ISSN 1022-8594
JUJS

Vol. 39, No. 2, pp.11-22

A Study on Monthly Maximum Wind Speed Probability Distributions at Hazrat Shahajalal and MAG Osmani International Airport of Bangladesh

Md. Moyazzem Hossain*

Department of Statistics, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh

Faruq Abdulla

Department of Statistics, Islamic University, Kushtia-7003, Bangladesh

Ajit Kumar Majumder

Department of Statistics, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh

Abstract

Wind speed is a fundamental atmospheric rate. Wind speed is caused by air moving from high pressure to low pressure, usually due to changes in temperature. Probability density functions (PDFs) have been used in literature to describe wind speed characteristics which include Weibull, Rayleigh, bimodal Weibull, Lognormal, Gamma and so on. This paper considers a data set on maximum sustained wind speed (Km/h) at Hazrat Shahjalal International Airport and Osmani International Airport, Bangladesh over the period January, 1982 to June, 2015. This paper attempts to determine the best fit wind speed distribution with statistical properties of the monthly maximum sustained wind speed (Km/h) of the Hazrat Shahjalal International Airport and Osmani International Airport. The higher value of R^2 and the lower values of K-S error, RMSE and Chi-square error indicate that GEV distribution is more accurate than other PDFs in modeling wind speeds of both locations.

Keywords: Wind speed, Probability density functions, likelihood method, Model selection, Bangladesh

Introduction

Wind speed is caused by air moving from high pressure to low pressure, usually due to changes in temperature (https://en.wikipedia.org/wiki/Wind_speed). Wind speed affects several activities like aviation; it is also the most important parameter when examining a place's potential in generating wind power (Zaharim, *et al.*, 2009 [1]). Also, Wind speed affects weather forecasting, aircraft and maritime operations, construction projects, growth and metabolism rate of many plant species, and countless other implications (Michael Hogan, 2010 [2]).

Hazrat Shahjalal International Airport is the largest airport in Bangladesh. Operated and maintained by the Civil Aviation Authority, Bangladesh, it is

*Author for correspondence: e-mail: mmhmm.justat@gmail.com

also used by the Bangladesh Air Force. The airport has an area of 1,981 acres (802 ha). The airport is located in Kurmitola and was originally 11 NM (20 km; 13 mi) north of the capital Dhaka (https://en.wikipedia.org/wiki/ Shahjalal_International_Airport). On the other hand, MAG Osmani International Airport is an international airport located 5 miles north-east of Sylhet in Bangladesh. The vast majority of passengers using the airport are expatriate Bangladeshis and their descendants from the Sylhet Division living in the United Kingdom

(https://en.wikipedia.org/wiki/Osmani_International_Airport).

Probability density functions (PDFs) have been used in literature to describe wind speed characteristics which include Weibull, Rayleigh, bimodal Weibull, Lognormal, Gamma and so on (Ravindra Kollu, *et al.*, 2012 [3]). Celik (2003) [4] made statistical analysis and summarized that Weibull model was better than Rayleigh model. Akdag and Bagiorgas, (2010) [5] discussed about the two component Weibull distribution and stated that Weibull - Weibull gave a goodfit. Tian Pau (2011) [6] used Rayleigh, Weibull and Gamma distribution and its generalized form. Gupta and Kundu (2010) [7] discussed about the generalized logistic distribution. Yilmaz, *et al.*, (2008) [8] 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. Thus, this paper attempt to determine the best fit wind speed distribution with statistical properties of the monthly maximum sustained wind speed (Km/h) of Hazrat Shahjalal International Airport and Osmani International Airport.

Methods and Materials

Data Source

This paper considers a data set on maximum sustained wind speed (Km/h) over the period January, 1982 to June, 2015 at Hazrat Shahjalal International Airport and Osmani International Airport, Bangladesh from the web address http://en.tutiempo.net/.

Probability Distributions

The primary tools to describe wind speed characteristics are probability density functions. Many PDFs have been proposed in recent past, but in present study Weibull, Lognormal, Gamma, and GEV are used to describe wind speed characteristics. Parameters defining each distribution function are calculated using maximum likelihood method.

Weibull (W) Distribution

Weibul distribution is named after Waloddi Weibull who described it in detail in 1951, although it was first identified by Fréchet (1927) [9] and first applied by Rosin and Rammler (1933) [10] to describe a particle size distribution. The Weibull function is commonly used for fitting measured wind speed probability distribution. The probability density function (PDF) and the cumulative distribution function (CDF) of the Weibull distribution with two parameters are given by (Weibull, 1951 [11]):

Probability density function (PDF):

$$f(v;k,c)=\frac{k}{c}\left(\frac{v}{c}\right)^{k-1}\exp\left[-\left(\frac{v}{c}\right)^{k}\right]\ ;\ 0<v<\infty, k>0, c>0$$

Cumulative distribution function (CDF):

$$F(v;k,c)=1-\exp\left[-\left(\frac{v}{c}\right)^{k}\right]\ ;\ 0<v<\infty, k>0, c>0$$

where, $'k'$ and $'c'$ are the shape and scale parameters respectively. The shape and scale parameters are calculated using the maximum likelihood method (Kececioglu, 2002 [12]) and an iterative technique such as Newton-Raphson technique and which are given by:

$$\hat{k}=\left[\frac{\sum_{i=1}^{n} v_i^k \ln(v_i)}{\sum_{i=1}^{n} v_i^k}-\frac{\sum_{i=1}^{n}\ln(v_i)}{n}\right]^{-1} \text{ and } \hat{c}=\left[\frac{\sum_{i=1}^{n} v_i^k}{n}\right]^{1/k}$$

where, v_i is the wind speed in time step i and n is the number of data points.

Lognormal (LN) Distribution

Lognormal distribution is a probability distribution of a random variable whose logarithm is normally distributed. The probability density function (PDF) and the cumulative distribution function (CDF) of the Lognormal distribution are given by (Johnson, *et al.*, 1994 [13]):

Probability density function (PDF):

$$f\left(v;\mu,\delta\right)=\frac{1}{v\sqrt{2\pi\delta^2}}\exp\left[\frac{-\left(\ln\left(v\right)-\mu\right)^2}{2\delta^2}\right]\;;\;0<v,<\infty,\mu\in R,\delta>0$$

Cumulative distribution function (CDF):

$$F\left(v;\mu,\delta\right)=\Phi\left(\frac{\ln\left(v\right)-\mu}{\delta}\right)\;;\;0<v,<\infty,\mu\in R,\delta>0$$

where, 'μ' and 'δ' are the mean and standard deviation of the normal random variable $\ln(v)$ respectively and 'Φ' is the standard cumulative distribution function. The mean and standard deviation are calculated using the maximum likelihood method and which do not need an iterative procedure are given by (Carta, *et al.*, 2009 [14]):

$$\hat{\mu}=\frac{\sum_{i=1}^{n}\ln\left(v_i\right)}{n}\quad\text{and}\quad\hat{\delta}=\sqrt{\frac{\sum_{i=1}^{n}\left[\ln\left(v_i\right)-\hat{\mu}\right]^2}{n}}$$

where, v_i is the wind speed in time step i and n is the number of data points.

Gamma (G) Distribution

Lancaster (1966) [15] quotes from Laplace (1836) [16] in which latter obtain a Gamma distribution. The probability density function (PDF) and the cumulative distribution function (CDF) of the gamma distribution are given by:

Probability density function (PDF):

$$f\left(v;a,b\right)=\frac{v^{a-1}}{b^a\Gamma\left(a\right)}\exp\left(-\frac{v}{b}\right);\;v>0,a>0,b>0$$

Cumulative distribution function (CDF):

$$F\left(v;a,b\right)=\int_0^v f\left(u;a,b\right)du=\frac{\gamma\left(a,\frac{v}{b}\right)}{\Gamma\left(a\right)}\;;\;v>0,a>0,b>0$$

where, $\gamma\left(a,\frac{v}{b}\right)$ is the lower incomplete gamma function. The parameters 'a' and 'b' are the shape and scale parameters respectively. The shape and scale parameters are calculated using the maximum likelihood method and which are given by:

$$\hat{a} \approx \frac{0.5}{\log \overline{x} - \overline{\log x}}\text{; } \log \overline{x} \geq \overline{\log x} \text{ and } \hat{b} = \frac{\overline{x}}{\hat{a}}$$

Generalized Extreme Value (GEV) distribution

GEV distribution is a flexible model that combines the Gumbel, Frechet and Weibull maximum extreme value distributions (Ying and Pandey, 2007 [17]). For, GEV

Probability density function (PDF):

$$f(v;\mu,\sigma,\xi) = \frac{1}{\sigma}\left[1+\xi\left(\frac{v-\mu}{\sigma}\right)\right]^{\left(-1/\xi\right)-1} \exp\left\{-\left[1+\xi\left(\frac{v-\mu}{\sigma}\right)\right]^{-1/\xi}\right\};\ \xi \neq 0$$

again, for $v > \left(\mu - \frac{\sigma}{\xi}\right)$ in the case $\xi > 0$, and for $v < \left(\mu - \frac{\sigma}{\xi}\right)$ in the case $\xi < 0$.

Cumulative distribution function (CDF):

$$F(v;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{v-\mu}{\sigma}\right)\right]^{-1/\xi}\right\};\ \xi \neq 0$$

For $1+\xi\left(\frac{v-\mu}{\sigma}\right) > 0$, where $\mu \in \mathbf{R}$ is the location parameter, $\sigma > 0$ the scale parameter and $\xi \in \mathbf{R}$ the shape parameter.

Generalized Gamma (3-P G) Distribution

The generalized gamma also known as three parameters gamma distribution is a continuous probability distribution with three parameters. For non-negative v, the probability density function of the generalized gamma is (Stacy, 1962 [18]):

Probability density function (PDF):

$$f(v;a,d,p)=\frac{\left(p/a^{d}\right)v^{d-1}e^{-\left(v/a\right)^{p}}}{\Gamma\left(d/p\right)};\ a>0, d>0, p>0$$

where, $\Gamma(.)$ denotes the gamma function.

Cumulative distribution function (CDF): $F(v;a,d,p)=\frac{\gamma\left(d/p,\left(v/a\right)^{p}\right)}{\Gamma\left(d/p\right)}$, where $\gamma(.)$ denotes the lower incomplete gamma function.

Three Parameters Lognormal (3-P LN) Distribution

The lognormal distribution derives its name from the relationship that exists between random variables V and $Y=\ln(V-a)$. If Y is distributed normally (b,c), then V is lognormal (a,b,c). Accordingly, the probability density function of V may be written as (Cohen and Whitten, 1980 [19]):

Probability density function (PDF):

$$f(v;a,b,c)=\frac{1}{(v-a)c\sqrt{2\pi}}\exp\left[\frac{-\left[\ln(v-a)-b\right]^{2}}{2c^{2}}\right];c^{2}>0, a<v<\infty\cdot$$

Cumulative distribution function (CDF):

$$F(v;a,b,c)=\int_{a}^{v}\frac{1}{(v-a)c\sqrt{2\pi}}\exp\left[\frac{-\left[\ln(v-a)-b\right]^{2}}{2c^{2}}\right]dv\ .$$

Goodness-of-fit Tests

Goodness-of-fit tests are used to check the accuracy of the predicted data using theoretical probability function. To evaluate the goodness-of-fit of the PDFs to the wind speed data, the KS, the R^2, the χ^2 and the RMSE were used.

Kolmogorov-Smirnov (KS) Error Test

The KS test computes the largest difference between the cumulative distribution function of the model and the empirical distribution function. The KS test statistic is defined as:

$$D = \max_{1 \le i \le n} \left| F_i - \hat{F}_i \right|$$

where, $\hat{F}_i$ is the predicted cumulative probability of the i^{th} observation obtained with the theoretical cdf and F_i is the empirical probability of the i^{th} observation are obtained with the Cunnane (1978) [20] formula:

$$F_i = \frac{i - 0.4}{n + 0.2}$$

where, $i = 1, \cdots, n$ is the rank for ascending ordered observations.

R^2 *Test*

The R^2 test is used widely for goodness-of-fit comparisons and hypothesis testing because it quantifies the correlation between the observed cumulative probabilities and the predicted cumulative probabilities of a wind speed distribution. A larger value of R^2 indicates a better fit of the model cumulative probabilities $\hat{F}$ to the observed cumulative probabilities F. The R^2 is defined as:

$$R^2 = \frac{\sum_{i=1}^{n} \left(\hat{F}_i - \bar{F} \right)^2}{\sum_{i=1}^{n} \left(\hat{F}_i - \bar{F} \right)^2 + \sum_{i=1}^{n} \left(F_i - \hat{F}_i \right)^2} \; ; \; \bar{F} = \frac{\sum_{i=1}^{n} \hat{F}_i}{n}$$

Chi-Square Error Test

Chi-Square test is used to assess whether the observed probability differs from the predicted probability. Chi-Square test statistic is defined as $\chi^2 = \sum_{i=1}^{n} \frac{\left(F_i - \hat{F}_i \right)^2}{\hat{F}_i}$.

Root Mean Squared Error (RMSE) Test

Root mean square error (RMSE) provides a term-by-term comparison of the actual deviation between observed probabilities and predicted probabilities. A lower value of RMSE indicates a better distribution function model. Root mean square error (RMSE) is defined as

$$RMSE = \left[\frac{\sum_{i=1}^{n} \left(F_i - \hat{F}_i \right)^2}{n} \right]^{1/2}.$$

Results and Discussion

The mean and standard deviation of observed maximum sustained wind speed for Station 419230 (Dhaka) are 25.96484 km/h and 22.25632 km/h, respectively with minimum 3.5 km/h and maximum 94.3 km/h. Whereas, for the station 418910 (Sylhet), the mean and standard deviation of observed maximum sustained wind speed are 27.40224 km/h and 21.55088 km/h, respectively. The minimum and maximum values of observed maximum sustained wind speed are 5.4 km/h and 103.5 km/h respectively for the station 418910 (Sylhet). That is the average maximum sustained wind speed in Sylhet is higher than Dhaka station.

The estimation of parameters of all the PDFs considered in this study were carried out using maximum likelihood method and computed parameter values of different PDFs used for all the two stations are presented in Table 1.

Table 1 Computed parameter values of different PDFs considered in this study.

PDF	Parameters	Estimates	
		HSIA, Dhaka	OIA, Sylhet
W	Shape (k)	1.296405	1.453077
	Scale (c)	28.281896	30.028883
LN	Mean (μ)	2.9475227	3.0663842
	Standard Deviation (δ)	0.7673038	0.6313268
G	Shape (a)	1.780474	2.374259
	Scale (b)	14.541727	11.322937
GEV	Shape (ξ)	0.5166852	0.4991436
	Scale (σ)	9.2498732	8.0903576

	Location (μ)	13.8419043	16.0611118
3-P G	Shape (d)	1.160726	1.426174
	Scale (a)	19.331254	15.092712
	Threshold (p)	3.452904	5.358748
3-P LN	Shape (c)	0.9073547	0.8327014
	Scale (a)	2.7629273	2.7341814
	Threshold (b)	2.3200703	4.8253193

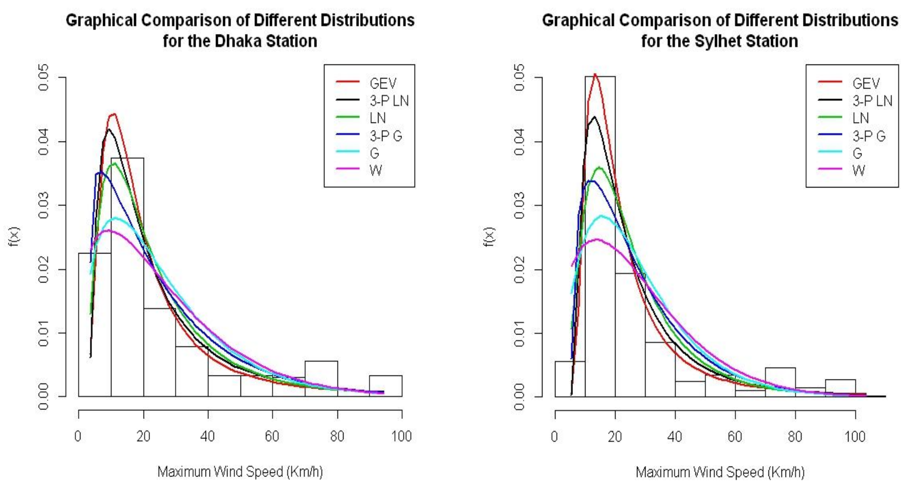
* HSIA = Hazrat Shahjalal International Airport and OIA = Osmani International Airport.

The statistical parameters for fitness evaluation of PDFs currently analyzed are presented in Table 2. Considering K-S error, χ^2 error and RMSE, the distribution functions Weibull, Lognormal, Gamma, three parameters Gamma and three parameters Lognormal have large errors indicating their inadequacy in modeling wind speeds of the both stations considered in this study. The higher value of R^2 and the lower values of K-S error, RMSE and Chi-square error indicate that GEV distribution is more accurate than other PDFs in modeling wind speeds of both locations.

Table 2 Values of Statistical tests for different distribution functions of OSIA, Dhaka and OIA, Sylhet.

PDF	**Values of Different Statistical Tests**							
	HSIA, Dhaka				**OIA, Sylhet**			
	$K-S$ Error	R^2	χ^2 Error	$RMSE$ Error	$K-S$ Error	R^2	χ^2 Error	$RMSE$ Error
W	0.13094	0.94173	5.27091	0.06800	0.14390	0.89413	9.29634	0.08882
LN	0.08927	0.98469	1.13112	0.03657	0.10855	0.95927	3.86296	0.05932
G	0.13425	0.94975	4.43101	0.06593	0.13922	0.91579	7.98422	0.08429
GEV	0.04868	0.99470	0.35711	0.02145	0.05765	0.99313	0.59696	0.02423
3-P G	0.10084	0.97292	2.17107	0.04720	0.11468	0.94517	4.82677	0.06735
3-P LN	0.06356	0.99239	0.50561	0.02540	0.07246	0.98265	1.53318	0.03814

The graphical comparisons of different distribution functions considered in this study and the histogram of the observed maximum sustained wind speed for HSIA, Dhaka and OIA, Sylhet are presented in Figure 1. As seen from Figure 1 and statistical parameters from Table 2, Generalized Extreme Value distribution (GEV) provided the best fit for the observed wind data for both locations.



(a) (b)

Figure 1: Graphical Comparison for different distribution functions of (a**)** HSIA, Dhaka and (b**) OIA,** Sylhet.

Conclusion

In recent years it has been investigated that the fitting of specific distribution to wind speed is required for use in practical application as air pollution modeling, estimation of wind loads on building and wind power analysis. So a model is required for wind speed distribution. Extensive literature search indicates that various parametric distribution models have been presented to estimate wind speed distributions. Wind speed data at HSIA, Dhaka and OIA, Sylhet were used in evaluating different pdfs to access their suitability. The Weibull, Lognormal, Gamma, three parameters Gamma and three parameters lognormal distribution functions have large errors indicating their inadequacy in modeling wind speeds of the both stations considered in this study. On the other hand, the higher value of R^2 and the lower values of K-S error, RMSE and chi square error indicate that GEV distribution is more accurate than other PDFs in modeling wind speeds of both Dhaka an Sylhet locations.

Reference

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

[2] Michael Hogan, C., *Abiotic factor, Encyclopedia of Earth*, eds Emily Monosson and C. Cleveland, National Council for Science and the Environment, Washington DC, 2010.

[3] Ravindra Kollu, Srinivasa Rao Rayapudi, SVL Narasimham and Krishna Mohan Pakkurthi, Mixture probability distribution functions to model wind speed distributions, *International Journal of Energy and Environmental Engineering*, 2012, **3**, 27. doi:10.1186/2251-6832-3-27.

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

[5] Akdag, S. A. and Bagiorgas , H. S., Use of two component Weibull mixtures in the analysis of wind speed in the Eastern Mediterranean, *Applied Energy*, 2010, **87**, 2566-2573.

[6] Tian Pau, C., Estimation of wind energy potential using different probability density functions, *Applied Energy*, 2011, **88(5)**, 1848-1856.

[7] Gupta, R. D. and Kundu, D., Generalized logistic distribution, *Journal of Applied Statistical Sciences*, 2010, **18**, 51-66.

[8] Yilmaz, V. and Celik, H. E., A Statistical Approach to Estimate The Wind Speed Distribution: The Case Of Gelibolu Region, *Doğuş Üniversitesi Dergisi*, 2008, *9(1)*, 122-132.

[9] Fréchet, Maurice, Sur la loi de probabilité de l'écart maximum, *Annales de la Société Polonaise de Mathematique*, Cracovie, 1927, **6**, 93–116.

[10] Rosin, P. and Rammler, E., The Laws Governing the Fineness of Powdered Coal, *Journal of the Institute of Fuel*, 1933, **7**, 29–36.

[11] Weibull, W., A statistical distribution function of wide applicability, *J. Appl. Mech.-Trans*, ASME, 1951, **18(3)**, 293–297.

[12] Kececioglu, D., *Reliability engineering handbook*, vol.1 and 2, Destech Pubblications, Pennsylvania, 2002.

[13] Johnson, Norman L., Kotz, Samuel, and Balakrishnan, N., *14: Lognormal Distributions*, *Continuous univariate distributions. Vol. 1*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics (2nd ed.), New York: John Wiley & Sons, 1994.

[14] Carta, J. A., Ramírez, P., and Velázquez, S., A review of wind speed probability distributions used in wind energy analysis case studies in the canary islands, *Renew Sustain Energy Rev*, 2009, **13(5)**, 933–955.

[15] Laplace, P. S., *Theorie Analytique des Probabilities*, 1836.

[16] Lancaster, H. O., Forerunners of the Pearson Chi-square, *Australian Journal of Statistics*, 1966, **8**, 117-126.

[17] Ying, A. and Pandey, M. D., The r largest order statistics model for extreme wind speed estimation, *J Wind Eng Ind Aerodyn*, 2007, **95(3)**, 165–182.

[18] Stacy, E. W., A Generalization of the Gamma Distribution, *Annals of Mathematical Statistics*, 1962, **33(3)**, 1187-1192.

[19] Cohen, A. C. and Whitten, B. J., Estimation in the Three-Parameter Lognormal Distribution, *Journal of the American Statistical Association*, 1980, **75**, 399- 404.

[20] Cunnane, C., Unbiased plotting positions – a review, *J Hydrol*, 1978, **37(3–4)**: 205–222.