2013, Scienceline Publication

Journal of Civil Engineering and Urbanism

Volume 3, Issue 1: 06-11 (2013) (Received: May 21, 2012; Accepted: August 22, 2012; Published: January 30, 2013) ISSN-2252-0430

Comparison of Four Distributions for Frequency Analysis of Wind Speed: A Case Study

R. Daneshfaraz[1], S. Nemati[2], H. Asadi[3] and M. Menazadeh[4]

[1] *Department of Civil Engineering, University of Maragheh, Maragheh, Iran*
[2] *Department of Water Engineering, University of Tabriz, Tabriz, Iran*
[3] *Department of Civil Engineering, Maragheh Branch, Islamic Azad University, Maragheh, Iran*

[*]Corresponding Author's E-mail address: daneshfaraz@yahoo.com

ABSTRACT: The increase in negative effects of fossil fuels on the environment has forced many countries to use renewable energy sources, especially wind energy. Wind speed is the most important parameter of the wind energy. Probability distributions are useful for estimating wind speed because it is a random phenomenon. This study analyzes wind speed frequencies using wind data from Urmia synoptic station in Iran. Four different distributions are fitted to the maximum annual wind from station, and parameters of the distributions are estimated using the method of maximum likelihood and the method of moments. Calculations are performed with *Mathematica*, a computer algebra system developed by Wolfram Research. The advantage of using this software is that the symbolic, numerical, and graphical computations can be combined and that all quantities can be accurately calculated; in particular, there is no need to resort to any approximate methods for the calculation of quantiles. There is a ready-to-use command for calculating quantiles from distributions that are built in *Mathematica*, while for other distributions they can be easily and accurately calculated by inverting the cumulative distribution functions or by solving nonlinear equations where the inversion is not possible. The best distribution is selected based on the root mean square error (RMSE), the coefficient of determination (R^2), and the probability plot correlation coefficient (PPCC). The results indicate that the best performance can be obtained by the truncated extreme value distribution.

Keywords: *Mathematica;* Probability distribution; Urmia; Wind energy; Wind speed frequency

ORIGINAL ARTICLE

INTRODUCTION

Nowadays more and more countries in the world have had to seek help to renewable resources, such as wind, solar and geothermal, not only to meet the increasing energy demand, but also for environmental reasons. Wind energy production is an attractive and feasible method employing renewable energy source without emission of pollutants.

Wind energy can be considered a green power technology as it has only minor impacts on the environment. Currently, wind energy is one of the fastest developing renewable energy source technologies across the globe. Countries all around the world are doing a thorough research on the specific energy that comes out by the uneven heating by the sun. Wind analysis gives remarkable information to researches involved in renewable energy studies.

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 in measuring wind data will yield the estimated energy collected that is close to the actual energy collected. The wind speed distribution, one of the wind characteristics, is of great importance 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. For this reason, an accurate determination of probability distribution of wind speed values is very important in evaluating wind speed energy potential of a region. Wind energy potential can be determined by wind measurements of a certain investigation region depending on years. And also extreme wind speed frequency estimation is usually important in many fields of environmental studies such as climatology, hydrology, developing wind energy facilities, agricultural management, and structure designing (Lopez, 1998; Gomes et al., 2003).

Many investigators have tried to fit different frequency distributions to wind data. Al Buhairi and Mahyoub (2006) found that, in recent years, many efforts have been made to construct an adequate model for the wind speed frequency distribution. In the literature, the Weibull distribution is commonly used in the practical studies related to the wind energy modeling (Stevens, 1979; Toure, 2005; Zhou et al., 2006). Auwera et al. (1980) used Weibull three-parameter model for estimating mean wind power densities. Weisser (2003) analyzes wind energy analysis of Grenada using the Weibull density function. Lun and Lan (2000) studied Weibull parameters using long-term wind observations. Seguro et al. (2000) estimated the parameters of Weibull wind speed distributions for wind energy analysis. Celik (2003) used Weibull distribution to estimate wind energy output of large- and small-scale turbines. Rehman et al. (1994) used Weibull parameters for wind speed distribution in Saudi Arabia.

Recently, Pandey and Sutherland (2003) fitted generalized Pareto distribution to peak-over-threshold extreme wind speed through bootstrapping. Holmes and Moriarty (1999) also suggested generalized Pareto distribution to fit the extreme wind speed in Australia. Recently, Zaharim et al. (2009) fitted gamma, lognormal and Weibull distributions to wind speed data in the east coast of Malaysia. The numerical and graphical results obtained from the specific statistics showed that the Weibull and gamma distributions, whose parameters are estimated using the maximum likelihood principle, provide the best fits for the year 2005 and 2006, respectively.

This study aims to find the best probability distribution to the annual maximum wind speed of Urmia, Iran and is the first study in Iran that investigates the best probability distribution of maximum wind speed. The study uses the *Mathematica* code developed for (Ghorbani et al., 2010); see http://users.utu.fi/ruskeepa/.

Probability Density Functions

Four different probability distributions are considered in this study. These distributions and their probability density functions are presented in Table 1. Only the lognormal, truncated extreme value, truncated logistic, Weibull distributions are used.

Since wind speed is always non-negative, it is more realistic to truncate the density functions so that they yield a domain that consists of only non-negative values. The truncation is done by simply dividing the original density function by a suitable constant, to make the integral of the truncated density function equal to one; the constant is given by $P(X \geq 0) = 1 - F(0)$.

Table 1. Probability distributions and their density functions

Distribution	PDF	Assumption	Domain
Lognormal	$\frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2}$	$\sigma > 0$	$x > 0$
Truncated extreme value	$\frac{1}{\beta}\left(1-e^{-e^{\frac{\alpha}{\beta}}}\right)^{-1} e^{-\frac{x-\alpha}{\beta}-e^{-\frac{x-\alpha}{\beta}}}$	$\alpha > 0, \beta > 0$	$x > 0$
Truncated logistic	$\frac{1+e^{-\frac{\mu}{\beta}}}{\beta} e^{-\frac{x-\mu}{\beta}}\left(1+e^{-\frac{x-\mu}{\beta}}\right)^{-2}$	$\mu > 0, \beta > 0$	$x > 0$
Weibull	$\frac{\alpha x^{\alpha-1}}{\beta^{\alpha}} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$	$\alpha > 0, \beta > 0$	$x > 0$

Comparison of Estimated Probability Density Functions

Many methods are available for estimating the parameters of the above distributions, such as least-squares, maximum likelihood, moments, weighted moments, linear moments, and entropy. Extensive details of these methods are already available in the literature (Singh, 1996) and, therefore, are not reported here. In this study, only two of these methods are employed: maximum likelihood estimation and the method of moments.

There is no specific reason for preferring these two methods against the others, except that they are simple and also sufficient for the purpose of this study. They are neither treated as superior to the other methods nor any effort is made compare with them.

The probability density functions thus fitted are compared using quantiles. Assuming that there are n number of observations, Cunnane's plotting positions are first calculated as: $p_i = (i-0.4)/(n+0.2)$ for $i = 1,\ldots,n$, where i is the order of the i^{th} observation arranged in ascending order and p_i is the probability of non-exceedance of the i^{th} observation estimated by the Cunnane's plotting position formula. For each of the density functions, the p_i-quantiles, given by X_{pi}, i=1,…,n, are calculated. These quantiles are then compared with the observed values, denoted as X_i, the i^{th} ordered value. Three statistical indicators are used to compare the computed and the observed quantiles (O'Donnell, 1985): (1) the root mean square error (RMSE), which is the square root of $\frac{1}{n}\sum_{i=1}^{n}(X_{pi}-X_i)^2$; (2) the coefficient of determination (R^2), which is the coefficient square of the correlation between the computed and the observed quantiles; and (3) the probability plot correlation coefficient (PPCC). The test uses the correlation *r* between the ordered X_i. The PPCC is a test statistic to measure linearity of the probability plot.

Critical values have been obtained for the normal and Gumbel distributions by Vogel (1986). The test has also been developed for the Weibull and the uniform distribution by Vogel & Kroll (1989) and for the extreme value distributions by Chowdhury et al. (1991). The critical values obtained for the normal distribution can also be used for the Log-normal distribution.

The correlation coefficient test statistic, *r*, is calculated between the ordered observed values and the inversed values of the cumulative distribution function of the fitted distribution. If the observed values conform to the applied distribution, the *r* statistic should be greater than the critical value for the selected significance level. The equations that give unbiased estimates of the inverse values have been obtained by different researchers. The *r* test statistic is defined as

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(m_i - \bar{m})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=}^{n} (m_i - \bar{m})^2}} \qquad (1)$$

in which x_i ($x_1 \leq ... \leq x_i \leq ... \leq x_n$) is the ordered ith value and $\bar{X}$ is the mean of observed values. Also,

$$m_i = F_x^{-1}(p_i) \qquad (2)$$

in which F(.) is the cumulative probability value and p_i is the value estimated by unbiased equations for different probability distributions.

DATA ANALYSIS AND RESULTS

In this study, wind speed frequency analysis is performed for Urmia synoptic station in Iran. The station is located in Iran north-west 37.55° N, 45.07° E with elevation of 1330 meters above sea level. Figure 1. shows the geographical location of the study area. The data considered for the wind speed frequency analysis are the annual maximum wind values.

Figure 2 shows the variations of wind values for the Urmia synoptic station (in the above order) and Table 2 presents some statistical parameters of wind speed. Wind speed frequency analysis is carried out using several distribution functions and with the two parameter estimation methods (i.e. MLE and MOM).

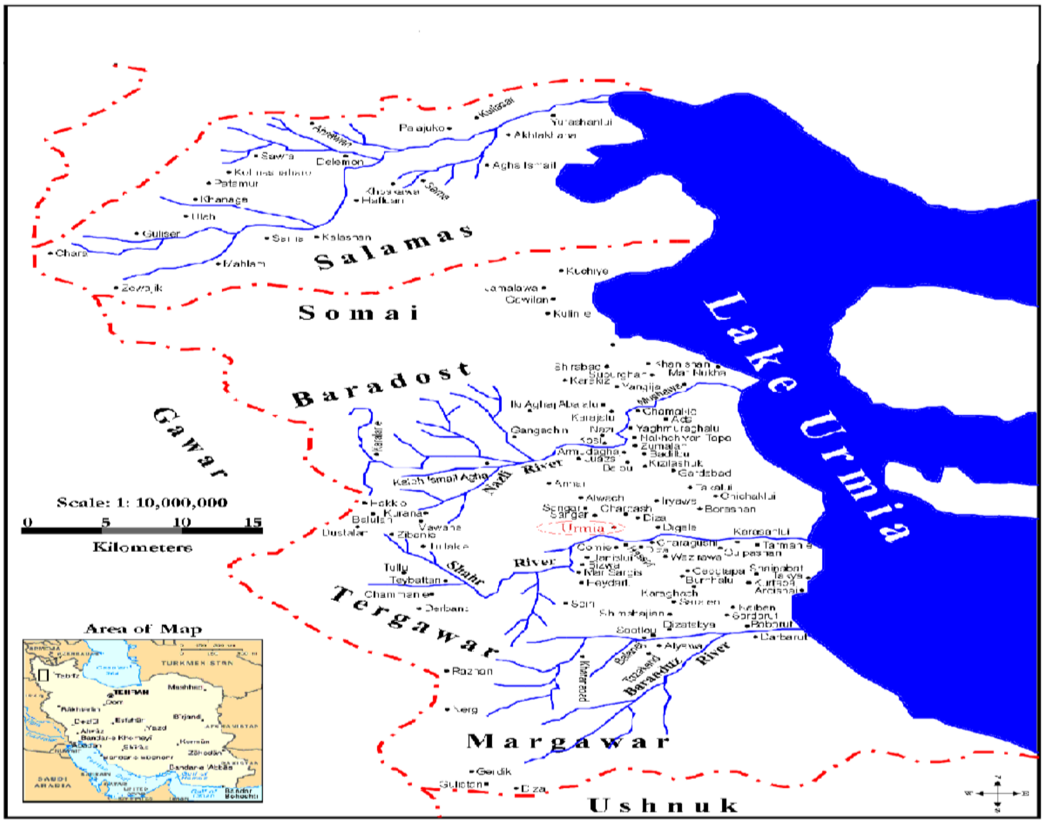

Figure 1. The geographical location of the study area

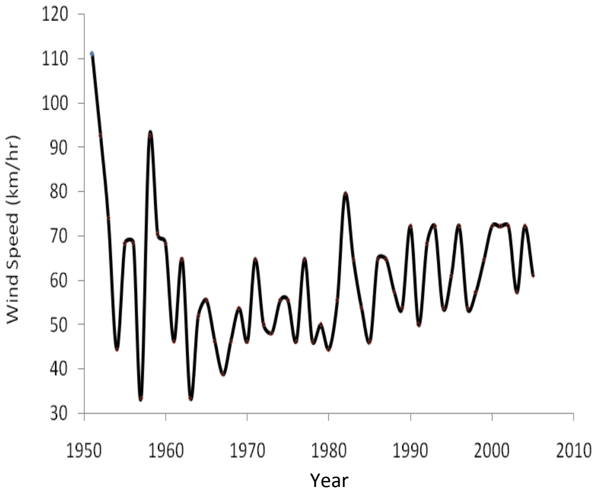

Figure 2. Maximum annual wind speed at the Urmia station

Table 2. Some statistical parameters of wind speed (km/hr) data used.

Statistical parameter	Value
Number	55
Average (km/hr)	60.070
Max (km/hr)	111
Min (km/hr)	33.3
Skewness (km^2/hr^2)	1.684
Standard Deviation (km/hr)	14.523

Comparison of distributions

Table 3. shows the estimated parameters of the distributions when the estimation is done with the maximum likelihood method and the method of moments.

To cite this paper: Daneshfaraz R., Nemati S., Asadi H., Menazadeh M. 2013. Comparison of Four Distributions for Frequency Analysis of Wind Speed: A Case Study. *J. Civil Eng. Urban*, 3(1): 06-11.
Journal homepage http://www.ojceu.ir/main/

Table 3. Estimated values of parameters

Distribution / Method	Log-normal	Truncated Extreme Value	Truncated Logistic	Weibull
MLE	μ=3.4529 σ=0.234122	α=28.8747 β=6.53236	μ =32.0111 β =4.26542	α =4.15335 β =35.5337
MOM	μ=3.45254 σ=0.236066	α =28.9741 β=6.06116	μ=32.453 β =4.31079	α =4.76378 β =35.4662

Table 4. Performance evaluation for selected distributions

Method	**Distribution**	**PPCC**	**R^2**	**RMSE**
MLE	Log-Normal	0.981665	0.963666	1.49065
	Truncated Extreme Value	**0.982969**	**0.966227**	**1.44048**
	Truncated Logistic	0.970186	0.94126	1.88497
	Weibull	0.958555	0.910433	2.3356
MOM	Log-Normal	0.981718	0.96377	1.48319
	Truncated Extreme Value	**0.982969**	**0.966227**	**1.44048**
	Truncated Logistic	0.970186	0.94126	1.88497
	Weibull	0.954166	0.910433	2.3356

Three goodness-of-fit methods (RMSE, R^2, and PPCC) are considered to select the best distribution. For the synoptic station, the results for selected distributions are presented in Table 4 for MLE and MOM. Figure 3. presents the best estimated density for the Urmia station obtained with these two parameter estimation methods.

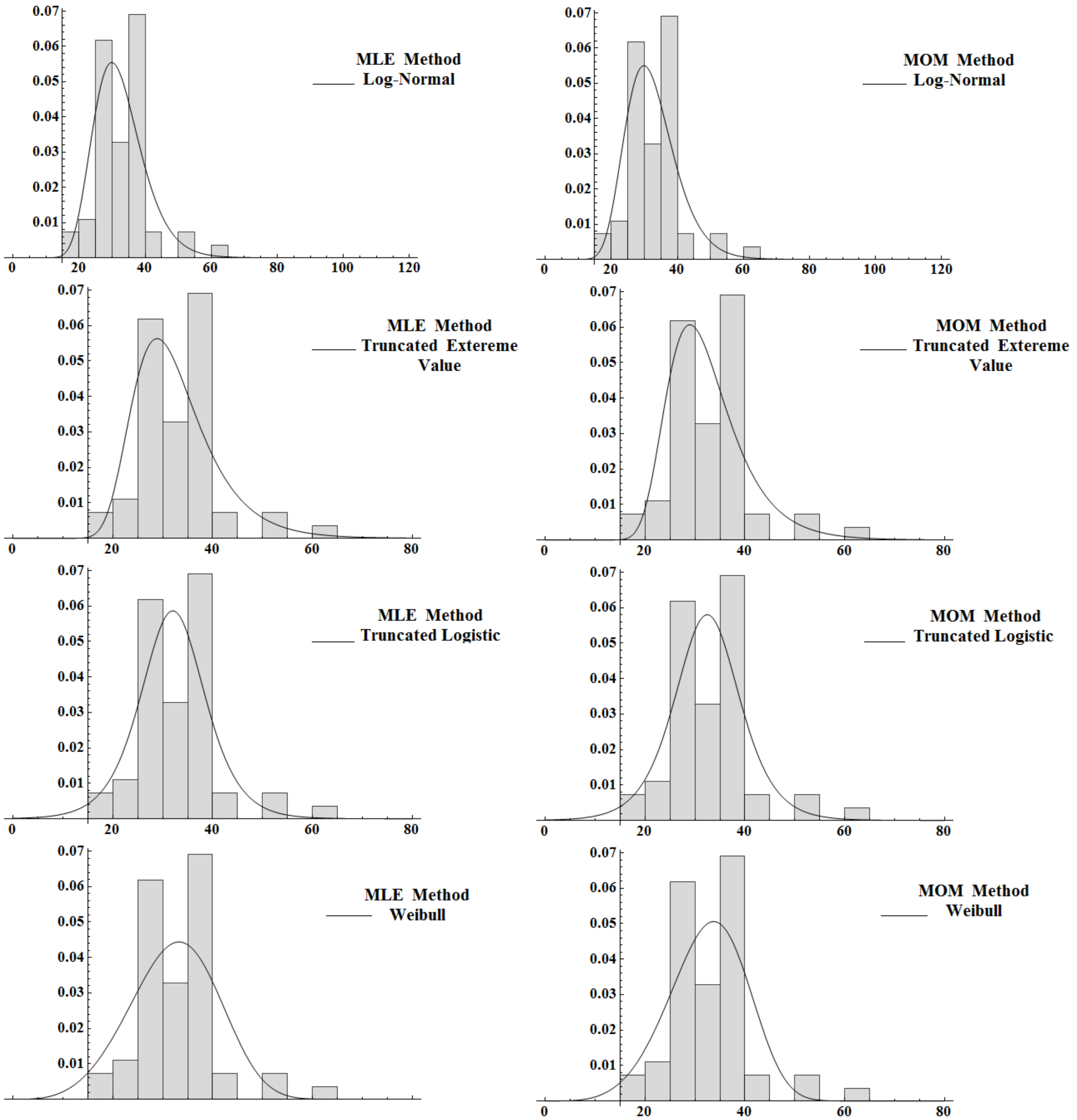

Figure 3. Histograms and the estimated probability density functions

To cite this paper: Daneshfaraz R., Nemati S., Asadi H., Menazadeh M. 2013. Comparison of Four Distributions for Frequency Analysis of Wind Speed: A Case Study. *J. Civil Eng. Urban*, 3(1): 06-11.
Journal homepage http://www.ojceu.ir/main/

With two methods, the truncated extreme value distribution is better than the other distributions according to all the three criteria (RMSE, R^2, and PPCC) and by both methods (MLE and MOM). Based on this result, it may be inferred that the truncated extreme value is generally suitable for synoptic stations.

Thus, the truncated extreme value distribution may be suggested as an appropriate distribution for synoptic Urmia station, and possibly for other Iranian conditions.Table 5 shows values that the wind speed exceeds with a given probability (i.e. quantiles) for MLE and MOM.

Table 5. Values that the wind speed exceeds with a given probability

Method	Criteria	Best Distribution	Wind speed quantiles							
			P=0.5	P=0.6	P=0.7	P=0.8	P=0.9	P=0.95	P=0.99	P=0.999
MLE	PPCC RMSE R2	Truncated Extreme Value	31.195	33.045	35.2228	38.0655	42.614	46.977	56.856	70.8401
MOM	PPCC RMSE R2	Truncated Extreme Value	31.195	33.045	35.2228	38.0655	42.614	46.977	56.856	70.8401

CONCLUSIONS

In this study, wind speed frequency analysis was performed for Iranian conditions. Maximum annual wind speed values observed at Tabriz synoptic station were studied. Four different probability distributions were fitted, and the method of maximum likelihood and the method of moments were used for parameter estimation. This study is also the first one where the software Mathematica was used for performing any type of wind speed frequency analysis. The results also indicate that, among the four different distributions, the truncated extreme value distribution is the most appropriate for the Urmia station.

The present study has important implications for wind speed frequency analysis for Iran in particular, and for regional climatology in general. Further, the use of Mathematica provides a new dimension to the wind speed frequency analysis. With the many challenges faced in using the existing methods (often due to difficulties in calculations) for the selection of the most appropriate probability distribution for a given region, the symbolic, numerical, and graphical capabilities of Mathematica together with its flexibility can go a long way. Future work will focus on advancing the use of Mathematica towards developing a more generalized and flexible framework for wind speed frequency analysis, details of which will be reported elsewhere.

Acknowledgement

The authors would like to thank Dr. Heikki Ruskeepää, University of Turku, Finland, for his help in preparing the manuscript.

REFERENCES

Auwera V, Meyer L. F, Malet L. M. (1980). The use of Weibull three-parameter model for estimating mean wind power densities. Journal of Apply Meteorology, 19: 819–825.

Buhairi Al, 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: 21 – 33.

Celik A.N. (2003). Energy output estimation for small-scale wind power generators using Weibull-representative wind data. Journal of Wind Engineering and Industrial Aerodynamics, 9: 1693–707.

Chowdhury J.U, Stedinger J.R, Lu L.H. (1991). Goodness-of-fit tests for regional generalized extreme value flood distributions. Journal of Water Resources Research, 27: 1765–1776.

Ghorbani M.A, Ruskeepää H, Singh V.P, Sivakumar B. (2010). Flood frequency analysis using *Mathematica*. Turkish J. Eng. Env. Sci, 34: 171–188.

Gomes L, Arrue J.L, Lopez M.V, Sterk G, Richard D, Gracia R, Sabre M, Gaudichet A, Frangi J.P. (2003). Wind erosion in a semiarid agricultural area of Spain: the WELSONS project catena. Catena, 52: 235–256.

Holmes J.D, Moriarty W.W. (1999). Application of the generalised Pareto distribution to extreme value analysis in wind engineering. Journal of Wind Engineering and Industrial Aerodynamics, 83: 1–10.

Lopez M.V. (1998). Wind erosion in agricultural soils: an example of limited supply of particles available for erosion. Catena, 33: 17–28.

Lun I.Y.F, Lan J.C. (2000). A study of Weibull parameters using long-term wind observations. Renewable Energy, 20: 145–53.

O'Donnell T. (1985). A direct three-parameter Muskingum procedure incorporating lateral inflow. Journal of Hydrology Science, 30: 494–495.

Pandey M.D, Sutherland H.J. (2003). Probabilistic analysis of list data for the estimation of extreme design loads for wind turbine components. A collection of the 2003 ASME Wind.

Rehman S, Halawani T.O, Husain T. (1994). Weibull parameters for wind speed distribution in Saudi Arabia. Solar Energy, 53: 473–479.

Seguro J.V, Lambert T.W. (2000). Estimation of the parameters of Weibull wind speed distributions for wind energy analysis. Journal of Wind Engineering and Industrial Aerodynamics, 85: 75–84.

Singh V.P. (1996). Entropy-based parameter estimation in hydrology. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Stevens D.E, Vorticity. (1979). Momentum and divergence budgets of synoptic-scale wave disturbances in the tropical eastern Atlantic. Mon. Wea. Rev, 107: 535–550.

Toure S. (2005). Investigations on the eigen-coordinates method for the 2-parameter Weibull distribution of wind speed. Renewable Energy, 30: 511–521.

Vogel R.M. (1986). The probability plot correlation coefficient test for the normal, lognormal and Gumbel distributional hypotheses. Journal of Water Resources Research, 22: 587–590.

Vogel R.M, Kroll C.N. (1989). Low-flow frequency analysis using probability plot correlation coefficients. J. Wat. Resour. Planning and Manag, 115: 338–357.

Weisser D. (2003). A wind energy analysis of Grenada: an estimation using the Weibull density function. Renewable Energy, 28: 1803–1812.

Zaharim A, Razali A.M, Abidin R.Z, Sopian K. (2009). Fitting of statistical distributions to wind speed data in Malaysia. European Journal of Scientific Research, 26: 6–12.

Zhou J, Xu F, Wang T, Cao A, Yin C. (2006). Cenozoic deformation history of the Qaidam Basin, NW China: Results from crossection restoration and implications for Qinghai-Tibet Plateau tectonics. Earth and Planetary Science Letters, 243: 195–210.