Research Journal of Applied Sciences, Engineering and Technology 6(10): 1774-1779, 2013
DOI:10.19026/rjaset.6.3902
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: October 22, 2012 Accepted: December 20, 2012 Published: July 20, 2013

Research Article

The Probability Distribution Model of Wind Speed over East Malaysia

[1]Nurulkamal Masseran, [1, 2]Ahmad Mahir Razali, [1, 2]Kamarulzaman Ibrahim,
[2, 3]Azami Zaharim and [2]Kamaruzzaman Sopian
[1]Centre for Modelling and Data Analysis (DELTA), School of
Mathematical Sciences, Faculty of Science and Technology,
[2]Solar Energy Research Institute (SERI),
[3]Head of Project Group of Renewable Energy Resources Analysis,
Policy and Energy Management, Renewable Energy Niche, Universiti
Kebangsaan Malaysia, 43600 UKM Bangi, Selangor D.E., Malaysia

Abstract: Many studies have found that wind speed is the most significant parameter of wind power. Thus, an accurate determination of the probability distribution of wind speed is an important parameter to measure before estimating the wind energy potential over a particular region. Utilizing an accurate distribution will minimize the uncertainty in wind resource estimates and improve the site assessment phase of planning. In general, different regions have different wind regimes. Hence, it is reasonable that different wind distributions will be found for different regions. Because it is reasonable to consider that wind regimes vary according to the region of a particular country, nine different statistical distributions have been fitted to the mean hourly wind speed data from 20 wind stations in East Malaysia, for the period from 2000 to 2009. The values from Kolmogorov-Smirnov statistic, Akaike's Information Criteria, Bayesian Information Criteria and R^2 correlation coefficient were compared with the distributions to determine the best fit for describing the observed data. A good fit for most of the stations in East Malaysia was found using the Gamma and Burr distributions, though there was no clear pattern observed for all regions in East Malaysia. However, the Gamma distribution was a clear fit to the data from all stations in southern Sabah.

Keywords: Goodness of fit, spatial pattern, wind energy, wind regime, wind speed distribution

INTRODUCTION

In wind turbine design and site planning, the probability distribution of wind speed becomes critically important in estimating the energy production (Morgan *et al*., 2011) It has been defined in engineering practice, the average wind turbine power, $\bar{P}_w$ associated with the Probability Density Function (PDF) of wind speeds, *X* is obtained from:

$$\hat{\bar{P}}_w = \int_0^\infty P_w(X) f(X)\, dX \tag{1}$$

where,
f (X) = The PDF of *X*
P_w (X) = The turbine power curve that is used to describe the power output of wind speed

Generally, P_w (X) is defined as a proportion of the area of the airstream, measured in a plane perpendicular to the direction of the wind speed:

$$P_w(X) = \frac{1}{2} A \rho X^3$$

where,
A = The area
ρ = A constant for air density

Morgan *et al*. (2011) stated that the largest uncertainty in estimation of $\bar{P}_w$ lies in the choice of wind speed PDF, f (X), since the turbine manufacturer can measure P_w (X) fairy accurate. Thus, the utilization of a more accurate wind speed PDF will minimize the uncertainty in wind resource estimates and improve the site assessment phase of planning.

Wind speed distribution has been explored successfully by several scientists with 2-parameter Weibull and Rayleigh distributions are often quoted as popular distribution for wind speed. However, several authors have indicated that the Weibull and Rayleigh distribution should not be used in a generalized way, as

Corresponding Author: Ahmad Mahir Razali, Centre for Modelling and Data Analysis (DELTA), School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor D.E., Malaysia

they fail to represent some wind regimes (Carta and Ramirez, 2007; Brano *et al*., 2011; Jaramillo and Borja, 2004; Safari, 2011). For example, Brano *et al.* (2011) investigated 7 probability density functions employed to describe wind speed frequency distributions: Weibull, Rayleigh, Lognormal, Gamma, Inverse Gaussian, Pearson type V and Burr. The Burr distribution was the most reliable statistical distribution. Jaramillo and Borja (2004) showed that the Mixed Weibull distribution is more appropriate than the 2-parameter Weibull distribution for regions where wind speed presents a bimodal PDF. Safari (2011) used 5 probability distribution functions to fit wind speed data from 4 wind stations in Rwanda: Weibull, Rayleigh, Lognormal, Normal and Gamma. His results showed that Weibull and Gamma were the most suitable distributions. More research on wind speed distribution has been conducted by Carta *et al*. (2008, 2009), Zhou *et al*. (2010) and Celik (2003) and etc.

Among the early works on wind energy research in Malaysia is the work by Sopian *et al*. (1995). They have analyzed data from 10 wind stations in Malaysia and used the Weibull distribution. Their results indicated that Mersing and Kuala Terengganu possess the best potential for wind energy development. Masseran *et al.* (2012) investigated the persistence of wind speed in Peninsular Malaysia using the stationary properties of time series and the wind speed duration curve. Their results revealed that Chuping wind station had the most persistent wind speeds, compared to other stations. However, data from Mersing wind station are the most persistent for the level of wind speed suitable for generating energy. In addition, Ong *et al.* (2011) noted that a 150 kW wind turbine, which was built in Terumbu Layang-Layang in 2005, demonstrated some success. However, Tenaga Nasional Berhad (TNB), which is the only electricity supplier in Malaysia, built two units of wind turbines at Pulau Perhentian. Additionally, the Ministry of Rural and Regional Development built 8 small units of wind turbines in Sabah and Sarawak for local communities (Ong *et al*., 2011).

In this study, we focus on determining the best statistical model that describes the wind regime in East Malaysia. This model provides vital information for the assessment of wind energy potential.

Study area, regional climate and data: East Malaysia is a country that lies entirely in the equatorial zone, situated on the island of Borneo, with a geographic coordinate of 2° 30' north latitude and 119° 30' east longitude. Throughout the year, East Malaysia experiences a wet and humid climate with daily temperatures ranging from 25.5 to 35°C. The wind that blows across East Malaysia is influenced by the northeast monsoon that occurs from November until March. East Malaysia is also influenced by the effect of sea breezes and land breezes, especially when the sky is clear. During the afternoon, sea breezes occur with a speed of 10 to 15 knots, while land breezes occur at night. The data used in this study consist of hourly wind speeds (km/h) from 20 wind stations across the country. The collection period was from January 2000 to November 2009. Data were obtained from the Department of Environment and Malaysian Meteorology Department.

METHODOLOGY

Wind speed probability distribution (Table 1): To describe the behavior of wind speed at a particular location, it is necessary to identify the distribution that best fits the data. Suitable distributions for each wind station were determined by fitting nine types of statistical distribution to the data: Weibull (WE), Burr (BR), Gamma (GA), Inverse Gamma (IGA), Inverse Gaussian (IGU), Exponential (EX), Rayleigh (RY), Lognormal (LN) and Erlang (ER). Here, ER is simply a special case of Gamma distribution, where the shape parameter is an integer. Table 1, lists the probability density functions with their respective cumulative distribution functions (Morgan *et al*., 2011; Carta and Ramirez, 2007; Carta *et al*., 2009; Zhou *et al*., 2010; Evans *et al*., 1993).

Maximum likelihood estimator (Table 2): In this study, parameter estimation for each model was performed using the maximum likelihood method. The Maximum Likelihood Estimator (MLE) for the parameters of the WE, GA, IGU, ER, IGA and BR distributions can be determined numerically using methods such as Newton-Rapson, scoring, EM algorithm, quasi-Newton and the Nelder-Mead method. In this study, the Nelder-Mead method was used as an optimization technique for determining the MLE (R Development Core Team, 2008). For other distributions, such as LN, RY and EX, the MLEs can be easily determined. After the parameter estimation process, several goodness of fit tests were used to determine the most suitable statistical distribution for the data from each wind station. Goodness of fit tests included Kolmogorov-Smirnov (KS), Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). In addition, the R^2 correlation coefficient was also used to evaluate the goodness of fit for each method. A large R^2 value indicates a good fit of the theoretical distribution to the data.

The Kolmogorov-Smirnov statistic (KS): To determine the suitable probability distribution of wind speed from each station, the Kolmogorov-Smirnov test was calculated by comparing the cumulative

Table 1: List of probability density functions and cumulative distribution function

Model	Probability Density Function (PDF)	Cumulative Distribution Function (CDF)
Lognormal (LN)	$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left[\frac{-\left(\ln(x)-\mu\right)^2}{2\sigma^2}\right]$	$F(x)=\frac{1}{2}+\frac{1}{2}erf\left[\frac{\ln(x)-\mu}{\sigma\sqrt{2}}\right]$ where, erf = Complementary error function
Weibull (WE)	$f(x)=\frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right]$	$F(x)=1-eksp\left[\left(-\frac{x}{\alpha}\right)^{\beta}\right]$
Rayleigh (RY)	$f(x)=\frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)$	$F(x)=1-\exp\left(-\frac{x^2}{2\sigma^2}\right)$
Exponential (EX)	$f(x)=\frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right)$	$F(x)=1-\exp\left(-\frac{x}{\theta}\right)$
Gamma (GA)	$f(x)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)$	$F(x)=\frac{\gamma\left(\alpha,\frac{x}{\beta}\right)}{\Gamma(\alpha)}$ where, γ ( ) = Lower incomplete gamma function.
Inverse Gaussian (IGU)	$f(x)=\left[\frac{\lambda}{2\pi x^3}\right]^{\frac{1}{2}}\exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}$	$F(x)=\Phi\left[\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}-1\right)\right]+e^{\frac{2\lambda}{\mu}}\Phi\left[-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}+1\right)\right]$ where, Φ ( ) = Standard normal distribution.
Burr (BR)	$f(x)=\frac{aqx^{a-1}}{b^a\left[1+(x/b)^a\right]^{1+q}}$	$F(x)=1-\left[1+\left(\frac{x}{b}\right)^{\alpha}\right]^{-q}$
Inverse Gamma (IGA)	$f(x)=\frac{\beta^p}{\Gamma(p)}x^{-p-1}\exp\left(-\frac{\beta}{x}\right)$	$F(x)=\frac{\Gamma\left(p,\beta/x\right)}{\Gamma(p)}$ where, numerator is upper incomplete gamma function

Table 2: The maximum likelihood estimator for all theoretical distributions

	Maximum Likelihood Estimator (MLE)		Maximum Likelihood Estimator (MLE)
LN	$\hat{\mu}=\frac{\sum_{i=1}^{n}\ln x_i}{n}$ and $\hat{\sigma}=\frac{\sum_{i=1}^{n}\left(\ln x_i-\hat{\mu}\right)^2}{n}$	GA	$\hat{\beta}=\frac{\overline{x}}{\alpha}$ and $\ln(\hat{\alpha})-\psi(\hat{\alpha})=\ln\left(\frac{1}{n}\sum_{i-1}^{n}x_i\right)-\frac{1}{n}\sum_{i=1}^{n}\ln x_i$
RY	$\hat{\sigma}=\sqrt{\sum_{i=1}^{n}x_i \Big/ 2n}$	EX	$\hat{\theta}=\overline{x}$
IGU	$\hat{\mu}=\overline{x}$ and $\hat{\lambda}=n\left[\sum_{i=1}^{n}x_i^{-1}-(\overline{x})^{-1}\right]^{-1}$	IGA	$\frac{np}{\beta}=\sum_{i=1}^{n}\frac{1}{x_i}$ and $n\ln\beta-n\psi(p)=\ln\sum_{i=1}^{n}x_i$
WE	$\hat{\beta}=\left[\left(\sum_{i=1}^{n}x_i^{\hat{\beta}}\ln x_i\right)\left(\sum_{i=1}^{n}x_i^{\hat{\beta}}\right)^{-1}-n^{-1}\sum_{i=1}^{n}\ln x_i\right]^{-1}$ and $\hat{\alpha}=\left[\left(\frac{1}{n}\right)\sum_{i=1}^{n}x_i^{\hat{\beta}}\right]^{\frac{1}{\hat{\beta}}}$	BR	$\frac{n}{a}+\sum_{i=1}^{n}\ln(x_i/b)=(1+q)\sum_{i=1}^{n}\ln(x_i/b)\left[\left(\frac{b}{x_i}\right)^a+1\right]^{-1}$ and $n=(1+q)\sum_{i=1}^{n}\left[\left(\frac{b}{x_i}\right)^a+1\right]^{-1}$ and $\frac{n}{q}=\sum_{i=1}^{n}\ln\left[\left(\frac{x_i}{b}\right)^a+1\right]$

distribution for the observed data to the cumulative distribution for the fitted data. The empirical distribution function F_n for n observations is defined as:

$$F_n(x)=\frac{1}{n}\sum_{i=1}^{n}I_{X_i\le x} \tag{2}$$

where, $I_{X_i}\le x$ is an indicator function. Indicator function equals to 1 if $X_i\le x$ and 0 otherwise. The Kolmogorov-Smirnov statistic for a given theoretical cumulative distribution function F (x) is given by:

$$D_n=\sup_x|F_n(x)-F(x)| \tag{3}$$

where, $\sup x$ is the supremum of the set of distances between F_n and F (x). If the sample comes from distribution F (x), then D_n will almost surely converges to 0. To implement Kolmogorov-Smirnov statistic test, the information about their theoretical cumulative distribution function, F (x) need to be known (Shorak and Wellner, 1986). The list of theoretical Cumulative Distribution Function (CDF) for each distribution is shown in Table 1.

Akaike's Information Criteria (AIC): The AIC is a tool used for model selection. It is defined in terms of an appropriate information criterion. The AIC offers a relative measure of the information lost when a given model is used to describe reality. The AIC uses a mechanism that assigns a score to each candidate model. The model with the minimum AIC value is selected as the best fit model. The AIC usually be used to measure the goodness-of-fit for a statistical model. The AIC is not a hypothesis test of the model; rather, it provides a means for comparison among models. The AIC general formula is given by:

$$\text{AIC}= -2\log(L)+2k \tag{4}$$

where,

L : The likelihood

k : The number of parameter in the fitted model (Burhnam and Anderson, 2002; Hirotugu, 1974)

The AIC acts as penalizes based on the log-likelihood criterion, affording a balance between a good fit and complexity. The model with the minimum AIC value is the preferred model. The AIC was used in this study because of its mathematical reason related to the maximum likelihood estimators (Claeskens and Hjort, 2010).

Bayesian Information Criteria (BIC): The BIC is another set of model selection criteria that chooses the candidate model with the highest probability, given the data. Gideon E. Schwarz developed the BIC using the Bayesian framework. The BIC uses the prior probabilities and the prior densities of all parameter vectors in the different models to select a model. It is closely related to the Akaike information criterion. The BIC is also known as Schwarz's Bayesian Criterion (SBC). The formula for BIC is given by:

$$\text{BIC}=-2\log(L)+k\log(n) \tag{5}$$

where,

L = The likelihood

k = The number of parameters

n = The number of observations in the fitted model

The BIC takes the form of a penalized log-likelihood function, where the penalty is equal to the logarithm of the sample size times the number of estimated parameters in the model (R Development Core Team, 2008; Claeskens and Hjort, 2010; McQuarrie and Tsai, 1998). The model with the minimum BIC value is selected. In this study, the BIC and AIC scores were compared with the results from the Kolmogorov-Smirnov test statistic.

Evaluating the goodness of fit: The R^2 correlation coefficient was used to evaluate the goodness of fit for each method. A larger R^2 value indicates a better fit of the theoretical distribution to the data. R^2 was used for goodness-of-fit comparisons because it quantifies the correlation between observed probabilities and the predicted probabilities from a distribution. The R^2 coefficient is determined as:

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{F}_i - \bar{F}\right)^2}{\sum_{i=1}^{n}\left(\hat{F}_i - \bar{F}\right)^2 + \sum_{i=1}^{n}\left(F_i - \hat{F}\right)^2} \tag{6}$$

where, $\bar{F} = \frac{\sum_{i=1}^{n}\hat{F}_i}{n}$. The estimated cumulative probabilities, $\hat{F}$, were derived from the cumulative distribution function of the proposed model. A large value of R^2 indicates a good fit of the model's cumulative probabilities, $\hat{F}$, to the empirical cumulative probabilities, F. The R^2 coefficient has been used by other researchers in similar studies to measure goodness of fit methods (Morgan *et al.*, 2011; Carta *et al.*, 2008, 2009).

RESULTS AND DISCUSSION

Table 3 provides the results from the goodness of fit statistics, the associated R^2 values and the selected

Table 3: The result of goodness of fit tests found based on Kolmogorov Smirnov test, Akaike's information criterion, Bayesian information criterion and the selected distribution (in bold and *italic*) for each station

	Goodness-of-fit method							Goodness-of-fit method					
St.	KS	***R^2 (%)***	AIC	***R^2 (%)***	BIC	***R^2 (%)***	St.	KS	***R^2 (%)***	AIC	***R^2 (%)***	BIC	***R^2 (%)***
1	***GA***	98.11	GA	98.11	GA	98.11	11	***BR***	98.36	GA	98.10	GA	98.10
2	***BR***	99.34	GA	99.20	GA	99.20	12	***GA***	96.97	WE	97.32	WE	97.32
3	***GA***	99.62	GA	99.62	GA	99.62	13	***GA***	98.28	GA	98.28	GA	98.28
4	***BR***	99.43	BR	99.43	BR	99.43	14	BR	99.46	***GA***	98.85	GA	98.85
5	WE	99.53	***BR***	99.54	WE	99.53	15	***GA***	99.29	GA	99.29	GA	99.29
6	WE	99.09	***GA***	99.90	GA	99.90	16	***GA***	98.98	GA	98.98	GA	98.98
7	***GA***	99.30	GA	99.30	GA	99.30	17	***BR***	99.25	BR	99.25	BR	99.25
8	***BR***	98.77	GA	98.66	GA	98.66	18	***GA***	99.48	GA	99.48	GA	99.48
9	***GA***	99.08	GA	99.08	GA	99.08	19	***ER***	99.38	WE	99.24	WE	99.24
10	***WE***	99.86	WE	99.86	WE	99.86	20	***GA***	97.13	GA	97.13	GA	97.13

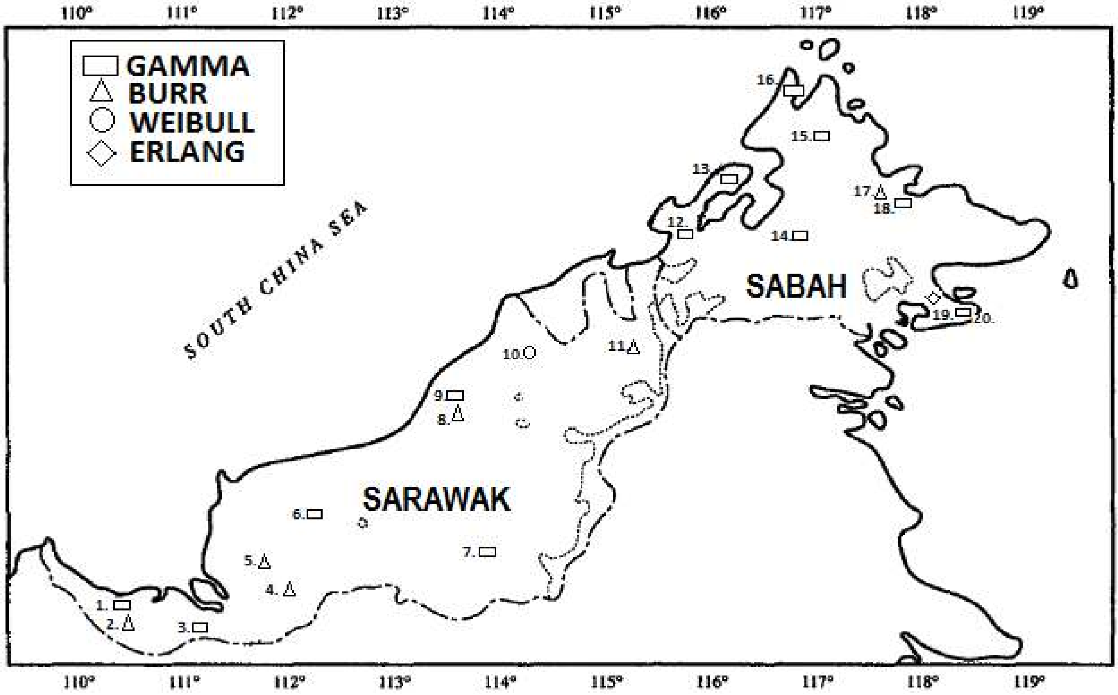

Fig. 1: The spatial distribution of wind speed over East Malaysia

distribution to describe the data for each station. Based on the results for all distributions and with respect to all goodness of fit methods, all R^2 values were found to be greater than 0.97, indicating that these distributions fit the data well. However, for the purpose of selecting the best distribution, we used the largest value of R^2. The Weibull, Gamma, Erlang and Burr distributions were found to be the most suitable for explaining the hourly mean speed in East Malaysia. The most frequent distribution was selected based on the highest number of stations that were successfully fit using that particular distribution. Based on results shown in Table 2, GA was the most frequently selected distribution: it provided the best fit to wind speed observations at 12 stations. The second most frequently selected distribution was BR. Six stations were successfully fitted with the BR distribution. This was followed by WE and ER, which were found to adequately fit the data observed at 1 station each. However, LN, EX, RY, IGA and IGU did not result in a good fit to the distribution of wind speed at all stations. A map of East Malaysia is provided in Fig. 1, which clearly shows the pattern of suitable statistical models that describe the wind regime. Data from most of the stations in the Sabah region (especially northern Sabah) were best fit to the Gamma distribution. A variety of different statistical distributions were observed for data from other regions.

CONCLUSION

The Gamma distribution is the distribution that most frequently adequately described the distribution of wind speed at the 20 stations considered in this study. A variety of different statistical distributions were observed in East Malaysia, except for northern Sabah, where the Gamma was the best fit distribution for all stations in that area. However, we suggest that a more comprehensive analysis needs to be conducted in the future. More stations should be included to obtain a better understanding of wind speed in East Malaysia before the effort is made to assess wind energy potential.

ACKNOWLEDGMENT

The authors are indebted to the staff of the Department of Environment and Malaysian Meteorology Department for providing data to make this study possible. This research would not have been possible without the sponsorship from Universiti Kebangsaan Malaysia and Ministry of Higher Education, Malaysia (grant number UKM-GUP-2011-213 and OUP-2012-064).

REFERENCES

Brano, V.L., A. Orioli, G. Ciulla and S. Culotta, 2011. Quality of wind speed fitting distribution for urban area of Palermo, Italy. Renew. Energ., 36: 1026-1039.

Burhnam, K.P. and D.R. Anderson, 2002. Model Selection and Multimodel Inference: A Practical Information-theoritical Approach. 2nd Edn., Springer-Verlang, New York.

Carta, J.A. and P. Ramirez, 2007. Analysis of two-component mixture Weibull statistics for estimation wind speed distribution. Renew. Energ., 32: 518-531.

Carta, J.A., P. Ramirez and C. Bueno, 2008. Joint probability densities function of wind speed and direction for wind energy analysis. Energ. Convers. Manage., 49: 1309-1320.

Carta, J.A., P. Ramirez and S. Velazquez, 2009. A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands. Renew. Sust. Energ. Rev., 13: 933-955.

Celik, A.N., 2003. Assessing the suitability of wind speed probability distribution based on power density. Renew. Energ., 28: 1563-1574.

Claeskens, G. and N.L. Hjort, 2010. Model Selection and Model Averaging. Cambridge University Press, New York.

Evans, M., N. Hastings and B. Peacock, 1993. Statistical Distributions. 2nd Edn., John Wiley and Son, New York.

Hirotugu, A., 1974. A new look at the statistical model identification. IEEE T. Automat. Contr., 19: 716-723.

Jaramillo, O.A. and M.A. Borja, 2004. Wind speed analysis in La Ventosa, Mexico: A bimodal probability distribution case. Renew. Energ., 29: 1613-1630.

Masseran, N., A.M. Razali, K. Ibrahim and W.Z. Wan Zin, 2012. Evaluating the wind speed persistence for several wind stations in Peninsular Malaysia. Energy, 37: 649-656.

McQuarrie, A.D.R. and C.L. Tsai, 1998. Regression and Time Series Model Selection. World Scientific Publishing, Singapore.

Morgan, E.C., M. Lackner, R.M. Vogel and L.G. Baise, 2011. Probability distributions for offshore wind speeds. Energ. Convers. Manage., 52: 15-26.

Ong, H.C., T.M.I. Mahlia and H.H. Masjuki, 2011. A review on energy scenario and sustainable energy in Malaysia. Renew. Sust. Energ. Rev., 15: 639-647.

R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Safari, B., 2011. Modeling wind speed and wind power distribution in Rwanda. Renew. Sust. Energ. Rev., 15: 925-935.

Shorak, G.R. and J.A. Wellner, 1986. Empirical Process with Application to Statistics. John Wiley and Sons, New York.

Sopian, K., M.Y. Hj. Othman and A. Wirsat, 1995. The wind energy potential in Malaysia. Renew. Energ., 6(8): 1005-1016.

Zhou, L., E. Erdem, G. Li and J. Shi, 2010. Comprehensive evaluation of wind speed distribution model: A case study for North Dakota sites. Energ. Convers. Manage., 51: 1449-1458.