Probabilistic Power Flow for Distribution Networks with Photovoltaic Generators

Zhouyang Ren, Wei Yan, *Member*, *IEEE*, Xia Zhao, Yiming Li, *Student Member*, *IEEE*, Juan Yu, *Member*, *IEEE*
State Key Laboratory of Power Transmission Equipment & System Security and New Technology
Chongqing University
Chongqing, China
Email:rzhouyang@gmail.com, cquyanwei@cqu.edu.cn, zx@cqu.edu.cn, yi29281-q@126.com, dianlixi@cqu.edu.cn

***Abstract*—Based on Monte Carlo technique, this paper develops a probabilistic power flow (PPF) algorithm to evaluate the influence of photovoltaic (PV) generation uncertainty on distribution networks. Not only the randomness, but also the correlation of PV power and the moments when PV generators start and stop producing power in a day are taken into account with the presented method using the theory of conditional probability and nonparametric kernel density estimation. The measured power data of photovoltaic generator in Oregon State, USA and 34 node distribution test network are used to demonstrate the application of the presented method in PPF analysis.**

***Index Terms*—correlation, Monte Carlo, photovoltaic generation, probabilistic power flow, random.**

I. Introduction

Solar photovoltaic (PV) generation has been increasingly developed in recent years across the world. Almost 30 GW of new solar PV capacity came into operation worldwide in 2011, thereby increasing the current global total PV capacity by 74% (70 GW) [1]. Power flow in distribution system with grid-connected PV systems changes randomly because of the uncertain output power of PV generators. It is unable to reflect the random change's impact on the operation of distribution system roundly by simply using traditional deterministic power flow analysis, but the problem is solved through probabilistic power flow (PPF) analysis.

Compared to the deterministic power flow, PPF characterizes the uncertainty in system information by describing the variation in terms of a suitable probability distribution. Work related to the probabilistic analysis in power system flow first appeared in 1974. Borkowska [2] first proposed the concept of PPF and implemented an algorithm based on convolution. The PPF is applied to analysis the power system containing PV in recent years. A cumulant based probabilistic power flow algorithm is applied for the distribution system containing PV generators with the assumption that the probability density function of PV output is Beta distribution [3]. The probabilistic model is acquired by combining the beta distribution of solar irradiance and the normal distribution of the forecast error of PV cell temperature, and then the cumulant method is applied to compute the cumulative probability of bus voltage magnitude and line flow [4]. The clearness index is presented to denote the impact of clouds to radiation in [5] and [6]. The probabilistic model of PV power is derived from the probabilistic density function of the clearness index proposed in [7] and the Monte Carlo technique is used to analysis the PPF of distribution system[5]. The probabilistic models of global and diffuse irradiation are established respectively in [6] and based on the functional relationship of PV output and irradiation, the probability density function is derived for PV power. The above probability models of PV do not consider the chronology of PV variation. The dynamic probability model of PV generation is presented in reference [8] Solar irradiation is supposed to be stochastic and regular. The sine function of time is adopted to reflect daily regularity. The random variates, which are assumed to obey certain parametric distribution and be independent with each other, are used to express the hindering effect of hourly cloud cover variation on solar irradiation. However, not only the chronology of PV generation is not completely considered, but also the moments that PV generators start and stop producing power in a day is considered to be deterministic.

This paper proposes a novel PPF method which taken into account the uncertainty of PV power and the moments that PV generators start and stop producing power in a day. The main contributions of the paper include the following:

- The nonparametric kernel density estimation theory, which has been applied to modeling the randomness of wind speed [9] and studies probability distribution only based on the data without any prior knowledge or assumption, is used to derive the probabilistic PV model.
- The chronological probabilistic model of PV power is developed by combining the conditional probability theory and nonparametric kernel density estimation.
- The joint probability distribution of the moments when PV generators start and stop producing power in

This work is supported in part by the National Natural Science Foundation of China (51177178) and the Fundamental Research Funds for the Central Universities in China (CDJZR10150004).

978-1-4799-1303-9/13/$31.00 ©2013 IEEE

a day are evaluated by nonparametric kernel density estimation technique.

- Based on the Monte Carlo technique, the method of PPF for distribution networks is developed and tested by 34 node test system.

The rest of the paper is organized as follows. The novel chronological probability model of PV generation is proposed in Section II. The joint probability distribution of the start and stop moments of PV output is established in Section III. The PPF method is developed based on Monte Carlo technique in Section IV. The proposed method is illustrated by the 34 node test distribution network in Section V, followed by conclusions in Section VI.

II. The Chronological Probability Model of PV Power

A. Nonparametric kernel density estimation theory for PV power

1) Univariate kernel density estimation of PV power

Let P_1, P_2,..., P_3 denote n samples of PV power p at t o'clock; thus, the real probability density function $f(p)$ can be estimated by the following kernel density function [10]:

$$f_h(p)=\frac{1}{nh}\sum_{i=1}^{n}k\left(\frac{p-P_i}{h}\right) \quad (1)$$

where h denotes the bandwidth, $k(*)$ denotes the kernel function, and n is the sample size. $k(*)$should be a symmetric single-peak probability density function.

If the sample size n is sufficiently large, then (1) will converge into $f(p)$[11]. From (1), the precision of the kernel density estimation depends on the selection of bandwidth h and kernel function $k(*)$. Bandwidth has little effect on the accuracy of kernel density estimation. Consequently, bandwidth h is crucial for the accurate estimation of $f(p)$.

The Gaussian function is selected as the kernel function in this paper, which is a popular choice recommended in mathematics books [11].The method in Reference [9] is introduced for optimum bandwidth selection, which only relies on the kernel function and measured data and does not contain any information about the true density of the whole population. The minimization problem of the optimum bandwidth hop in an analytical expression is as follows:

$$\begin{aligned}\min\ \frac{1}{n^2h\sqrt{\pi}}\sum_{i=1}^{n}\sum_{j=1}^{n}\Bigg\{&\frac{1}{2}\exp\left[\frac{-(P_i-P_j)^2}{4h^2}\right]\\&+\frac{1}{4}\exp\left[\frac{-(P_i-P_j)^2}{16h^2}\right]-\frac{2}{\sqrt{10}}\exp\left[\frac{-(P_i-P_j)^2}{10h^2}\right]\Bigg\}\end{aligned} \quad (2)$$

2) Multivariate kernel density estimation of PV power

Multivariate kernel density estimation theory is introduced to develop the chronology probabilistic model of PV power at multiple hours because multivariate probability distribution cannot be obtained by using univariate kernel estimation, which is only suitable in deriving the density function of PV power at certain moments.

Let vector $\boldsymbol{P}=[p_1, p_2, \ldots, p_d]$ represent the PV hourly power output and d the number of hours in a day. Assuming that n day samples are present and the ith day sample is $\boldsymbol{P}_i=[P_{i1}, P_{i2},\ldots, P_{id}]$, we can denote the multivariate kernel estimation of $f(\boldsymbol{P})$ as follows [12]:

$$f_H(P)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\det(H)}K\left[H^{-1}\left(P-P_i\right)\right] \quad (3)$$

where H is the $d*d$ symmetric positive definite matrix and $K(*)$ denotes a multivariate kernel function operating on d arguments. To simplify calculation, H is assumed a diagonal matrix H=diag $[h_1, h_2,\ldots, h_d]$ and $K(*)$ is replaced by the following multiplicative kernel function:

$$K(u)=k(u_1)\cdots k(u_2)k(u_d) \quad (4)$$

where $k(*)$ is a univariate kernel function and is also supposed to be Gaussian kernel function. Hence, (3) is simplified as follows:

$$\begin{aligned}f_H(P)=\frac{1}{n}\sum_{i=1}^{n}\Bigg\{&\frac{1}{h_1h_2...h_d}k(\frac{p_1-P_{i1}}{h_1})*\\&k(\frac{p_2-P_{i2}}{h_2})\cdots k(\frac{p_d-P_{id}}{h_d})\Bigg\}\end{aligned} \quad (5)$$

Many effective methods are available in calculating the optimum bandwidth of univariate kernel density estimation. However, the development of methods for optimum bandwidth selection for multivariate kernel density estimation has a slow pace. Mathematically and in applications, the rule of thumb method taking multivariate normal distribution as the reference distribution is used to select optimum bandwidth. However, this method is only optimal for reference probability density functions and will fail for multimodal densities for instance [13]. Hence, the Cross-Validation (CV) [12] method, which requires no assumption or reference distribution, is applied to select the optimum bandwidth in this paper.

The estimation error of $f_H(\boldsymbol{P})$ can be represented by the integrated squared error (ISE). The bandwidth H that minimizes the ISE is the optimum bandwidth.

$$\begin{aligned}\min ISE(H) &= \int[f_H(P)-f(P)]^2dP\\&=\int f_H^2(P)dP-2\int f_H(P)f(P)dP+\int f^2(P)dP\end{aligned} \quad (6)$$

In (6), the first term can be easily calculated from the data, and the last term does not depend on H and can be ignored as far as minimization over H is concerned. Hence, only the second term of (6) is unknown and must be estimated. Thus,

$$E(f_H(P)) = \int f_H(P)f(P)dP \quad (7)$$

Here, $E(f_H(\boldsymbol{P}))$ can be replaced by a leave-one-out estimator, which is also the unbiased estimation of $E(f_h(\boldsymbol{P}))$.

$$\hat{E}(f_H(P))=\frac{1}{n}\sum_{i=1}^{n}f_{h,-i}(P_i) \quad (8)$$

where,

$$f_{h,-i}(P)=\frac{1}{n-1}\sum_{j=1\ j\neq i}^{n}\frac{1}{\det(H)}K\left[H^{-1}\left(P-P_i\right)\right] \quad (9)$$

By substituting the unbiased estimation of $E(f_h(\boldsymbol{P}))$ in (8) into (6) and ignoring the last term in (6), we can obtain following analytical expression:

$$\min ISE(H)=\int f_H^2(P)dP-2\int f_H(P)f(P)dP$$
$$=\int f_H^2(P)dP-\frac{2}{n(n-1)\det(H)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}K\left[H^{-1}\left(P_j-P_i\right)\right] \quad (10)$$

By choosing the Gaussian function as kernel function and substituting (5) into (10), we can transform (10) into the following:

$$\min ISE(H)=$$
$$\frac{\det(H)^{-1}}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{\prod_{m=1}^{d}\frac{1}{2\sqrt{\pi}}\exp[-\frac{1}{4}(\frac{p_{mj}-p_{mi}}{h_m})^2]\}-$$
$$\frac{2\det(H)^{-1}}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1\ j\neq i}^{n}\{\prod_{m=1}^{d}\frac{1}{\sqrt{2\pi}}\exp[-\frac{1}{2}(\frac{p_{mj}-p_{mi}}{h_m})^2]\} \quad (11)$$

This is a non-constraint optimization problem. The interior point method is adopted to solve the optimum bandwidth H.

The joint probability model of the PV power output at multiple hours can be obtained by multivariable kernel estimation. However, the multivariate kernel density estimation theory is usually not applied if $d \geq 5$ [12] because the amount of measured data and computational effort of this technique increases rapidly with the variate dimension d. The number of hours that the PV generator produces power in a day is always higher than 5. Hence, the multivariate kernel estimation theory is not suitable in establishing the PV probability model directly.

B. Chronological probability model of PV power

To take advantage of high precision characteristic and avoid excess demand for measured data and computation of multivariable kernel density estimation theory, this paper establishes the chronological probability model of PV power at multiple hours by combining the multivariate kernel estimation and the conditional probability theory.

Let vector $\boldsymbol{P}=[p_1, p_2,\ldots, p_d]$ represent the PV hourly power output and d the number of hours in a day. Assuming that n day samples are present and the ith sample is $\boldsymbol{P}_i=[P_{i1}, P_{i2},\ldots, P_{id}]$, we can assume that the PV outputs at adjacent hours are relevant to each other. Take the power output at i o'clock p_i as an example, we can denote that the value of p_{i-1} has an impact on the probability distribution of p_i. According to the conditional probability theory, the conditional probability density function $f(p_i|p_{i-1})$ is as follows [14]:

$$f(p_i\mid p_{i-1})=\begin{cases}\dfrac{f(p_{i-1},p_i)}{f(p_{i-1})} & f(p_{i-1})\neq 0\\ 0 & f(p_{i-1})=0\end{cases} \quad (12)$$

where $f(p_{i-1}, p_i)$ is the joint probability density function of p_i and p_{i-1}.

The conditional probability density function of the hourly PV output can be acquired successively by (12), such as $f(p_2|p_1)$, $f(p_3|p_2)$,…, $f(p_d|p_{d-1})$. Consequently, the probability distributions of the hourly PV output consider the correlation between PV powers at adjacent hours. Hence, the stochastic and chronological characteristic of PV output can be correctly presented. Moreover, $f(p_{i-1})$ and $f(p_{i-1}, p_i)$ are the key points in determining the accuracy of the proposed chronological probability model. Univariate and multivariate (the dimension of the variate is two) kernel density estimation are adopted to evaluate $f(p_{i-1})$ and $f(p_{i-1}, p_i)$ with high precision. Let $f_h(p_{i-1})$ and $f_H(p_{i-1}, p_i)$, which are shown in (13) and (14), be the kernel estimation of $f(p_{i-1})$ and $f(p_{i-1}, p_i)$; thus,

$$f_h(p_{i-1})=\frac{1}{nh}\sum_{j=1}^{n}k(\frac{p_{i-1}-P_{j,i-1}}{h}) \quad (13)$$

$$f_H(p_{i-1},p_i)=$$
$$\frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_ih_{i-1}}k(\frac{p_i-P_{j,i}}{h_i})k(\frac{p_{i-1}-P_{j,i-1}}{h_{i-1}}) \quad (14)$$

where h is the bandwidth of the univariate kernel density estimation $f_h(p_{i-1})$ and h_i, h_{i-1} is the bandwidth of multivariate kernel estimation $f_H(p_{i-1}, p_i)$. By substituting (13) and (14) into (12), we can obtain the kernel estimation of $f(p_{i-1}, p_i)$.

The stochastic and chronological characteristic of PV power is simulated successfully by the proposed method, and the dimension of the involved multivariate kernel estimation is only two. The new method is easy to implement with less computation and has no excess demand for measured data of PV generator compared with the PV model, which is based on the multivariate kernel estimation. The modeling method is also used to acquire the probabilistic model of load in this paper.

III. The Joint Probability Model of the Start and Stop Moments of the PV output

PV generation depends on solar radiation. Thus, power is only produced during daytime because of alternating day and night. The start and stop moments of the PV output in a day are uncertainty.

Assuming that the start and stop moments of PV output are t_s and t_e, and T_s and T_e are the measured data of t_s and t_e, respectively, then $\boldsymbol{T}_s=[T_{s1}, T_{s2},\ldots, T_{sn}]$, $\boldsymbol{T}_e=[T_{e1}, T_{e2},\ldots, T_{en}]$. Moreover, $f(t_s, t_e)$ is the joint probability density function of t_s and t_e,. Based on the multivariate kernel estimation theory with dimension of the variate being two, the kernel density estimation function of $f(t_s, t_e)$ is acquired as follows:

$$f_H(t_s,t_e)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_1h_2}k(\frac{t_s-T_{si}}{h_1})k(\frac{t_e-T_{ei}}{h_2}) \quad (15)$$

where h_1, h_2 is the bandwidth of $f_H(t_s, t_e)$ that can be solved by the CV method. The kernel function $k(*)$ is also chosen as the Gaussian kernel function.

IV. THE PPF FOR DISTRIBUTION NETWORKS BASED ON MONTE CARLO TECHNIQUE

Once the statistical models are defined in terms of probability density function, Monte Carlo simulations involves repeating the simulation process using in each simulation a particular set of values of the hourly PV output and load generated in accordance with the corresponding probability density function. The PPF calculation method of distribution networks containing PV generation is developed based on the Monte Carlo technique. The procedure is as follows and the flow chart is shown in Fig.1.

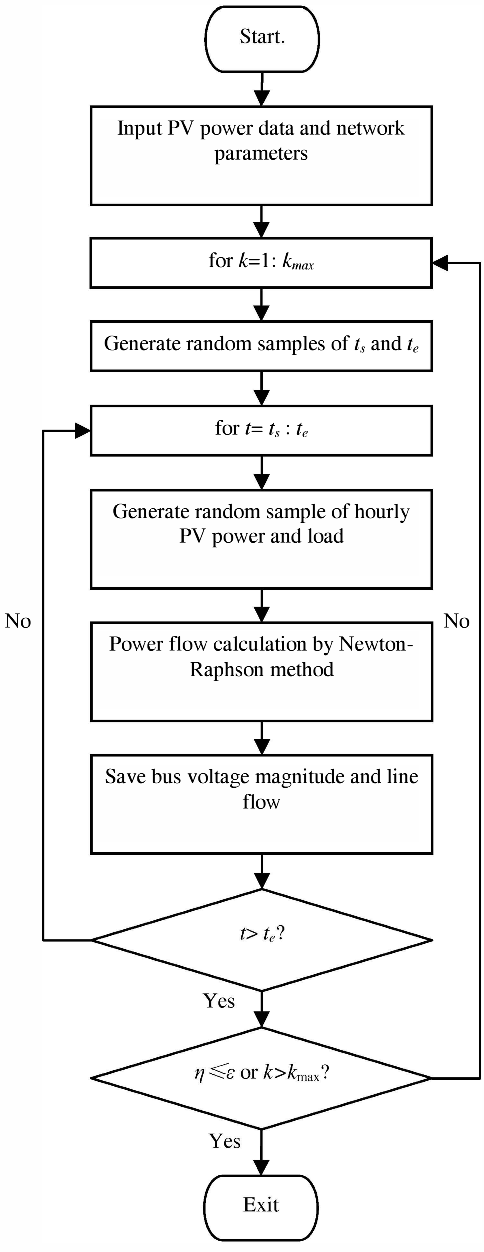


Figure 1. Flow Chart of Probabilistic Power Flow Analysis

1) Input the measured data of PV power, the electrical and geometrical parameters of the distribution network. Initialize the maximum and minimum iteration time k_{max}.

2) Based on the proposed joint probabilistic model of t_s and t_e in Section III, generate random sample of t_s and t_e using the rejection sampling method [15].

3) Generate the hourly PV power and load during time interval [t_s , t_e] using the chronological probability model of PV power in section II.

4) The Newton-Raphson method is used to solve the power flow with the hourly stochastic sample of PV power and load.

5) Save the result of every power flow calculation, such as bus voltage and line flow.

6) The coefficient of variance η is taken as convergence criterion. If η is less than the given precision ε or k is up to k_{max}, exit; otherwise, let k=k+1 and proceed to Step 2.

V. CASE STUDY

The proposed method for probabilistic power flow has been implemented using MATLAB and is tested on the 34 node distribution network [16] (Fig. 2). The hourly measured power data of PV generator in Ashland, United States are selected for the case study [17].

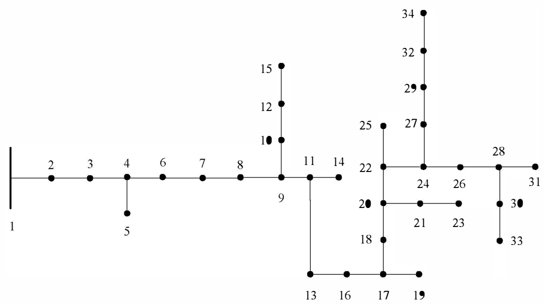
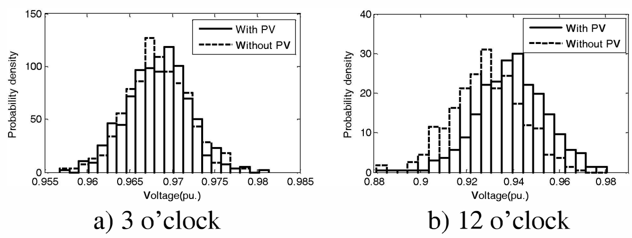


Figure 2. 34 Node Distribution Network



a) 3 o'clock b) 12 o'clock

Figure 3. Histogram of the Voltage Magnitude at Node 26

The 90kW PV generator is connected to the system at node 26. The proposed method in this paper is used to analysis the PPF of 34 node distribution network. The histogram of voltage magnitude of node 26 at 3, 12 o'clock are shown in Fig. 3. There is almost no change after the PV generator is installed at node 26 for the voltage at 3 o'clock since the PV generator produces no power during at night. However, at 12 o'clock, the voltage probability distribution becomes much different if the PV generator is connected to the network. The probability in interval [0.94, 0.98] is increased. Hence, it is demonstrate that PV can improve the voltage level during the daytime.

Fig. 4 and 5 show the hourly probabilistic distribution curves of line flow through line 24-26 with or without PV generator respectively. It can be observed that PV generation has no impact on the line flow at night. However, the probability of line flow locating [0.05, 0.1] is decreased, and the probability locating [0, 0.05] is largely increased during the daytime. Hence, the line flow is only affected by the PV generation during the daytime.

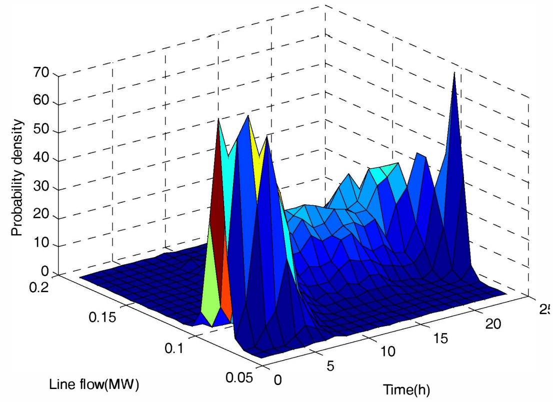


Figure 4. Histogram of Line Flow through Line 24-26 without PV

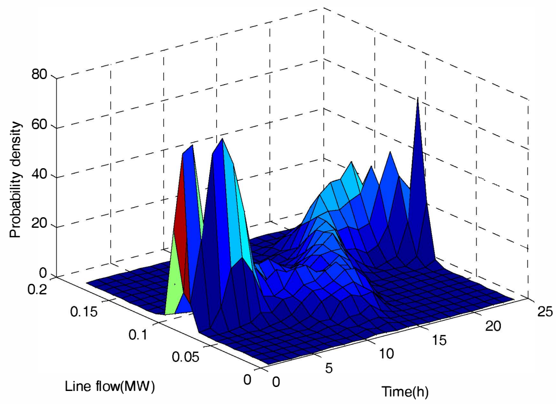


Figure 5. Histogram of Line Flow through Line 24-26 with PV

VI. CONCLUSIONS

Based on Monte Carlo technique, this paper develops a probabilistic power flow algorithm to evaluate the influence of PV generation uncertainty on distribution networks. Not only the randomness, but also the correlation of PV power and the moments when PV generators start and stop producing power in a day are considered with the proposed method. The 34 node distribution test network is applied to demonstrate the application of the presented probabilistic power flow method. The test results indicate that PV generation can decrease the line flow and improve the voltage level during the daytime only.

REFERENCES

[1] E. Mohamed. (2012, June). Renewable 2012 global status report. Renewable Energy Policy Network for the 21st Century. [Online]. Available: http://www.ren21.net/REN21Activities/Publications/GlobalStatusReport/GSR2012/tabid/79218/Default.aspx.

[2] B. Borkowska, "Probabilistic load flow," IEEE trans. Power App, vol. 3, pp. 752-759, 1974.

[3] CS Wang, HF Zheng, YH Xie, "Probabilistic power flow containing distributed generation in distribution system," Automation of Electric Power System, vol. 29, pp.39-44, Dec. 2005.

[4] M Fan, V Vittal, GT Heydt, R Ayyanar, "Probabilistic Power Flow Studies for Transmission Systems with Photovoltaic Generation Using Cumulants," IEEE trans. on power system, vol 27, pp:2251-2261, Nov. 2012.

[5] Conti S, Raiti S, "Probabilistic load flow using Monte Carlo techniques for distribution networks with photovoltaic generators," Solar Energy, vol. 81, pp. 1473-1481, 2007.

[6] F.J. Ruiz-Rodriguez, J.C. Hernández, F Jurado, "Probabilistic load flow for photovoltaic distributed generation using the Cornish–Fisher expansion," IET Renewable Power Generation, vol. 6, pp. 129-138, 2012.

[7] K G T Hollands, R G Huget, "A probability density function for the clearness index, with applications," Solar Energy, vol. 30, pp. 195-209, 1983.

[8] K Yu, YJ Cao, XY Chen, CX Guo, H Zheng, "Dynamic Probability Power Flow of district Grid Containing Distributed Generation," Proceedings of the CSEE, vol. 31, pp. 20-25, Jan. 2011.

[9] ZL Qin, WY Li, XF Xiong, "Estimating wind speed probability distribution using kernel density method," Electric Power Systems Research, vol. 81, pp. 2139-2146, 2011.

[10] E Parzen, "On estimation of a probability density function and model," Annals of mathematical statistics, vol. 33, pp. 1065-1076, 1962.

[11] B.W. Silverman, Density Estimation for Statistics and Data Analysis, U.K: Chapman& Hall, 1986.

[12] W.K. Härdle, M. Müller, S. Sperlich, A. Werwatz, Nonparametric and Semiparametric Models, New York: Springer, 2004.

[13] V A Epanechnikov, "Nonparametric Estimation of a Multidimensional Probability Density," Theory of Probability and Application, vol.14, pp. 153-158, 1969.

[14] P.G. Hoel, S.C. Port, C.J. Stone, Introduction to probability theory, Boston: Houghton Mifflin, 1971, pp.153-157.

[15] W Hörmann, J. Leydold, G. Derflinger, Automatic Nonuniform Random Variate Generation, Berlin: Springer, 2004, pp. 16-20.

[16] R M Ciric,F. A. Padilha, I. Denis, "Observing the performance of distribution systems with embedded generators," European transactions on electrical power, vol.14, pp.347-359, 2004.

[17] The hourly PV power data. [Online]. Available: http://solardata.uoregon.edu/SelectArchival.html.