclusion in section V.

II. QUANTILE REGRESSION

A. Theory of quantile regression

Considering a random variable Y, all the properties of it can be described by its distribution function, as

$$F(y) = P(Y \leq y) \quad (1)$$

The definition of τ quantile function $Q(\tau)$ is written as

$$Q(\tau) = inf\{y: F(y) \geq \tau\} \ , \quad 0< \tau <1 \quad (2)$$

It can be seen that the ratio of which is located below the quantile function $Q(\tau)$ is τ, while the ratio of which is located beyond the quantile function $Q(\tau)$ is 1- τ.

Sponsors: National High Technology Research and Development Program (863 Program) (No.2007AA05Z428) and Smart Grid National High Technology Research and Development Program (863 Program) (No.SQ2010AA0523193001)

Given the "Test Function" $\rho(u)$

$$\rho(u) = \left(\tau - l_{(u)}\right)u = \begin{cases} \tau u & u \geq 0 \\ (\tau - 1)u & u < 0 \end{cases} \quad (3)$$

Where $l_{(u)}$ is indicative function and $0< \tau <1$.

Expression (3) show that τ is equivalent to the slope of the line and Test Function itself is piecewise and non-negative.

To facilitate integration, the Test Function can be rewritten as $\rho(u) = \tau u l_{(u\geq 0)} + (\tau - 1)u l_{(u<0)}$ (4)

When $u = y - \delta$, Function (4) can be rewritten as

$$\rho(y - \delta) = \tau(y - \delta)l_{(y-\delta\geq 0)} + (\tau - 1)(y - \delta)l_{(y-\delta<0)} \quad (5)$$

The τ quantile regression of Y can be down to find δ which can minimize $E\big(\rho(y - \delta)\big)$, namely $min_{\delta\in R}\, E\big(\rho(y - \delta)\big)$.

Taking the expectation of Function (5), then the integral of it is as $min_{\delta\in R}\, E\big(\rho(y - \delta)\big)$

$$= (\tau - 1)\int_{-\infty}^{\delta}(y - \delta)dF(x) + \tau\int_{\delta}^{+\infty}(y - \delta)dF(x) \quad (6)$$

The deviation of above function for τ is as

$$0 = (1 - \tau)\int_{-\infty}^{\delta} dF(x) - \tau\int_{\delta}^{+\infty} dF(x) = F(\delta) \quad (7)$$

As distribution function, F is a monotone increasing function, on a certain interval any element from aggregate {y: $F(\delta) = \tau$ } can be achieved to minimize $E\big(\rho(y - \delta)\big)$. By the definition of quantile, there is only one $\hat{y}$ obtained when $Q(\tau) = \hat{y}$.

In a similar way, for the linear conditional mean function $E(Y|X = x) = x'\beta$, a general linear conditional quantile function can be described as $Q(\tau|x) = x'\beta(\tau)$, where $0<\tau<1$.

For a random sample of variable $Y\{y_1, y_2, \dots, y_n\}$, sample linear quantile regression of general quantile requires satisfying Function (8), as

$$min_{\beta\in R}\sum_i \rho\tau\big(y_i - x_i{'}\beta(\tau)\big) \quad (8)$$

$$= \min_{\beta\in R}\left[\sum_{i:y_i\geq x_i{'}\beta(\tau)} \tau\left|y_i - x_i'\beta(\tau)\right| + \sum_{i:y_i<x_i{'}\beta(\tau)} (1 - \tau)\left|y_i - x_i'\beta(\tau)\right|\right]$$

Where $0 < \tau < 1$.

Parameter estimate can be obtained by solving

$$\hat{\beta} = argmin\sum_{i=1}^{n}\rho_\tau\left(y_i - x_i'\beta(\tau)\right) \quad (9)$$

For any $0 < \tau < 1$, only one $\hat{\beta}(\tau)$ can be achieved as τ quantile.

Currently, there are 3 recognized effective algorithms for computing quantiles, and in the statistical software such as SAS, R and Splus package, quantile regression can be achieved [9].

B. *Characteristic of quantile regression*

Quantile regression was first proposed by Koenker and Bassett (1978) [10]. It is an extension of least squares, and uses several quantile functions to estimate the overall model. Median regression is a special case of the quantile regression, using symmetric weights to solve the residual minimization problem. While, other conditional quantile regression need to be solved by minimizing the non-symmetric weights residuals. As quantile regression method estimates parameters by minimizing weighted sum of absolute errors, there are some advantages as follow:

- It shows strong robustness, because it is not necessary for quantile regression model to assume the distribution of random disturbance;Continuous functions are not used to describe the relationship between mean and variance of dependent variable, which brings about good flexibility[11];
- Different types of explanatory variables can be introduced to improve the accuracy of quantile estimate, which can diversify the shape of conditional distribution of the local sample;
- Any quantile can be estimated. And it is able to provide with more detailed information of variable conditional distribution, rather than the mean information given by traditional regression model [12].

III. UNCERTAINTY ANALYSIS MODEL

A. *Model input*

Considering the process of wind power prediction, there are two sources of uncertainty, that is, the input of prediction model and the model itself (model structure and model parameters, etc).

To take the popular Numerical Weather Prediction-based prediction model as example, the uncertainty of wind power prediction primarily stems from NWP errors and prediction model itself. However, the uncertainty level from NWP errors would be cube-fold enlarged because of the nonlinear transformation of wind speed to wind power by power curve. Therefore, NWP data should be involved into the input of uncertainty model, including wind speed, wind direction, temperature, pressure, relative humidity. Moreover, being the output of prediction model and foundation of uncertainty analysis, prediction power partly reflects characteristic and uncertainty of prediction model. It is reasonable for prediction wind power to be one of the input vectors of uncertainty model. Finally, because the uncertainty analysis model calculates possible fluctuation of prediction power under a certain confidence level, it should also contain confidence level into the input group.

B. *Model structure*

The uncertainty analysis model is a 8-dimentional input and 1-dimentional output model involving uncertainty from both the input of prediction model and model itself. Input vectors are wind speed, sine of wind direction, cosine of wind direction, pressure, temperature, relative humidity, prediction wind power and confidence level, while output vector is risk index. As mentioned above, the structure of uncertainty analysis model is as Figure. 1:

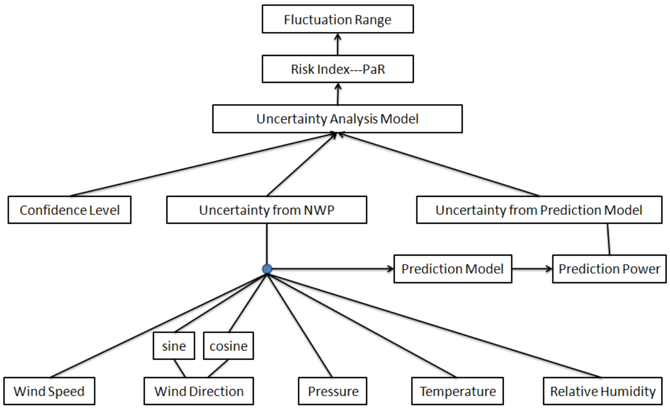

Figure 1. Uncertainty Analysis Model Structure

C. *Risk assessment index*

Risk assessment index of wind power prediction termed Prediction at Risk (PaR) represents the maximum error of predicted wind power that may occur at a given confidence level. PaR is able to simplify uncertainty of wind power prediction brought by a variety of uncertain elements using a simple figure. The essence of PaR is the quantile value of relative error of predicted power under given confidence level. As mentioned in 3.1, this quantile value of prediction error is closely related to every uncertainty source. Therefore, PaR = $e_{max} = Q(\tau)$.

For the predicted wind power P_{prep} obtained through short-term wind power prediction method and actual power value P_{act}, relative predicted error at time of t is $e_t = (P_{prep} - P_{act})/Cap$, where Cap is rated capacity of wind turbine. Then, distribution function can be described as:

$$F(e) = Prob(e \leq e_t) \tag{10}$$

For any $0 < \tau < 1$, the quantile function of prediction error $Q(\tau)$ can be defined as the minimum value which satisfies $F(e) \geq \tau$.

$$Q(\tau) = inf\{e: F(e) \geq \tau\} \tag{11}$$

Therefore, the definition of quantile function is as $Prob(Q(\tau) \geq e) = \tau$ ($0 < \tau < 1$). It means that under confidence level of τ, possible upper fluctuation percentage is $Q(\tau)$. When $Prob(Q(1-\tau) \geq e) = 1-\tau$, that is, $Prob(Q(1-\tau) \leq e) = \tau$. Its physical meaning is that under confidence level of τ, possible lower fluctuation percentage is $Q(1-\tau)$.

Therefore, according to the theory of quantile regression, quantile function could be described as:

$$Q(\tau) = \beta_0 + \beta_{ws}x_{ws} + \beta_{sin}x_{sin} + \beta_{cos}x_{cos} + \beta_{pres}x_{pres} + \beta_{tem}x_{tem} + \beta_{hum}x_{hum} + \beta_{prep}x_{prep} \tag{12}$$

Where, β_i is the vector of model parameters; x_{ws} represents the wind speed vector; x_{sin}, x_{cos} represent sine and cosine of wind direction; x_{pres} denotes pressure; x_{tem} denotes temperature; x_{hum} indicates relative humidity; x_{prep} indicates predicted wind power.

IV. CASE STUDY

A. *Data*

Taking a wind farm in northern China as example, the actual wind power data in 2006 and its predicted value data which are obtained by short-term prediction method RBF are used in this paper. Resolution of the whole data is 15min. Data from 1st July to 31th Sep. 2006 are used for developing and training the models. Data from Nov. 2006 are used to test the uncertainty analysis results.

B. *Process*

1) Normalization: To normalize input vectors for simplifying calculation except sine and cosine of wind direction and relative humidity which have already ranged between 0 and 1. The normalization formula is as follow:

$$x_{0-1} = \frac{x - min(x)}{max(x) - min(x)} \tag{13}$$

2) Iteration: To set confidence level τ for upper limit model and $1-\tau$ for lower limit model to calculate all the unknown coefficients $\beta(\tau)$ in each model by iteration;

3) Calculation of PaR: To put unknown input data and confidence level into developed model to obtain PaR value of upper limit and lower limit respectively which are the maximum errors of predicted power;

4) Calculation of fluctuation: To Multiply PaR which is maximum predicted error with corresponding predicted power to transform it into fluctuation value of upper limit and lower limit respectively;

5) Calculation of power limit: To to obtain upper limit and lower limit under τ by following formulas:

$$upper\ limit = Predicted\ power \times (1 + Q(\tau)) \tag{14}$$

$$lower\ limit = Predicted\ power \times (1 + Q(1-\tau)) \tag{15}$$

C. *Results and discussion*

Validity Check: at confidence level of 85%, 81.25 per cent of actual power points located between the upper limits of predicted power and the lower limits of predicted power, while 91.25per cent located below the upper limits. At confidence level of 75%, 69.56 per cent of actual power value located within the upper and lower limits of prediction power, while 86.33 per cent located below the upper limits.

Table 1shows every coefficient for upper and lower power limits at 85% and 75% confidence level. Table 2 presents mean value of uncertainty analysis results at two confidence level.

Figure 2 to 4 depict test results of Nov. 12th in 2006. Figure 3 shows that with the increase of the confidence level, the range of power fluctuations increases; at confidence level of 85%, average fluctuation value is 567.62kW, while at the confidence level of 75%, it is 451.61kW. That is because higher confidence level needs larger possible fluctuation to response to higher risk.

TABLE I. MODEL COEFFICIENT RESULTS

β \ τ	0.85	0.15	0.75	0.25
β_0	-4.78910	0.24067	-5.24907	0.23266
β_1	-0.06625	0.63043	-0.06438	0.72264
β_2	-0.20004	0.00756	-0.14492	0.00607
β_3	-0.04141	0.00581	-0.02446	0.00196
β_4	7.53470	0.00554	7.91360	0.00404
β_5	0.00318	-3.62371	0.00179	-3.36587
β_6	-0.09760	-9.23633	-0.05902	3.4979 2
β_7	0.39065	-0.02918	0.21372	-0.03017

TABLE II. UNCERTAINTY ANALYSIS RESULTS

Confidence level	Mean Value of Uncertainty Information			
	τ	*PaR*	*Upper limit (kW)*	*Lower limit (kW)*
85%	0.85	0.37247	1032.574	
	0.15	0.20181		171.145
75%	0.75	0.29514	916.566	
	0.25	0.17443		212.205

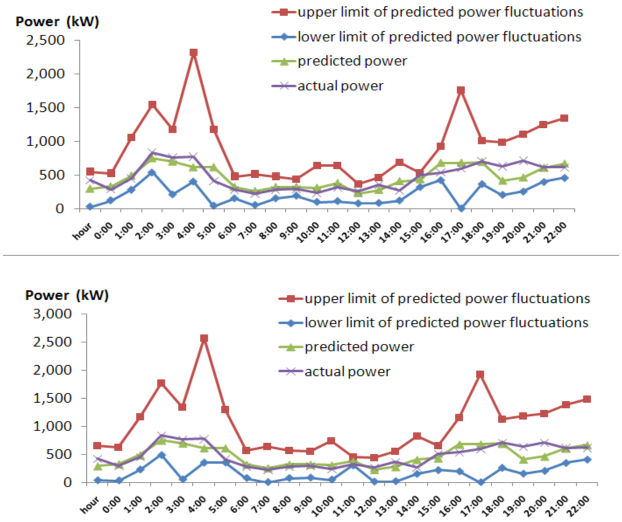

Figure 2. Upper/Lower Power Limits and Actual Power Value at Confidence Level of 85% (up) and 75% (low)

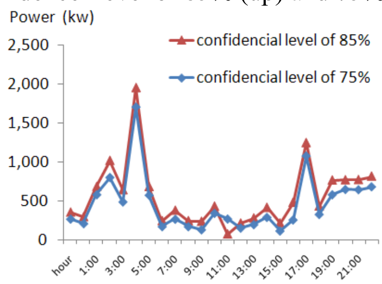

Figure 3. Fluctuations of Predicted Power at Confidence Level of 85% and 75%

V. CONCLUSION

- PaR (Predict at Risk), a risk assessment index for wind power prediction is proposed. This index can be used for quantitative assessment of the largest wind power prediction error at any confidence level.
- An uncertainty analysis model of wind power prediction has been established based on quantile regression. At given confidence interval, parameter estimation method was applied. Considering both the uncertainty from NWP and prediction model, NWP and predicted power from RBF model are taken as the input of the quantile regression model. This model can provide a possible fluctuation range of predicted wind power output at given confidence interval.
- It is unnecessary to hypothesize the error distribution of predicted wind power in this proposed model. Therefore it has a higher possibility to reflect the actual forecast error distribution on a more accurate level.
- The results of a test case for a wind farm in north China show that the proposed index and model can be used to effectively and practically analyze the uncertainty of wind power prediction.

ACKNOWLEDGMENT

This work is supported by two Chinese National High-Tech Research and Development projects (2007AA05Z428 and SQ2010AA0523193001).

REFERENCES

[1] Claus Krog Ekman. On the synergy between large electric vehicle fleet and high wind penetration – An analysis of the Danish case. Renewable Energy, Volume 36, Issue 2, February 2011, p546-553

[2] Andrew Kusiak*, Wenyan Li. Short-term prediction of wind power with a clustering approach. Renewable Energy, Volume 35 (2010), p2362-2369

[3] Lange M, Heinemann D. Relating the uncertainty of short-term wind speed predictions to meteorological situations with methods from synoptic climatology. Proceedings of European Wind Energy Conference & Exhibition, Madrid, 2003

[4] Han Shuang, Yang Yongping, Liu Yongqian. Study of Short-Term Wind Power Prediction Method. PhD thesis of Engineering in Power Engineering in North China Electric Power University, 2008.6

[5] Pinson P, Kariniotakis G. On-line assessment of prediction risk for wind power production forecasts. Wind Energy, 2004; 7: 119–132

[6] H.A Nielsen, H. Madsen, T.S. Nielsen. Using Quantile Regression to Extend an Existing Wind Power Forecasting System with Probabilistic Forecasts. Wind Energy, 2006; 9:95-108

[7] Gregor Giebel, Lars Landberg. Short-term (30min-72hrs) prediction for power output from wind farms. http://powwow.risoe.dk/Bestpracticeworkshop.htm.

[8] George N. Kariniotakis, Member, IEEE and Pierre Pinson, Uncertainty of Short-term Wind Power Forecasts - A Methodology for On-line Assessment, 8th International Conference on Probabilistic Methods Applied to Power Systems, Iowa State University, Ames, Iowa, September 12-16, 2004

[9] Li Yu'an. Introduction to Quantile Regression and Application. Statistics & Information Forum, Vol.21 No.3, 2006.5:89-96

[10] Chen Jianbao, Ding Junjun. Review of Quantile Regression Technology. Statistics & Information Forum, Vol.23 No.3, 2008.3:89-96

[11] Koenker R, gassett G W. Regression quantiles [J]. Econornetrica, 1978, 46: 33—50

[12] Wang Xinyu, Zhao Shaojuan. Quantile Regression Based Risk Measurement in the Chinese Stock Markets. Journal of China University of Mining & Technology. Vol.37 No.3 May 2008