Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Probability density function selection based on the characteristics of wind speed data

To cite this article: N Y Yürüşen and Julio J Melero 2016 *J. Phys.: Conf. Ser.* **753** 032067

Journal of Physics: Conference Series **753** (2016) 032067 doi:10.1088/1742-6596/753/3/032067

Probability density function selection based on the characteristics of wind speed data

N Y Yürüşen and Julio J Melero

CIRCE-Universidad de Zaragoza, C/ Mariano Esquillor 15, 50018, Zaragoza, Spain

E-mail: nyildirim@fcirce.es and melero@fcirce.es

Abstract.

The probabilistic approach has an important place in the wind energy research field as it provides cheap and fast initial information for experts with the help of simulations and estimations. Wind energy experts have been using the Weibull distribution for wind speed data for many years. Nevertheless, there exist cases, where the Weibull distribution is inappropriate with data presenting bimodal or multimodal behaviour which are unfit in high, null and low winds that can cause serious energy estimation errors. This paper presents a procedure for dealing with wind speed data taking into account non-Weibull distributions or data treatment when needed. The procedure detects deviations from the unimodal (Weibull) distribution and proposes other possible distributions to be used. The deviations of the used distributions regarding real data are addressed with the Root Mean Square Error (RMSE) and the annual energy production (AEP).

1. Introduction

Wind Energy practitioners use wind speed data in order to evaluate the available power at a selected site. The most common approach is to use the Weibull distribution to represent the wind speed data due to its simplicity. Wind data collection is performed in an outdoor environment usually with not favorable conditions for the measurement. In most of the cases, achievement on 100% data availability is not possible due to harsh environmental conditions and the limitations of the equipments. In practice, wind speed datasets are collections of values consisting on 10 minutes of averaged raw wind speed observations. This methodology implies that collected datasets are discrete by nature [1]. Nevertheless, it must be kept in mind that wind speed is assumed to fit continuous distributions, as a continuous random variable in wind statistics science [2].

Most of the authors tend to use the frequently cited probability density functions such as Rayleigh [3] and Weibull [4] distributions which were first applied to wind speed data by Davenport, while he was working on wind tunnel testing and structural loading in 1960's [5], [6]. Indeed the Weibull distribution is very appropriate when dealing with unimodal, univariate variables as is often the case with most of the wind speed data [7–10]. The suitability of the application of the Weibull distribution to wind speed data was demonstrated by Tuller in 1980's [11]. The fulfillment of special conditions related to orthogonal components of wind velocity guaranteed the fit of the Weibull distribution to wind speed data. In general terms, measured wind speed must follow a circular pattern or in another words be congruent with circular normal distribution.

Two main reasons have been reported for using the Weibull distribution in wind energy: it is a positively skewed probability density distribution which favours moderate wind speeds; and it has a relatively easy estimation requirements with only two parameters [12].

When the nature of wind speed data is taken into account, regional, climatic, seasonal and diurnal effects can be observed. These effects must be accounted during the statistical processing of data. Such is the case of some regions and sites where for several reasons, wind speed data behaves as a bimodal and sometimes a multimodal distribution. This is due to the effect of local winds, including the Etesian winds in the Eastern Mediterranean (Turkey and Greece) [13], the combination of mountain and sea winds such as seen in a Mexico site [14], and the channeling as in Butler Grade region [15], among others.

Limitations of the Weibull distribution for wind speed data are also criticised in various studies as it is not good enough to capture null, low and high wind speeds, which causes overestimation or underestimation in terms of power production [16–19]. Weaknesses of Weibull to explain tails are frequently emphasised, and the issues related to diurnality and annual seasonality are already referred by literature [20]. Furthermore, fit between wind speed data and Weibull does not perform well in comparison with other probability density distributions, as shown in [21] and [22] where 59 and 5 distributions, respectively, where compared. In a recent study the Weibull distribution was used to fit clear bi-modal input data just indicating the superiority of Weibull on Rayleigh [23]. This was done after a detailed literature review where the authors clearly gained awareness of the limitations of the Weibull distribution, just for the sake of simplicity.

Wind speed data are simplified by measuring as averaged, treating as continuous variable, grouping in 0.5 or 1 m/s bins and neglecting at the tails of Weibull, which result in a certain estimation error when performing energy evaluations. It is a well known fact that application of the Weibull distribution to wind speed data can cause systematic overestimation in long-term energy studies on a global scale [12] due to wind speed's cubic effect on power [24].

The Weibull distribution is broadly used in wind energy to represent wind speed data even knowing its limitations. Commercial softwares such as WAsP and WindSim tend to show their simulation results and wind statistics with Weibull distribution [25] [23]. Even in the the International Electrotechnical Commission (IEC) standards, the Weibull distribution is suggested, as the default tool, for wind power production estimation and safety modelling [26], [27].

Nevertheless, still some studies of other distributions applied to wind speed data can be found in the literature, although these cases are out numbered. As an example, Jung [21] compares the goodness of fit of several probability distributions concluding that Log-normal distribution gives a better fit than Weibull, using also two parameters. Other studies claim that Mixture Weibull distribution is superior to traditional Weibull, specially for bi-modal wind regimes [22], [18], but having as disadvantage the increased number of parameters and the model complexity. Celik also provides an evaluation tool for selecting the best probability distribution [22].

In this paper a time-efficient compact methodology to enhance the goodness of fit, based on characteristics of input data, is provided. A procedure is presented with well defined guidelines for the task of processing wind speed data and with the capability to distinguish between unimodal and multimodal data distributions. The Classical Weibull probability density function (PDF) approach is combined with data subdivision into several unimodal Weibull PDF's and the multi-modality approach using more sophisticated PDF's such as the Beta Exponentiated Power Lindley (BEPL) [28] distribution and the Mixture Weibull. The figures of merits are the annual energy production (AEP) and the root mean square error(RMSE) results.

2. Approach and Methods

The developed procedure is as follows. First, data are inspected using Silverman test [29] and visual inspection in order to detect the type of distribution to be used (uni/multi modal). Then in case of unimodal distribution, classical Weibull PDF approach is used. On the contrary, when data present multimodal behaviour, two possibilities are proposed. If the origin of the multimodal behaviour can be determined, input data are processed through a transformation procedure finishing into several unimodal Weibull PDF's. This is done with the help of the other meteorological variables, temperature, pressure, humidity, wind direction, seasonality and diurnal variability. In the case that the origin of the multi-modaltiy can not be clearly found, the recently announced Beta Exponentiated Power Lindley (BEPL) [28] and Weibull mixture distributions are used as probability density functions in order to determine which one fit best. Mathematical expressions for the Weibull and the Weibull mixture distributions [30] are given in (1) and (2) respectively:

$$Weibull(v;\alpha,\beta) = (\frac{\alpha}{\beta}) * (\frac{v}{\beta})^{\alpha-1} * exp(-((\frac{v}{\beta})^{\alpha}) \tag{1}$$

$$Mixture\ Weibull(v;\alpha_1,\beta_1,\alpha_2,\beta_2) = mix * f(v;\alpha_1,\beta_1) + (1-mix) * f(v;\alpha_2,\beta_2) \tag{2}$$

where α is the shape, β is the scale, mix is the mixture factor and v represents the wind speed variable. The Beta Exponential Power Lindley [28] is given by equation (3):

$$\begin{aligned} BEPL(v;\alpha,\beta,\omega,a,b) = \frac{\alpha * \beta^2 * \omega}{B(a,b) * (\beta+1)} * (1+v^{\alpha}) * v^{\alpha-1} * exp(-\beta * v^{\alpha}) * \\ (1-(1+\frac{\beta * v^{\alpha}}{1+\beta}) * exp(-\beta * v^{\alpha}))^{\omega * a-1} * \\ (1-(1-(1+\frac{\beta * v^{\alpha}}{(1+\beta)}) * exp(-\beta * v^{\alpha}))^{\omega})^{b-1} \end{aligned} \tag{3}$$

where α is the shape, β is the scale and v represents the wind speed variable, as before, and a, b, and ω are the BEPL parameters. Finally, $B(a,b)$ is the well known *Beta* function shown in equation (4):

$$B(a,b) = \frac{\Gamma(a) * \Gamma(b)}{\Gamma(a+b)} \text{ ,with } \Gamma(x) = (x-1)! \tag{4}$$

The selection of the model parameters of equations (1), (2) and (3) is a crucial step in order to obtain a good representation of the stochastic variable, wind speed in our case. These parameters are obtained using the Maximum Likelihood estimation model. This model has been proven superior among others in various studies but it has to be taken into account that this claim is valid in average because the terrain and the weather characteristics can cause differences [31–35]. The R system [36] with the help of the packages *fitdistrplus* [37] and *bbmle* [38] were used for the estimation of the parameters.

The performance of the method is evaluated using the annual energy production (AEP) of a known wind turbine. Here, the Energcon E-48 power curve has been used [39]. Wind speed values have been obtained from the actual data and from the fitted distribution functions are combined with the wind turbine power curve for the estimation of AEP values. These values are then compared using the absolute differences and the Root Mean Square Error (RMSE).

Figure 1 presents the flowchart of the proposed methodology. This process can produce three possible outcomes as a solution to the misfit occurrence between wind speed data and the classical uni-modal Weibull PDF.

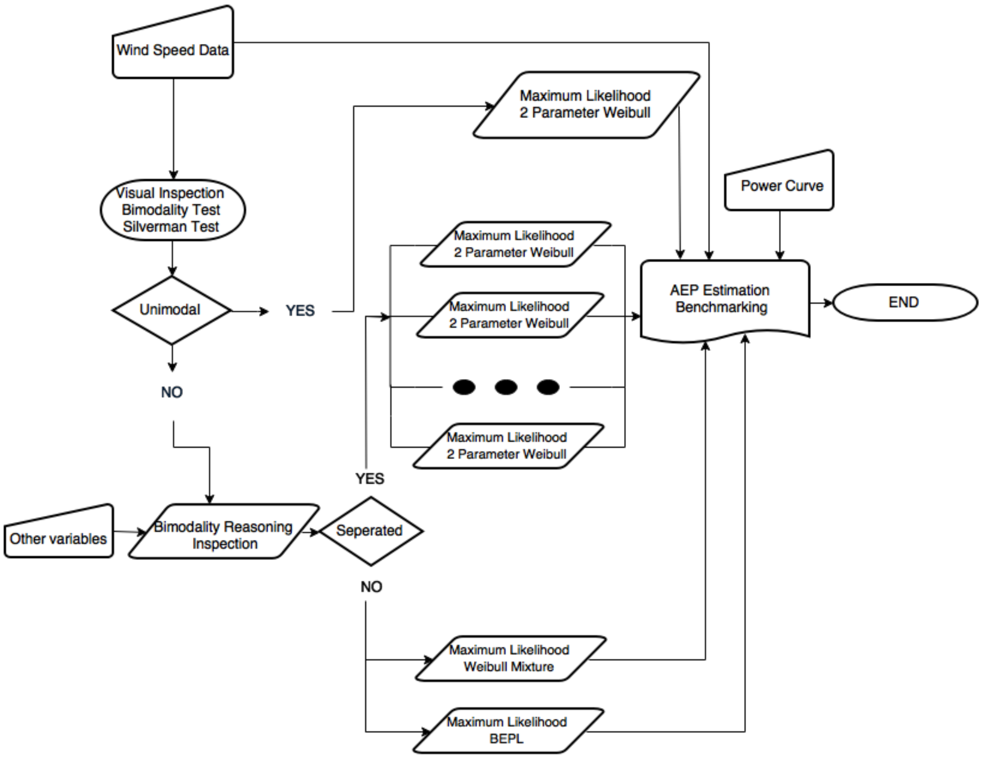

Figure 1. Flow Chart of Algorithm

3. Results

As a case study we used a training data-set covering 1 year of data for wind speed, wind direction and pressure. The wind speed distribution is shown in Figure 2. Multi-modality for this dataset can be detected by visual inspection without performing any detailed test. Anyway, it would also be possible to come across with challenging datasets where multi-modality detection is not possible using only visual inspection. Then, some statistical steps for the detection of multi-modality were added to the methodology. The reader should remind that when the data is subsetted, total observation number changes. In another words the change on sample space occurs, because of that y axis maximum density limits can vary between different inspections related to subset type.

Bimodality test results are consistent with the outcome of the visual inspection for the wind speed data. The smaller P-Values indicate bimodality occurrence in data set. In our case, test resulted with 0 value. Figure 3 is given as an illustrative example, to understand results of bimodality test with the typical bi-modal and uni-modal input data.

- Bi-Modal input and bimodality test result P-Value=0
- Uni-Modal input and bimodality test result P-Value=0.5

The second test is the Silverman test. The difference between the Silverman and the bimodality tests comes from the Silverman test's flexibility to multi-modality (higher degree than 2 modes) and the definition of the null hypothesis type. The significance value is taken as 0.05 to test validity of null hypothesis. The Silverman test results are given in below. The first test indicates uni-modality with its null hypothesis definition and it is rejected. The second test indicates bi-modality which can not be rejected.

- Silvermantest: Testing the hypothesis if the number of modes is <= 1, the P-Value is 0

- Silvermantest: Testing the hypothesis if the number of modes is <= 2, the P-Value is 0.3243243

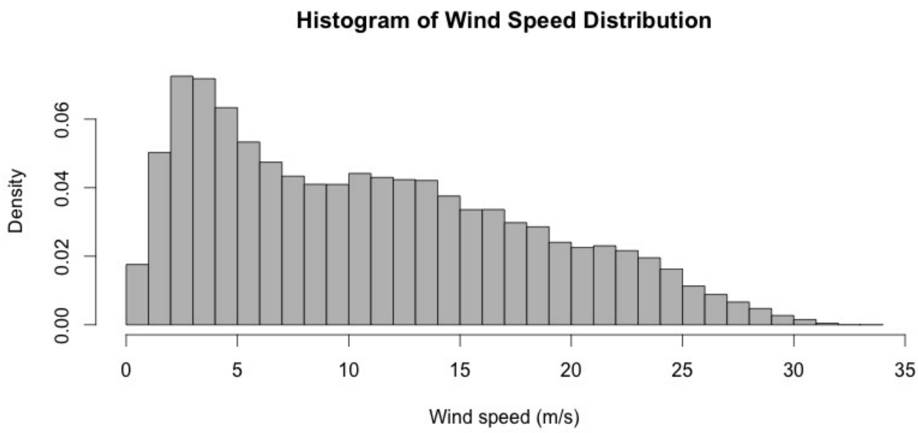

Figure 2. Raw wind speed data distribution

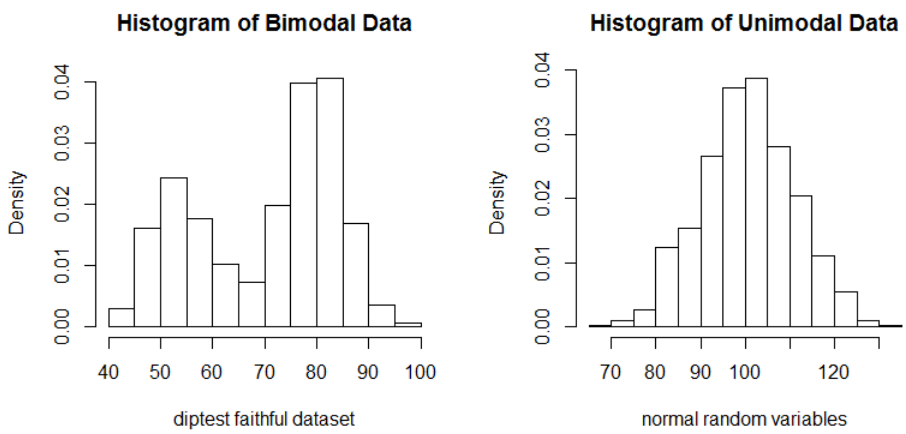

Figure 3. Uni-Modal vs Bi-Modal Data

These results lead to Bi-modality reasoning inspection using other recorded variables as wind direction, pressure and day time (mainly day/night). In the next paragraphs, the inspection of the raw data, taking into account the influence of the wind direction, the atmospheric pressure and the day time variables is presented.

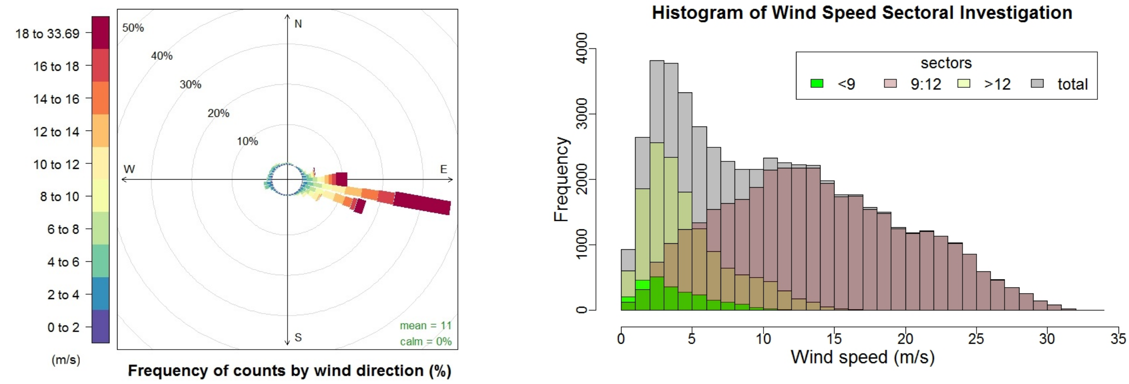

Figure 4. Wind rose and sectoral frequencies

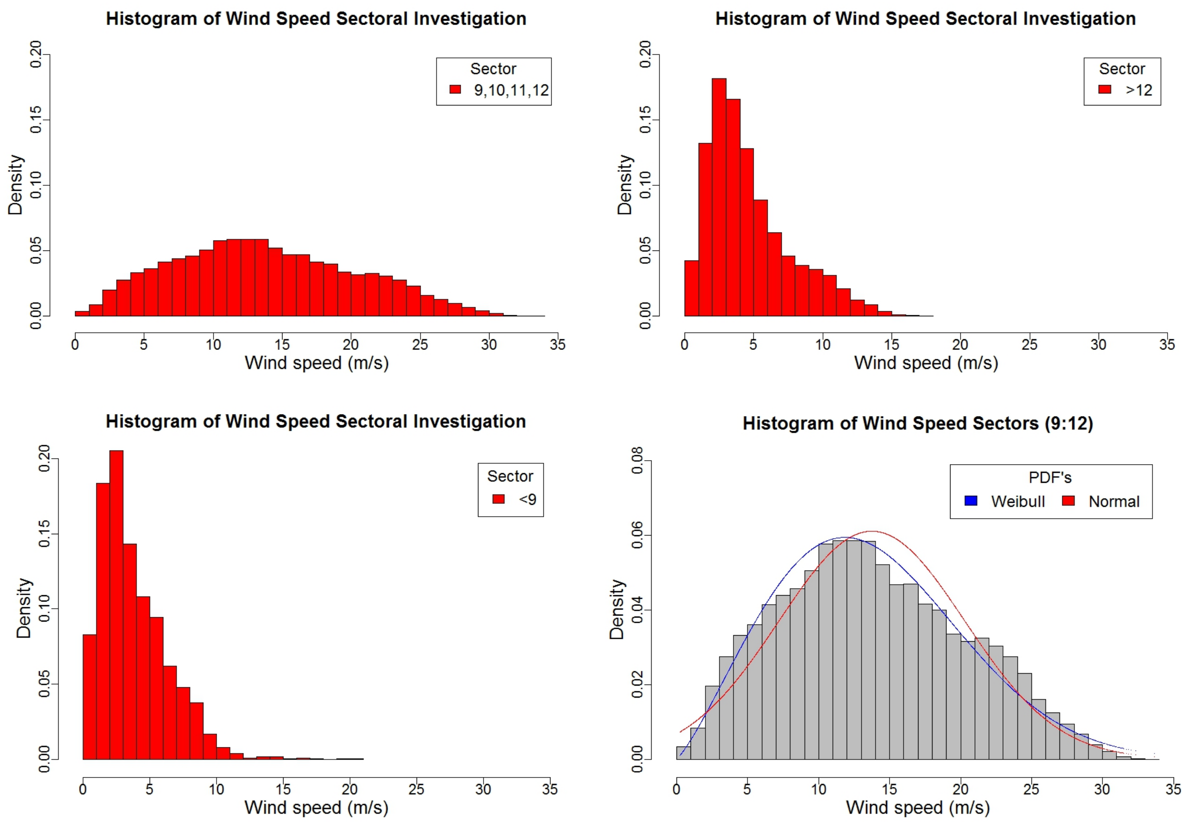

Figure 5. Separation due to directional data

Directional inspection: The wind speed dataset was divided into 36 wind direction sectors. Taking into account the wind rose, showed in Figure 4, three sub-sets were constructed to separate data as shown in Figure 4 and Figure 5 where it can be seen that bi-modality disappears in some degree with this filtering. The cumulative frequency histogram is plotted to illustrate the effects of wind speed subsets via sectors, intersection of different sectors causes the occurrunce of multi modality.

In Figure 5, wind speed subsets with positively skewed histograms fit well to Weibull distribution. An example comparison is added to figure for the leading sector subset. This division seem normally distributed however when the goodness of fit comparison is performed Weibull distribution gives better fit than normal distribution.

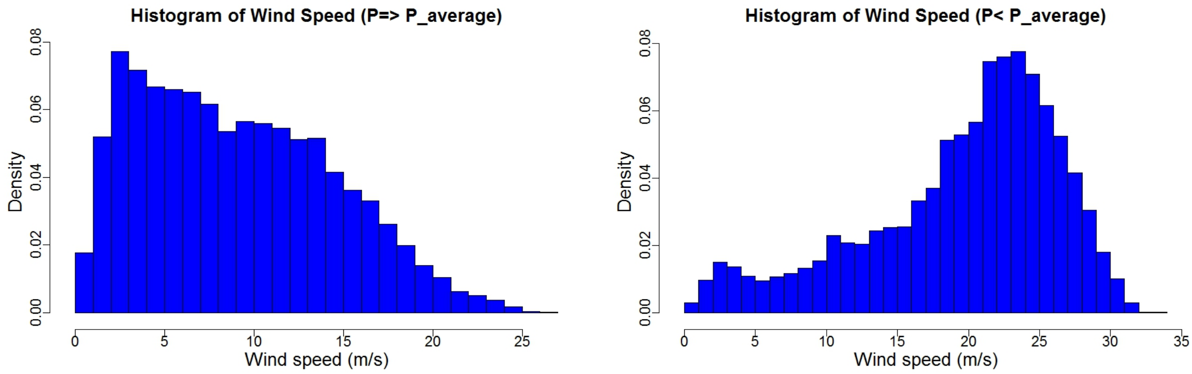

Figure 6. Inspection based on pressure

Pressure based inspection: Figure 6 shows the effect of the atmospheric pressure anomalies into wind speed. The mean value of the measured atmospheric pressure was used in the sub-setting process.

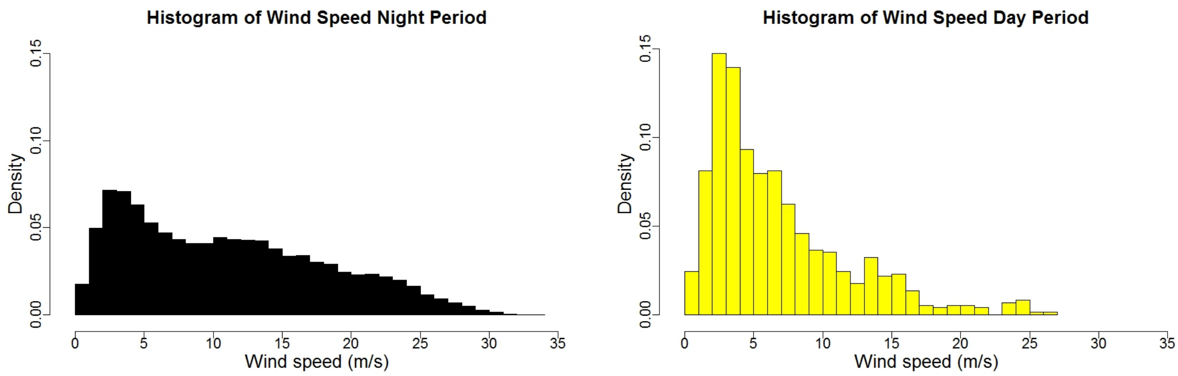

Figure 7. Inspection based on diurnal effects

Diurnal inspection: Herein day period is defined as the time interval between sun-rise and sun-set in order to sub-setting wind speed data. In case of day period, bi-modality disappears however in case of night period bi-modality still remains as shown in Figure 7.

Although multi-modality was cleaned to a certain degree, the third phase of the proposed methodology is presented in the following.

Table 1. Relative AEP difference and RMSE obtained with different PDF's

	Weibull	BEPL	MixW	Wdivided
Relative AEP difference	5.73%	2.45%	0.85%	0.82%
RMSE	21.81%	9.33%	3.23%	3.11%

Table 1, shows the bechmark between Weibull, Beta Exponentiated Power Lindley (BEPL), Mixture Weibull (MixW) and 3 Weibull distributions (Wdivided = Wind speed dataset is divided into 3 sub-sets due to wind direction sectors to clean multi-modality) is presented. It can be seen that in this case of study, the Weibull approach can be improved significantly using both, the Mixture Weibull distribution or the division of data into three Weibull PDF's. Figure 8 presents the comparison of the PDF's fits to actual raw data. It can be seen how the traditional Weibull fails not only with the number of modes but also in the fitting of the body and the tail of the distribution.

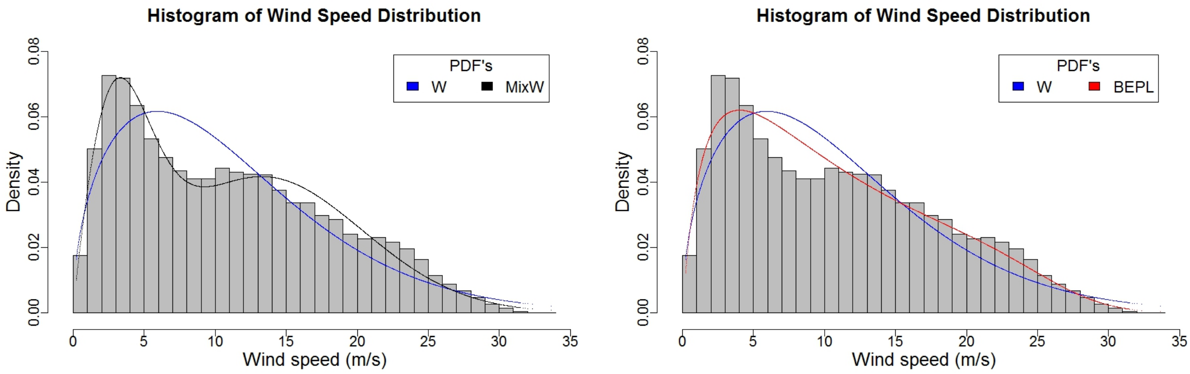

Figure 8. Comparison of Weibull and Mixture Weibull fits (left) and Weibull and BEPL fits (right) with wind speed raw data.

4. Conclusion

Weibull PDF approach is used systematically to process wind data. Nevertheless, there exist cases where this approach is not good enough for obtaining accurate power production and AEP results. A combined approach using multi-modal PDF's and/or multi-modality cleaning, separating input data into several Weibull PDF's, has been proposed. Results obtained with both methods are similar and present a significant improvement when compared with the Weibull approach, so the multi-modal methodology can be used even in the case of not finding the root origin of data behavior. The multi-modality cleaning approach is performed with sub-setting of the wind speed, based on the wind direction data, the atmospheric pressure anomalies and the wind speed changes on diurnal periods. However this sub-setting approach can be performed with different variables and in different data cases. Regarding the usage of multi-modal PDF's, model parameter estimation is still the most challenging issue. We experienced also this phenomena. Model parameter estimation process of multi-modal PDF's is quite time-consuming due to the computational complexity. Nevertheless, the input data is the decision making mechanism for the selection of the distribution type. The high computational complexity or the amount of the required input variables determine the level of accuracy.

Another outcome of the results is the importance of the data collection, not only wind speed and direction but also the other meteorological parameters such as humidity, pressure and temperature have a significant place in wind energy applications. The authors desire to encourage wind farm owners to continue on data gathering of all possible meteorological variables to track anomalies and prevent losses or increase the accuracy in estimations during the whole lifetime of a wind turbine.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 642108.

References

[1] Li X L 2011 *Green Energy Basic Concepts and Fundamentals* (London: Springer)
[2] Zhang M H 2015 *Wind Resource Assessment and Micro-siting: Science and Engineering* (John Wiley & Sons)
[3] Rayleigh L 1880 *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **10** 73–78
[4] Weibull W 1951 *Journal of applied mechanics* **103**
[5] Isyumov N 2012 *Journal of Wind Engineering and Industrial Aerodynamics* **104** 12–24
[6] Davenport A G 1964 *ICE Proceedings* vol 28 (Thomas Telford) pp 187–196
[7] Shu Z, Li Q and Chan P 2015 *Applied Energy* **156** 362–373
[8] Celik A N 2003 *Journal of Wind Engineering and Industrial Aerodynamics* **91** 693–707
[9] Fyrippis I, Axaopoulos P J and Panayiotou G 2010 *Applied Energy* **87** 577–586
[10] Altunkaynak A, Erdik T, Dabanlı İ and Şen Z 2012 *Applied Energy* **92** 809–814
[11] Tuller S E and Brett A C 1984 *Journal of Climate and Applied Meteorology* **23** 124–134
[12] Fant C, Gunturu B and Schlosser A 2016 *Applied Energy* **161** 565–573
[13] Akdağ S, Bagiorgas H and Mihalakakou G 2010 *Applied Energy* **87** 2566–2573
[14] Jaramillo O and Borja M 2004 *Renewable Energy* **29** 1613–1630
[15] Draxl C, Clifton A, Hodge B M and McCaa J 2015 *Applied Energy* **151** 355–366
[16] Van Donk S, Wagner L E, Skidmore E L and Tatarko J 2005 *Transactions of the American Society of Agricultural Engineers,ASAE* **48** 503–510
[17] Carta J and Ramirez P 2007 *Renewable Energy* **32** 518–531
[18] Carta J A, Ramirez P and Velazquez S 2009 *Renewable and Sustainable Energy Reviews* **13** 933–955
[19] Greene S and Morrissey M 2011 *Advanced Wind Resource Characterization and Stationarity Analysis for Improved Wind Farm Siting* (InTech)
[20] Drobinski P and Coulais C 2012 *arXiv preprint arXiv:1211.3853*
[21] Jung C 2016 *Energies* **9** 344
[22] Celik A, Makkawi A and Muneer T 2010 *Journal of renewable and sustainable energy* **2** 013102
[23] Ally C, Bahadoorsingh S, Singh A and Sharma C 2015 *Renewable and Sustainable Energy Reviews* **51** 863–874
[24] Morrissey M L, Cook W E and Greene J S 2010 *Journal of atmospheric and oceanic technology* **27** 1153–1164
[25] WindSim 2010 *Energy* URL `http://windsim.com/`
[26] IEC 61400-12-1: Power performance measurements of electricity producing wind turbines 2005
[27] IEC 61400-1: Wind turbines part 1: Design requirements 2005
[28] Pararai M, Warahena-Liyanage G and Oluyede B O 2015 *Theoretical Mathematics & Applications* **5**
[29] Silverman B W 1981 *Journal of the Royal Statistical Society. Series B (Methodological)* 97–99
[30] Kollu R, Rayapudi S R, Narasimham S and Pakkurthi K M 2012 *International Journal of Energy and Environmental Engineering* **3** 1–10
[31] Chang T P 2011 *Applied Energy* **88** 272–282
[32] Azad A K, Rasul M G and Yusaf T 2014 *Energies* **7** 3056–3085
[33] Saxena B K and Rao K V S 2015 *Renewables: Wind, Water, and Solar* **2** 1–11
[34] Kantar Y M and Şenoğlu B 2008 *Computers & Geosciences* **34** 1900–1909
[35] Costa Rocha P A, de Sousa R C, de Andrade C F and da Silva M E V 2012 *Applied Energy* **89** 395–400 ISSN 03062619 URL `http://dx.doi.org/10.1016/j.apenergy.2011.08.003`
[36] R Core Team 2015 *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing Vienna, Austria URL `https://www.R-project.org/`
[37] Delignette-Muller M L and Dutang C 2014 *J. Stat. Softw* **64** 1–34

Journal of Physics: Conference Series **753** (2016) 032067 doi:10.1088/1742-6596/753/3/032067

[38] Bolker B 2016 *Tools for General Maximum Likelihood Estimation* R Foundation for Statistical Computing URL `https://cran.r-project.org/web/packages/bbmle/bbmle.pdf`
[39] Enercon 2015 *Enercon Product Overview* URL `http://www.enercon.de/`