Identification of Suitable Probability Density Function for Wind Speed Profiles in Power System Studies

Mohammed Alzubaidi
School of Engineering
RMIT University
Melbourne, Australia
S3583733@student.rmit.edu.au

Kazi N. Hasan
School of Engineering
RMIT University
Melbourne, Australia
kazi.hasan@rmit.edu.au

Lasantha Meegahapola
School of Engineering
RMIT University
Melbourne, Australia
lasantha.meegahapola@rmit.edu.au

***Abstract*— The uncertainty in the prediction of wind power generation can be reduced by accurately identifying the probability distribution for modelling wind speed, which is the most influential parameter in calculating wind power. This paper aims to identify the most appropriate probability density function (PDF) for wind speed to use in power system studies, particularly for low wind speed, where the most commonly used Weibull distribution unable to produce a satisfying representation. Therefore, this paper has tested various probability density functions (PDF), which include Rayleigh, Weibull, Gamma, Lognormal, Normal, Inverse Gaussian, Generalized Extreme Value and Exponential distributions, to identify an accurate PDF for modelling low wind speed data. Root mean square error (RMSE) and the coefficient of determination (R^2) are used as the measure of accuracy. The obtained results have indicated that the Gamma distribution (followed by Generalized Extreme Value distribution) provides the best representation for modelling wind speed data. This result has been further verified by performing probabilistic power flow simulation in the IEEE-30 bus test system, where the accuracy of the voltage profile PDF follows the same trend similar to wind profile PDF.**

Keywords*— *Probabilistic analysis, probability distribution, voltage profile, wind speed.

I. Introduction

Environmental concerns, higher demand and sustainable power sources have driven the power energy sources to be shifted to renewable energy sources (RESs), such as wind power and solar power. The wind power generation can be considered as the most significant and widely used RES other than synchronous generator based hydropower generation [1]. However, the intermittency of wind generators has produced various uncertainties in the power system networks, which can strongly affect power system security and reliability [2]. Therefore, these uncertainties in the power system networks need to be modelled and evaluated by probabilistic methods to maintain the stability and reliability of renewable-rich power systems [3].

In the probabilistic assessment of power system studies, it is essential to identify an accurate probability distribution function (PDF) to appropriately represent the uncertainties of the input parameters [3]. The randomness and intermittency of the wind generation can be described by the probability distribution, which can lead to enhancing and improving the analysis of the renewable-rich future power systems. The output of wind power is significantly based on wind speed characteristics, which is dependent on the variation in weather conditions. Based on the previous work, the probability distribution for representing wind speed has been employed to various aspects of the power system stability, which include small-disturbance stability [4, 5], voltage stability [6, 7], transient stability assessment [8, 9] and frequency stability [10, 11]. Also, the probability distribution is applied for power quality evaluation [12] and power-flow assessment [13, 14].

In the literature, it was found that the Weibull distribution is the most widely used probability density function (PDF) for fitting the wind speed data [6, 9, 15-17]. Additionally, the Normal distribution is applied to represent wind speed data in [14, 18]. The different probability distribution functions are also used in various power system studies for representing wind speed, which includes joint Gaussian distribution [10, 11], Gamma distribution [19], lognormal distribution [19, 20] and Burr distribution [20, 21]. However, most of these studies used Weibull distribution to model wind speed without taking into account the fact that the Weibull distribution cannot represent the wind speed under low wind conditions [22, 23]. Thus, this paper will identify an appropriate PDF for representing wind speed data that have low wind speeds.

This paper will further validate the result of selecting a proper PDF for low wind speed by performing a probabilistic power flow on the IEEE-30 bus test network. Additionally, root mean square error (RMSE) and the coefficient of determination (R^2) are applied to evaluate the accuracy of selecting a suitable PDF for fitting the dataset. In this study, various PDFs are employed to represent the low wind speed data, which involves Rayleigh, Weibull, Gamma, Lognormal, Normal, Inverse Gaussian, Generalized Extreme Value and Exponential distributions.

The remainder of the paper is organized into four sections: Section II describes the criteria for selecting a suitable probability density function. The implementation of the proposed methodology is presented in Section III, and Section IV shows the simulation results and discussion. Section V concludes the paper with a summary and future works.

II. Theoretical Analysis

The variation of wind speed can play a vital role in calculating the output of wind power. Hence, the wind speed parameter is the most influential parameter that can affect wind power, which can lead to providing a more reliable and accurate prediction of the power system response.

A. The criteria for selecting proper probability distributions

The choice of an accurate PDF for representing any dataset is based on some crucial factors, which include [24, 25]:

- **Discrete and continuous distributions**: the probability distributions can be distinguished by their property of being continuous or discrete. In a discrete distribution, a random variable can be from one of a set of finite values (identifiable). In a continuous distribution, a random variable can take a value from a set which has infinite values, as this set is continuous. In this distribution, a variable can take any value from the specific range. Based on the previous definition of distribution types, the data of wind speed follows the continuous distribution due to taking any value in a particular range.

- **Bounded and unbounded distributions:** in a bounded distribution, the distribution is constrained from both sides, the distribution would lie between two values. In an unbounded distribution, the distribution is not constrained from any side, this basically means that the distribution can be from negative infinity to positive infinity. Therefore, the wind speed data are bounded between the maximum and minimum values, which always cannot become below zero.

- **Unimodal and multimodal distributions:** the mode of the uncertain variable can be defined as the value of the highest peak of the probability distribution. Therefore, the data that only have one peak are represented by the unimodal distributions, while the data that have many peaks are described by the multimodal distributions. In this study, the wind speed data consider following the unimodal distribution.

- **Univariate and multivariate distributions:** in the univariate distributions, the distribution is only based on a single random variable, which means that the variable is independent. On the other hand, the multivariate distributions represent a group of variables, which are probabilistically linked to each other in some way. These probabilistic links can be created by using various correlation methods. In this paper, the wind speed data consider to be independent, and therefore, they follow the univariate distribution.

B. *Probability Distribution Function*

The wind speed data can be modelled by employing different PDFs. In this paper, the most popular continuous distribution functions will be used to model wind speed data, which involve; Rayleigh (RAY), Weibull (WEI), Gamma (GAM), Lognormal (LN), Normal (NORM), Inverse Gaussian (IG), Generalized Extreme Value (GEV) and Exponential (EXP) distribution. The equations of these PDFs are described in Table I [26, 27].

C. *Evaluation Criteria*

There are different goodness-of-fit measures that can be implemented to assess the accuracy of the selected probability distribution. In this paper, the RMSE and R^2 criteria are applied to evaluate the suitability of the probability distributions for fitting wind speed data [21]. The RMSE and R^2 values can be calculated by applying (1) and (2), respectively.

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}(xi-yi)^2}{n}} \tag{1}$$

$$R^2=1-\frac{\sum_{i=1}^{n}(xi-yi)^2}{\sum_{i=1}^{n}(xi-zi)^2} \tag{2}$$

In (1) and (2), the i represent the index value (i = 1, 2, 3, …, n), n is the length of the wind speed data, xi is the original probability, yi is the predicted probability that calculated from different PDFs and zi is the mean of the original dataset and it can be calculated as in (3).

$$zi=\frac{1}{n}\sum_{i=1}^{n} xi \tag{3}$$

Table I. Probability distribution functions and their equations.

PDF	Equation	Parameters
RAY	$f(x;c)=\frac{x}{c^2}\,e^{-(\frac{x^2}{2c^2})}$	c, scale parameter
WEI	$f(x;k,c)=\frac{k}{c}(\frac{x}{c})^{k-1}\,e^{-(\frac{x}{c})^k}$	k, shape parameter c, scale parameter
GAM	$f(x;k,c)=\frac{x^{k-1}}{c^k\,\Gamma(k)}\,e^{-\frac{x}{c}}$	k, shape parameter c, scale parameter
LN	$f(x;\mu,\sigma)=\frac{1}{x\sigma\sqrt{2\pi}}\,e^{-\frac{(ln(x)-\mu)^2}{2\sigma^2}}$	μ, the mean σ, standard deviation
NORM	$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ, the mean σ, standard deviation
IG	$f(x;\mu,k)=\sqrt{\frac{k}{2\pi x^3}}\,e^{-\frac{k(x-\mu)^2}{2\mu^2 x}}$	μ, the mean k, shape parameter
GEV	$f(x;u,k,c)=\frac{1}{c}[1-\frac{k}{c}(x-u)]^{\frac{1}{k}-1}\,e^{[-[1-\frac{k}{c}(x-u)]^{\frac{1}{k}}}$	u, the location parameter k, shape parameter c, the scale parameter
EXP	$f(x;\mu)=\frac{1}{\mu}e^{-\frac{x}{\mu}}$	μ, the mean

III. Probabilistic Modelling of Analytical Approach

A. *Probabilistic Power Flow*

The number of uncertainties is increasing in power system networks due to the increased integration of intermittent RESs and the continuous variability of system loads [28]. The probabilistic methods can accurately model these uncertainties by using the PDFs to represent the uncertain system variables. In this paper, the Monte Carlo (MC) simulation technique, which is the most extensively used for probabilistic simulation, is implemented to probabilistically model uncertain system input parameters [29]. As the MC simulation considers as a numerical solution, it is involved in a repeated random sampling of system uncertainties. Thus, large datasets for uncertain variables can be obtained, and therefore, the probability distribution of an unknown probabilistic entity can be identified. In this simulation, identifying an accurate PDF for fitting uncertain variables is essential, which can lead to accurate results for the probabilistic modelling. Hence, The suitability of the chosen probability distributions is a vital part of the probabilistic methods [3].

B. *Methodology*

Fig.1 shows the research methodology for this study, which can be classified into two main points, which includes:

- Modelling wind speed data to eight PDFs.
- Performing probabilistic power flow for the original dataset and generated samples that follow the eight PDFs.

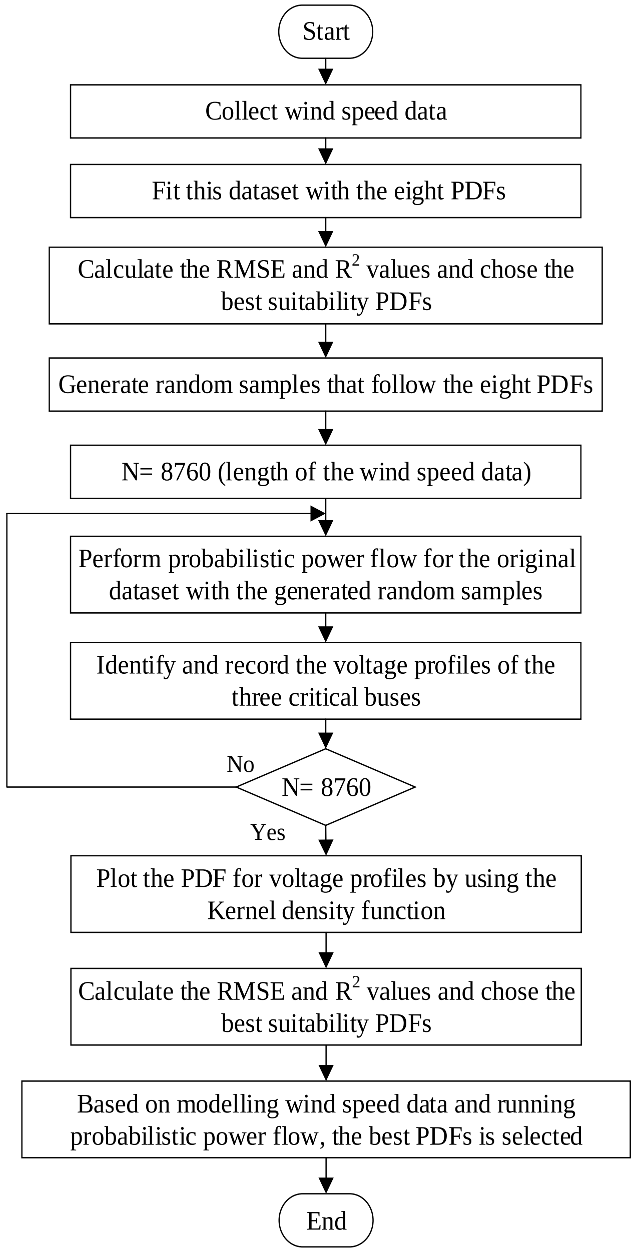

Fig. 1: The flowchart of probabilistic modelling of wind speed data.

C. Wind Speed Data

In this case study, the average hourly wind speed data for one year is obtained from Blue Creek station in the USA [30]. This dataset has a low mean wind speed, which is 2.90 ms^{-1}.

D. System Description

In this study, the IEEE-30 bus test network, which is shown in Fig. 2, is used for power system simulation. Three synchronous generators, which are located at bus 13, 22 and 23, are replaced by wind farm generators. This study has used MATLAB and MATPOWER Toolbox [31] as the software platform for modelling wind speed data and running probabilistic power flow, respectively.

The actual wind speed data and the generated samples that follow the eight PDFs are used to perform the probabilistic power flow in order to validate the selection of the best PDFs for representing low wind speed data. After performing the probabilistic power flow, the voltage profiles of three critical busbars are recorded, which are bus 4, bus 7 and bus 8. Additionally, the kernel density function is applied to plot the PDF for the voltage profiles, and then, the RMSE and R^2 values are calculated to evaluate the accuracy of the selected PDF for representing wind speed data.

E. Uncertainty Modelling

In this study, the system load and wind generators are considered as the uncertain system parameters. The variation of the load is commonly modelled by the Normal distribution. Besides, the continuous variability of system loads is considered to be $3\sigma = 5\%$ of the μ and this variation in the system load represents the load forecasting error, which happens over 24 hours [18]. Also, the output of wind power is calculated based on the eight PDFs that represent the wind speed data, as discussed before.

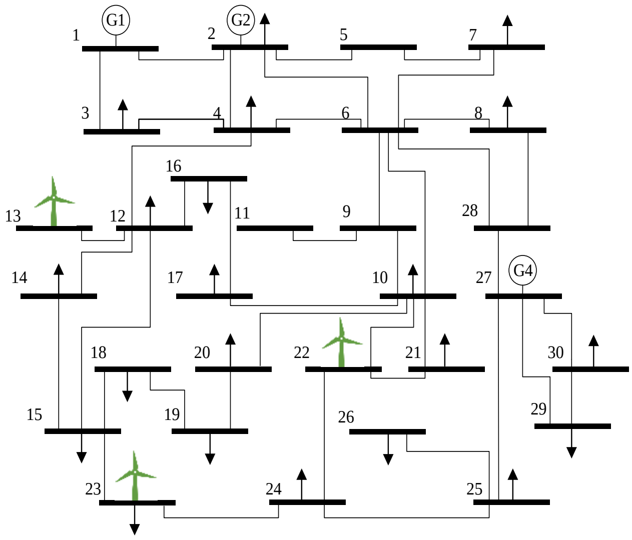

Fig. 2: The IEEE 30-bus test network.

IV. RESULTS AND DISCUSSION

A. Fitting wind speed data

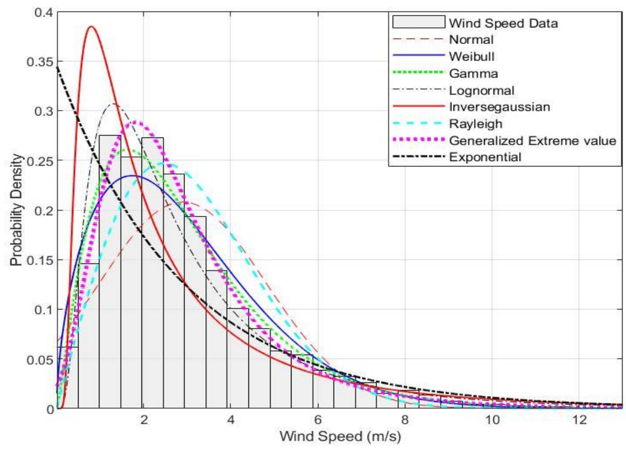

Fig. 3: The histogram of wind speed data fitted with the eight PDF.

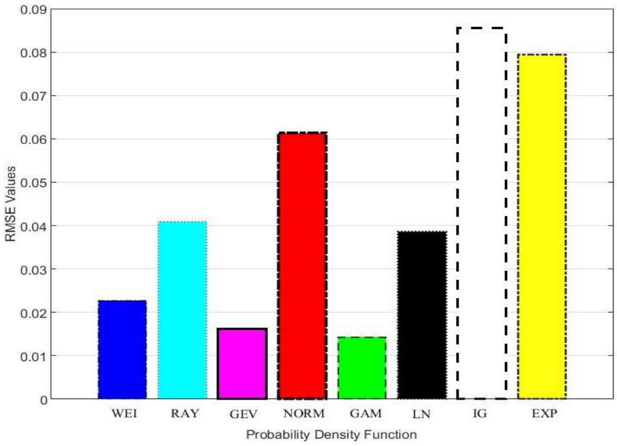

Fig. 4: The RMSE values of the eight PDFs.

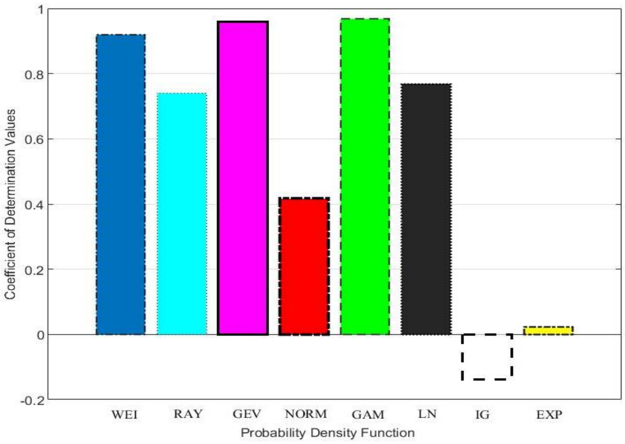

Fig. 5: The R^2 values of the eight PDFs.

Table II. The estimated parameters of the eight PDFs

Probability Density Functions (PDF)	**Parameters for the dataset used in this study**
Rayleigh	c= 2.4564
Weibull	k= 1.5894
	c= 3.2375
Gamma	k= 2.2889
	c= 1.2650
Lognormal	μ= 0.8291
	σ= 0.7540
Normal	μ= 2.8955
	σ= 1.9192
Inverse Gaussian	μ= 2.8955
	k= 2.5793
Generalized Extreme Value	u= 1.9660
	k= 0.1331
	c= 1.2870
Exponential	μ= 2.8955

The estimated parameters of the eight PDFs for fitting the wind speed data are shown in Table II. Fig. 3 presents the histogram of wind speed data fitted with the eight PDFs. It can be seen from Fig. 3 that the Gamma distribution (green dotted line) provides the best representation for wind speed data, followed by the Generalized Extreme Value and Weibull distribution, respectively. On the other hand, the Exponential and Inverse Gaussian distributions show the worst PDF for fitting wind speed data, respectively.

The evaluation of the PDFs is presented based on RMSE and R^2 criteria, which are shown in Fig. 4 and Fig. 5, respectively. The lowest value of the RMSE and the highest value of the R^2 indicate the best performance of the PDF for modelling wind speed data. This result also proves that the Gamma distribution (followed by Generalized Extreme Value distribution) is the best PDF for representing wind speed data, which has the lowest RMSE value and highest R^2 value.

B. Probabilistic power flow and voltage profile analysis

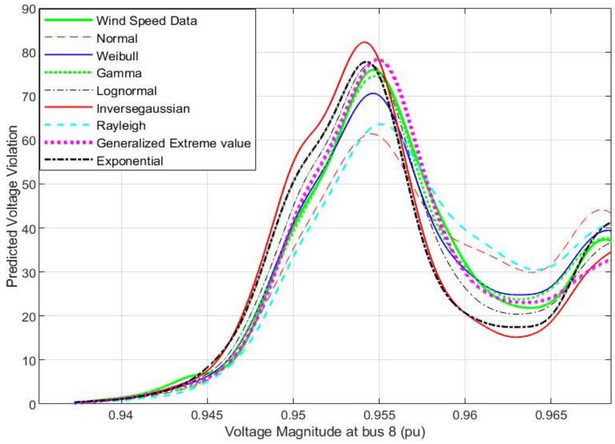

Fig. 6: The voltage profiles of the original dataset and the generated random datasets.

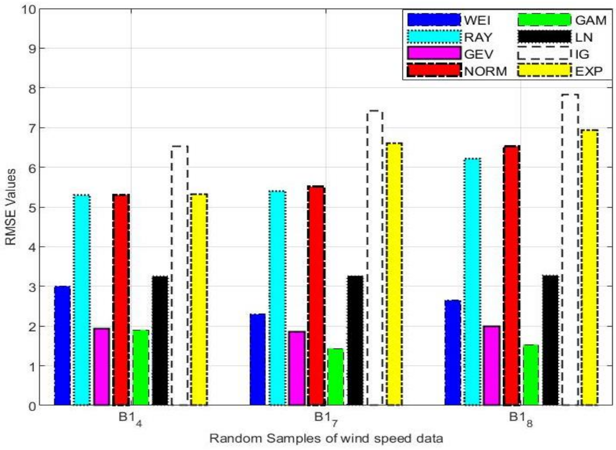

Fig. 7: The RMSE values of the eight PDFs for the voltage profiles of the three critical busbars.

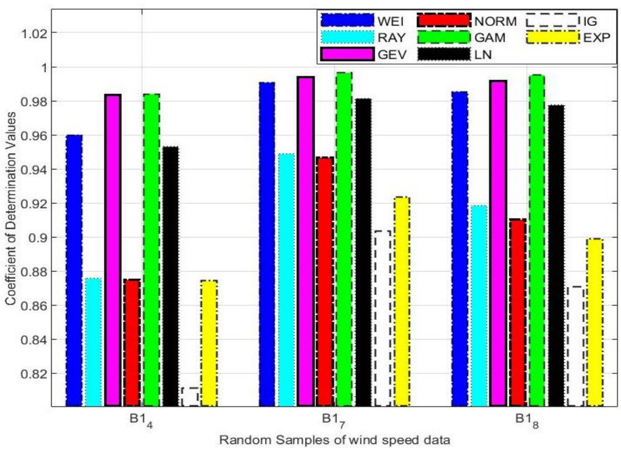

Fig. 8: The R^2 values of the eight PDFs for the voltage profiles of the three critical busbars.

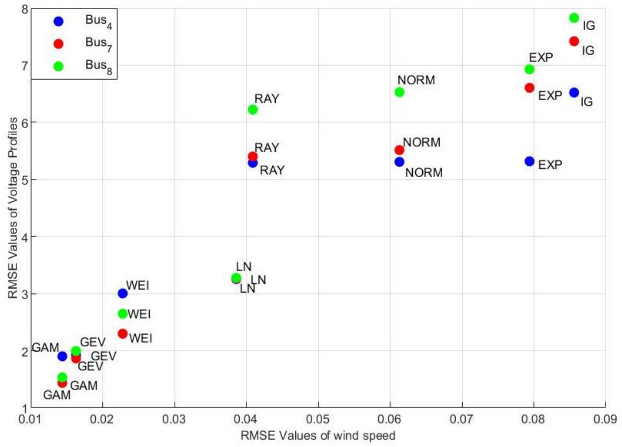

Fig. 9: Scatter plot of RMSE values of fitting wind speed data and voltage profiles analysis of the three critical busbars.

In order to verify the best PDF that can represent low wind speed data, as discussed above, the probabilistic power flow has been carried out with hourly wind speed data over a one-year period. This simulation was performed in MATPOWER for the original dataset and the generated random samples that follow the eight PDFs, which lead to 8760 simulations. Fig. 6 shows the PDFs of voltage profiles of bus 8 for the original dataset and the generated random datasets over the number of actual voltage violations (which are below 0.95 p.u.). Additionally, the voltage profiles for the three critical busbars are recorded, and then the RMSE and R^2 values are calculated to validate the suitability of the selected PDF, as shown in Fig. 7 and Fig. 8, respectively.

It can be observed from Fig. 6 that the Gamma (green dotted line) and Generalized Extreme Value (pink dotted line) distributions are the best PDFs that follow the obtained result from the original wind speed data (solid green line), respectively. In contrast, the Exponential and Inverse Gaussian distributions provide the worst PDFs for wind speed modelling that follow the actual wind speed data. This result has further confirmed by calculating the RMSE and R^2 values for the three critical busbars, as exhibited in Fig. 7 and Fig. 8, respectively. This result is also consistent with the result of the best PDFs for fitting wind speed data in the previous section.

The RMSE values of the voltage profiles and modelling of wind speed are presented in Fig. 9. This result can clearly show that the Gamma distribution (followed by Generalized Extreme Value distribution) provides the best PDF for representing the low wind speed data. Conversely, the Inverse Gaussian distribution (followed by Exponential distribution) presents the least accurate PDF. Fig. 9 also indicates that the voltage profile analysis can be significantly impacted by the selected probability distribution for representing the uncertain system parameters. That is why it is essential to identify a proper PDF for modelling uncertain system parameters.

V. Conclusions and Future Work

Identification of an appropriate probability distribution to represent the wind speed variation is essential to accurately model and reduce the uncertainty of the prediction of wind power, as the wind speed is the most influential parameter that can determine wind power output. Furthermore, the probabilistic power flow is performed in the IEEE30 bus test network to verify the suitability of selecting the appropriate PDFs for representing wind speed data.

The simulation results suggest that the Gamma distribution is the best PDF for fitting low wind speed data, whereas the Inverse Gaussian distribution (followed by Exponential distribution) are the worst. The next stage of our research will extend this study to include the impact of the locations, duration and resolution parameters on selecting an appropriate PDF for representing wind speed data on power system studies.

References

[1] S. Ali, S.-M. Lee, and C.-M. Jang, "Statistical analysis of wind characteristics using Weibull and Rayleigh distributions in Deokjeok-do Island – Incheon, South Korea," *Renewable Energy*, vol. 123, pp. 652-663, 2018.

[2] J. V. Milanovic, "Probabilistic stability analysis: the way forward for stability analysis of sustainable power systems," *Philos Trans A Math Phys Eng Sci*, vol. 375, no. 2100, Aug 13 2017.

[3] K. N. Hasan, R. Preece, and J. V. Milanović, "Existing approaches and trends in uncertainty modelling and probabilistic stability analysis of power systems with renewable generation," *Renewable and Sustainable Energy Reviews*, vol. 101, pp. 168-180, 2019.

[4] S. Q. Bu, W. Du, and H. F. Wang, "Investigation on Probabilistic Small-Signal Stability of Power Systems as Affected by Offshore Wind Generation," *IEEE Transactions on Power Systems*, vol. 30, no. 5, pp. 2479-2486, 2015.

[5] K. N. Hasan, R. Preece, and J. V. Milanovic, "Priority Ranking of Critical Uncertainties Affecting Small-Disturbance Stability Using Sensitivity Analysis Techniques," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2629-2639, 2017.

[6] B. Qi, K. N. Hasan, and J. V. Milanovic, "Identification of Critical Parameters Affecting Voltage and Angular Stability Considering Load-Renewable Generation Correlations," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 2859-2869, 2019.

[7]S. Q. Bu, W. Du, and H. F. Wang, "Probabilistic Analysis of Small-Signal Rotor Angle/Voltage Stability of Large-Scale AC/DC Power Systems as Affected by Grid-Connected Offshore Wind Generation," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3712-3719, 2013.

[8] P. N. Papadopoulos and J. V. Milanovic, "Probabilistic Framework for Transient Stability Assessment of Power Systems With High Penetration of Renewable Generation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3078-3088, 2017.

[9] S. O. Faried, R. Billinton, and S. Aboreshaid, "Probabilistic evaluation of transient stability of a wind farm," *IEEE Transactions on Energy Conversion*, vol. 24, no. 3, pp. 733-739, 2009.

[10] L. Wu and D. Infield, "Power system frequency management challenges – a new approach to assessing the potential of wind capacity to aid system frequency stability," *IET Renewable Power Generation*, vol. 8, no. 7, pp. 733-739, 2014.

[11] L. Wu and D. G. Infield, "Towards an Assessment of Power System Frequency Support From Wind Plant—Modeling Aggregate Inertial Response," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2283-2291, 2013.

[12] M. Amiri, B. Bagen, and A. M. Gole, "Probabilistic analysis of the effect of wind speed variations on power quality of power systems," in *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2016, pp. 1-6.

[13] M. Aien, M. Rashidinejad, and M. Fotuhi-Firuzabad, "On possibilistic and probabilistic uncertainty assessment of power flow problem: A review and a new approach," *Renewable and Sustainable Energy Reviews*, vol. 37, pp. 883-895, 2014.

[14] M. Aien, M. Fotuhi-Firuzabad, and F. Aminifar, "Probabilistic Load Flow in Correlated Uncertain Environment Using Unscented Transformation," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2233-2241, 2012.

[15] H. Huang, C. Chung, K. W. Chan, and H. Chen, "Quasi-Monte Carlo based probabilistic small signal stability analysis for power systems with plug-in electric vehicle and wind power integration," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3335-3343, 2013.

[16] H. Ahmadi and H. Ghasemi, "Maximum penetration level of wind generation considering power system security limits," *IET Generation, Transmission & Distribution,* vol. 6, no. 11, pp. 1164-1170, 2012.
[17] M. Negnevitsky, D. H. Nguyen, and M. Piekutowski, "Risk Assessment for Power System Operation Planning With High Wind Power Penetration," *IEEE Transactions on Power Systems,* vol. 30, no. 3, pp. 1359-1368, 2015.
[18] K. N. Hasan, R. Preece, and J. V. Milanovic, "The Influence of Load on Risk-Based Small-Disturbance Security Profile of a Power System," *IEEE Transactions on Power Systems,* vol. 33, no. 1, pp. 557-566, 2018.
[19] Z. Qin, W. Li, and X. Xiong, "Generation System Reliability Evaluation Incorporating Correlations of Wind Speeds With Different Distributions," *IEEE Transactions on Power Systems,* vol. 28, no. 1, pp. 551-558, 2013.
[20] Y. Li, W. Li, W. Yan, J. Yu, and X. Zhao, "Probabilistic Optimal Power Flow Considering Correlations of Wind Speeds Following Different Distributions," *IEEE Transactions on Power Systems,* vol. 29, no. 4, pp. 1847-1854, 2014.
[21] S. A. Akdağ and A. Dinler, "A new method to estimate Weibull parameters for wind energy applications," *Energy Conversion and Management,* vol. 50, no. 7, pp. 1761-1766, 2009.
[22] K. Philippopoulos, D. Deligiorgi, and G. J. I. J. o. G. E. Karvounis, "Wind speed distribution modeling in the greater area of Chania, Greece," vol. 9, no. 2, pp. 174-193, 2012.
[23] Y. M. Kantar, I. Usta, I. Arik, and I. Yenilmez, "Wind speed analysis using the extended generalized Lindley distribution," *Renewable Energy,* vol. 118, pp. 1024-1030, 2018.
[24] M. K. Ochi, *Applied probability and stochastic processes.New York: Wiley, 1990.*
[25] S. J. Fletcher, "Univariate Distribution Theory," in *Data Assimilation for the Geosciences*, 2017, pp. 29-124.
[26] T. B. M. J. Ouarda *et al.*, "Probability distributions of wind speed in the UAE," *Energy Conversion and Management,* vol. 93, pp. 414-434, 2015.
[27] V. Lo Brano, A. Orioli, G. Ciulla, and S. Culotta, "Quality of wind speed fitting distributions for the urban area of Palermo, Italy," *Renewable Energy*, vol. 36, no. 3, pp. 1026-1039, 2011.
[28] M. Xiang, J. Yu, Z. Yang, Y. Yang, H. Yu, and H. He, "Probabilistic power flow with topology changes based on deep neural network," *International Journal of Electrical Power & Energy Systems,* vol. 117, 2020.
[29] R. Preece, K. Huang, and J. V. Milanovic, "Probabilistic Small-Disturbance Stability Assessment of Uncertain Power Systems Using Efficient Estimation Methods," *IEEE Transactions on Power Systems,* vol. 29, no. 5, pp. 2509-2517, 2014.
[30] AgriMet, *"Historical Dayfile Data Access,"* Available: https://www.usbr.gov/pn/agrimet/webaghrread.html: 2019.
[31] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. J. I. T. o. p. s. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," vol. 26, no. 1, pp. 12-19, 2010.