Journal of Statistics: Advances in Theory and Applications
Volume 13, Number 1, 2015, Pages 77-88

A STUDY OF WIND SPEED PROBABILITY MODELS: A CASE OF SIR SERETSE KHAMA INTERNATIONAL AIRPORT MONTHLY MAXIMUMS

KESAOBAKA MOLEBATSI and K. S. MADHAVA RAO

Department of Statistics
University of Botswana
Botswana
e-mail: raom@mopipi.ub.bw

Abstract

Wind speed affects aviation and wind energy production, among others, it is a random variable and hence statistical procedures are relevant when analysing it. One needs to approximate wind speed probability distribution empirically so as to make any evidence based conclusions regarding the characteristics of wind speed. Monthly maximum observations for the years 2008 to 2010 from the Sir Seretse Khama International Airport-Gaborone, Botswana are used. Both Weibull and lognormal distributions are fitted to the data under maximum likelihood estimation, while Kolmogorov-Smirnov and Anderson-Darling tests are used to test the goodness of fit of these distributions. The distributions provided good fit on the monthly wind speed maximums. The results proved conclusively that monthly maximums of wind speed follow Weibull and lognormal distributions.

2010 Mathematics Subject Classification: 62-07, 62P12.

Keywords and phrases: wind speed, monthly maximums, maximum likelihood estimation, goodness of fit.

Received February 13, 2015

1. Introduction

Wind speed affects several activities like aviation; it is also the most important parameter when examining a place's potential in generating wind power and a good application of wind speed studies (Zaharim et al. [14]) among others. Andrews [2] states that wind is among the most important influencing factors of wild land fire as fire behaviour is strongly affected by wind speed and direction. Wind speed is random; hence empirical statistical methods are useful in estimating it. In literature, the Weibull distribution is commonly used in the practical studies related to the wind energy modelling (Auwera et al. [3]; Rehman et al. [6]; Lun and Lan [5]; Seguro and Lambert [7]; Ulgen and Hepbasli [11]; Weisser [12]; Celik [4]; Stevens et al. [8]; Toure [10]; Zhou et al. [15]). However, Zaharim et al. [14] applied Burr, lognormal and Frechet to data sets for a specific location in Pahang, Malaysia and using different statistical tests determined which distribution provided a better fit. Al Buhairi and Mahyoub [1] found that in recent years, many efforts have been made to construct an adequate model for the wind speed frequency distribution. Statistically, any probability distribution may be examined rather than just assuming that it follows some distribution. This therefore implies that we cannot always accept the wind speed distributions as either Weibull probability density function or any other distribution; moreover, Yilmaz et al. [13] argued that in most studies fitting of data set to Weibull distribution was not examined even though this assumption was made. Different probability distributions should be investigated and incorporated to the analyses. This therefore becomes an issue to be addressed at least in the case of Botswana. This paper is aimed at determining the statistical properties of Botswana's wind speed distribution by using the data from the Sir Seretse Khama International Airport.

Sir Seretse Khama International Airport is located about ten kilometres north of Gaborone, the capital city of Botswana in Southern Africa with 24 32.21S, 025 55.08E coordinates. The airport acts as an aviation link to a number of Southern African cities like Johannesburg, Harare and other local cities like Francistown and Maun. To ensure safe aviation, wind speed among other weather indicators is monitored hourly and the data is readily available from the Department of Meteorological Services, Botswana (http://www.mewt.gov.bw/DMS) for processing.

The objectives of the study are to come up with probability density and distribution functions of wind speed at the Sir Seretse Khama International Airport (SSKIA), Gaborone-Botswana and to estimate the parameters of this distribution. The rest of the paper is organized as follows. Section 2 deals with the methodology dealing with the description of the two distributions to be fitted and the associated theory of goodness of fit. Section 3 summarizes the conclusions and provides a brief discussion of the results. Some concluding remarks are offered in Section 4.

2. Methodology

This study uses monthly maximums of wind speed (from January 2008 to December 2010) at the study area. Two probability distributions, namely, Weibull and lognormal distributions (both with two parameters) were tested to see if they can describe the data. Kolmogorov-Smirnov (KS) and the Anderson-Darling (AD) goodness of fit tests were used to examine the adequacy of the fitted models.

2.1. Weibull distribution

The Weibull probability density and distribution functions are, respectively, defined as follows:

$$f(x;\, c,\, k) = \left(\frac{k}{c}\right)\left(\frac{x}{c}\right)^{k-1} \exp\left[-\left(\frac{x}{c}\right)^{k}\right]; \quad x,\, c,\, k > 0, \tag{1}$$

$$F(x;\, c,\, k) = 1 - \exp(-\,c^{-k}x^{k}); \quad x,\, c,\, k > 0, \tag{2}$$

where 'c' and 'k' are the scale and shape parameters, respectively.

2.2. Lognormal distribution

The lognormal probability density and distribution functions are, respectively, defined as follows:

$$f(x;\ \mu,\ \delta) = \frac{1}{x\sqrt{2\pi\delta^2}} \exp\left\{-\frac{1}{2\delta^2}(\ln x - \mu)^2\right\}; \quad x,\ \delta > 0,\ \mu \in R, \tag{3}$$

$$F(x;\ \mu,\ \delta) = \phi\left(\frac{\ln x - \mu}{\delta}\right); \quad x,\ \delta > 0,\ \mu \in R, \tag{4}$$

where 'μ' is the mean and 'δ' is the standard deviation of the distribution for the normal random variable ln x and ϕ is the standard normal cumulative distribution function.

2.3. Maximum likelihood estimation of parameters

If $x_1, x_2, \ldots, x_n$ are independent and identically distributed random samples from a population with probability density function or probability mass function $f(x|\theta)$, the likelihood function of the sample vector $\underline{x} = (x_1, \ldots, x_n)$ is defined by

$$L(\underline{x}|\theta) = f(x_1|\theta) f(x_2|\theta) \ldots f(x_n|\theta). \tag{5}$$

The maximum likelihood principle enables us to take as our estimator of θ that value (say, $\hat{\theta}$) within the admissible range of θ which makes the likelihood function as large as possible. That is, we choose $\hat{\theta}$ so that for any admissible value θ;

$$L(\underline{x}|\hat{\theta}) \geq L(\underline{x}|\theta). \tag{6}$$

Unless otherwise specified, we assume that θ may take any real value in an interval and ignore factors in L not involving θ (Stuart et al. [9]). If the likelihood function is twice differentiable function of θ throughout its range, stationary values of the likelihood function within the admissible range of θ will, if they exist, be given by the roots of

$$\frac{\partial}{\partial\theta} L(\theta|\underline{x}) = 0. \tag{7}$$

The sufficient (though not necessary) for any of these stationary values $(\text{say } \hat{\theta})$ to be a local maximum is that

$$\frac{\partial^2}{\partial\theta^2} L(\theta|\underline{x}) < 0. \tag{8}$$

According to the researchers cited above, it is easier to work with the logarithm of the likelihood function than with the function itself.

2.4. Weibull distribution

It can be shown that the maximum likelihood estimators of the shape factor '*k*' and scale factor '*c*' are obtained by using the following two equations:

$$\hat{k} = \left(\frac{\sum_{i=1}^{k} x_i^k \ln x_i}{\sum_{i=1}^{n} x_i^k} - \frac{\sum_{i=1}^{n} x_i^k \ln x_i}{n} \right)^{-1}, \tag{9}$$

$$\hat{c} = \left(\frac{\sum_{i=1}^{n} x_i^k}{n} \right)^{1/k}, \tag{10}$$

where x_i is the wind speed in time step i and n is the number of non-zero wind speed data points. Equations (9) and (10) must be solved by any method of the iterative procedures. Care must be taken to ensure that Equations (9) and (10) must be applied only to non-zero wind speed data points. It can be shown that the second order derivatives are negative hence satisfying the necessary and sufficient condition to maximize the likelihood function. The Newton-Raphson iterative procedure was used to solve Equations (9) and (10).

2.5. Lognormal distribution

Following the maximum likelihood procedure, it can be shown that the estimates of 'μ' and 'δ' are given by

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \ln x_i}{n}, \tag{11}$$

$$\hat{\delta} = \sqrt{\frac{\sum_{i=1}^{n} (\ln x_i - \hat{\mu})^2}{n}}. \tag{12}$$

Further, it can also be shown that the second order derivative is negative hence satisfying the necessary and sufficient condition of maximizing the likelihood function.

2.6. Goodness of fit tests

Kolmogorov-Smirnov (KS) test

The Kolmogorov-Smirnov test statistic is based on the maximum difference between the sample cumulative distribution function (CDF) and the hypothesized CDF. To test a completely specified $H_0 : F = F_0$, let $F_n(x)$ be the theoretical cumulative distribution function and $S_n(x)$ be the empirical distribution function, where

$$S_n(x_i) = \frac{i}{n}, \quad i = 0, 1, \ldots, n. \tag{13}$$

The Kolmogorov-Smirnov test statistic is given by

$$D_n = \sup_x |F_0(x) - S_n(x)|. \tag{14}$$

We reject the null hypothesis at a level of significance α, when the p-value,

$$P(D_n \geq D_{\text{calculated}}) < \alpha.$$

Anderson-Darling (AD) test

This test gives higher weights to the tails than the KS test and the test statistic is given by

$$AD = \left[\sum_{x} \frac{1 - 2i}{n} \left\{\ln\left(1 - \exp\left(- Z_{(i)}\right)\right) - Z_{(n+1-i)}\right\} - n\right], \tag{15}$$

where $$Z_{(i)} = \left[\frac{x_{(i)}}{\hat{c}}\right]^{\hat{k}}. \tag{16}$$

The observed significance level (OSL) probability is now used for testing the adequacy of Weibull distribution. If $\mathrm{OSL} < 0.05,$ then the Weibull assumption is rejected at 5% level of significance. The OSL formula is given by

$$OSL = \frac{1}{\left\{1 + \exp\left[- 0.1 + 1.24 \ln\left(AD^*\right) + 4.48\left(AD^*\right)\right]\right\}}, \tag{17}$$

where

$$AD^* = \left(1 + \frac{0.2}{\sqrt{n}}\right) AD. \tag{18}$$

2.7. Data description

The study uses monthly maximum wind speed for the years 2008 to 2010 at study site Sir Seretse Khama International Airport, Gaborone, Botswana. The data were collected from the Department of Meteorological Services, located in Gaborone and are shown in Table 1.

Table 1. Monthly maximum wind speed

Year	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
2008	20	26	17	28	23	25	19	25	27	26	26	14
2009	21	30	22	21	19	17	16	30	25	26	23	30
2010	23	26	19	22	20	19	19	20	29	26	29	22

3. Results and Discussions

The descriptive statistics for the data are given in Table 2.

Table 2. Descriptive statistics of monthly wind speed maximums and its log for years 2008 to 2010

Variable	*n*	Min	Max	Mean	SD	Skewness
Wind	36	14	30	23.06	4.30	1.93
Log	36	2.64	3.4	3.12	0.19	1.17

The table above shows the descriptive statistics of wind speed and its corresponding natural log to be used when fitting the lognormal distribution. For the period of study, five missing observations were recorded and were estimated by the averages of their corresponding months. Minimums of 14 and 2.64 are reported while maximums are 30 and 3.40 knots for wind speed and its logarithm, respectively. Graphical displays are shown in Figure 1, which show observations and their corresponding probability distributions. Both the histograms indicate that wind speed/ log wind speed distributions are asymmetric in shape.

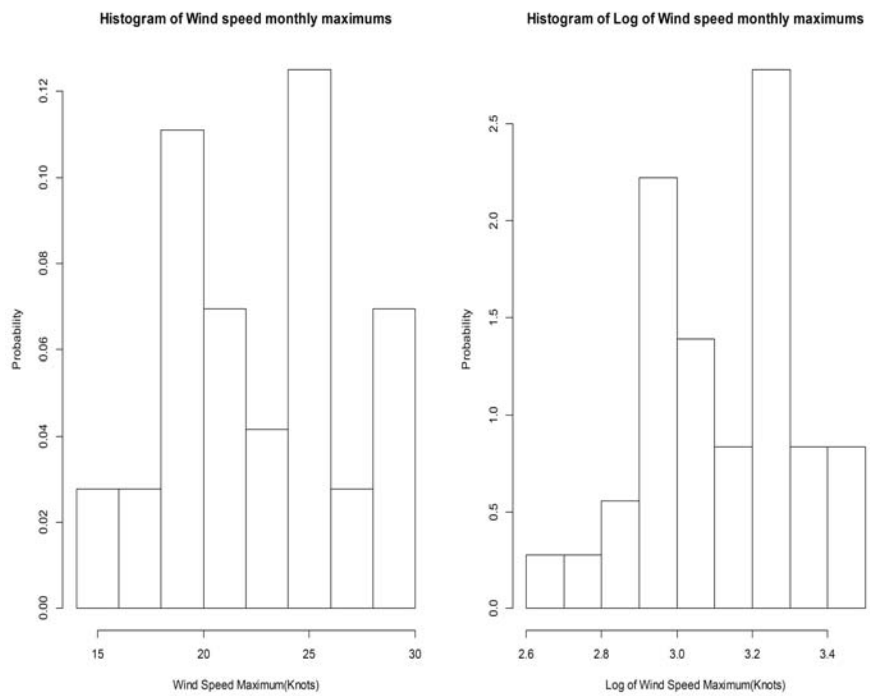

Figure 1. Histogram of recorded wind speed maximums and its log for years 2008 to 2010.

The maximum likelihood estimates are used for their generally good properties and are presented in Table 3. Next, we construct the Q-Q plots (Figure 2) to check if the hypothesized distributions provide reasonable fits for the data and then perform the discussed goodness of fit procedures on the data and the models.

Table 3. Maximum likelihood parameter estimates and their standard errors for the two distributions

Distribution	Parameter	Estimate (SE)
Weibull	Scale	24.82 (0.70)
	Shape	6.22 (0.82)
Lognormal	SD	0.19 (0.02)
	Mean	3.12 (0.03)

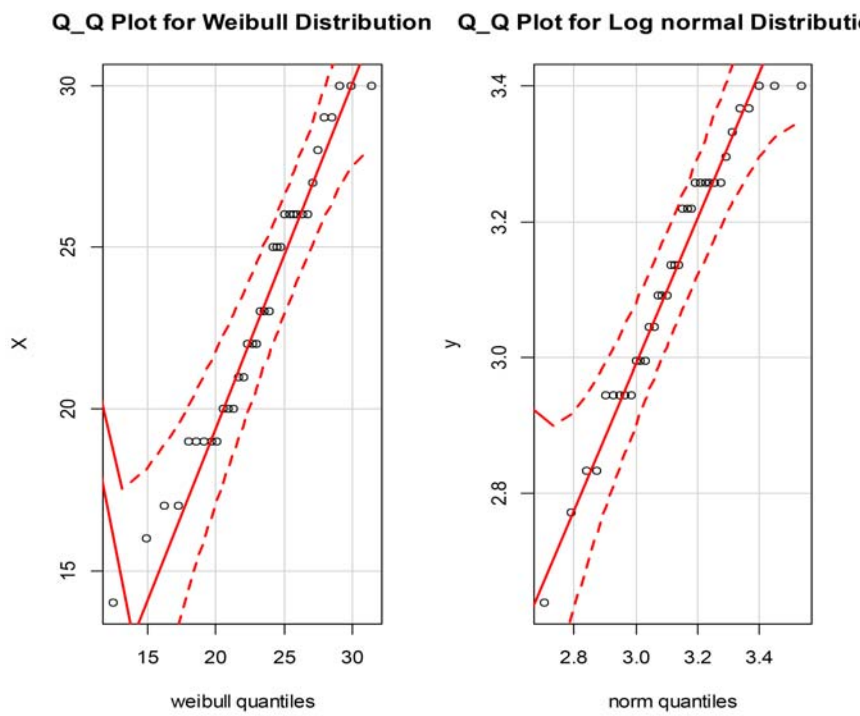

Figure 2. Q-Q Plots for the two probability distributions.

We observe that generally, the theorised quartiles and empirical ones do not seem to be equal for the two distributions; however, they all lie within the 95% confidence envelope. The Weibull seem to be the best among the two. It is apparent that these two distributions can be fitted for wind speed data. The goodness of fit tests, namely, Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) were performed and their corresponding p values and conclusions are presented in Table 4. We observe that Weibull and lognormal distributions can be used to estimate monthly maximum wind speed observations as the two tests accepted the null hypothesis that data come from the two hypothesized distributions.

Table 4. Goodness of fit test

Distribution	***p*-values**		
	KS	**AD**	**Conclusion**
Weibull	0.8347	0.7414	Fit is good
Lognormal	0.4528	0.7211	Fit is good

The theoretical Weibull and the lognormal fitted distributions are given by

$$f(x) = 3.99\left(\frac{x}{6.22}\right)^{23.82} \exp\left[-\left(\frac{x}{6.22}\right)^{24.82}\right]; \quad x > 0, \tag{19}$$

$$f(x) = \frac{1}{0.2687\sqrt{\pi}x} \exp\left\{-\frac{1}{(0.0722)}(\ln x - 3.12)^2\right\}; \quad x > 0. \tag{20}$$

4. Conclusion

The need for studying wind speed distribution is manifold. Wind speed affects several activities like aviation; evaluating wind power generation potential and so on. Further, it has been well established that wind is among the most important influencing factors of wild land fire as fire behaviour is strongly affected by wind speed and direction. Therefore, it becomes all the more important to make any evidence based conclusions regarding the characteristics of wind speed. In this paper, we investigate the validity of two probability distribution models, namely, the Weibull and lognormal for monthly maximum wind speed observations recorded at the Sir Seretse Khama International Airport-Gaborone, Botswana for the years 2008 to 2010. The distributions provided good fit on the monthly wind speed maximums. It would be interesting to see if the same conclusions are obtained for wind speed maximums recorded at other two important international airports of the country, Kasane and Maun. Further, subject to availability, it would be worthwhile to expand the scope of the study to include more recent data on wind speed monthly maximums.

References

[1] Al Buhairi and H. Mahyoub, A statistical analysis of wind speed data and an assessment of wind speed energy potential in taiz-yemen, Assiut University Bulletin of Environmental Research 9 (2006), 21-33.

[2] P. L. Andrews, Modelling wind adjustment factor and mid flame wind speed for Rorthermel's surface fire spread model, Gen. Tech. Rep. RMRS-GTR-266. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station (2012), 39-45.

[3] V. Auwera, L. F. Meyer and L. M. Malet, The use of the Weibull three parameter model for estimating mean wind power densities, Journal of Applied Meteorology 19 (1980), 819-825.

[4] A. N. Celik, A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey, Renewable Energy 29 (2003), 593-604.

[5] I. Y. F. Lun and J. C. Lan, A study of Weibull parameters using long-term wind observations, Renewable Energy 20 (2000), 145-153.

[6] S. Rehmann, T. O. Halawani and T. Husain, Weibull, parameters for wind speed distribution in Saudi Arabia, Solar Energy 53 (1994), 473-479.

[7] J. V. Seguro and T. W. Lambert, Modern estimations of the parameters of the Weibull wind speed distribution for wind energy analysis, Journal of Wind Engineering and Industrial Aerodynamics 85 (2000), 75-84.

[8] M. Stevens and P. T. Smulders, The estimations of the parameters of the Weibull wind speed distribution for wind energy utilization purposes, Wind Energy (1979), 132-145.

[9] Stuart, A. Ord and S. Arnold, Kendall's Advanced Theory of Statistics, Volume 1 & 2A: Classical Inference and the Linear Model (6th ed.), London: Holder Arnold, 1999.

[10] S. Toure, Investigation on the eigen-coordinates method for the 2-parameter Weibull distribution of wind speed, Renewable Energy 30 (2005), 511-521.

[11] K. Ulgen and A. Hepbasli, Determination of Weibull parameters for wind energy analysis of Izmir, Turkey, International Journal of Energy Research 6 (2002), 495-506.

[12] D. A. Weisser, Wind energy analysis of Grenada: An estimation using the 'Weibull' density function, Renewable Energy 28 (2003), 1803-1812.

[13] V. Yilmaz, H. Aras and H. E. Celik, Statistical analysis of wind speed data, Eskisehir Osmangazi University, Xvii (2005).

[14] A. Zaharim, S. K. Najid, A. M. Razali and K. Sopian, Analysing Malaysian wind speed data using statistical distribution, Universiti Kebngsaan Malaysia (2009), 363-370. Proceedings of the 4th IASME/WSES International Conference on Energy & Environment (EE'09).

[15] W. Zhou, H. Yang and Z. Fang, Wind power potential and characteristic analysis of the Pearl River Delta region, China, Renewable Energy 31 (2006), 739-753.

■