Fitting of Statistical Distributions to Wind Speed Data

[1,2]AHMAD MAHIR RAZALI, [1]ROZAIMAH ZAINAL ABIDIN, [2]AZAMI ZAHARIM and [2]KAMARUZZAMAN SOPIAN

[1]School of Mathematical Sciences,
Faculty of Science and Technology
Universiti Kebangsaan Malaysia,
43600 Bangi, Selangor D.E., MALAYSIA

[2]Solar Energy Research Institute
Universiti Kebangsaan Malaysia, 43600 UKM Bangi
Selangor D. E., MALAYSIA

Abstract: -This paper investigates the probability distribution of wind speed data recorded in Faculty of Engineering, University Kebangsaan Malaysia. The wind speed data represented in the form of frequency curves show the shape of a potential model. The two-parameter Weibull distribution and lognormal distribution are adopted in this study to fit the wind speed data. The scale and shape parameters were estimated by using maximum likelihood method. The goodness-of-fit tests based on the empirical distribution function (EDF) are conducted to show that the distribution adequately fits the data. It is found from the hypothesis test that, although the two distributions are all suitable for describing the probability distribution of wind speed data, the two-parameter Weibull distribution is more appropriate than the lognormal distribution.

Keywords: -Weibull distribution, lognormal distribution, wind speed data, maximum likelihood method

1 Introduction

Wind energy has been used for centuries for navigation and agriculture. Recently, wind energy has been receiving a lot of attention because of the focus on renewable energies. The effective utilization of wind energy entails having a detailed knowledge of the wind characteristics at the particular location. The distribution of wind speeds is important for the design of wind farms, power generators and agricultural applications like irrigation. Accurate information about wind speed is important in determining best sites for wind turbines. Wind speeds must also be measured by those concerned about dispersion of airborne pollutants. A number of studies in recent years have investigated the fitting of specific distribution to wind speed for use in such practical application as air pollution modeling, estimation of wind loads on building and wind power analysis. Several distributions have received the most attention for example [2,3,5]. Several mathematical models have been used to study wind data.

In this article, we model the wind speeds at a station in the Faculty of Engineering, University Kebangsaan Malaysia. Two different functions have been used to model the wind speed, the two-parameter Weibull distribution and the lognormal distribution. Both models have been chosen because of their frequent use in the literature.

2 Wind Speed Measurement

The data for this study was obtained from the Engineering Faculty of Universiti Kebangsaan Malaysia. The data was captured using a cup anemometer in meter per second. The data was collected daily starting 1st June 2007 until 30th September 2007. However, there are some missing data for several days in July and August. The average wind speed measurement in July and September are 0.001471363 m/s and 0.001621368 m/s will be placed for missing data. Figure 1 gives the histogram of the wind speeds at this station. The distribution is skewed to the right and therefore can be modeled by using the Weibull distribution and lognormal distribution.

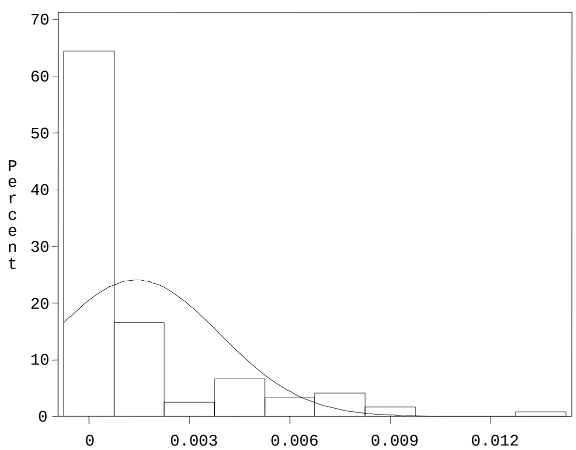


Figure 1 Histogram of wind speed data.

3 Methodology of Analysis

3.1 Weibull distribution

The Weibull distribution (named after the Swedish physicist W. Weibull, who applied it when studying material strength in tension and fatigue in the 1930s) provides a close approximation to the probability laws of many natural phenomena. It has been used to represent wind speed distributions for application in wind loads studies for some time. In recent years most attention has been focused on this method for wind energy applications not only due to its greater flexibility and simplicity but also because it can give a good fit to experimental data. The Weibull distribution function, which is a three-parameter function, but for wind speed, it can be expressed mathematically as

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

and the corresponding cumulative distribution is

$$F(x) = 1 - exp\left[-\left(\frac{x-\gamma}{\alpha}\right)^{\beta}\right] \tag{2}$$

where;
$x \geq 0$ is the wind speed (m/s in this study)
$\beta > 0$ is a shape parameter
$\alpha > 0$ is a scale parameter (m/s)
and γ is a location parameter.

In this article, we assume the two parameter Weibull distribution by setting the location parameter γ, equal to zero. Therefore Eqs. (1) and (2), the three-parameter model becomes the two-parameter model. Then, we have

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right] \tag{3}$$

and the corresponding cumulative distribution is

$$F(x) = 1 - exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right] \tag{4}$$

3.2 Lognormal distribution

The lognormal distribution may be one of the most versatile distributions. It has been seen to have applications in many fields, such as agriculture, entomology, economics, geology, industry and quality control. In terms of life testing and reliability, the lognormal distribution is known as a serious competitor to the Weibull distribution. The respective probability density

ISBN: 978-960-6766-71-8

ISSN: 1790-5095

function for the three parameter lognormal distribution is

$$f(x)=\frac{1}{(x-\gamma)\sigma\sqrt{2\pi}}\exp\left[\frac{-\left[\ln\left(\frac{x-\gamma}{\mu}\right)\right]^2}{2\sigma^2}\right] \tag{5}$$

where;

$x \geq 0$ is the wind speed (m/s in this study)

$\sigma > 0$ is a shape parameter

$\mu > 0$ is a scale parameter (m/s)

and γ is a location parameter.

In this article, we assume the two parameter lognormal distribution by setting the location parameter γ, equal to zero. Therefore Eq. (5), the three-parameter model becomes the two-parameter model. Then, we have

$$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left[\frac{-\left[\ln\left(\frac{x}{\mu}\right)\right]^2}{2\sigma^2}\right] \tag{6}$$

3.3 Parameter Estimation

For the present work, the two-parameter Weibull and lognormal parameters have been estimated by using a statistical software package (SAS) and by applying the maximum likelihood technique to wind speed data recorded in the engineering faculty station. The result appears in Table 1.

MLE for Weibull distribution

The maximum likelihood estimator for the shape and scale parameters are defined by the equation

$$\hat{\alpha}=\left[\left(\frac{1}{n}\right)\sum_{i=1}^{n} x_i^{\hat{\beta}}\right]^{\frac{1}{\hat{\beta}}} \tag{7}$$

$$\hat{\beta}=\frac{n}{\frac{1}{\hat{\alpha}}\sum_{i=1}^{n} x_i^{\hat{\beta}}\log x_i-\sum_{i=1}^{n}\log x_i} \tag{8}$$

MLE for Lognormal distribution

The maximum likelihood estimator for the shape and scale parameters are defined by the equation

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}\ln x_i \tag{9}$$

$$\hat{\sigma}^2=\frac{1}{n-1}\sum_{i=1}^{n}\left[\ln(x_i)-\hat{\mu}\right]^2 \tag{10}$$

Where x_i are the generated data sample and n the total number of sample in the data set. Eq. (8) being an implicit equation, has to be solved by an iterative process.

3.4 Goodness of Fit

In order to verify the goodness-of-fit of the distribution model to wind speed data observations, the EDF goodness of fit test should be conducted. The Kolmogorov-Smirnov *D* statistic, the Anderson-Darling statistic, and the Cramer-von Mises statistic are based on the empirical distribution function (EDF). We determine whether to reject the null hypothesis by examining the *p*-value that is associated with a goodness-of-fit statistic. When the *p*-value is less than the predetermined critical value, reject the null hypothesis and conclude that the data did not come from the specified distribution. Usually, the confidence level is taken to be 90% and thus *p*-value = 0.10. Besides that the probability plot can also be conducted. It is important to examine other statistics and plots to make a final assessment of normality.

ISBN: 978-960-6766-71-8

ISSN: 1790-5095

4 Results and Discussion

The result appears in Table 1 shows the estimated scale and shape parameter of two distribution models by using maximum likelihood method in which the location parameter value is 0. Figure 2 shows a comparison of probability densities of the two-parameter Weibull distribution and Lognormal distribution. The distributions of both the two-parameter Weibull distribution and lognormal distribution are apparently skewed to the right, which is consistent with the calculated result of skewness of 1.5751 for the wind speed data even though the distributions show a bit distinct skewness.

Figure 3 shows a comparison of the cumulative distribution of empirical distribution, the two-parameter Weibull distribution and lognormal distribution of the wind speed data. It can be seen from Figure 3 that the two-parameter Weibull distribution are closer to empirical distribution that was obtained on the basis of the measured wind speed data, while the lognormal distribution deviates from the empirical distributions.

Table 1 Values of the estimated parameters of the two distribution models.

Distribution models	Scale parameter	Shape parameter
The two-parameter Weibull model	0.002474	0.858535
Lognormal	-6.67852	1.449919

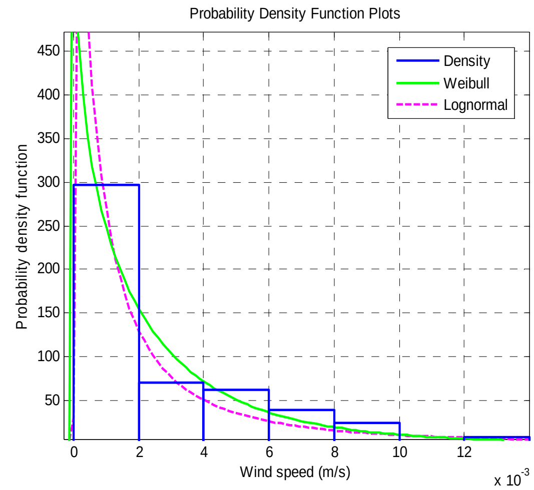


Figure 2 Comparison of different probability densities of wind speed data

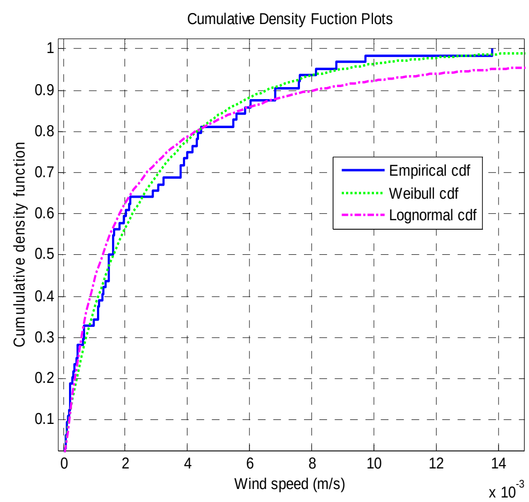


Figure 3 Comparison of different cumulative densities of wind speed data

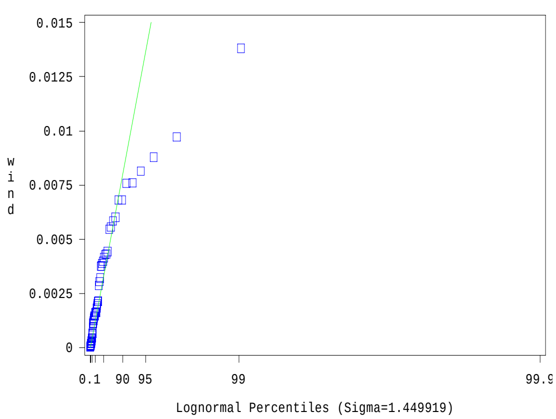
Table 2 The *p*-value of goodness-of-fit test of the two distributions model

Distributions	The two-parameter Weibull	Lognormal
Kolmogorov-Smirnov	0.250	0.010
Cramer Von Mises	0.250	0.015
Anderson Darling	0.250	0.012
Chi-square	0.301	0.106

ISBN: 978-960-6766-71-8

ISSN: 1790-5095

Table 2 shows a list of the EDF tests and chi-square test available for the two-parameter Weibull distribution and lognormal distribution. At the p-value 0.10 significance level, all four tests support the conclusion that the two-parameter Weibull distribution with scale parameter = 0.002474 and shape parameter = 0.858535 provides a good model for the distribution of wind speed data. While for lognormal distribution with scale parameter= -6.67852 and shape parameter = 1.449919, the p-value for the EDF test are all less than 0.10, indicating that the data does not support a lognormal model. Figure 4 shows that parameter estimates of the two-parameter Weibull distribution are more adequate compared with lognormal distribution.



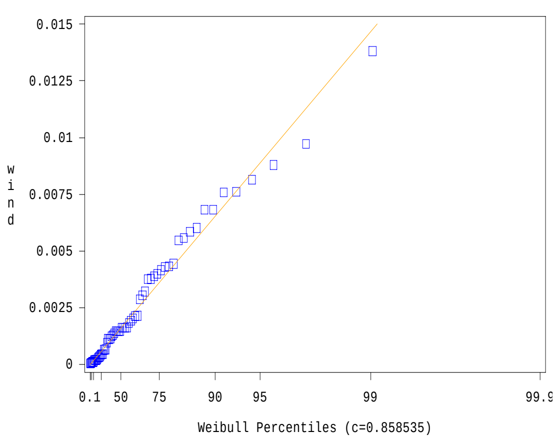


Figure 4 Comparison of probability plots of wind speed data

5 Conclusion

This paper investigates a comparative assessment of methods for wind speed data recorded in the Engineering Faculty, Universiti Kebangsaan Malaysia by using two different models, namely the two-parameter Weibull distribution and lognormal distribution. It was found from the goodness of-fit test that the two-parameter Weibull distribution is better than the lognormal model.

References:

[1] Chen, C. (2006). Test of fit for the three-parameter lognormal distribution, *Computational Statistics & Data Analysis*, **50**, 1418-1440.

[2] Dorvlo, A.S.S. (2001). Estimating wind speed distribution, *Energy Conversion and Management* **43**, 231-2318

[3] Gracia, A., Torres, J.L., Prieto, E., and DeFrancisco, A. (1998). Fitting wind speed distribution: A case study, *Solar Energy* **2**, 139-144.

[4] Hennessey, J.P.Jr. (1997). Some aspects of wind power statistics. Journal of Applied Meteorology **16**(2), 19-128.

[5] Muruthy, D.N.P., Xie, M., and Jiang, R. (2004). *Weibull Models*. John Wiley & Sons, Hoboken, New Jersey.

[6] Xiao, Y.Q., Li, Q.S., Li, Z.N., Chow, Y.W., and Li, G.Q. (2006). Probability distribution of extreme wind speed and its occurrence interval, *Engineering Structures* **28**, 1173-1181.

ISBN: 978-960-6766-71-8

ISSN: 1790-5095