Statistical Modelling of Long-Term Wind Speed Data

S.M. Lawan*[1], W.A.W.Z. Abidin[2], W.Y. Chai[1], A. Baharun[1] and T. Masri[1]

[1]*Faculty of Engineering, Universiti Malaysia, Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia.*
[2]*Faculty of Computer Science and Information Technology, Universiti Malaysia, Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia*

***Corresponding Email:** 13010004@siswa.unimas.my*

ABSTRACT

The attention of most countries of the world has been shifted towards reducing the occurrences of greenhouse gasses, developing of renewable energy and energy efficiency towards building a sustainable energy in the near future. Wind energy as one of these renewable is perhaps the most suitable, clean and environmental friendly. In modeling wind speed, Weibull function is the most widely adopted model in the scientific literatures, however, other statistical functions are also need to be considered and judged their suitability based on certain criteria. In this study, five statistical models were selected for modeling of Miri wind speed data for a period of ten years. Distribution Function (PDF) and Probability (PP) plots are employed to verify the Goodness of fit (GOF) for the   distributions. Lastly, graphical and GOF outcomes are compared, suggesting that, Lognormal and Gamma distributions are found to be most appropriate as compared to the Weibull, Rayleigh and Erlag distributions.

Keywords: Renewable energy, Wind energy, Distribution model, Malaysia, Miri.

INTRODUCTION

As a consequence of present day energy requirement and increasing environmental awareness, it's become very important to complement our energy base with clean and renewable sources. Speedy advancement of human population and industrialization, the energy needs are growing exponentially, causing an increase in environmental pollution. In certain places, power generation is not often enough to satisfy even the minimum requirement, causing social problems in those areas.

For this reason, progression in the wind energy technology has produced it substitute to traditional energy resources in recent times. Similar to this development, wind energy conversion systems (WECS) make a tremendous participation in everyday life in producing countries, where one third of the world's people live without electricity. Wind is certainly one of the prospective alternative energy sources that may be harnessed in a small, medium and commercial ways.

Malaysia has a really good prospective of renewable energy. Lately, wind energy conversion is given a considerable concern in Malaysia. Considering the fact that the country is situated in the tropical location in between the latitudes $1°N$ and $7°N$ and also between the longitudes $100°E$ and $120°E$ and its weather conditions is influenced by the monsoons, the potential for wind energy technology in Malaysia is extremely reliant on the accessibility of the wind resource that ranges with particular location[1].

Choosing the most appropriate distribution of wind speed is essential during the technical feasibility stage. Consequently, regionalized study, according to statistical perspective is essential to model wind speed behaviour. A number of scientific studies have examined the modeling of suitable wind speed distribution, for instance, the 2-parameter Lognormal distribution performs best for estimating extreme wind speeds, but still gives estimates with significant error[2]. A long term hourly average wind speed data was adopted and analyzed in three different weather location in Hong Kong[3], the two parameter Weibull has been applied in modeling the distribution of wind speed, Based on the data, it was observed that 2 parameters Weibull varied for these weather station over a broad range period.

A study conducted at Gelibolu region, the authors reported that, among the ten preference distributions were compared for the goodness of fit tests. At the conclusion of the comparability, Weibull has been identified to become the perfect distribution that represents wind data. This final decision has been given at 95% confidence level and 5% significance level[4]. In contrast, four distributions for frequency analysis of wind speed reported in[5], the research analyzed wind speed frequencies utilizing available wind data observed at 10m from Urmia synoptic station in Iran. All the distribution models are fitted to the maximum annual wind from station, and parameters of the distributions are estimated using the method of maximum likelihood and the method of moments. Weibull, Lognormal and Gamma distribution were applied for modeling the wind speed in the east of Malaysia, they concluded their findings that the outcome of the study demonstrate that Weibull and gamma distribution appears to satisfy and fit the wind speed data used[6].

Furthermore, diverse researchers all over the world presented the beneficial uses of distinct distributions to wind speed data, and judged the GOF using Anderson Darling (AD), and Kolmolgorov-Smirnoff (KS) statistic; for instance, Weibull function in[7,8], Rayleigh model has been used in[9,10], wind speed modeling based on Lognormal[11]. An extensive review on probability distribution functions can be found in[12,13].

The objective of this paper is to find the most acceptable distribution (s) for modeling the wind speed data of Miri. Therefore, the specific aim of this paper is to compare the most widely used models in the literatures Weibull, Rayleigh, Gamma, Lognormal and Erlang by employing various GOF qualifying measures.

METHODOLOGY

Weibull Model

The Weibull distribution is a special case of Pierson class III distribution. In Weibull distribution, the variances in wind velocity are characterized by the two functions; (1) the probability density function and (2) the cumulative distribution function. The probability density function p (v) indicates the fraction of time (or probability) from which the wind is at a given velocity v. The Weibull probability distribution model is given by:

$$p(v) = \frac{k}{c}\left(\frac{v}{c}\right)^{k-1} \exp\left\{-\left(\frac{v}{c}\right)\right\}^{k} \quad v>0, \ k,c >0$$

…...(1)

c is the scale parameter in m/s and k stand for shape parameter which has no unit. Thus the cumulative probability function

$$P(v) = \int_0^\infty p(v)dv = 1 - \exp\left\{-\left(\frac{v}{c}\right)\right\}^{k}$$

………...(2)

Rayleigh Model

The Rayleigh pdf is a particular case of the Weibull probability distribution function when k =2.

The cumulative distribution function is given by the following equation:

$$p(v) = \frac{2v}{c^2} \exp - [v/c]^2$$

………………..(3)

Gamma Model

The PDF and CDF of the two parameter gamma distribution function with shape and scale parameter is given as:

$$f(x : \alpha, \beta) = \frac{x^{\alpha-1}}{\beta\Gamma(\alpha)}\exp\left[-\frac{x}{\beta}\right]$$

x>0, a,b>0

………………….. (4)

$$F(x) = \frac{\Gamma(x/\beta)^\alpha}{\Gamma(\alpha)}$$

……....................(5)

Lognormal Model

The lognormal distribution and cumulative function are defined by the following equations:

$$f(x : \mu, \delta) = \frac{1}{x\delta\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{1n-\mu}{\delta}\right)^2\right]$$

x>0, s>0,8<μ8 …………(6)

$$F(x) = \phi\left(\frac{1nx - \mu}{\delta}\right)$$

………................(7)

Where Ø is the Laplace integral.

Erlang Model

The Erlang distribution (or γ-Erlang distribution) is a probability distribution formulated by A. K. Erlang. It is actually a unique case of the Gamma distribution. A Gamma (α, β) distribution is equal to an Erlang (γ, β) distribution with α = γ, when a is an integer, contrary to the Gamma distribution, the Erlang does offer a cumulative distribution function.

$$F(x) = 1 - \exp\left(\frac{-x}{\beta}\right)\sum_{i=0}^{m-1}\frac{x^i}{\beta^i i!}$$

………...(8)

Goodness of Fit (GOF)

In order to verify the suitability of each of the distributions, the goodness of fit of estimate from the above distributions has been assessed based on; Kolmogorov-Smirnov (KS), Anderson Darling (AD) and Chi-Squared (CS), presented by the following relationships, accordingly

Kolmogorov-Smirnov (KS)

The data are made up of a randomly sample X_1, X_2,..., X_n of dimension and related to some unknown distribution function, denoted by *F(x)*, and the sample is a random sample

$$D = Max(D^+, D^-)$$

………………...(9)

$$D^+ = Max(\frac{1}{n} - F(x_i))$$

………………….(10)

$$D^- = Max(F(x_i) - \frac{i-1}{n})$$

………………….(11)

Anderson-Darling

AD test is generally used to compare the fit of an observed CDF to an expected CDF. This test gives higher weights to the tails than KS test and the test statistic is given by,

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{N}(2i-1)[\ln F(x_i) + \ln(1 - F(x_{n-1+i}))]$$
$$\dots\dots(12)$$

Chi-Square Test

$$\chi^2 = \sum_{i=1}^{k}\frac{(O_i - E_i)^2}{E_i}$$
$$\dots\dots\dots\dots\dots(13)$$

Where O_i denotes the observed frequency and E_i denotes the expected frequency and is calculated by:

$$E_i = F(x_2) - F(x_1)$$
$$\dots\dots\dots\dots\dots(14)$$

Where F denotes the CDF of the distribution under consideration and x_1, x_2 are the lower and upper limits for bin i.

RESULTS AND DISCUSSION

The summary of the descriptive statistics for the long term daily average wind speed data is shown in Table 1. It is revealed that the wind speed varies from 0.4-4.8m/s and the average annual wind speed was found to be 2.02m/s at 10m, this condition is likewise examined by coefficient of kurtosis and skewness, the values obtained indicates that all the distribution of wind speeds through an asymmetric tail increasing in the direction of positive values. In the same manner, this indicates a relatively peaked distribution toward positive. (See table 1.)

The histograms give a visual exhibit of wind speed data, which is displayed in Fig. 1. These also reveal that the distribution of wind speed is really skewed to non-negative, i.e. the distribution curve is asymmetric being extended out to the right.

An additional significant characteristic is the fact that none of the wind speed values are less than zero m/s; as a result the lower bound of distributions is zero. These final results advocate that positively skewed distributions could possibly be employed to model the observed wind speed data.

Weibull, Rayleigh, Gamma, Lognormal and Erlang models have been applied for modelling the wind speed values, the parameters of each model were estimated using maximum likelihood method and the results are presented in table 2. To compare and contrast the probabilistic models, a number of goodness of fit: Kolmogrove-Smirnov test[14], Simirnov[15] Anderson Darling (AD)[16] and Pearson's Chi-Squared (CS) test[17] were applied to assess the most effective model.

Table 2 also indicates the p-values of these GOF tests. The outcomes of KS, AD and CS test evidently showed that gamma and Lognormal are well fitted, and then Weibull and Erlag distributions. Furthermore, the statistical values of Rayleigh distribution are not within acceptable range as reported in[18], showing that the wind speed data does not fit the model.

CONCLUSION

Five different statistical distribution models have been used to model the long term wind speed data of Miri which was measured at 10m height starting from 2003 till 2012. The effectiveness of each model was judged based on AD, KS and CS at the 5 % level of significance. Gamma and Lognormal models are found to fit the observed data more consistently compared with other models analysed.

ACKNOWLEDGEMENTS

REFERENCES

1. A. A. Rahman, N. A. Bakar, F. Hanaffi, A. Khamis, and U. Teknikal, "Study of Renewable Energy Potential in Malaysia," pp. 170–176, 2011.

2.   E. C. Morgan, M. Lackner, R. M. Vogel, and L. G. Baise, "Probability distributions for offshore wind speeds," *Energy Convers. Manag.*, vol. 52, no. 1, pp. 15–26, Jan. 2011.

3.   I. Y. . Lun and J. C. Lam, "A study of Weibull parameters using long-term wind observations," *Renew. Energy*, vol. 20, no. 2, pp. 145–153, Jun. 2000.

4.   V. Yilmaz and H. E. Ç. E. L. İ. K, "A Statistical Approach To Estimate The Wind Speed Distribution : The Case of Gelibolu Region 2 . *Wind Speed Data*," vol. 9, no. 1, pp. 122–132, 2008.

5.   R. Daneshfaraz, S. Nemati, H. Asadi, and M. Menazadeh, "Comparison of Four Distributions for Frequency Analysis of Wind Speed : A Case Study," vol. 3, no. 1, pp. 6–11, 2013.

6.   A. Zaharim, "Wind Speed Analysis in the East Coast of Malaysia," vol. 32, no. 2, pp. 208–215, 2009.

7.   A. N. Syafawati, I. Daut, S. S. Shema, M. Irwanto, Z. Farhana, N. Razliana, and C. Shatri, "Potential of Wind and Solar Energy using Weibull and Hargreaves Method Analysis," no. June, pp. 6–7, 2011.

8.   D. A. Fadare and O. State, "A Statistical Analysis of Wind Energy Potential in Ibadan, Nigeria, Based on Weibull Distribution Function.," vol. 9, no. 1, pp. 110–119, 2008.

9.   S. A. Ahmed and H. O. Mahammed, "A Statistical Analysis of Wind Power Density Based on the Weibull and Ralyeigh models of ' Penjwen Region ' Sulaimani / Iraq," vol. 6, no. 2, pp. 135–140, 2012.

10. Y. M. Atwa and E. F. El-Saadany, "Probabilistic approach for optimal allocation of wind-based distributed generation in distribution systems," *IET Renew*. Power Gener., vol. 5, no. 1, p. 79, 2011.

11. J. Lee, G. Park, E. Kim, Y. Kim, and I. Lee, "Wind Speed Modeling based on Artificial Neural Networks for Jeju Area," vol. 5, no. 2, pp. 81–88, 2012.

12. J. a. Carta, P. Ramírez, and S. Velázquez, "A review of wind speed probability distributions used in wind energy analysis," Renew. *Sustain. Energy Rev.*, vol. 13, no. 5, pp. 933–955, Jun. 2009.

13. S. M. Lawan, W. A. W. Z. Abidin, W. Y. Chai, A. Baharun, and T. Masri, "(WRA) Techniques , Wind Energy Potential and Utilisation in Malaysia and Other Countries," ARPN,vol. 8, no. 12, pp. 1039–1053, 2013.

14. O. Renaud and M.-P. Victoria-Feser, "A robust coefficient of determination for regression," J. Stat. Plan. Inference, vol. 140, no. 7, pp. 1852–1862, Jul. 2010.

15. A. Tuzuner, "Wind speed modeling and energy production simulation with Weibull sampling," 2008 *IEEE Power Energy Soc. Gen. Meet. - Convers. Deliv. Electr. Energy* 21st Century, pp. 1–6, Jul. 2008.

16. V. N. Dike, M. Sc, C. T. Chineke, D. Ph, O. K. Nwofor, and U. K. Okoro, "Estimating Wind Energy Potential in Some Coastal Cities of the Niger-Delta Region of Nigeria .," vol. 12, no. 1, pp. 598–604, 2011.

17. Z. O. Olaofe and K. A. Folly, "Statistical Analysis of the Wind Resources at Darling for Energy Production," vol. 2, no. 2, 2012.

18. H. Bludszuweit, J. A. Domínguez-navarro, and A. Llombart, "*Statistical Analysis of Wind Power Forecast Error*," vol. 23, no. 3, pp. 983–991, 2008.

Table 1. Descriptive statistics of ten years wind speed

Statistic value		Percentile value	
Sample size	3653	Min	0.4
Range	4.4	5%	1.2
Mean	2.0287	10%	1.4
Variance	0.31814	25% (Q1)	1.7
Std. Deviation	0.56403	50% (Median)	2
Coef. of Variation	0.27803	75% (Q3)	2.3
Std. Error	0.00933	90%	2.7
Skewness	0.88988	95%	3.1
Excess Kurtosis	2.3076	Max	4.8

Table 2. Fitting and goodness of fit summary

Distribution	Parameters	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
Erlang	α=12 β=0.15682	0.17814	4	153.14	4	584.75	4
Gamma	α =12.936 β =0.15682	0.07424	1	20.81	1	372.78	1
Lognormal	α =0.28344 β =0.66872	0.08813	2	24.954	2	373.58	2
Rayleigh (2P)	α =1.2197 β =0.39857	0.21796	5	248.74	5	1591.1	5
Weibull	α =4.4375 β =2.2219	0.10512	3	48.946	3	458.0	3
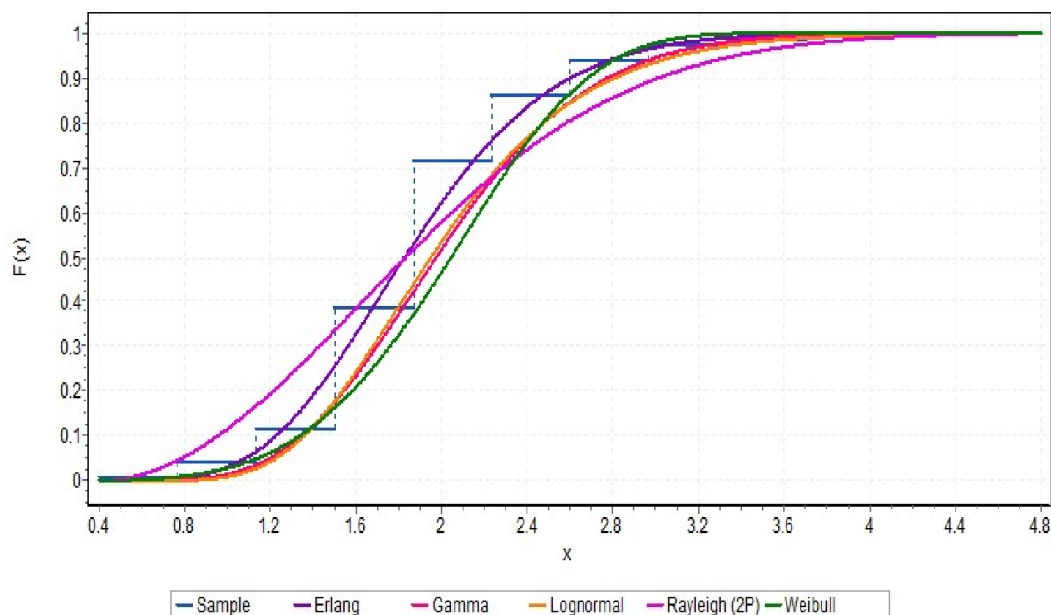
Figure 1. Cumulative distribution of wind speed