Modeling of Wind Power Generation for Probabilistic Power Flow Analysis

Tung Le-Duc*, Dinh Duong Le, and Nhi Thi Ai Nguyen

Abstract— *In this paper, a comprehensive framework for estimating the probability distribution of wind power generation is proposed. The proposed approach is mainly based on the combination of data pre-processing and clustering techniques to form an appropriate distribution characterizing the stochastic nature of the wind for the considered wind farm site. The proposed model of wind is tested on real wind data at different wind farm sites showing a good performance in comparison with popular probability distributions for wind speed and wind power. In order to illustrate the applicability of the proposed wind distribution, probabilistic power flow calculation and analysis are carried out on the modified IEEE 57-bus test power system.*

Keywords— **Wind, probability distribution, clustering, probabilistic power flow.**

1. INTRODUCTION

Renewable energy, especially wind, has attracted global attention in recent years. Wind is considered as one of the cleanest renewable energy resources. However, in addition to the environmental benefits and its contribution to providing electricity to meet the increasing demand of the load, wind power generation causes several difficulties in power system planning, operation, and security assessment due to its highly uncertain, time-varying, and intermittent characteristics [1]. Wind power generation has added additional uncertainty to power systems. In order to manage uncertainties in planning, operation, and security assessment of the system with integrated wind resources, probabilistic approach has been introduced and can serve as an effective tool in which uncertainty from wind is represented by a probability distribution.

Wind power probability distribution can be estimated from either wind power data or wind speed data. Wind power data are sometimes neither available nor reliable [1]. In such a case, wind speed distribution is first obtained and then wind power distribution is drived via a power curve, i.e., wind power-wind speed relationship.

Many types of probability distributions have been used in the literature to model wind speed [2, 3]. In general, a distribution may has only one peak, two peaks, three or more peaks corresponding to unimodal, bimodal, multimodal distribution, respectively [4, 5]. For a unimodal distribution, it can be characterized by a generic distribution function such as Weibull, Beta, Gamma, Rayleigh, Lognormal, etc. Among them, the Weibull distribution is the most widely used and accepted distribution. However, in practice wind speed at different wind farms can have different patterns, thus the Weibull distribution cannot represent well for some wind farms. Generally, several different distributions need to be estimated to eventually select the most appropriate one for the considered wind data. In case the wind speed distribution follows a multimodal shape of its histograms, the aforementioned distributions cannot characterize the wind speed well and mixture distributions need to be used. Several mixture distributions have been introduced including mixture of two Gaussian distributions, mixture of two Weibull distributions, mixture of Gaussian and Weibull distributions and so on [2]. Estimation of parameters for a mixture distribution is much more complicated than that for a unimodal distribution.

When wind power data are available, wind power distribution can be estimated directly from these data. In practice, the output power of a wind farm is often distributed according to complicated shapes and less obeying the common distributions as mentioned above. Moreover, probability distribution regularity of the power output is usually poor. Therefore, in general fitting power output of a wind farm to a distribution function are often more challenging than that for wind speed.

In order to overcome the above-mentioned difficulties, in this paper we propose a framework for estimation of wind probability distribution in which clustering techniques are used to form a discrete distribution for wind data. The proposed model can be applied for both wind speed and wind power data. In addition, it is suitable for any type of distribution including unimodal, bimodal, and multimodal distributions.

In this paper, the wind power probability distribution obtained above is used for the purpose of calculation and analysis of the power system by a probabilistic power flow approach. We make use of the technique proposed in [6] which can effectively account for the discrete probability distribution of wind together with other types

Tung Le-Duc is with School of Electrical Engineering, Hanoi University of Science and Technology, 1 Dai Co Viet St., Ha Noi, Vietnam.

Dinh Duong Le and Nhi Thi Ai Nguyen are with Faculty of Electrical Engineering, The University of Danang – University of Science and Technology, 54 Nguyen Luong Bang St., Da Nang city, Vietnam.

*Corresponding author: Tung Le-Duc; Phone: +84-943842803; E-mail: tung.leduc1@hust.edu.vn.

of probability distributions which represent uncertainties in the power system such as from loads, contingencies due to random branch and generating unit outages.

The proposed model of wind is tested on real wind speed and wind power observed at different wind farms in Italy while probabilistic power flow calculation and analysis are carried out on modified IEEE 57-bus test power system.

The rest of the paper is outlined as follows: Section 2 presents the proposed framework for estimating distribution of wind power generation. In Section 3, the probabilistic power flow is described. Simulations results are presented and discussed in Section 4. Concluding remarks are provided in Section 5.

2. PROPOSED FRAMEWORK FOR ESTIMATING DISTRIBUTION OF WIND POWER GENERATION

In this section, a methodology developed to effectively estimate probability distribution of wind speed and wind power generation based on observed wind data at a wind farm is presented. As discussed above, probability distribution of wind power output can be estimated directly from wind power data or indirectly via wind speed data as shown in Fig. 1.

It is worth noticing that the proposed framework can be appropriately used for both wind speed and wind power data and for either a unimodal or a multimodal distribution. This is a particularly attractive feature of the proposed model that makes it widely applicable to various types of wind data in practice.

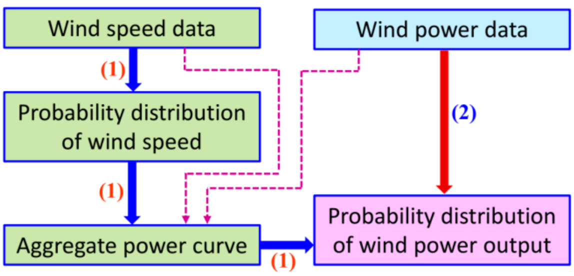

Fig.1. Flowchart of two ways for estimating wind power generation distribution.

The proposed approach for estimation of wind probability distribution is mainly based on the combination of data pre-processing and clustering techniques.

First, outlier detection [7, 8] is performed to find values which significantly deviate from the underlying wind data distribution. Detecting outliers is of major importance for the next steps of the estimation because outliers can cause errors and decrease accuracy in the estimation of the distribution function. The methods of identifying and eliminating outliers can be found in [7, 8], including the following groups of methods: Z-Score or Extreme Value Analysis, Probabilistic and Statistical Modeling, Linear Regression Models, Proximity Based Models, Information Theory Models, etc. Generally, each method has its own advantages and disadvantages. In this paper, we apply the method used effectively in practice presented in [9] in which outliers are detected and removed from the data set based on its distribution function and significance level.

Next, we make use of a clustering technique to form all impulses for a discrete distribution characterizing the stochastic nature of the wind. Clustering partitions wind data into distinct groups (or clusters), then data points in each group are used to build an impulse for the discrete distribution. The probability of each cluster (impulse) is calculated proportionally to the total number of data points.

In the literature, many clustering techniques have been developed [10]. Among them, K-means is one of the most popular ones. The method is easy to execute in practice; however, there are several drawbacks. It is stochastic and does not guarantee to converge to the global optimum solution for clustering. The result obtained is highly dependent on the position of the initial cluster centroids. In addition, it is difficult to select optimal number of clusters for K-means algorithm.

Clustering algorithms divides the set of wind data into distinct clusters by minimizing the dissimilarities between different clusters and maximizing the similarities among members within the same clusters. As a result, the problem can be considered as an optimization one. Hence, it can be solved by using optimization algorithms such as Genetic Algorithms (GA) [11, 12], Particle Swarm Optimization (PSO) [13, 14], and Differential Evolution (DE) [15–17]. Different from K-means, the above algorithms provide globally optimal solution and among them, DE is simple to implement and requires little or no parameter tuning so we propose to use it in this paper.

In order to evaluate the performance of the discrete distribution proposed in this paper compared to the common distributions discussed above, criteria such as Chi-squared Statistic (χ^2), Kolmogorov-Smirnov (K-S), Coefcient of Determination (R^2), Akaike Information Criterion (AIC), Bayseian Information Criterion (BIC), Log-likelihood (LL), etc., can be used [18–20].

As previously mentioned, the proposed estimation method can be applied to both wind speed as well as wind power data; however, we prefer to estimate wind power distribution using wind speed data via an aggregate power curve of the wind farm because of its flexibility in practical application. In such a case, an aggregate power curve for the entire wind farm is needed for mapping wind speed into wind power. There are several techniques to fit the power curve of a single wind turbine or an entire wind farm, which can be classified into parametric and nonparametric methods [21]. Nevertheless, in this paper, we aim to develop a new methodology for estimating the wind distribution rather than studying techniques to fit the power curve. Instead, we adopt the method of bins [1] using measurement data of wind power–wind speed pairs of a wind farm site.

3. PROBABILISTIC POWER FLOW CALCULATION

In this section, we present the technique developed in [6] that can effectively account for different types of probability distributions representing uncertainties in the power system from nodal power injections such as loads, wind power generation and contingencies due to random branch and generating unit outages.

The well-known basic power flow equations can be expressed by matrix form as:

$$\boldsymbol{w} = g(\boldsymbol{x}) \tag{1}$$

$$\boldsymbol{z} = h(\boldsymbol{x}) \tag{2}$$

where,

- $\boldsymbol{w}$: vector of nodal power injections;
- $\boldsymbol{x}$: vector of state variables;
- $\boldsymbol{z}$: vector of line power flows;
- $g(\boldsymbol{x})$: the power flow equations;
- $h(\boldsymbol{x})$: the functions to compute line power flows.

After solving a conventional deterministic power flow for the system to obtain solution denoted as $\bar{\boldsymbol{x}}$, Taylor series expansion is used to linearize the power flow equations around the solution point gives:

$$\Delta\boldsymbol{x} = \mathbf{S}|_{\bar{\boldsymbol{x}}}\, \Delta\boldsymbol{w} \tag{3}$$

$$\Delta\boldsymbol{z} = \mathbf{T}|_{\bar{\boldsymbol{x}}}\, \Delta\boldsymbol{w} \tag{4}$$

where,

$\mathbf{S}|_{\bar{\boldsymbol{x}}}$: the inverse of the Jacobian matrix, computed at $\bar{\boldsymbol{x}}$;

$\mathbf{T}|_{\bar{\boldsymbol{x}}}$: the sensitivity matrix of power flows with respect to nodal power injections, computed at $\bar{\boldsymbol{x}}$.

At this point, each element of $\boldsymbol{w}$, $\boldsymbol{x}$ and $\boldsymbol{z}$ is considered as the realization of the random variable associated with each nodal power injection, state variable and power flow, respectively.

Based on the linear relationships in (3) and (4), cumulant-based probabilistic power flow can be adopted. In [6], discrete and continuous distributions of input random variables are separately treated. First, self and joint cumulants of nodal power injections for both discrete and continuous distributions are calculated. Next, cumulants of state variables and line power flows are obtained by using (3) and (4). Then, Von Mises method is used to represent discrete part, while series expansion method such as Gram-Charlier is used to approximate continuous part. Eventually, the probability distributions of output random variables are constructed by combining the resulting approximation of both discrete and continuous parts [6].

4. SIMULATION RESULTS AND DISCUSSIONS

In this section, we apply the proposed approach to estimate probability distribution for wind data observed at different wind farm sites in Italy. We use hourly wind speed and wind power data from September 2011 to August 2012. Wind speed data are used for estimation of wind speed distribution, whereas wind power-wind speed pairs are used for estimating the aggregate power curve for the wind farm site.

Wind speed at different wind farms has different patterns depending on the climate, weather conditions at the sites. In this study, we choose three wind farms with three different types of distribution to test the performances of the proposed distribution in comparison with popular distributions for wind speed, i.e., Weibull, Gamma, Rayleigh, Loglogistic, Lognormal, Nakagami, Generalized Extreme Value (GEV), etc.

In Fig. 2, wind speed distributions for wind farm site 1 are estimated by using the proposed model and eight popular distributions. It can be seen from the figure that the histograms of the wind speed show a clear peak, so most probability distributions provide good results as shown in Table 1. In Fig. 2, to make it clear, we plot only three most suitable common distributions for the data and the proposed distribution. In Table 1, the distribution functions are arranged in order of accuracy from highest to lowest based on criteria AIC and LL. Both Fig. 2 and Table 1 show that the proposed distribution function gives very accurate estimation results.

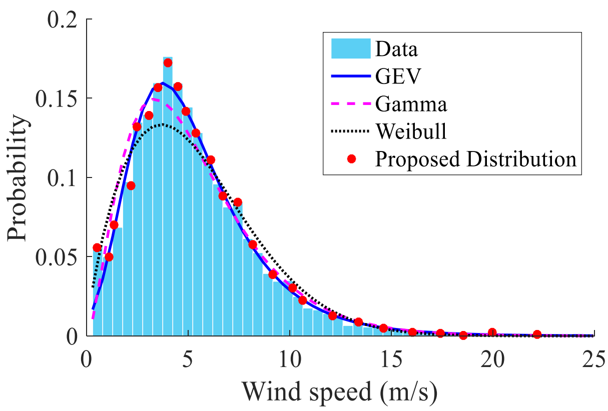

Fig.2. Estimated wind speed distributions for wind farm site 1.

Table 1. Comparison of performances of estimated distributions for wind farm site 1

Rank	Distribution	AIC ($\times 10^4$)	LL ($\times 10^4$)
1	Proposed Distribution	4.270	-2.136
2	GEV	4.291	-2.145
3	Gamma	4.298	-2.149
4	Weibull	4.315	-2.157
5	Nakagami	4.328	-2.164
6	Loglogistic	4.336	-2.168
7	Rayleigh	4.343	-2.172
8	Lognormal	4.398	-2.199
9	Normal	4.520	-2.260

For wind farm site 2, the peak of wind speed is not very clear, and its probability distribution regularity is poorer than that of wind speed at site 1. Therefore, estimation by using common distribution functions gives less accurate results than the estimation for wind speed at site 1. However, the proposed distribution function still gives very good results as shown in Fig. 3 and Table 2.

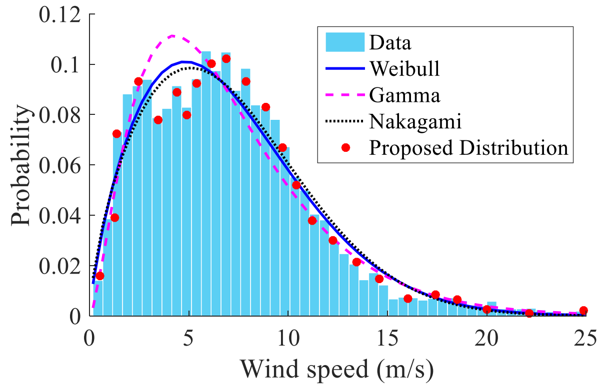

Fig.3. Estimated wind speed distributions for wind farm site 2.

Table 2. Comparison of performances of estimated distributions for wind farm site 2

No.	Distribution	AIC ($\times 10^4$)	LL ($\times 10^4$)
1	Proposed Distribution	4.351	-2.215
2	Weibull	4.815	-2.407
3	Gamma	4.820	-2.410
4	Nakagami	4.821	-2.410
5	GEV	4.843	-2.421
6	Rayleigh	4.844	-2.422
7	Loglogistic	4.905	-2.452
8	Lognormal	4.917	-2.458
9	Normal	4.991	-2.495

Different from two cases above, the wind speed at wind farm site 3 has a distributed shape obeying a bimodal distribution. In this case, the above common distributions cannot be used because they are just suitable for a unimodal distribution. Nevertheless, the proposed distribution maintains good performance compared to a most popular bimodal distribution, i.e., Gaussian mixture as indicated in Table 3 and Fig. 4.

Table 3. Comparison of performances of estimated distributions for wind farm site 3

No.	Distribution	AIC ($\times 10^4$)
1	Proposed Distribution	4.401
2	Gaussian mixture	4.665

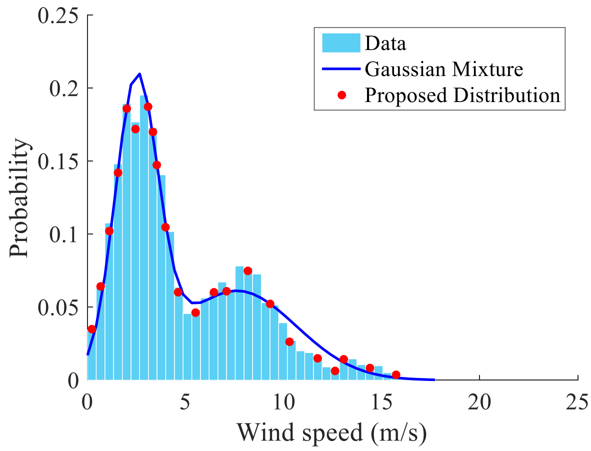

Fig.4. Estimated wind speed distributions for wind farm site 3.

For application in the calculation and analysis of power systems, the distribution function of wind power output is required. Based on wind power-wind speed pairs at site 2, for example, aggregate power curve for the site can be obtained as in Fig. 5 [1], then wind power output distribution for site 2 is achieved as in Fig. 6 as well.

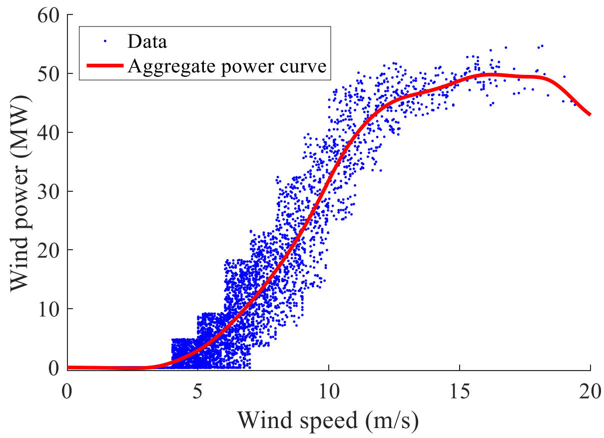

Fig.5. Aggregate power curve estimated for wind farm site 2.

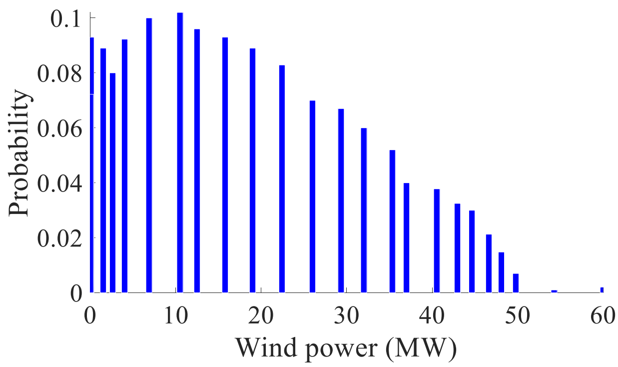

Fig.6. Wind power output distribution for wind farm site 2.

To illustrate the applicability of the proposed distribution, the wind power distribution of a wind farm site, i.e., site 2, obtained is then used as an input for probabilistic power flow calculation and analysis to assessing the security of the modified IEEE 57-bus test power system (see Fig. 7). The test system is modified by adding wind farm 2 with a rated capacity of 60 MW to bus 39 and a 40 MW solar PV power plant to bus 45.

In this test, we consider uncertainties of wind power generation, solar PV power generation and loads. For the sake of simplicity and without loss of generality, load at each bus is represented by a Gaussian distribution with its expected value equal to the base value and standard deviation of 10% of the expected value. Solar PV power output at bus 45 is assumed to have Weibull distributions with scale and shape parameters equal to 20 and 2, respectively, while distribution of wind power output at bus 39 is provided in Fig. 6.

We carry out cumulant-based probabilistic power flow method (denoted as CPPF) in [6] together with Monte Carlo simulation (MCS) with 10.000 samples. Both these approaches can easily take into account the proposed distribution of wind as well as other types of distributions representing uncertainties in the power system.

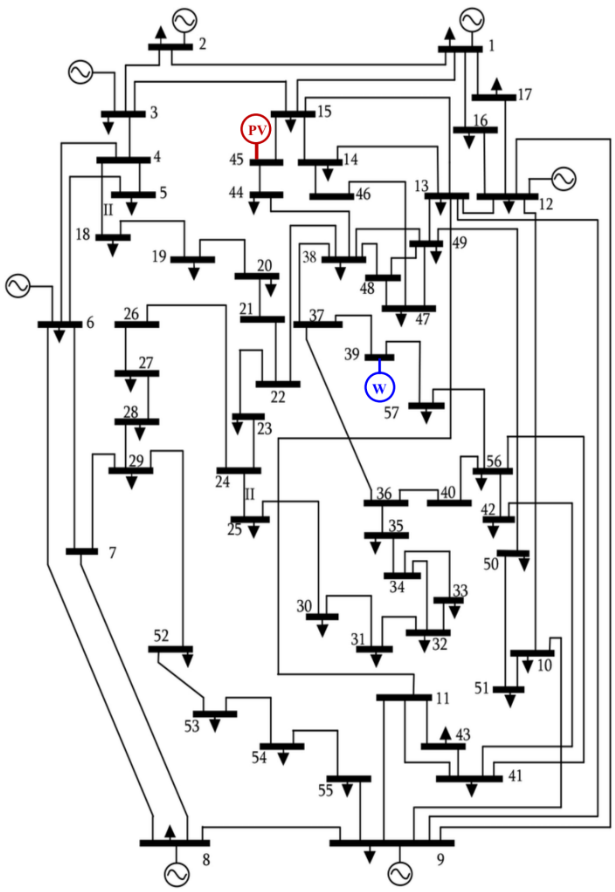

Fig.7. Modified IEEE 57-bus test power system.

The obtained results in terms of probability distributions of power flows and voltages allow assessing security of the system. For example, Fig. 8 and Fig. 9 show probability distributions for power flow through line 39-37 and voltage at bus 45, respectively. Probabilistic security assessment can be carried out to evaluate the probability of line overloading, over-/under-voltage, etc. In Fig. 8, suppose that the upper limit of the power flow of line 39-37 is 60 MVA (the vertical line in Fig. 8), the probability being greater than the limit is 3.2%. In this system, voltages at all buses are within the operating range (i.e., [0.95, 1.05] p.u.).

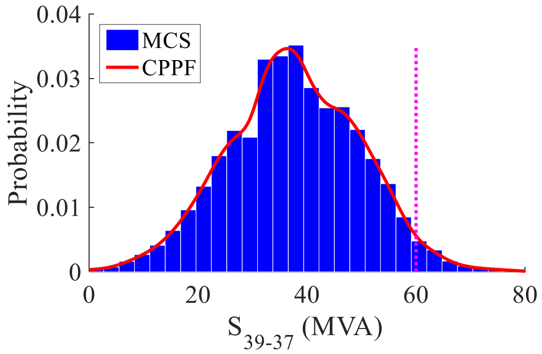

Fig.8. Distribution of power flow through branch 39-37.

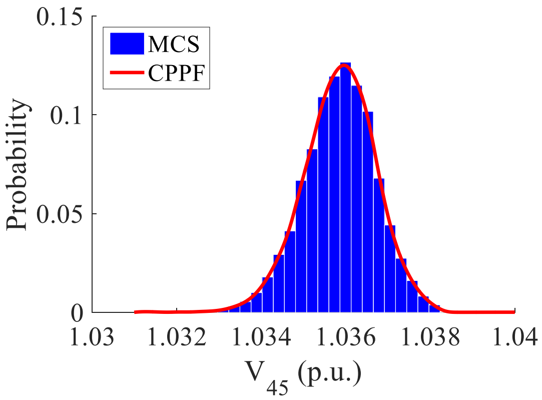

Fig.9. Distribution of voltage at bus 45.

5. CONCLUSIONS

In this paper, we propose a methodology to estimate the probability distribution of wind power generation. The proposed discrete probability distribution is built based on the combination of data pre-processing and clustering techniques. It can be appropriately used for both wind speed and wind power data and either a unimodal or a multimodal distribution. This is a particularly attractive feature of the proposed model allowing it widely applicable to various types of wind data in practice. The proposed model of wind is tested on different types of distributions of wind data at different sites. The obtained results demonstrate that our developed distribution outperforms other probability distributions widely used to estimate wind speed in the literature such as Weibull, Gamma, Rayleigh, Loglogistic, Lognormal, Nakagami, Generalized Extreme Value (GEV), etc., for unimodal distributions and Gaussian mixture for bimodal distributions. Probabilistic power flow calculation and analysis using both the cumulant-based method and Monte Carlo simulation are carried out on the modified IEEE 57-bus test power system to illustrate the applicability of the proposed wind distribution.

ACKNOWLEDGMENT

This work was supported by The University of Danang, University of Science and Technology, code number of Project: T2019-02-50.

REFERENCES

[1] Le, D.D.; Gross, G.; Berizzi, A. 2015. Probabilistic modeling of multisite wind farm production for scenario-based applications. *IEEE Trans. Sustain. Energy*. 6(3): 748-758.

[2] Wang, J.; Hu, J.; Ma, K. 2016. Wind speed probability distribution estimation and wind energy assessment. *Renewable and Sustainable Energy Reviews*. 60: 881-899.

[3] Sohoni, V.; Gupta, S.; Nema, R. 2016. A comparative analysis of wind speed probability distributions for wind power assessment of four sites. *Turk J Elec Eng & Comp Sci.* 24: 4724-4735.

[4] Davies, T.M. 2016. *The Book of R: A First Course in Programming and Statistics*. San Francisco.

[5] Peck, R. 2014. *Statistics: Learning from Data*. Boston, MA, USA: Brooks/Cole.

[6] Le, D.D.; Berizzi, A; Bovo, C. 2016. A probabilistic security assessment approach to power systems with integrated wind resources. *Renewable Energy*. 85: 114-123.

[7] Aggarwal, C.C. 2017. *Outlier Analysis*. Springer.

[8] Ben-Gal, I. 2005. *Outlier detection, Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*. Kluwer Academic Publishers.

[9] Grubbs, F.E. 1969. Procedures for Detecting Outlying Observations in Samples. *Technometrics*. 11(1): 1-21.

[10] Gan, G; Ma, C; Wu, J. 2007. *Data Clustering: Theory, Algorithms, and Applications*. Philadelphia, PA, USA: SIAM.

[11] Falkenauer, E. 1998. *Genetic Algorithms and Grouping Problems.* New York, NY, USA: John Wiley & Sons, Inc.

[12] Bandyopadhyay, S.; Maulik, U. 2002. Genetic clustering for automatic evolution of clusters and application to image classification. *Pattern Recognit*. 35: 1197-1208.

[13] Vander Merwe, D.W.; Engelbrecht, A.P. 2003. Data clustering using particle swarm optimization. In *Proceedings of the 2003 Congress on Evolutionary Computation*. Canberra, Australia.

[14] Kuo, R.; Zulvia, F. 2013. Automatic clustering using an improved particle swarm optimization. *J. Ind. Intell. Inf.* 1: 46-51.

[15] Das, S.; Suganthan, P.N. 2011. Differential evolution: A survey of the state-of-the-art. *IEEE Trans. Evol. Comput*. 15: 4–31.

[16] Cai, Z.; Gong, W.; Ling, C.X.; Zhang, H. 2011. A clustering-based differential evolution for global optimization. *Appl. Soft Comput*. 11: 1363-1379.

[17] Kuila, P.; Jana, P.K. 2014. A novel differential evolution based clustering algorithm for wireless sensor networks. *Appl. Soft Comput.* 25: 414-425.

[18] Huber-Carol, C.; Balakrishnan, N.; Nikulin, M. S.; Mesbah, M. 2002. *Goodness-of-Fit Tests and Model Validity*. Springer.

[19] Ingster, Yu.I.; Suslina, I.A. 2003. *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. Springer.

[20] Rayner, J.C.W.; Thas, O.; Best, D.J. 2009. *Smooth Tests of Goodness of Fit*. Wiley.

[21] Shokrzadeh, S.; Jozani, M.J.; Bibeau, E. 2014. Wind Turbine Power Curve Modeling Using Advanced Parametric and Nonparametric Methods. *IEEE Transactions on Sustainable Energ.* 5(4): 1262-1269.