Send Orders for Reprints to reprints@benthamscience.ae

The Open Electrical & Electronic Engineering Journal, 2014, *8*, 501-507

Open Access

A Stat istical Model for Wind Power Forecast Error Based on Kernel Density Estimation

Liyang Liu*, Junji Wu and Shaoliang Meng

School of Energy and Power Engineering, Nanjing University of Science and Technology, Xuanwu District, Nanjing, 210094, China

Abstract: Wind power has been developed rapidly as a clean energy in recent years. The forecast error of wind power, however, makes it difficult to use wind power effectively. In some former statistical models, the forecast error was usually assumed to be a Gaussian distribution, which had proven to be unreliable after a statistical analysis. In this paper, a more suitable probability density function for wind power forecast error based on kernel density estimation was proposed. The proposed model is a non-parametric statistical algorithm and can directly obtain the probability density function from the error data, which do not need to make any assumptions. This paper also presented an optimal bandwidth algorithm for kernel density estimation by using particle swarm optimization, and employed a Chi-squared test to validate the model. Compared with Gaussian distribution and Beta distribution, the mean squared error and Chi-squared test show that the proposed model is more effective and reliable.

Keywords: Wind power; Forecast error, Probability density function, Kernel density estimation, Optimal bandwidth.

1. INTRODUCTION

As the environmental issues catch the attention of public, many countries are beginning to increase their investments in clean energy. As a mature technology, wind power has been widely used in recent years, especially in China, the total installed capacity of wind power is close to 90 gigawatts by 2014. Up to now, a large number of wind farms have been connected to the power grids, which brings about many negative effects, for example, the fluctuations of wind power make the power system more vulnerable and increase the difficulty of electric power dispatching. In addition, the power system must keep a large reserve capacity to cope with the fluctuations, which will eventually lead to higher costs. The forecast error costs usually can reach as much as 10% of the total incomes from generated energy [1]. As a result of this, many wind power forecast methods have been developed to handle these problems [2, 3]. A correct forecast method can improve the stability of the power system and reduce the operation costs and largely reduce the effect of power fluctuations, so it's very helpful to optimize the usage of wind energy. Currently, various forecasting methods are being used and developed, ranging from basic persistence methods to complex statistical models [4]. But so far none of these methods have proven to be precise enough to be equal to actual values. Therefore, the forecast error is inevitable. Many researchers have paid much attention to the forecast methods, but they ignored to study the forecast error. This paper focuses on the modeling of wind power forecast error, the probability density function (PDF) of forecast error can be found by analyzing the error data, and the statistical law of errors can be obtained in this process. This will help improve the accuracy of wind power forecasting.

Actually, only few papers have studied the wind power forecast error. The forecast error was generally assumed to be a Gaussian distribution [5, 6], based on this assumption literature [7] proposed an optimization method for spinning reserves, and in [8], the Gaussian distribution for wind power forecast error is used for system risk management. However, the Gaussian assumption has been challenged in some works [9-10]. Literature [10] points out that it's difficult to find a proper definition for the forecast error PDF and the error PDF is fat-tailed with variable kurtosis, so it cannot be modeled with Gaussian distribution. In the meanwhile, a new PDF, a so-called Beta distribution, is proposed. The analysis of error data shows that the Beta distribution is more suitable than Gaussian distribution. But in some cases, Beta distribution is still not sufficiently heavy-tailed to model the leptokurtosis of the error data, because it will cause a lot of errors. In [4], the Levy skew alpha-stable distribution is proposed as a well-suited statistical model to describe the heavy-tailed character of the error data. This model can provide a more appropriate approach than Gaussian distribution and Beta distribution. There are still few methods for the modeling of the forecast error such as Cauchy distribution [11], copula theory [12], Gamma distribution [13] and mixed distribution [14]. Most of the mentioned methods are based on parametric statistics, which assume that the error data fulfill some kind of known PDF and then try to use this PDF to fit the data. But in fact, the assumed PDF will not fit the actual data in a strict manner, which will lead to significant fitting error.

1874-1290/14 2014 Bentham Open

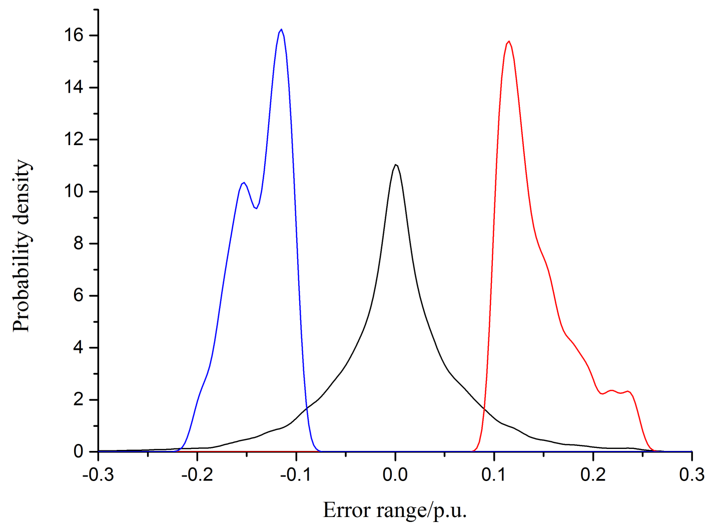

Fig. (1). Different Types of PDF Obtained by KDE.

In this paper, a new forecast error model based on kernel density estimation (KDE) was proposed. KDE is a non-parametric way to estimate the PDF of a random variable. Unlike the parametric statistics, the non-parametric statistics make no assumptions about the probability distributions of the variables assessed. It uses such variables as the training data and the PDF obtained by this method is closer to the real PDF than the parametric way is. When using the KDE method, we should note that one of the most important things is to select a suitable bandwidth, which is a free parameter and has a strong influence on the estimating of results. An adaptive bandwidth selection method is also presented in this paper, which can automatically seek the optimal bandwidth by considering the goodness and smoothness of the fitting curve. Finally, the proposed model is tested by using the forecast error data from a real wind farm, and it will be compared with the Gaussian distribution and Beta distribution. And in the data analysis section, Chi-squared test is used to calculate the fitting goodness of these models.

2. MATHEMATICAL MODEL

2.1. Kernel Density Estimation

Kernel density estimation can be applied to analyze the forecast errors. Let $(X_1,X_2,X_3,\dots X_n)$ be an independent variables from the error data and the PDF of the variables can be represented as $f(x)$. Its kernel density estimator is $\hat{f}(x)$ [15-16],

$$\hat{f}(x)=\frac{1}{N\cdot h}\sum_{i=1}^{N}K\left(\frac{x-X_i}{h}\right) \tag{1}$$

where N is the count of the variables, and h is a smoothing parameter, called bandwidth. $K(\cdot)$ is the kernel function, a non-negative function that integrates to one and has mean zero. It must satisfy the Eq.(2), $u=(x-X_i)/h$. The commonly-used kernel functions are as follows: uniform kernel, Gaussian kernel, triangle kernel, Epanechikov kernel and Biweight kernel. Theoretically speaking, if the N tends to be infinite, no matter what kind of kernel functions is chosen, a reliable PDF can be obtained. But in most cases, N is a limited value. Therefore, different kernel function will have different results even though the difference is not significant.

$$\begin{cases}\int K(u)\mathrm{d}u=1\\ \int uK(u)\mathrm{d}u=0\\ \int u^2K(u)\mathrm{d}u\neq 0\end{cases} \tag{2}$$

The kernel function used in this paper is Gaussian kernel.

$$K(u)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right) \tag{3}$$

Then the KDE expression is

$$\hat{f}(x)=\frac{1}{N\cdot h\sqrt{2\pi}}\sum_{i=1}^{N}\exp\left[-\frac{(x-X_i)^2}{2h^2}\right] \tag{4}$$

Fig. (**1**) shows that the PDF curves are obtained by KDE, their sharps are quite different. So the KDE model is very adaptable and can deal with different types of the error data.

2.2. Bandwidth Selection

An appropriate bandwidth is the key to the successful application of the KDE algorithm. If the bandwidth is too small, the fitting goodness will be better. Otherwise, the fit smoothness will be worse and the fitting curve will produce a lot of noise; on the contrary, if the bandwidth is too large, the fitting goodness will be getting worse and the fitting curve will be smoother, it will lose some density characteristics. So choosing an appropriate bandwidth is very important. Eq.(5) is the Mean Squared Error (MSE) of the KDE, which is an important indicator of bandwidth. In order to get the optimal bandwidth, the MSE should be as small as possible.

$$\mathrm{MSE}(h_n)=\mathrm{E}[(\hat{f}(x)-f(x))^2] \tag{5}$$

The MSE can be divided into two parts: the bias and the variance. Those two parts are contradictory and cannot be reduced at the same time. When the kernel function is determined, the bandwidth is the main factor for MSE.

$$\mathrm{MSE}(h_n)=[\mathrm{Bias}(x)]^2+\mathrm{Var}(x) \tag{6}$$

Due to the fact that the PDF is continuous, Mean Integrated Squared Error (MISE) is usually used to compute the precision of KDE, which is the sum of MSE.

$$\begin{aligned}\mathrm{MISE}(h)&=\mathrm{E}[\int(\hat{f}(x)-f(x))^2\mathrm{d}x]\\&=\int\left\{[\mathrm{Bias}(x)]^2+\mathrm{Var}(x)\right\}\mathrm{d}x\end{aligned} \tag{7}$$

The bias in Eq. (6) and Eq. (7) can be expressed by Eq.(8). Assuming that the PDF is $f(x)$, its second derivative is $f''(x)$, where $m_i(K)=\int u^i K(u)\mathrm{d}u$, and i is a positive integer.

$$\begin{aligned}\mathrm{Bias}(x)&=\mathrm{E}[\hat{f}(x)]-f(x)\\&=\frac{1}{2}f''(x)m_2(K)h_n^2+o(h_n^2)\end{aligned} \tag{8}$$

The variance is

$$\mathrm{Var}(x)=\frac{1}{nh_n}f(x)R(K)+o(\frac{1}{nh_n}) \tag{9}$$

where $R(K)=\int K^2(u)\mathrm{d}u$.

Substituting Eq.(8) and Eq.(9) into Eq. (6), Eq.(10) can be obtained as follows. The $o(\cdot)$ is higher order infinitesimal and it is usually ignored.

$$\begin{aligned}\mathrm{MSE}(h)=&\frac{h_n^4}{4}(f''(x))^2(m_2(K))^2+\frac{R(K)}{nh_n}f(x)\\&+o(\cdot)\end{aligned} \tag{10}$$

Using the extreme value theory, it can obtain that the optimal bandwidth is

$$h_{\mathrm{opt}}\approx n^{-\frac{1}{5}}\left\{[m_2(K)f''(x)]^{-2}f(x)R(K)\right\}^{\frac{1}{5}}+o(\cdot) \tag{11}$$

In theory, the minimum of MSE is shown by Eq. (12).

$$\begin{aligned}\mathrm{MSE}(h_{\mathrm{opt}})=&\frac{5}{4}\left\{f^2(x)[R(K)]^2m_2(K)f''(x)\right\}^{\frac{2}{5}}n^{-\frac{4}{5}}\\&+o(\cdot)\end{aligned} \tag{12}$$

But in fact, Eq. (11) cannot be directly used to obtain the optimal bandwidth, because the $f(x)$ and the $f''(x)$ are unknown. A numerical solution shall be found to solve this problem.

A Particle Swarm Optimization (PSO) based method for optimal bandwidth was proposed in this paper. The steps of this method are as follows:

Step 1: Initialize positions and associate velocity of a number of particles randomly. Each particle represents a potential solution. h_i is the position and $h_i\in H$, where H is the domain of the bandwidth. v_i is the velocity. i is the index number of the particles, $i\in[1,n]$ where n is particle count.

Step 2: Use Eq.(1) to obtain the $\hat{f}(x)$, and let $\hat{f}(x)$ replace the $f(x)$, then use Eq.(7) and Eq.(10) to evaluate the MISE of all particles.

Step 3: Compare the $P_{\mathrm{best}}(i)$ of every particle with its current MISE. $P_{\mathrm{best}}(i)$ is the best position (bandwidth) of a particle in history, which makes the MISE minimum.

Step 4: Determine the current minimum MISE value in the whole population and its coordinates. If the current minimum of MISE is better than that of the G_{best}, assigning the new position to G_{best}. G_{best} is the best position (bandwidth) of all particles in history.

Step 5: Update positions and velocities by using Eq.(13), where ω is the inertia weight. c_1 and c_2 are two positive constants. r_1 and r_2 are two random numbers in the range [0,1] and k is the iteration number [17].

$$\begin{cases}v_i^{(k+1)}=\omega\cdot v_i^k+c_1\cdot r_1\cdot(P_{\mathrm{best}}(i)^k-h_i^k)\\\qquad\quad+c_2\cdot r_2\cdot(G_{\mathrm{best}}{}^k-h_i^k)\\h_i^{(k+1)}=h_i^k+v_i^{(k+1)}\end{cases} \tag{13}$$

Step 6: Repeat from Step 2 to Step 5 until a prespecified number of iterations is completed. Then the G_{best} is the optimal bandwidth.

2.3. Chi-squared Test

Chi-squared test (χ^2 test) is the most widely used method to test the fitting goodness, which can be applied to any univariate distribution. Based on this, the cumulative distribution function (CDF) can be calculated. The χ^2 test is defined for the hypothesis:

H_0: The data follow a specified distribution;

H_1: The data do not follow the specified distribution.

For the χ^2 fitting goodness computation, the data are divided into m bins and the test statistic, which is defined as

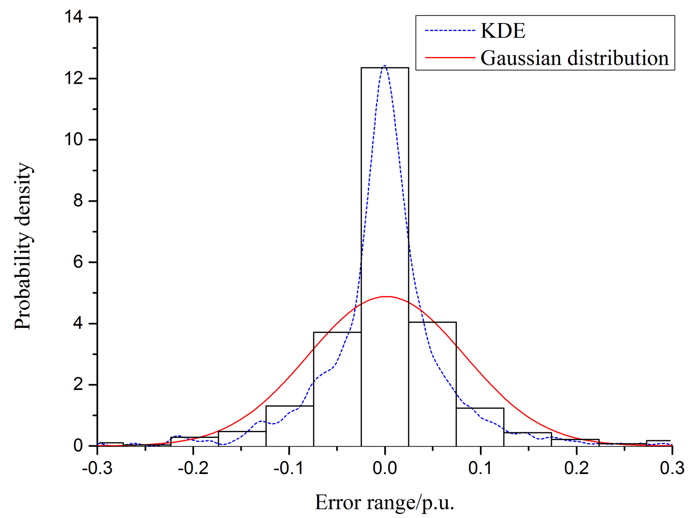

Fig. (2). Error Data with Time Interval of 15 Minutes.

$$\chi^2 = \sum_{i=1}^{m} \frac{(P_i - \hat{P}_i)^2}{\hat{P}_i} \tag{14}$$

where P_i is the observed frequency for bin i and $\hat{P}_i$ is the expected frequency for bin i. The expected frequency is calculated by

$$\hat{P}_i = N \cdot \left(\hat{F}(x_{i,\text{upper}}) - \hat{F}(x_{i,\text{lower}}) \right) \tag{15}$$

where $\hat{F}(\cdot)$ is the CDF. It's the integral of $\hat{f}(x)$. $x_{i,\text{upper}}$ is the upper limit for class i while $x_{i,\text{lower}}$ is the lower limit for class i. N is the sample size.

The hypothesis that the data are from a population with the specified distribution is rejected if

$$W = \{\chi^2 > \chi^2_\alpha(N-1)\} \tag{16}$$

where α is the significance level, W is the critical region. So the χ^2 test is a useful tool to test the validity of the obtained distributions.

3. ERROR DATA ANALYSIS

The forecast error data sets from a real wind farm are used to conduct experiments in this study. The wind farm is located in north China, and the data sets were collected in April 2013. In this section, KDE algorithm will be employed to analyze the error data sets in different time scales (from minutes to hours), and will compare the results with Gaussian distribution and Beta distribution.

Fig. (**2**) analyses the error data of the first week in March 2013, which contains two PDF curves and a histogram. The time interval is 15 minutes. The optimal bandwidth obtained by PSO is 0.008. The horizontal axis is the error range (in per unit) and the vertical axis is the probability density. The histogram is the real density of the error data. As it can be seen from Fig. (**2**), the main error range is from -0.2p.u. to 0.2p.u. The Gaussian distribution does not fit the histogram very well, because it loses some important features of the error data. But the KDE is much more suitable for data sets with heavy-tailed character and its fitting goodness is much better than Gaussian distribution.

In order to study the different time scales of the error data, Fig. (**3**) shows the error density with a time interval of 1 hour. The optimal bandwidth is 0.01. The result from Fig. (**3**) is more similar to Fig. (**2**). The PDF obtained by KDE is very flexible and fits the error data very well. In the Fig. (**3**), the KDE curve is close to the histogram, which means that it approaches to the real PDF of the error data sets.

Table **1** is the comparison between KDE and Gaussian distribution. The MSE of KDE is lower than that of Gaussian distribution, which indicates that the KDE is more accurate. Let the significance level $\alpha = 0.05$, when $N > 30$. The value of the critical region is usually larger than 13. So the KDE will pass the χ^2 test, but the Gaussian distribution will be rejected. The χ^2 statistic of Gaussian distribution is much larger than that of KDE, so the Gaussian distribution is not suitable for modeling of the forecast error in this case.

As the Beta distribution is ranging from 0 to 1, it cannot estimate the negative errors, which is a big flaw for forecast error modeling. It usually divides the error range into several intervals. For the purpose of comparison, Fig. (**4**) is the error interval ranging from 0.05p.u. to 0.15p.u. of the error data in April 2013 and the time interval is 15 minutes. Fig. (**4**) shows that the shapes of the KDE and the Beta distribution are similar, but there are still some differences between them. The KDE can preserve the details of the PDF, and its fitting goodness is better than Beta distribution.

Fig. (**5**) is the error interval from 0.15p.u. to 0.25p.u. and the time interval is 1 hour. From the Fig. (**4**), it can be

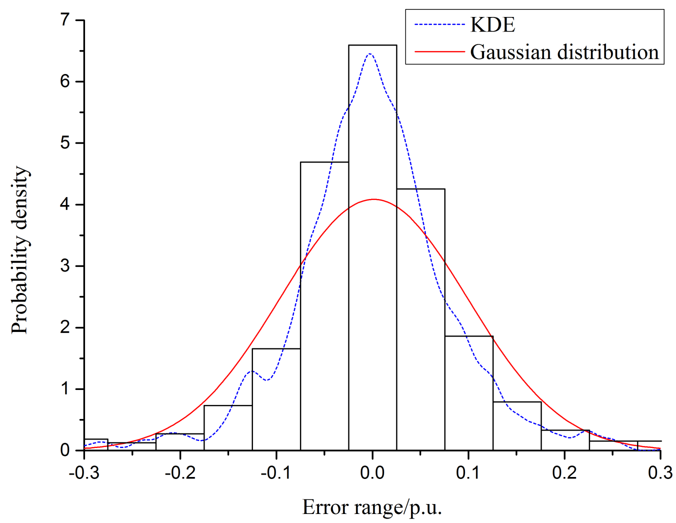

Fig. (3). Error Data with Time Interval of 1 Hour.

Table 1. The Comparison Between KDE and Gaussian Distribution.

Cases	PDF	MSE	χ^2 statistic
Fig.2	KDE	0.1532	1.3282
	Gaussian distribution	1.0276	32.2213
Fig.3	KDE	0.1054	0.8725
	Gaussian distribution	0.9328	21.8576

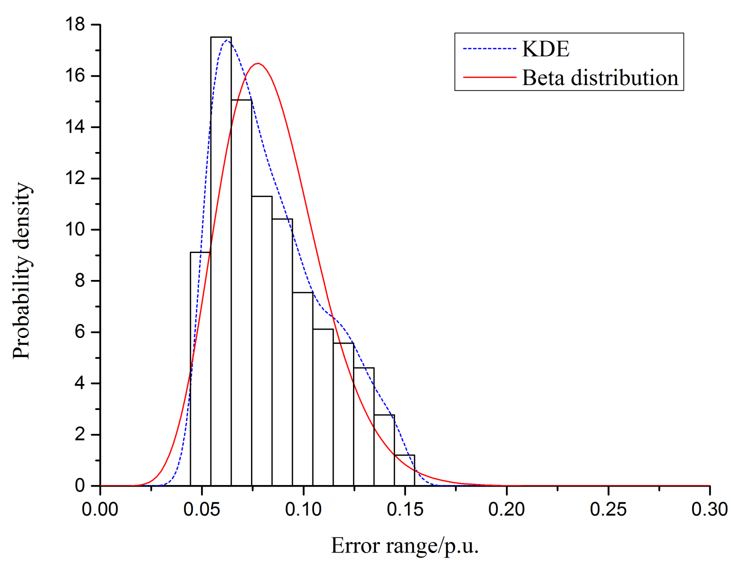

Fig. (4). Error Range from 0.05p.u. to 0.15p.u. with Time Interval of 15 Minutes.

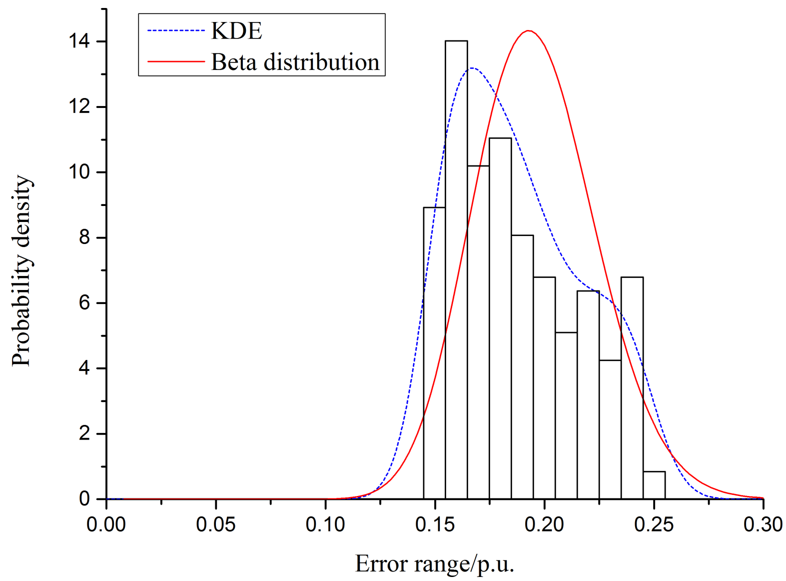

Fig. (5). Error Range from 0.15p.u. to 0.25p.u. with Time Interval of 1 Hour.

Table 2. The Comparison Between KDE and Beta Distribution.

Cases	PDF	MSE	χ^2 statistic
Fig.4	KDE	0.1041	1.1362
	Beta distribution	0.9363	20.7536
Fig.5	KDE	0.1135	1.8169
	Beta distribution	1.2143	36.7546

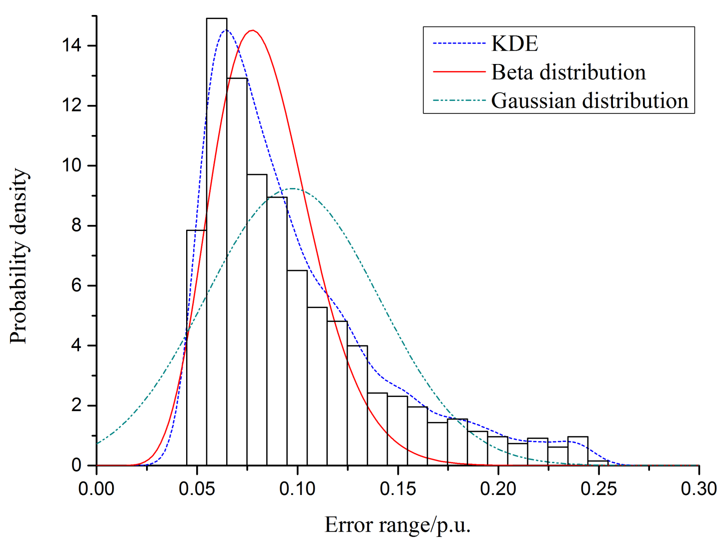

Fig. (6). Error Range from 0.05p.u. to 0.25p.u. with Time Interval of 15 Minutes.

observed that the KDE has a better fitting goodness. But the Beta distribution is obviously inappropriate.

The results of comparison between KDE and Beta distribution are shown in Table **2**. All the statistical indicators of KDE are better than those of the Beta distribution. The results show that the KDE model is more reliable and accurate.

Finally, the three methods are compared in Fig. (**6**). The fitting goodness of Gaussian distribution is the worst, Beta

Table 3. The Comparison Among Three Methods.

Case	PDF	MSE	χ^2 statistic
Fig.6	KDE	0.1017	1.0436
	Gaussian distribution	1.0894	29.2952
	Beta distribution	0.6261	16.3833

distribution is better than Gaussian distribution, and obviously, KDE is the best. As shown in Fig. (**6**), Gaussian distribution is not suitable for the PDF with high kurtosis. Although Beta distribution fits the kurtosis very well, it fails to fit the part of fat tail. Instead, The KDE provides a better solution for the PDF with high kurtosis and fat tail, it can presents different styles of PDF. Table **3** is the comparison among three methods, the MSE of KDE is about 90% lower than that of Gaussian distribution, and 80% lower than Beta distribution. χ^2 statistics show that the fitting goodness of KDE is far better than the others.

CONCLUSION

In this paper, KDE model was proposed for the modeling of wind power forecast error. Because of its flexibility and adaptability, it can deal with the error data sets in different time scales. The results of this study show that the PDF obtained by KDE model is very close to the real values, and its fitting goodness is always better than that of parametric statistic models (such as Gaussian distribution and Beta distribution). The proposed model provides a new way to estimate the PDF of forecast error, through which the statistics law of the forecast error can be obtained. It may help improve the forecasting algorithms and wind farm control strategies.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGMENTS

This work is supported by National Natural Science Foundation of China (No.51405235).

REFERENCES

[1] A Fabbri, T.G.S Roman, J.R Abbad and V.H.M Quezada,"Assessment of the Cost Associated With Wind Generation Prediction Errors in a Liberalized Electricity Market," *IEEE Transactions on Power Systems*, vol.20, pp.1440-1446, 2005.

[2] L Ma, S Luan,C Jiang, H Liu and Y Zhang, "A review on the forecasting of wind speed and generated power," *Renewable and Sustainable Energy Reviews,* vol.13, pp.915-920, 2009.

[3] A.M Foley, P.G Leahy, A Marvuglia and E.J McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable Energy*, vol.37, pp.1-8, 2012.

[4] K Bruninx and E Delarue, "A Statistical Description of the Error on Wind Power Forecasts for Probabilistic Reserve Sizing," *IEEE Transactions on Sustainable Energy*, vol.5, pp.995-1002, 2014.

[5] R Doherty and M O'Malley, "A new approach to quantify reserve demand in systems with significant installed wind capacity," *IEEE Transactions on Power Systems*, vol.20, pp.587-595, 2005.

[6] T.K.A Brekken, A Yokochi and A.V Jouanne, "Optimal energy storage sizing and control for wind power applications," *IEEE Transactions on Sustainable Energy*, vol.2, pp.69–77, 2011.

[7] M Black and C Strbac, "Value of bulk energy storage for managing wind power fluctuations," *IEEE Transactions on Energy and Conversion*, vol.22, pp.2077-2084, 2007.

[8] K Methaprayoon, C Yingvivatanapong, W.J Lee and J.R Liao. "An integration of ANN wind power estimation into unit commitment considering the forecasting uncertainty," *IEEE Transactions on Industry Applications*, vol.43, pp.1441-1447, 2007

[9] S Bofinger, A Luig and H.G Beyer, "Qualification of wind power forecasts," in *Proceedings of Global Wind Power conference*, Paris, France, 2002, pp.1-5.

[10] H Bludszuweit, J.A. D.Navarro, and A Llombart, "Statistical Analysis of Wind Power Forecast Error," *IEEE Transactions on Power Systems*, vol.23, pp.983-991, 2008.

[11] M.B Hodge and M Milligan, "Wind power forecasting error distributions over multiple timescales," in *General Meet of IEEE Power and Energy Society*, Detroit, USA, 2011, pp.1-8.

[12] N Zhang, C Kang, Q Xia and J Liang, "Modeling Conditional Forecast Error for Wind Power in Generation Scheduling," *IEEE Transactions on Power Systems*, vol.29, pp.1316-1324, 2014.

[13] N Menemenlis, M Huneault and A Robitaille, "Computation of dynamic operating balancing reserve for wind power integration for the time-horizon 1-48 hours," *IEEE Transactions on Sustainable Energy*, vol.3, pp.692-702. 2012.

[14] J Wu, B Zhang, H Li and Z Li, "Statistical distribution for wind power forecast error and its application to determine optimal size of energy storage system," *International Journal of Electrical Power and Energy Systems*, vol.55, pp.100-107, 2013.

[15] M.L Hazelton, "An optimal local bandwidth selector for kernel density estimation," *Journal of Statistical Planning and Inference*, vol.77, pp.37-50, 1999.

[16] T Alberts and R.J Karunamuni, "A semiparametric method of boundary correction for kernel density estimation," *Statistics and Probability Letters*, vol.61, pp.287-298, 2003.

[17] Y Shi and R.C Eberhart, "A modified particle swarm optimizer," in *Proceedings of Congress on Evolutionary Computation*, Washington D.C, USA, 1998, pp.591-600.

Received: October 16, 2014 Revised: December 23, 2014 Accepted: December 31, 2014