The Suitability Of Statistical Distribution In Fitting Wind Speed Data

AZAMI ZAHARIM[1,2], SITI KHADIJAH NAJID[1], AHMAD MAHIR RAZALI[2] AND KAMARUZZAMAN SOPIAN[2]

[1] Fundamental Engineering Unit
Faculty of Engineering and Built Environment
[2]Solar Energy Research Institute (SERI)
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor D.E.
MALAYSIA

azaminelli@gmail.com, khadijahnajid@gmail.com,
mahir@ukm.my,ksopian@vlsi.eng.ukm.my

Abstract:-Wind energy has been used for navigation and agriculture. Recently, wind energy is given a lot of attention because of the focus on renewable energy. Wind energy growth in Asia is currently on the rise. Both India and China are leading with more installed capacity and manufacturing facilities. In Malaysia, wind energy conversion is also given a serious consideration. The potential for wind energy generation in Malaysia depends on the availability of the wind resource that varies with specific location. This paper deals with how to model or how to fit several probability distribution models to Malaysia wind speed data available. As usually described in the literature concerning efforts to develop an adequate statistical model for wind speed, there are a few statistical models discussed such as Weibull Distribution and Rayleigh Distribution. In the literature, it is a common procedure to compare these functions to determine which one fits the measured distribution best. The result from a simple descriptive statistics shows that Weibull distribution might be the probability distribution that can fit the data well.

Key-Word:- Wind speed, wind speed distribution, Weibull

1 Introduction

Wind is a source of renewable energy which is clean, efficient and offers many advantages to human beings. One of its many advantages is through the energy conversion in the form of wind power into useful forms of electricity which in between involves kinetic energy. Wind turbines which used wind power is developed by water pumping windmills, and later replaced by steam engine, need a constant wind speed to generate electricity. Wind energy can be considered a green power technology as it has only minor impacts on the environment. Any means impact might be because of the process of the production of the wind energy itself, but so far it has not made any major impacts.

Wind which is actually a form of solar energy is one of a kind that researchers put efforts in addressing the challenges to greater use. Winds are caused by the heating of the atmosphere by the sun, and also because of the rotation of the earth. Malaysia is one of the countries that lie in the equatorial zone and its climate is very much influenced by the monsoons. Malaysia has vast solar and wind resources available for energy generation through renewable energy technologies. Thus, continuous efforts should be done in order to utilize this kind of renewable energy. Commercialization of renewable energy technologies is needed especially for rural, social and economic development.

Currently, the wind energy is one of the fastest developing renewable energy source technologies across the globe. Wind energy is an alternative clean energy source compared to fossil fuel, which pollute the lower layer of the atmosphere. It has the advantage of being suitable to be used locally in rural and remote areas. The increasing demand for energy supply coupled with limited energy resources creates an urgency to find new solutions for this energy shortage.

Nowadays, wind analysis gives remarkable information to researches involved in renewable energy studies. The use of wind energy can significantly reduce the combustion of fossil fuel and the consequent emission of carbon dioxide. Supplementing our energy base with clean and renewable sources of energy has become imperative due to the present days' energy crisis and growing environmental consciousness. Knowledge of the statistical properties of wind speed is essential for predicting the energy output of a wind energy conversion system. Because of the high variability in space and time of wind energy, it is important to verify that the analyzing method used for the measuring wind data will yield the estimated energy collected that is close to the actual energy collected. Mahyoub H. Al Buhairi (2006) found that in recent years, many efforts have been made to construct an adequate model for the wind speed frequency distribution. The distribution of wind energy at different wind speeds is commonly known as the wind power density which is calculated by multiplying the power of each wind speed with the probability of each wind speed. A. N. Celik (2003) summarized that in the field of engineering, the wind speed distribution functions are ultimately used to be able to correctly model the wind power density, not the wind distribution itself. Therefore, the most important criterion of the ability as how successful it is to predict the measured wind power density, not the wind speed distribution.

Various distributions have been used by past researchers on the efforts of utilizing wind potential, and there are several methods to calculate the parameters of specific wind speed distribution. The most commonly used is through the graphical method. In addition, to check for the accuracy of the specific distribution, ones have to apply two or more methods to the given data.

A statistical distribution commonly used for describing measured wind speed data is the Weibull distribution. A review of the methods found in statistical literature for the purpose of estimation of the parameters in Weibull distribution is given, with a special emphasis on the efficiency of the different methods. From this review, the most appropriate method for a given application can be chosen. Maximum likelihood estimators should be used due to their large sample efficiency. However, they require an iterative minimization. K. Condradsen et al. (1983) recommended closed form estimators when there are few observations (say, less than 25) are the least-squares estimators. However, in this paper, observations are about a year of daily data of a particular place and it involved. Hence the general conclusion is that maximum likelihood is the appropriate method for the parameter estimation.

2 Wind Energy Technology

Wind is the movement of air caused by forces acting on the atmosphere which include the uneven heating of the earth's surface by the sun and the earth rotation. The uneven heating of the earth depends on the amount of energy absorbed by its surface, which varies with the amount of energy striking the surface. This amount depends on the angle of the sun (which varies seasonally) and on the type of surface (land or water) being struck. Uneven heating of the earth's surface causes the atmosphere above the surface to be unevenly heated. Because air that is hot is at a higher pressure than air that is cold, an imbalance is created by the different in pressure; air flows from a hot, high-pressure region to a cool, lower-pressure region. This is actually wind.

Figure 1 is a simplified version of airflow, showing two columns of air (warm and cold), with arrows (airflow) moving from high to low pressure regions.

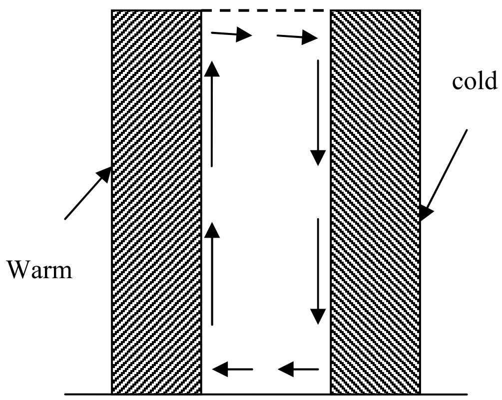

FIGURE 1 Simplified version of airflow

The second major force causing the wind is the rotation of the earth. Figure 2 shows the earth's rotation, giving a clear picture of the Coriolis force. This rotation has two effects; a direct acceleration of the air and a deflection (caused by the Coriolis force) of the wind's direction. The Coriolis force is an example of a fictitious force (or *pseudo force*), because it does not appear when the motion is expressed in an inertial frame of reference, in which the motion of an object is explained by the real impressed forces, together with inertia. In a rotating frame, the Coriolis force, which depends on the velocity of the moving object, and centrifugal force, which does not depend on the velocity of the moving object, are needed in the equation to correctly describe the motion.

Perhaps the most commonly encountered rotating reference frame is the Earth. Freely moving objects on the surface of the Earth experience a Coriolis force, and appear to veer to the right in the northern hemisphere, and to the left in the southern. As air moves from high to low pressure in the northern hemisphere, it is deflected to the right by the Coriolis force. In the southern hemisphere, air moving from high to low pressure is deflected to the left by the Coriolis force.

The amount of deflection the air makes is directly related to both the speed at which the air is moving and its latitude. Therefore, slowly blowing winds will be deflected only a small amount, while stronger winds will be deflected more. Likewise, winds blowing closer to the poles will be deflected more than winds at the same speed closer to the equator. The Coriolis force is zero right at the equator. Since Since Malaysia lies mainly in the equatorial region, it never experience Coriolis force.

Other factors account for variability in the flow of wind near the ground. Two of these are terrain and surface roughness. Together with the uneven solar heating of the atmosphere, the terrain and roughness can force the air to change direction and to flow at extremely varied speeds.

Terrain has a major effect on the speed and direction of wind as it follows a course of least resistance. Furthermore, the shape of the cliff determines the magnitude and location of the turbulent flow area, which changes dramatically.

Speed also can be enhanced by a topographical feature like a cliff, if the area of turbulence is avoided. The side of a hill can be regions of accelerated airflow as the winds are divided by the hill. Since winds seek the path of least resistance, gaps and gorges also tend to affect wind direction and velocity. Because the land surface gives off its heat more quickly than a water mass, a small-scale of the uneven heating of the earth's surface' takes place daily near a body of water. Since air tends to flow from an area of high pressure to an area of low pressure, this sets up local-scale winds that usually blow in one of two directions only, depending on the time of day. Surface roughness is the term applied to site-specific conditions that affect the wind. Surface friction causes a wind speed to decrease. A greater surface roughness will give a greater impact on the wind speed and turbulence.

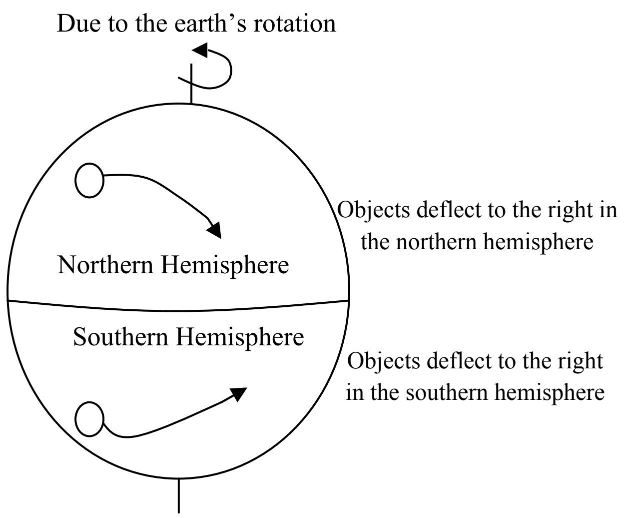

FIGURE 2 The earth's rotation

3 Energy in the wind

The energy in the wind varies as the cube of the wind speed. Therefore, if you double the wind speed, the power available in the wind represented is by $2^3=8$; or to state it in another way, the power of a windstream increases by a factor of eight for each doubling of the wind speed.

The energy potential of the wind is also related to the air. This means that at higher latitude, where the air is thinner, less energy is available at a certain wind speed .

4 Wind Speed Data

The data used in this survey is wind data obtained from a yearly published book at Pemerhatian Cuaca Harian Pusat Pengajian Sosial, Pembangunan & Persekitaran (PPSPP), Fakulti Sains Sosial & Kemanusiaan (FSSK), Universiti Kebangsaan Malaysia (UKM). The data of this study consisted of daily wind speed measured in meter per second. Measured at 3 meters height at Kuala Lumpur International Airport (KLIA), Sepang, it covered over a year surface wind speed data which is from January until December 2006. From the data collected, all missing data which were assumed defective were omitted.

5 Wind Speed Probability Distribution Function

A few distributions have predominantly been used for fitting the measured wind speed data. However, this paper will discuss only on Weibull and Rayleigh distribution. The wind speed distribution, one of the wind characteristics, is of great importance for not only for structural and environmental design and analysis, but also for the assessment of the wind energy potential and the performance of wind energy conversion system as well. Over the last two decades, many researchers have been devoted to develop an adequate statistical model to describe wind speed frequency distribution. E. Kavak Akpinar and S. Akpinar (2004) summarize that the Weibull and Rayleigh functions are commonly used for fitting the measured wind speed probability distribution. Corotis et al. (1978) found that the Rayleigh distribution is better than the Weibull distribution. However, Hennessey J. (1977) found that the energy output calculated using wind speeds derived from Rayleigh distribution was within 10% of those derived from the Weibull distribution. Justus et al. (1976) found that Weibull distribution gave the best fit to wind speed data from more than 100 stations in the United States National Climate Centre. While, in Denmark, ***Petersen et al. (1981) found that the Weibull distribution gave the an excellent fit to the wind speed distribution.*** A.S.S Dorvlo (2002) found that in using the Weibull distribution, the parameters estimated using the Chi Square method gave a better estimates for the parameters compared to other methods, as once indicated by the Kolmogrov-Smirnov statistic.

Weibull models have been used in many different applications and for solving variety of problems from many different disciplines. The Weibull function is a two parameter functions which are the shape parameter β and scale parameter α. The probability of the Weibull distribution function is given by

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

Where $f(x)$ is the probability of observing wind speed x, α is the scale parameter and β is the shape parameter. The corresponding cumulative probability function of the Weibull distribution is

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

The mean and variance of the data are first calculated using the data. Then, the following approximation can be used to calculate the Weibull parameters α and β:

$$\beta = \left(\frac{\sigma}{\bar{x}}\right)^{-1.086} \quad (1 \le \beta \le 10), \quad (3)$$

$$\alpha = \frac{\bar{x}}{\Gamma(1 + 1/\beta)} \quad (4)$$

where the average wind speed $\bar{x}$ is

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (5)$$

the variance σ^2 of wind velocity recordings

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (6)$$

and the gamma function of (x) (standard formula) is the calculated as

$$\Gamma(x) = \int_0^{\infty} e^{-u} u^{x-1} du \quad (7)$$

In circumstances where the shape parameter is equal to 2, the Weibull distribution is referred as Rayleigh distribution. Rayleigh distribution can be simplified as

$$f(x) = \frac{2x}{\alpha^2} exp\left[\left(-\frac{x}{\alpha}\right)^{\beta}\right] \quad (8)$$

The cumulative probability distribution for the Rayleigh distribution is given by

$$F(x) = 1 - exp\left[-\left(\frac{2x}{\alpha^2}\right)\left(\frac{x}{\alpha}\right)^2\right] \quad (9)$$

The scale and shape parameter were estimated using the method of maximum likelihood (MLE) as follows:

$$\frac{\sum_{i=1}^{n}\left(x_i^{\hat{\beta}} \ln x_i\right)}{\sum_{i=1}^{n} x_i^{\hat{\beta}}} - \frac{1}{\hat{\beta}} - \frac{1}{n}\sum_{i=1}^{n} \ln x_i = 0 \quad (10)$$

$$\hat{\beta} = \left[\left(\sum_{i=1}^{n} x_i^{\hat{\beta}} \ln x_i\right)\left(\sum_{i=1}^{n} x_i^{\hat{\beta}}\right)^{-1} - n^{-1}\sum_{i=1}^{n} \ln x_i\right]^{-1} \quad (11)$$

After fitting data with one or more models, you should evaluate the goodness of fit. As is common in statistical literature, the term goodness of fit is used here in several senses: A "good fit" might be a model that your data could reasonably have come from, given the assumptions of least-squares fitting in which the model coefficients can be estimated with little uncertainty that explains a high proportion of the variability in your data, and is able to predict new observations with high certainty.

There are many types of goodness of fit, it can be either the sum of squares due to error (SSE) R-square Adjusted R-square Root mean squared error (RMSE). But, unfortunately, this paper will only use RMSE and chi square error only. Root Mean Squared Error (RMSE) is also known as the fit standard error and the standard error of the regression. It is an estimate of the standard deviation of the random component in the data, and is defined as

$$RMSE = s = \sqrt{MSE} \quad (12)$$

where MSE is the mean square error or the residual mean square

$$MSE = \frac{SSE}{v} \quad (13)$$

Where SSE is the statistic measures the total deviation of the response values from the fit to the response values. It is also called the summed square of residuals and is usually labeled as SSE.

$$SSE = \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2 \quad (14)$$

v indicates the number of independent pieces of information involving the n data points that are required to calculate the sum of squares. It is the residual degrees of freedom which is defined as the number of response values *n* minus the number of fitted coefficients *m* estimated from the response values

$$v = n - m \quad (15)$$

The chi square distribution is a special case of the gamma distribution where b = 2 in the equation for gamma distribution below.

$$\frac{(n-1)s^2}{\sigma^2} - \chi 2(n-1) \quad (16)$$

The chi square distribution gets special attention because of its importance in normal sampling theory. If a set of n observations is normally distributed with variance σ^2 and s^2 is the sample standard deviation, then

$$y = f(x|a,b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{\frac{x}{b}} \quad (17)$$

6 Result And Discussion

In this study, wind speed data in the Kuala Lumpur International Airport (KLIA) were analyzed. Descriptive statistics is obtained in order to summarize the whole data. The main results obtained from the present study can be summarized as follows:

TABLE 1 Descriptive Statistics of KLIA data in 2006

Statistics	Estimates (m/s)
Mean	1.665
Standard deviation	0.4701
Skewness	1.073
Kurtosis	0.128
Minimum value	0.5
Maximum value	3.9
Sample size	365

TABLE 2 Monthly wind characteristics at KLIA for the year 2006

Parameters	V (m/s)	σ (m/s)	β (m/s)	α
JAN	1.5645	0.4054	1.7199	4.1843
FEB	1.9429	0.7559	2.1854	2.7760
MAR	1.5516	0.5421	1.7348	2.8696
APR	1.3533	0.3866	1.4908	4.0918
MAY	1.5194	0.3745	1.6601	4.7681
JUNE	1.5467	0.3203	1.6767	5.2511
JUL	2.0871	0.3845	2.2476	5.9783
AUG	1.7774	0.4931	1.9630	3.6749
SEPT	1.7500	0.3937	1.9067	4.7969
OCT	1.6000	0.2683	1.7120	6.7651
NOV	1.6167	0.2627	1.7305	5.8528
DEC	1.6839	0.4532	0.4532	3.6862

TABLE 3 Wind characteristics at KLIA for the year 2006

Parameters	V_m (m/s)	σ (m/s)	β (m/s)	α
	0.4701	1.6649	1.8392	3.5392

The descriptive statistics for the data set is shown in Table 1. The total number of observation is 365, intending that there is no defective value or missing observation. The mean wind speed is 1.665 m/s with a standard deviation of 0.4701 m/s. The kurtosis which shows that the peak is narrower than the normal distribution gives a value of 0.128 m/s. The monthly wind characteristics derived including the parameter estimates for KLIA are shown in Table 2. The parameters β and α can be obtained by using equation (3) and (4). From this table, July has shown the highest wind speed at 2.0871. While Table 3 shows the KLIA wind speed for the whole year 2006.

Figure 3 and 4 show the probability density function (PDF) and cumulative density function (CDF) respectively. Both Figures 3 and 4 below show the comparison between the Weibull and Rayleigh distribution to the actual data. It is clearly shown in Figure 4 that the Weibull distribution is closer to the empirical distribution obtained from the basis of the measured wind speed data.

TABLE 4 Parameter estimates for Weibull and Rayleigh distribution

Distributions	Parameter estimates

	Scale	*shape*
Weibull	1.78	0.47
Rayleigh	1.22	2

Parameter estimation for both the Weibull and its transformation; Rayleigh distribution are done. These shape and scale parameter values are obtained by solving equation (10) and (11). Parameter estimates is done using Maximum likelihood estimator (MLE) method.

Table 5 shows the goodness of fit test of the two distributions. As can be seen from the table, the results have shown that the RMSE and chi-square values of the Weibull distribution are lower than the values obtained by the Rayleigh distribution. Hence, the test support the conclusion Weibull distribution provides a better model for distribution of wind speed for this KLIA data sets.

TABLE 5 Goodness of fit test

	Weibull	**Rayleigh**
RMSE	0.6680	0.7346
Chi square	0.0535	0.0879

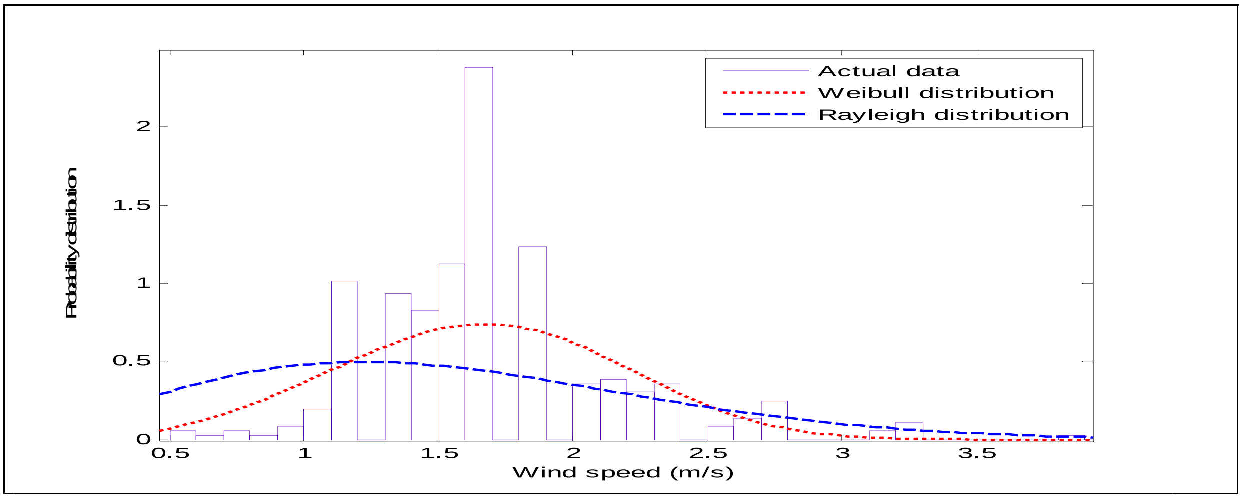

FIGURE 3 Probability distribution for the year 2006

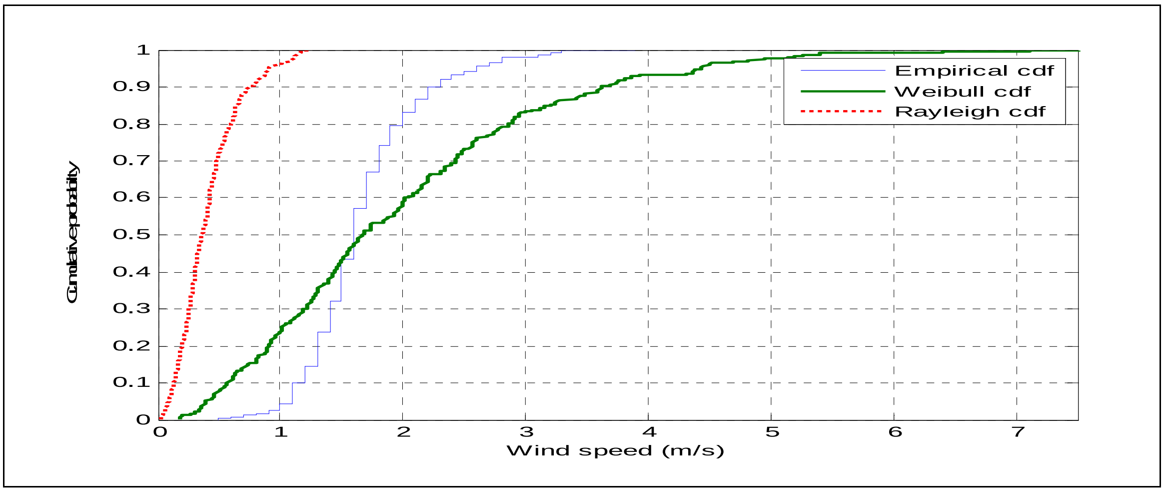

FIGURE 4 Cumulative distributions for the year 2006

7 Conclusion

This paper investigates a comparative assessment between two statistical distributions; Weibull and Rayleigh using a one year period data sets from the international airport of Malaysia, KLIA. It was found in this paper that both Weibull and Rayleigh gives a close approximation indicate that the Rayleigh is a special case of the Weibull distribution in which the shape parameter takes the value of two. However, the Weibull approximation was found to be the accurate distribution according to its lower value of RMSE and chi square.

Substituting both the scale and shape parameter for the Weibull distribution in Table 5 into equation 1, we get the final Weibull model for this data set;

$$f(x) = \frac{1.78}{0.47}\left(\frac{x}{0.47}\right)^{0.78} \exp\left[-\left(\frac{x}{0.47}\right)^{1.78}\right]$$

References

[1] Ahmad Mahir Razali, Rozaimah Zainal Abidin, Azami Zaharim and Kamaryzzaman Sopian. 2008. Modelling wind speed data via two parameter weibull. *Proceedings of Seminar on Engineering Mathematics*, 27 - 29 June 2008. Cameron Highlands, Pahang.

[2] Akpinar, E. Kavak and Akpinar, S. 2004. An assessment on seasonal analysis of wind energy characteristics and wind turbine characteristics. *Energy Conversion and Management* 46; 1848 - 1867

[3] Al Buhairi, Mahyoub H. 2006. A statistical analysis of wind speed data and an assessment of wind energy potential in taiz-yemen. *Ass. Univ. Bull. Environ. Res.* (9) 2 : 21 – 33

[4] Celik, A.N. 2003. Assessing the suitability of wind speed probability distribution functions based on wind power density. *Renewable Energy* 28: 1563 - 1574.

[5] Condradsen, K., Nielsen L. B., Prahm L. P. 1984. Review of Weibull statistics for estimation of wind speed distributions. *Journal of Climate and Applied Meteorology* 23: 1173 - 1183

[6] Corotis RB, Sigi AB, Klein J. 1978. Probability models of wind velocity magnitude and persistence. *Solar energy* 20: 483 – 493

[7] Dorvlo, A.S.S. 2002. Estimating wind speed distribution. *Energy*

Conversion and Management 43: 2311 – 2318

[8] Hennessey J. Some aspects of wind power statistics. 1977. *Journal of Applied Meteorology*. 16 (2): 119 – 128

[9] Justus, C.G., Hargraves, W.R and Yalcin (1976). Nationwide assessment of potential output from wind generators. *Journal of Applied Meteorology* 15: 673 - 678

[10] Kamaruzzaman Sopian, Mohd Yusof Hj. Othman, Rosli Zulkifli and Ahmad Shazli Abdul Wahab. 1999. Wind energy potential on Malaysian islands: A case study of Tioman, Redang and Perhentian Island. *Proceedings of the World Renewable Energy Congress, Malaysia*. 493 – 497.

[11] M. Cellura, G. Cirrincione, A. Marvuglia, A. Miraoui. 2008. Wind speed spatialestimation for energy planning in Sicily: Introduction and statistical analysis. *Renewable Energy* 33: 1237 – 1250

[12] Menon, M.V. 1963. Estimation of the shape and scale parameter of the Weibull distribution. *Technometrics* 5: 175-182.

[13] Millard, S.P. 2002. *Environmental Stats For S-Plus*. New York: Springer-Verlag, Inc.

[14] Muruthy, D.N.P., Xie, M. & Jiang, R. 2004. *Weibull Models*. New Jersey: John Wiley & Sons, Inc.

[15] Ramı´rez, Pene´lope and Carta, Jose´ Antonio. 2006. The use of wind probability distributions derived from the maximum entropy principle in the analysis of wind energy: A case study *Energy Conversion and Management* 47: 2564–2577

[16] Rozaimah Zainal Abidin. 2007. Penyuaian taburan kebarangkalian bagi data halaju angin. Tesis Sarjana Sains (Statistik). Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia.