Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2016) 24: 4724 – 4735

doi:10.3906/elk-1412-207

A comparative analysis of wind speed probability distributions for wind power assessment of four sites

Vaishali SOHONI*, **Shivcharan GUPTA**, **Rajeshkumar NEMA**
Department of Electrical Engineering, Maulana Azad National Institute of Technology, Bhopal, India

Received: 31.12.2014 • **Accepted/Published Online:** 03.09.2015 • **Final Version:** 06.12.2016

Abstract: In this paper, five probability distribution functions are employed to fit the wind speed data from four different geographical locations in the world in a preliminary analysis. These wind regimes are selected such that they represent wide ranges of mean wind speeds and present different shapes of wind speed histograms. The wind speed data used for modelling consist of 10-min average SCADA data from three US wind farms and hourly averages recorded at a weather station in Canada. Out of the five, three functions, namely Weibull, Rayleigh, and gamma, which provide a better fit to the data, are selected to carry out further analyses. This study investigates the ability of these functions to match different statistical descriptions of wind regimes. Parameter estimation is done by the method of moments, and models are evaluated by root mean square error and R square methods. The suitability of PDFs to predict the wind power densities and annual energy production using manufacturers' power curve data at three of the selected sites is analysed. Power curves extracted from actual data of one wind farm using novel four- and five-parameter logistic approximations are also introduced here for energy analyses.

Key words: Wind speed distribution, probability distribution function, method of moments, wind power density

1. Introduction

Wind energy is clean, renewable, and one of the fastest growing alternative energy sources. However, the intermittent nature of wind and increasing penetration of wind energy into power systems can have adverse effects on the reliability and stability of power systems. With the increase in wind power installations all over the world, development of models, methods, and computing tools that can help in minimising these adverse effects has gained significance. Wind power production is highly dependent on the wind speed encountered at a site. Accurate assessment of wind energy potential at a candidate site requires detailed knowledge of wind characteristics of the location, and becomes a challenging task because of the highly unpredictable nature of wind. A function that can represent the wind speed data conveniently is often required in several wind power based applications [1]. Wind speed at a site varies randomly and its variation in a certain region over a period of time can be represented by different probability distribution functions (PDFs). Two-parameter Weibull distribution is the most commonly used and accepted distribution as it is found to be a fairly accurate PDF for most of the wind regimes found in nature and also it is simple to use [2]. However, it is not suitable for certain wind regimes, such as those having high frequencies of null winds and for short time horizons [3,4]. Another distribution Rayleigh PDF has also been used in many applications, owing to its simplicity and ability of describing wind speed with sufficient accuracy when little detail is available about the wind characteristics of

*Correspondence: vaishalisohonibpl@gmail.com

a site [5]. In [6,7] it has been noted that a three-parameter Weibull distribution can represent the wind speed data more accurately than the two-parameter Weibull function because of the addition of a third parameter. This additional location parameter establishes a lower bound, which is taken to be zero in the two-parameter Weibull distribution; however, the parameter estimation in this method becomes difficult. A hybrid density function that takes into account the null wind speeds was suggested by [8]. Mixture distributions have been used to fit the bimodal shape of wind speed histograms [9–11]. A number of other PDFs have also been proposed by various researchers to characterise wind speed frequency distributions; they include gamma, lognormal, inverse Gaussian, beta, Burr, Wakeby, kappa, and hybrid distributions [12–16]. In [17], two flexible families of distributions, the skewed generalised error and skewed t, have been proposed for the description of wind speed. Both of these distributions were found to have the flexibility of accommodating the shape of the wind speed data, which included some well-known distributions as special cases. In [16,18], different distribution functions are analysed for their suitability in offshore applications and in [18] Johnson S_E distribution has been introduced for the first time. In [13], seven different distributions were analysed to identify suitable distribution for an urban area and the Burr distribution recently applied to wind speed problems [12] was found suitable for urban applications. Six probability density functions, namely Weibull, Rayleigh, gamma, lognormal, inverse Gaussian, and maximum entropy principle (MEP) derived PDFs, were evaluated [19] using six goodness-of-fit criteria for five representative sites in North Dakota. Statistical distributions have been analysed in the literature for examining differences between the results for day and night, between various seasons, and for monthly and yearly data [17,20]. The suitability of these distribution functions for a particular application depends upon a number of factors such as the type of wind regime, availability of data, and recording intervals [3,16]. Moreover, a number of methods for estimating the parameters of these distributions and model evaluation criteria are given in the literature [1,21,22]. These methods have varying degrees of accuracy and complexity. Selection of an appropriate PDF to characterise wind speed data is of critical importance during the planning and operation stage of a wind based system and helps in improving the performance of the system. Further research is therefore required for evaluating the applicability of various distributions to different wind regimes and wind power problems. The selection of modelling methods and evaluation metrics is also crucial and should be given due importance. This paper presents a critical evaluation of Weibull, Rayleigh, and gamma distributions to describe four different wind regimes that characterise different statistical and geographical descriptions (Table 1). Additionally, the two-parameter lognormal and inverse Gaussian distributions are also fitted to the datasets but are not used in further analyses. The three selected functions for the subsequent study are examined on the basis of quality of fit using root mean square error (RMSE) and R square statistical indicators and prediction of average annual wind power density (WPD) and annual energy production (AEP) for sites A, B, and C.

A number of PDFs have been used in the literature to represent wind speed data of selected locations around the world. This study presents a new approach for analysis of PDFs compared to the earlier studies. The wind data series included here has been measured at four different sites. These sites are selected such that they characterise wide ranges of yearly mean wind speeds and present different shapes of histograms, so that the suitability of the selected PDFs to describe the varying characteristics of wind speed data can be analysed. Moreover, most of the previous studies evaluate the PDFs only on the basis of goodness of fit with the actual data set. As these distributions will ultimately be used in wind energy applications it is not appropriate to judge their suitability on the basis of goodness of fit parameters alone but it should also be examined how successfully these distributions can be employed for a particular application. Although the traditionally used Weibull and Rayleigh functions have been used for wind power potential evaluation of selected sites in some

Table 1. Data for wind speed modelling.

Data set	Geographical features of sites		Wind speed data description				
	Latitude and Longitude	Altitude	Mean	SD	Skewness	TI	Nulls
A	31.19°N 102.24°W	849.5 m	7.854 m/s	3.242 m/s	0.609	0.4255	-
B	33.84°N 116.54°W	140.6 m	2.392 m/s	1.96 m/s	2.296	0.8193	-
C	40.34°N 105.51°W	2625.2 m	11.728 m/s	8.685 m/s	1.06	0.7405	-
D	54.28°N 112.97°W	664 m	10.549 km/h	6.68 km/h	0.845	0.6113	54

SD: Standard deviation, TI: Turbulent intensity

studies [3,5,6,21], application and comparisons of other PDFs for wind energy assessment are also required by the wind industry. The three selected functions are therefore analysed here on the basis of their ability to predict the wind power potential and estimation of the energy at the selected sites. The earlier works compare predicted AEPs using the fitted PDFs with those calculated by the time series data for the candidate sites. In the study presented here, the data of actual energy being produced at wind farms A, B, and C are available; hence it was possible to compare the predicted AEPs with the annual energy actually produced at the wind farms. The energy assessment was done by using the manufacturer's power curve of a Vestas V 90 turbine. Ten turbines of this model are actually installed at each of these farms. Wake effect of the turbines is neglected. The turbines in a wind farm may produce less power due to the wake effect and underperformance of some turbines; therefore, instead of the manufacturer's curve, using the power curve derived from the actual data of wind farm can give better results in energy estimation [23]. Hence, in addition to the manufacturer's curve, power curves derived from the actual data of wind farm site A are also used for energy analyses. Novel four- and five-parameter logistic approximations [4,24] for describing the power curve have been used in this study to extract the power curves.

2. Data

Four sites that cover low, medium, and high mean wind speed ranges and represent different geographical locations have been considered for modelling. Datasets A, B, and C are 10-min average wind speeds obtained from SCADA (supervisory control and data acquisition system) data of NREL (National Renewable Energy Laboratory) wind farms measured at 100 m height and the data set D is hourly wind speed averages recorded at 10 m height at Alberta (Canada) weather station. Each of the datasets covers a period of 1 year. Other details of the wind speed data are given in Table 1. The corresponding 10-min average power output values recorded by the SCADA system of the three wind farms (datasets A, B, C) are also utilised to calculate the actual energy produced at these wind farms in the year. The mean wind speed m, standard deviation of wind speed s, turbulence intensity T, and skewness G have been calculated by using the following expressions [5,16,25,26]:

$$m = \frac{1}{N}\sum_{i=1}^{N} v_i \tag{1}$$

$$s = \left[\frac{1}{N-1}\sum_{i=1}^{N}(v_i - m)^2\right]^{\frac{1}{2}} \tag{2}$$

$$T = \frac{s}{m} \tag{3}$$

$$G = \frac{1}{N} \times \frac{\sum_{i=1}^{N} (v_i - m)^3}{s^3} \tag{4}$$

where v_i is the value of the ith wind speed and N is the number of wind speed records in the year.

3. Modelling methodology

Knowledge of wind characteristics of a site is required for wind resource assessment, cost optimisation studies, energy assessment, and siting [27]. Several PDFs, parameter estimation methods, and model evaluation criteria for wind speed modelling have been proposed in the literature. The modelling methodology is decided according to the availability of data, the complexity of methods, and the desired accuracy.

3.1. Probability distribution functions

The following distributions are used in this study for modelling the wind speed. The PDFs and the expressions used for parameter estimation of these distributions are given in Table 2.

Table 2. Expressions of statistical distributions and parameter estimation.

Distribution function	PDF $f(\nu)$	Parameters	
2 P Weibull [2,3]	$\frac{k}{c}\left(\frac{v}{c}\right)^{k-1} exp\left(-\frac{v}{c}\right)^{k}$	$k = \left(\frac{s}{m}\right)^{-1.086}$	$c = \frac{m}{\Gamma(1+1/k)}$
Rayleigh [28,32]	$\frac{2v}{c^2}\exp\left(-\frac{v}{c}\right)^2$	$k = 2$	$c = \frac{2}{\sqrt{\pi}}m$
2 P gamma [20]	$\frac{v^{\eta-1}}{\beta^{\eta}\Gamma(\eta)}\exp\left(-v/\beta\right)$	$\beta = \frac{s^2}{m}$	$\eta = \frac{m^2}{s^2}$
2 P lognormal [20]	$\frac{1}{v\beta\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{ln(v)-\alpha}{\beta}\right)^2\right]$	$\alpha = ln\left[\frac{m}{\sqrt{1+s^2/m^2}}\right]$	$\beta = \sqrt{ln\left(1+\frac{s^2}{m^2}\right)}$
2 P inverse Gaussian [32]	$\frac{\beta}{2\pi v^3} exp\left[-\frac{\beta(v-\alpha)^2}{2v\alpha^2}\right]$	$a = m$	$\beta = \frac{m^3}{s^2}$

k- shape parameter, c- scale parameter, Γ- gamma function, α, β, η- parameters of distributions

3.1.1. Two-parameter Weibull distribution

The two-parameter Weibull distribution is the most widely used distribution in wind power applications. It is a versatile PDF, is simple to use, and is found to be accurate for most of the wind regimes encountered in nature. However, in some studies it is reported that it is not able to represent the wind speed data accurately for certain applications, which include regimes with high frequency of nulls, low and high wind speeds, bimodal distributions, and short time horizons [3,4,16].

3.1.2. Rayleigh distribution

Rayleigh distribution is a special case of Weibull distribution in which the shape parameter is taken as $k = 2$. This function has also been used by many researchers due to its simplicity as there is only one parameter to be evaluated and is useful when not much information about the site is available. Although found to be suitable for some wind regimes [23], this function has less flexibility because it has a single model parameter.

3.1.3. Two-parameter gamma distribution

The two-parameter gamma distribution has been used for wind speed modelling in some studies. Its applicability to model low wind speeds is reported in [28].

3.1.4. Two-parameter lognormal distribution

The two-parameter lognormal distribution can be used if the variable is such that its logarithm has a normal distribution and has been used in some studies to represent wind speed data [23].

3.1.5. Inverse Gaussian distribution

The inverse Gaussian distribution has been advocated as an alternative to the three-parameter Weibull distribution in [29], for describing low probabilities of low wind speeds and because of the simplicity of parameter estimation.

In this study, the two-parameter Weibull, Rayleigh, gamma, two-parameter lognormal, and inverse Gaussian distributions are fitted to the wind speed data of the chosen sites. A detailed analysis of Weibull, Rayleigh, and gamma PDFs for describing the four different wind regimes is carried out. These models are assessed on the basis of goodness of fit criteria and their suitability for use in wind resource assessment and energy prediction. The analyses have been carried out in MATLAB.

3.2. Parameter estimation

Several methods have been used in the literature to determine the parameters of various distributions, out of which the graphical method, the method of moments, and the maximum likelihood method are the most commonly used [21,22,30]. The graphical method has the advantage of simplicity and has been used extensively in earlier studies. However, the accuracy of parameter estimation in this method is not good. The maximum likelihood method has good accuracy and is more robust than the above method but it requires iterative methods to obtain the parameter values. The method of moments equates a certain number of statistical moments of the sample with the corresponding population moments. Thus if μ_k' = E[Xk'] is the kth moment about the origin of a random variable X, and whenever it exists, the corresponding kth sample moment is given by $m_k^{'} = \frac{1}{n}\sum_{i=1}^{n} x_i^k$; then the estimator of μ_k by the method of moments is m$_k$'. This method is computationally easy and provides explicit estimators of the parameters [31]. The method of moments has been selected in this paper to calculate the parameters of the distributions as it has a fair degree of accuracy, the results are consistent, and the calculation of parameters is simple [32].

3.3. Model evaluation

Before using a particular function for a particular application, it is important to determine whether it is able to represent the actual wind speed distribution of the site appropriately. A visual comparison of the fitted distribution with the histogram of sample wind speed is a commonly used and convenient method [23,32]. A number of statistical indicators to measure the goodness of fit have been used in various works, which include RMSE, R^2, chi square, Kolmogorov–Smirnov, and Anderson–Darling tests [3,6,18]. The performance evaluation of these PDFs for estimation of wind energy potential is done in some studies. These works use manufacturers' power curves for energy assessment [3,9,33]. In this paper, the fitted PDFs are superimposed on the wind speed histograms of measured data and a visual comparison is done for a primary evaluation. The suitability of fit of the selected three functions is also evaluated on the basis of RMSE and R^2 criteria and their ability to predict the wind power potential and estimation of the energy at the selected sites. Energy evaluation has been done here by using the power curve data from the manufacturers. Two power curve models derived from the actual SCADA data of wind farm A are also used for estimation of annual energy. As the manufacturers' curves are created under standard conditions [34–36] they may not represent the realistic conditions of the site under consideration. In addition, the manufacturers' curves are suitable for predicting the power output of a single turbine of a specific type. In a big wind farm a number of turbines are spread over a wide area and the power produced by turbines with identical specifications can differ [37] due to variation in wind speed and direction encountered by different turbines and the wake effect of turbines, which causes reduced wind speeds at the turbines that operate in the wake of other turbines. The difference in power outputs can also be due to factors such as wear and tear, aging, and dirt or ice deposition on blades. Power curves derived from the actual data of wind speed and power measured from the turbines can incorporate the actual conditions at the wind farms. This study also compares the energies estimated using the manufacturer's and derived power curves. The criteria used to analyse the performance of these functions and the expressions used are given below.

3.3.1. Root mean square error

The RMSE between the actual probabilities and fitted probabilities is calculated by [3]

$$RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}(p_i - f_i)^2\right]^{\frac{1}{2}} \tag{5}$$

where p_i and f_i are the probability values of actual data and the fitted PDF respectively at the ith wind speed.

3.3.2. R^2 (coefficient of determination)

The R^2 between the actual data and fitted probabilities is calculated by [18]

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(p_i f_i)^2}{\sum_{i=1}^{N}(p_i \bar{p}_i)^2} \tag{6}$$

where p_i bar is the mean of actual probability.

3.3.3. Wind power density (WPD)

The average power density for the year is calculated by [5]

$$WPD = \sum_{i=1}^{N}\frac{1}{2}\rho v_i^3 f(v_i) \tag{7}$$

where $f(v_i)$ is the probability of wind speed at the ith speed value. The standard value of air density ρ is taken as $\rho = 1.225$ kg/m^3 [1]. The percentage error between the time series power density and that calculated by the fitted PDF is calculated.

3.3.4. Annual energy production

The energy calculations for data sets A, B, and C were carried out by using the manufacturer's curve data for a Vestas V 90 wind turbine. Ten turbines of this model are installed at each of these wind farms. The annual energy production predicted for each turbine with these fitted distributions is calculated by

$$E = \sum_{i=1}^{N} P_i f(v_i) T_i \tag{8}$$

where P_i is the power output of the turbine at ith wind speed obtained from the manufacturer's curve data and T_i is the hours for which the wind blows at the ith hour. The annual energy E_p predicted for ten turbines of these wind farms is then calculated (wake effect is neglected). The actual annual energy E_a produced in the wind farms is calculated from the 10-min averaged power output recorded by the SCADA system at the wind farms and the percentage error between annual energy calculated from these fitted PDFs and the actual annual energy produced is calculated.

In addition to the manufacturers' curve data, energy estimation is also done by extracting power curves, using the four-parameter logistic (4PL) and five-parameter logistic (5PL) approximations. The 4PL approximation is given by [34]

$$P = d + \frac{(a-d)}{1 + \left(\frac{v}{c}\right)^{b}} \tag{9}$$

The 5PL approximation is given by [24]

$$P = d + \frac{(a-d)}{\left(1 + \left(\frac{v}{c}\right)^{b}\right)^{g}} \tag{10}$$

where a is the minimum asymptote, b is the hill slope, c is the inflection point, d is the maximum asymptote of 4PL and 5PL functions, and g is the asymmetry factor of the 5PL function. The parameters of these functions are obtained using the least squares method by genetic algorithm optimisation using MATLAB.

4. Results and discussion

Figure 1 shows five PDFs, namely the two-parameter Weibull, Rayleigh, gamma, two-parameter lognormal, and the inverse Gaussian, fitted to the wind speed values of dataset A. Graphically it can be observed that the Weibull PDF yields the best fit. The Rayleigh and gamma distributions match the histogram to a lesser degree, whereas the two-parameter lognormal and the inverse Gaussian functions provide the poorest fits. Similarly, these five PDFs were also fitted to other wind regimes and it was observed that the two-parameter lognormal and the inverse Gaussian functions performed worst for all of the selected datasets; therefore, they were not used for further analyses. Figures 2 to 5 show the histograms of wind speeds at the four sites and the fitted Weibull, Rayleigh, and gamma PDFs superimposed on them. It can be seen from the figures that these sites present different shapes of histograms. The parameter values obtained for these distributions and the fitting accuracies based on RMSE and R^2 criteria are given in Table 3. It can be seen that both statistical indicators gave similar results in all cases. The Weibull PDF gives the least fitting error for datasets A and D. This is also verified from Figures 2 and 5 that for sites A and D the Weibull PDF is fitting best to the actual data. It is worthwhile noting here that site A characterises a good and site D a moderate mean wind speed value. The statistical tests show that the two-parameter gamma distribution is the best fit for dataset B, which is a site

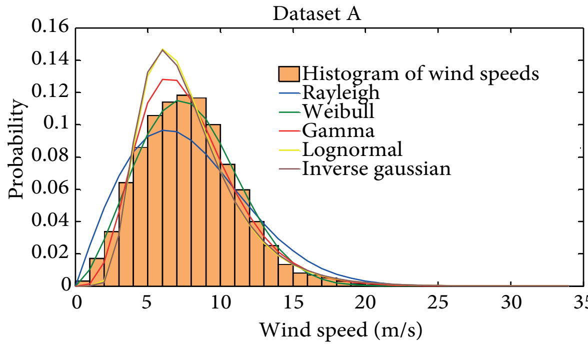

Figure 1. Histogram of dataset A fitted with Weibull, Rayleigh, gamma, two-parameter lognormal, and inverse Gaussian PDFs.

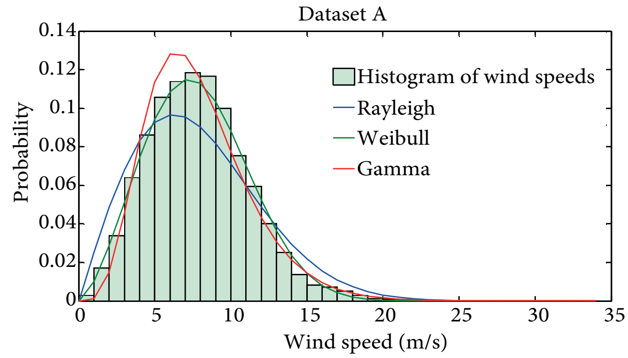

Figure 2. Histogram of dataset A fitted with Weibull, Rayleigh, and gamma PDFs.

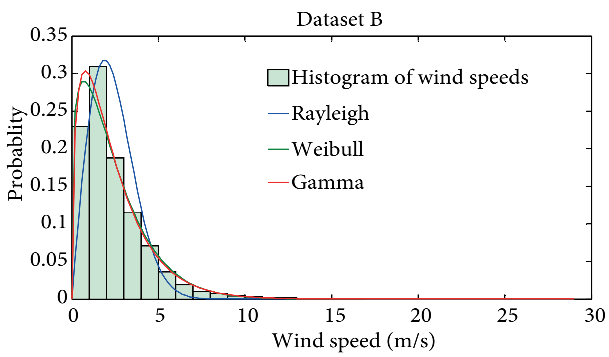

Figure 3. Histogram of dataset B fitted with Weibull, Rayleigh, and gamma PDFs.

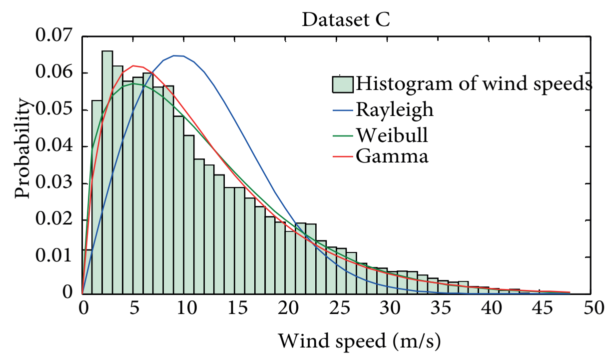

Figure 4. Histogram of dataset C fitted with Weibull, Rayleigh, and gamma PDFs.

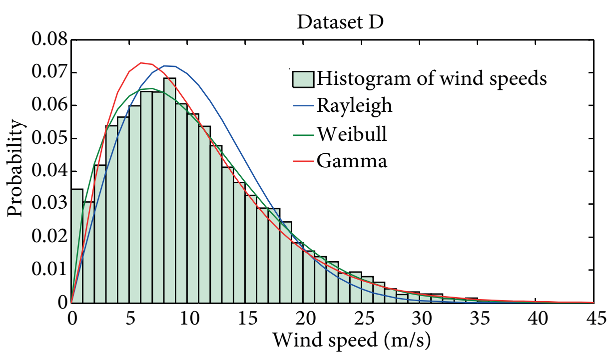

Figure 5. Histogram of dataset D fitted with Weibull, Rayleigh, and gamma PDFs.

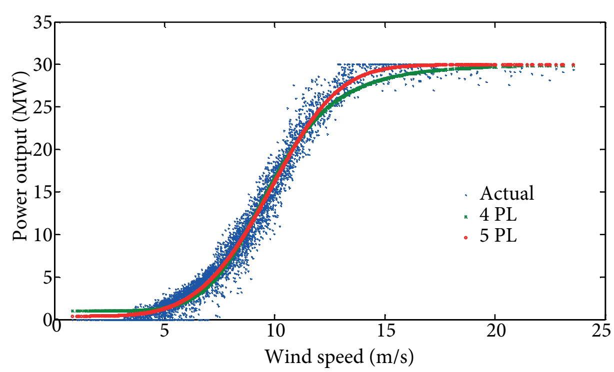

Figure 6. Power curves extracted from actual wind speed and power curve data site by four- and five-parameter logistic approximations.

with very low mean wind speed, and also for dataset C, which has a high mean wind speed value. However, the Weibull PDF also gives fairly accurate results for both sites. The Rayleigh PDF gives a very poor performance for site C and it is a poor fit for the other three sites also. These results also show that the Weibull PDF was the best fit for sites A and D, which have a lower value of skewness coefficient, and the gamma PDF was best for sites B and C, which have higher skewness values. The performance of these three PDFs for assessing the wind energy potential of the three wind farm sites was also analysed and the results are summarised in

Table 4. The Rayleigh PDF produced the maximum error among the three PDFs for all the sites and produced significant errors in wind power assessment for sites B and C. Overall, the Weibull and gamma PDFs resulted in less WPD and AEP errors. These three PDFs, namely Weibull, Rayleigh, and gamma, are ranked on the basis of four different evaluation criteria for the three sites in Table 5. It can be seen that the gamma function produced the least WPD and AEP errors for site A among the three functions, whereas Weibull was ranked number 1 based on RMSE and R^2 criteria. It can be said that evaluating these distribution functions based on the goodness of fit criteria alone is not sufficient. These criteria should be used for identifying suitable distributions before a detailed analysis is done. As these fitted PDFs are used for different applications by the wind industry, the performance of these PDFs for specific applications like wind power prediction should also be evaluated. The results show that there is an underestimation of power density in general. The percentage errors mostly show an overestimation of energy, which might be due to the wake effect and underperformance of some turbines in the wind farm. The power curves derived from actual data of the wind farm using 4PL and 5PL approximations are shown in Figure 6. Energy assessments using the manufacturer's curve and derived power curves are compared in Table 6. It is worthwhile noting here that the curves derived from the actual data of this site result in remarkably less errors in energy assessment compared to the manufacturer's curve. This might be due to consideration of the wake effect and other site specific factors in the derived curves that are not accounted for in the manufacturer's curve.

Table 3. Comparison of Weibull, Rayleigh, and gamma PDFs.

Data set	Type of distribution function								
	Weibull			Rayleigh			Gamma		
	Parameters	RMSE	R^2	Parameters	RMSE	R^2	Parameters	RMSE	R^2
A	k = 2.529 c= 8.849	0.0027	0.9954	k = 2 c = 8.859	0.0112	0.9244	$\beta = 1.422$ $\eta = 5.523$	0.0061	0.9779
B	k= 1.241 c= 2.564	0.0127	0.9722	k = 2 c = 2.698	0.0292	0.8537	$\beta = 1.605$ $\eta = 1.489$	0.0101	0.9823
C	k= 1.385 c= 12.846	0.0042	0.9574	k = 2 c = 13.229	0.0124	0.6365	$\beta = 6.432$ $\eta = 1.823$	0.0039	0.9636
D	k= 1.641 c= 11.785	0.0032	0.9805	k = 2 c = 11.89	0.0065	0.9191	$\beta = 4.233$ $\eta = 2.49$	0.0053	0.9468

Table 4. Comparison of Weibull, Rayleigh, and gamma PDFs for wind power density and energy assessment.

Data set	Time series WPD (w/m^2)	Actual AEP (Gwh)	Type of distribution function					
			Weibull		Rayleigh		Gamma	
			AEP (Gwh) and % error	WPD (w/m^2) and % error	AEP (Gwh) and % error	WPD (w/m^2) and % error	AEP (Gwh) and % error	WPD (w/m^2) and % error
A	471.91	81.98	92.87 13.28%	464.23 –1.62%	93.97 14.61%	566.23 19.98%	88.91 8.45%	477.41 1.16%
B	35.86	5.67	6.16 8.65%	31.33 –12.61%	2.07 –63.46%	15.99 –55.38%	6.392 12.73%	32.83 –8.45%
C	3039.42	107.56	120.77 12.28%	2860.16 –5.89%	153.18 42.41%	1885.02 –37.98%	119.07 10.75%	2843.16 –6.45%

Table 5. Ranking of Weibull, Rayleigh, and gamma PDFs based on four different criteria for sites A, B, and C.

Site	PDF	RMSE	R^2	WPD	AEP
A	Weibull	1	1	2 (–)	2 (+)
	Rayleigh	3	3	3 (+)	3 (+)
	Gamma	2	2	1 (+)	1 (+)
B	Weibull	2	2	2 (–)	1 (+)
	Rayleigh	3	3	3 (–)	3 (–)
	Gamma	1	1	1 (–)	2 (+)
C	Weibull	2	2	1 (–)	2 (+)
	Rayleigh	3	3	3 (–)	3 (+)
	Gamma	1	1	2 (–)	1 (+)

(+) shows an overestimation and (–) shows an underestimation

Table 6. Comparison of AEPS from manufacturers' curve and 4PL and 5PL power curve approximations derived from SCADA data site A.

Actual AEP (GWh)	Type of distribution	Manufacturer's curve		4 PL curve		5 PL curve	
		AEP (GWh)	% error	AEP (GWh)	% error	AEP (GWh)	% error
81.98	Weibull	92.87	13.28%	85.2	3.92%	85.84	4.7%
	Rayleigh	93.97	14.61%	87.42	6.63%	88.02	7.36%
	Gamma	88.91	8.45%	80.85	1.37%	81.66	0.39%

5. Conclusion

This paper presented a novel approach for analysing wind speed probability distribution functions based on their suitability to describe wind regimes with different statistical features and their capability of assessing wind energy potential for these sites. This study also introduced four-parameter and five-parameter logistic approximations to derive power curves from SCADA data of a wind farm for use in energy analyses. Five probability distribution functions were initially fitted to wind speed samples selected from four different topographical locations. A detailed analysis of two-parameter Weibull, Rayleigh, and gamma distributions was then carried out. For each selected location the most suitable distribution to describe the wind speed data was identified. The Weibull distribution was found the best option for two sites with moderate values of mean wind speeds and the gamma distribution performed best for low and high wind speed sites. The Weibull function fitted well to wind speed data with less skewness, whereas data that presented highly skewed histograms were represented better by the gamma function. These two distributions also gave good results for assessing the wind energy potential of these sites. The Rayleigh function was not suitable for any of the selected locations. Choice of appropriate function for describing wind speed frequency distribution is a crucial requirement. With the growth of the wind power industry all over the world, most of the high wind speed areas with flat terrains have been exploited. Recently, sites with low and medium wind speeds, complex terrains, urban environments, and offshore areas are being identified for wind power installations. Further research should focus on the applicability of various functions to match the wind speed encountered in these areas, so that maximum utilisation of the wind potential can be ascertained.

References

[1] Mathew S. Wind Energy: Fundamentals, Resource Analysis and Economics. Berlin, Germany: Springer, 2006.

[2] Justus CG, Hargraves WR, Mikhail A, Graber D. Methods for estimating wind speed frequency distributions. J Appl Meteorol 1978; 17: 350-353.

[3] Akdag S, Dinler A. A new method to estimate Weibull parameters for wind energy applications. Energ Convers Manage 2009; 50: 1761-1766.

[4] Kusiak A, Zheng H, Song Z. On line monitoring of power curves. Renew Energ 2009; 34: 1487-1493.

[5] Johnson GL. Wind energy systems. Manhattan, KS, USA: Electronic Edition, 2001.

[6] Auwera V, Meyer LF, Malet LM. The use of the Weibull three parameter model for estimating mean wind power densities. J Appl Meteorol 1980; 19: 819-825.

[7] Stewart D, Eseenwanger O. Frequency distribution of wind speed near the surface. J Appl Meteorol 1978; 17: 1633-1642.

[8] Tackle E, Brown J. Note on the use of Weibull statistics to characterize wind speed data. J Appl Meteorol 1978; 17: 556-559.

[9] Carta JA, Ramirez P. Use of finite mixture distribution models in the analysis of wind energy in the Canarian Archipelago. Energ Convers Manage 2007; 48: 281-291.

[10] Jaramillo O, Borja M. Wind speed analysis in La Ventosa, Mexico: a bimodal probability distribution case. Renew Energ 2004; 29: 1613-1630.

[11] Carta JA, Mentado D. A continuous bivariate model for wind power density and wind turbine energy output estimations. Energ Convers Manage 2007; 48: 420-432.

[12] Usta I, Kantar YM. Analysis of the Burr type XII distribution for estimation of wind speed distributions. In International Science, Social Science, Engineering and Energy Conference; 2012, Bangkok: Proceedings - International Science and Engineering 2013. pp. 302-308.

[13] Brano V, Orioli A, Ciulla G, Culoota S. Quality of wind speed fitting distributions for the urban area of Palermo. Renew Energ 2011; 36: 1026-1039.

[14] Justus C, Hargraves W, Walcin A. Nationwide assessment of potential output from wind powered generators. J Appl Meteorol 1976; 15: 673-678.

[15] Carta JA, Ramirez P, Velazquez V. Influence of the level of fit of a probability density function to wind-speed data on the WECS mean power output estimation. Energ Convers Manage 2008; 49: 2647-2655.

[16] Morgan EC, Lackner M, Vogel RM, Bais LG. Probability distributions for offshore wind speeds. Energ Convers Manage 2011; 52: 15-26.

[17] Usta I, Kantar YM. Analysis of some flexible families of distributions for estimation of wind speed distributions. Appl Energ 2012; 89: 355-367.

[18] Soukissian T. Use of multi-parameter distributions for offshore wind speed modelling: the Johnson S_E distribution. Appl Energ 2013; 111: 982-1000.

[19] Zhou J, Erdem E, Li G, Shi J. Comprehensive evaluation of wind speed distribution models: a case study for North Dakota sites. Energ Convers Manage 2010; 51: 1449-1458.

[20] Morgan VT. Statistical distributions of wind parameters at Sydney, Australia. Renew Energ 1995; 6: 39-47.

[21] Seguro JV, Lambert TW. Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. J Wind Eng Ind Aerod 2000; 85: 75-84.

[22] Costa Rocha PA, De Sousa RC, de Andrade CF, Da Silva MEV. Comparison of seven numerical methods for determining Weibull parameters for wind energy generation in the northeast region of Brazil. Appl Energ 2012; 89: 395-400.

[23] Olaofe Z, Folly K. Wind energy analysis based on turbine and developed site power curves: a case study of Darling city. Renew Energ 2013; 53: 306-318.

[24] Lydia M, Selvakumar AI, Suresh Kumar S, Prem Kumar E. Advanced algorithms for wind turbine power curve modeling. IEEE T Sustain Energ 2013; 4: 827-835.

[25] Burton T, Sharpe D, Jenkins N, Bossanyi. Wind Energy Handbook. West Sussex, UK: Wiley, 2001.

[26] Jay Kerns G. Introduction to Probability and Statistics Using R. 1st ed. ISBN: 978-0-557-24979-4, 2011.

[27] Manwell J, McGowan G, Rogers A. Wind Energy: Explained Theory, Design and Application. 2nd ed. West Sussex, UK: Wiley, 2009.

[28] Olaofe Z, Folly K. Statistical Analysis of the Wind Resources at Darling for Energy Production. International Journal of Renewable Energy Research 2012; 2: 250-261.

[29] Bardsley WE. Note on the use of the inverse Gaussian distribution for wind energy applications. J Appl Meteorol 1980; 19: 1126-1130.

[30] Gebizlioglu OL, Senoglu B, Kantar YM. Comparison of certain value-at-risk estimation methods for the two parameter Weibull loss distribution. J Comput Appl Math 2011; 235: 3304-3314.

[31] Kantar YM, Senoglu B. A comparative study for the location and scale parameters of the Weibull distribution with given shape parameter. Computers and Geosciences 2008; 34: 1900-1909.

[32] Carta JA, Ramirez P, Velazquez V. A review of wind speed probability distributions used in wind energy analysis: case study in Canary Islands. Renew Sust Energ Rev 2009; 13: 933-955.

[33] Chang TP. Estimation of wind energy potential using different probability density functions. Appl Energ 2011; 88: 1848-1856.

[34] Rodbard D. Statistical quality control and routine data processing for radioimmunoassay and immune radiometric assays. Clin Chem 1974; 20/10: 1255-1270.

[35] International standard IEC 61400-12-1, 2005, Wind turbines- Power performance measurement of electricity producing wind turbines.

[36] Lydia M, Selvakumar AI, Suresh Kumar S, Prem Kumar E. A comprehensive review on wind turbine power curve modelling techniques. Renew Sust Energ Rev 2014; 30: 452-460.