INTERNATIONAL JOURNAL OF TREND IN RESEARCH AND DEVELOPMENT, VOLUME 2(2), ISSN: 2394-9333
WWW.IJTRD.COM

Comparison of Probability Distributions for Extreme Value Analysis of Wind Speed – A Case Study

N. Vivekanandan
Central Water and Power Research Station, Pune, Maharashtra
E-mail: anandaan@rediffmail.com

Abstract — **Extreme Value Analysis (EVA) of wind speed plays an important role in estimating the design values of the wind load-effect on structures for any structural design. This can be carried out by fitting of probability distributions to the series of Annul Maximum Wind Speed (AMWS) data. This paper illustrates the adoption of Gumbel (EV1), Frechet (EV2), 2-parameter Log Normal (LN2) and Log Pearson Type-3 (LP3) distributions for estimation of Extreme Wind Speed (EWS) using AMWS recorded at Delhi. For determination of parameters of EV1, EV2, LN2 and LP3 distributions, Method of Moments (MoM) and Maximum Likelihood Method (MLM) are used. In addition to MoM and MLM, method of least squares and Order Statistics Approach (OSA) are also used for determination of parameters of EV1 and EV2 distributions. Goodness-of-Fit (GoF) tests viz., Anderson-Darling and Kolmogorov-Smirnov are applied for checking the adequacy of fitting of probability distributions to the recorded data. Diagnostic test (D-index) is used for the selection of suitable probability distribution for EVA of wind speed. Based on GoF and diagnostic test results, the study suggests the EV1 (OSA) is better suited probability distribution for estimation of EWS for Delhi.**

Keywords — **Anderson-Darling, D-index, Kolmogorov-Smirnov, Probability Distribution, Wind Speed**

I. INTRODUCTION

Structures are designed with the intention of safely withstanding ordinary and extreme wind loads over the entire intended economic lifetime. The wind pressures on a structure are a function of the characteristics of the approaching wind, the geometry of the structure under consideration, and the geometry and proximity of the structures upwind. Because of the many uncertainties involved, the maximum wind loads experienced by a structure during its lifetime, may vary widely from those assumed in design. Therefore, accurate estimation of the occurrence of Extreme Wind Speed (EWS) for a particular return period is an important factor for the design purposes. Such estimate is expressed in terms of the quantile value (X_T), viz., the maximum wind speed which is exceeded, on average, once every T-year, the return period. Generally, the 10000-year return period Mean+1σ (where Mean denotes the estimated EWS (X_T) and σ the Standard Error (SE)) value of EWS is considered for design purposes as per Atomic Energy Regulatory Board (AERB) guidelines [1]. This can be achieved by fitting of probability distribution to the recorded Annual Maximum Wind Speed (AMWS) data.

Number of probability distributions like Gumbel (EV1), Frechet (EV2), 2-parameter Log Normal (LN2) and Log Pearson Type-3 (LP3) are widely used for EVA of wind speed. In accordance with theory of probability distributions, EV1 and EV2 are classified as family of **e**xtreme **v**alue distributions whereas LN2 falls in a family of normal distribution and LP3 falls in a family of Gamma distribution [2]. Standard parameter estimation procedures viz., Method of Moments (MoM), Maximum Likelihood Method (MLM), Method of Least Squares (MLS) and Order Statistics Approach (OSA) are used based on the applicability of a particular distribution. In the recent past, number of studies has been carried out by different researchers on adoption of probability distributions for EVA of wind speed [3-6]. Kunz et al. [7] compared the Gamma and Generalized Pareto (GPA) distributions for estimation of EWS and concluded that GPA provides better estimates than Gamma. Morgan et al. [8] applied Extreme Value, Gamma and Normal family of probability distributions for estimation of EWS using the 10-minute wind speed observations recorded at 178 ocean buoy stations around North America. They have found that the LN2 distribution yielded the best estimate of EWSs, but still exhibited large errors. El-Shanshoury and Ramadan [9] applied EV1 distribution to estimate EWS for Dabaa area in the north-western coast of Egypt. Lee et al. [10] applied Gumbel and Weibull probability distributions for estimation of EWSs using the Korea wind map. They have observed that the Gumbel distribution gives better results than the Weibull. Daneshfaraz et al. [11] carried out the wind speed frequency analysis adopting LN2, truncated extreme value, truncated logistic and Weibull probability distributions and found that the truncated extreme value is the most appropriate distribution for Urmia synoptic station in Iran. Lawan et al. [12] evaluated the suitability of five different statistical distributions through GoF tests and found that the Gamma and LN2 distributions are better suited for modelling wind speed data of Miri, Malaysia.

Generally, when different distributional models are used for EVA, a common problem that arises is how to determine which distribution fits best for a given set of data. This can be answered by formal statistical procedures involving Goodness-of-Fit (GoF) and diagnostic tests; and the results are quantifiable and reliable than those from the empirical procedures. Qualitative assessment is made from the plots of the recorded and estimated EWS. For the quantitative assessment on EWS within in the recorded range, GoF tests such as Anderson-Darling (A^2) and Kolmogorov-Smirnov (KS) are applied. A diagnostic test of D-index is used for the selection of a suitable probability distribution for EVA of wind speed. The objective of the paper is to compare the four probability distributions used for EVA of wind speed and to identify the best suitable distribution for estimation of EWS through Goodness-of-Fit (GoF) and diagnostic tests. The procedures involved in EVA of wind speed adopting four probability distributions, computation of GoF tests statistic and D-index are briefly described in the following sections.

II. METHODOLOGY

In this paper, EV1, EV2, LN2 and LP3 distributions are used for EVA of wind speed. The Probability Density Function (PDF) and quantile estimator (X_T) of these distributions are presented in Table 1.

TABLE 1. PDF AND QUANTILE ESTIMATOR OF PROBABILITY DISTRIBUTIONS

Distribution	PDF	Quantile estimator
EV1	$f(X;\alpha,\beta)=\dfrac{e^{-(X_i-\alpha)/\beta}e^{-e^{-(X_i-\alpha)/\beta}}}{\beta}$, β>0	$X_T=\alpha+Y_T\beta$
EV2	$f(x;\beta,\gamma)=\dfrac{\gamma}{\beta}\left(\dfrac{\beta}{X}\right)^{\gamma+1}e^{-\left(\frac{X}{\beta}\right)^{-\gamma}}$, β>0	$X_T=\beta e^{(Y_T/\gamma)}$
LN2	$f(X;\mu_Y,\sigma_Y)=\dfrac{1}{\sqrt{2\pi}\sigma_Y X}\exp\left(-\dfrac{(\ln(X)-\mu_Y)^2}{2\sigma_Y^2}\right)$,-∞<X<∞,σ>0	$X_T=e^{\mu_Y+K_P\sigma_Y}$
LP3	$f(X;\alpha,\beta,\gamma)=\dfrac{1}{\beta X\Gamma\gamma}\left(\dfrac{\ln(X)-\alpha}{\beta}\right)^{\gamma-1}e^{-\left(\frac{\ln(X)-\alpha}{\beta}\right)}$,β,γ>0	$x_T=\mathrm{Exp}((\alpha+\beta\gamma)+K_P\beta\sqrt{\alpha})$

In Table 1, μ_y and σ_y are the mean and standard deviation of the log-transformed series of recorded data. α, β and γ are the location, scale and shape parameters of the distributions respectively. For EV1 and EV2 distributions, the reduced variate (Y_T) corresponding to the return period (T) is defined by $Y_T=-\ln(-\ln(1-(1/T)))$. K_P is the frequency factor corresponding to the probability of exceedance and Coefficient of Skewness (C_S) viz., C_S =0.0 for LN2 whereas C_S is based on the log transformed series of the recorded data for LP3 [13].

A) Goodness-of-Fit Tests

Generally, A^2 statistic is applied for checking the adequacy of fitting of EV1 and EV2 distributions. The procedures involved in application of A^2 statistic for LN2 and LP3 are more complex though the utility of the test statistic is extended for checking the quantitative assessment. In view of the above, KS test is widely applied for the purpose of quantitative assessment. Theoretical descriptions of GoF tests statistic are as follows:

A^2 statistic is defined by:

$$A^2=(-N)-(1/N)\sum_{i=1}^{N}\left\{\begin{matrix}(2i-1)\,\mathrm{Ln}(Z_i)+\\(2N+1-2i)\,\mathrm{Ln}(1-Z_i)\end{matrix}\right\} \quad \ldots (1)$$

Here, $Z_i=F(X_i)$ for i=1,2,3,…,N with $X_1<X_2<\ldots X_N$, $F(X_i)$ is the Cumulative Distribution Function (CDF) of i[th] sample (X_i) and N is the sample size [14].

KS statistic is defined by:

$$KS=\underset{i=1}{\overset{N}{\mathrm{Max}}}\,(F_e(X_i)-F_D(X_i)) \quad \ldots (2)$$

Where, $F_e(X_i)$ is the empirical CDF of X_i and $F_D(X_i)$ is the computed CDF of X_i. In the present study, Weibull plotting position formula is used for computation of empirical CDF and presentation of results in the form of probability plots. The theoretical value A^2 and KS statistic for different sample size (N) at 1% and 5% significance level is available in the technical note on "Goodness-of-Fit Tests for Statistical Distributions" by Charles Annis [15].

Test criteria: If the computed values of GoF tests statistic given by the distribution is less than that of theoretical values at the desired significance level (either at 5% or 1%) then the distribution is found to be acceptable for EVA of wind speed at that level.

B) Diagnostic Test

The selection of a suitable probability distribution for EVA of wind speed is performed through D-index test [16], which is defined as below:

$$\text{D-index} = \left(1/\overline{X}\right)\sum_{i=1}^{6}\left|X_i - X_i^*\right| \quad \ldots (3)$$

Here, $\overline{X}$ is the average value of the recorded data whereas X_i (i= 1 to 6) and X_i^* are the six highest recorded and corresponding estimated values by different PDFs. The distribution having the least D-index is considered as better suited distribution for estimation of EWS.

III. APPLICATION

In this paper, EVA of wind speed data is carried out to estimate the expected EWS for different return periods adopting four EV1, EV2, LN2 and LP3 distributions. MoM, MLM, MLS and OSA are used for determination of parameters of EV1 and EV2 distributions whereas MoM and MLM for LN2 and LP3 distributions. Hourly wind speed data (with missing values) recorded at Delhi for the period 1969-2012 is used. The series of AMWS is extracted from the hourly data and used for EVA. From the scrutiny of the wind speed data, it is observed that the data for the twelve years (1974, 1979-81, 1983-1988, 1990 and 2004) are missing. So, the data for the missing years are imputed by the series maximum value of 85 km/hr as per AERB guidelines and the entire data set is used for EVA. Table 2 gives the descriptive statistics of the recorded and log-transformed series of AMWS for Delhi.

TABLE 2. DESCRIPTIVE STATISTICS OF AMWS

Series of wind speed data	Statistical parameters (SD: Standard Deviation)			
	Average (km/hr)	SD (km/ hr)	C_S	C_K
Recorded	66.6	15.4	-0.007	-1.584
Log-transformed	4.171	0.239	-0.201	-1.386

IV. RESULTS AND DISCUSSIONS

Based on the parameter estimation procedures of EV1, EV2, LN2 and LP3 distributions, as given in the text book of "Flood Frequency Analysis" by Rao and Hameed, computer codes are developed in FORTRAN language and used for EVA of wind speed. These programs compute the distribution parameters, estimates of EWSs with standard error for different return periods, GoF tests statistic and D-index. The estimated EWS (X_T) with Standard Error (SE) computed from EV1 and EV2 distributions (using MoM, MLM, MLS and OSA) are given in Tables 3 and 4. Similarly, the estimated EWS with SE computed from LN2 and LP3 distributions (using MoM and MLM) are given in Tables 5. The plots of recorded and estimated EWS by EV1, EV2, LN2 and LP3 distributions are presented in Figures 1 to 4 respectively.

From Tables 3 and 5, it is noticed that there is no appreciable difference between the estimated EWS when MoM and MLM is used for determination of parameters of EV1 and LN2 distributions. Also, from Table 4, it is noticed that the MLM give higher estimates when compared with the other three parameter estimation methods of EV2. The fitted curves using LP3 (MoM) distribution indicates the estimated EWSs are higher when compared with the corresponding values of LP3 (MLM) for return periods from 2-year to 10000-year.

TABLE 3. ESTIMATED EWS WITH SE BY EV1 DISTRIBUTION

Return period (year)	EWS (km/hr) with SE							
	MoM		MLM		MLS		OSA	
	X_T	SE	X_T	SE	X_T	SE	X_T	SE
2	64.1	2.1	64.1	2.1	64.2	2.4	65.1	1.6
5	77.7	3.6	77.6	3.5	80.0	4.0	75.2	2.5
10	86.7	4.8	86.5	4.7	90.4	5.5	81.9	3.3
20	95.4	6.1	95.0	6.0	100.4	7.0	88.3	4.2
50	106.6	7.8	106.1	7.7	113.3	8.9	96.7	5.2
100	115.0	9.0	114.4	8.9	123.0	10.4	102.9	6.0
200	123.4	10.3	122.7	10.2	132.7	11.8	109.1	6.9
500	134.4	12.0	133.6	11.9	145.4	13.9	117.3	7.9
1000	142.8	13.3	141.9	13.1	155.0	15.4	123.5	8.8
2000	151.1	14.6	150.1	14.4	164.7	16.8	129.7	9.7
5000	162.1	16.4	161.0	16.1	177.4	18.8	137.9	10.7
10000	170.5	17.6	169.2	17.4	187.0	20.3	144.1	11.5

TABLE 4. ESTIMATED EWS WITH SE BY EV2 DISTRIBUTION

Return period (year)	EWS (km/hr) with SE							
	MoM		MLM		MLS		OSA	
	X_T	SE	X_T	SE	X_T	SE	X_T	SE
2	62.3	1.6	62.4	1.6	62.4	10.0	63.3	1.6
5	77.0	3.2	80.3	3.3	79.7	17.0	74.5	3.1
10	88.5	4.9	94.9	5.2	93.8	23.0	83.1	4.5
20	101.3	6.9	111.4	7.6	109.6	29.1	92.2	6.3
50	120.5	10.5	137.0	12.0	134.2	37.1	105.5	9.2
100	137.3	14.0	160.1	16.3	156.1	43.2	116.7	11.9
200	156.3	18.2	186.9	21.8	181.5	49.3	129.0	15.1
500	185.5	25.3	229.2	31.4	221.4	57.5	147.3	20.2
1000	211.1	32.1	267.4	40.8	257.3	63.7	162.9	24.7
2000	240.2	42.5	312.1	55.1	299.0	69.9	180.0	31.8
5000	285.0	53.8	382.6	72.3	364.7	78.1	205.5	38.8
10000	324.3	66.6	446.5	91.5	423.8	84.3	227.2	46.6

TABLE 5: ESTIMATED EWS WITH SE BY LN2 AND LP3 DISTRIBUTIONS

Return period (year)	EWS (km/hr) with SE							
	LN2				LP3			
	MoM		MLM		MoM		MLM	
	X_T	SE	X_T	SE	X_T	SE	X_T	SE
2	64.8	2.3	64.8	2.3	65.3	4.8	56.9	4.5
5	79.2	3.3	79.1	3.2	79.4	6.5	71.7	6.3
10	88.0	4.3	87.7	4.2	87.6	8.2	80.5	8.2
20	96.0	5.3	95.6	5.2	94.7	10.1	88.3	10.2
50	105.9	6.7	105.3	6.6	103.2	12.5	97.7	12.9
100	113.0	7.8	112.3	7.7	109.1	14.4	104.3	15.0
200	120.0	8.9	119.2	8.7	114.7	16.3	110.7	17.1
500	129.0	10.4	128.0	10.2	121.7	18.8	118.7	19.9
1000	135.7	11.6	134.6	11.4	126.8	20.7	124.6	22.1
2000	142.3	12.9	141.1	12.6	131.7	22.5	130.3	24.2
5000	150.6	13.5	150.0	12.4	137.1	24.9	136.6	26.6
10000	157.7	13.6	156.1	13.4	142.0	27.0	141.0	29.0

A) Analysis Based on GoF Tests

For assessing the adequacy of fitting of EV1, EV2, LN2 and LP3 distributions, GoF tests statistic values are computed from Eqs. (1) and (2), and the results are presented in Table 6.

TABLE 6: COMPUTED AND THEORETICAL VALUES OF A^2 AND KS STATISTIC BY EV1, EV2, LN2 AND LP3 DISTRIBUTIONS

Distri-bution	Computed values of							
	A^2				KS			
	MoM	MLM	MLS	OSA	MoM	MLM	MLS	OSA
EV1	3.047	2.685	1.678	9.423	0.186	0.188	0.157	0.229
EV2	2.678	1.638	1.715	8.061	0.177	0.142	0.147	0.215
LN2	1.736	1.951	-	-	0.186	0.185	-	-
LP3	1.718	1.944	-	-	0.175	0.187	-	-

From Table 6, it may be noted that the computed values of A^2 by EV1, EV2, LN2 and LP3 distributions are greater than the theoretical values of 1.038 at 1% level of significance, and at this level, all four distributions are not acceptable for modelling AMWS. Also, from Table 6, it may be noted that the computed values of KS statistic by these four probability distributions are lesser than of its theoretical value of 0.238 at 1% level of significance, and at this level, these four distributions are acceptable for modelling AMWS when MoM and MLM is applied for determination of parameters of distributions in addition to MLS and OSA for EV1 and EV2.

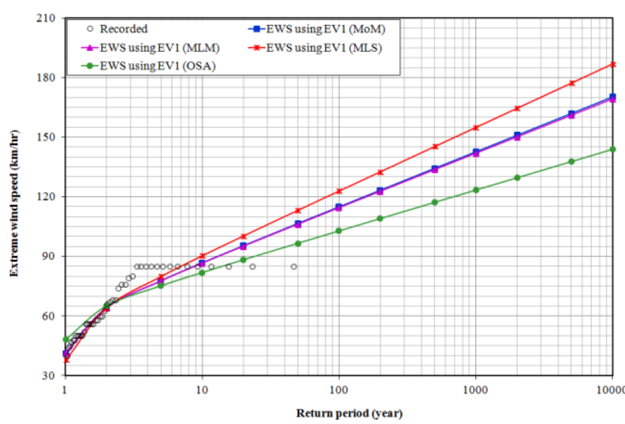

FIGURE 1: PLOTS OF RECORDED AND ESTIMATED EWS BY EV1 DISTRIBUTION (USING MOM, MLM, MLS AND OSA) FOR DELHI

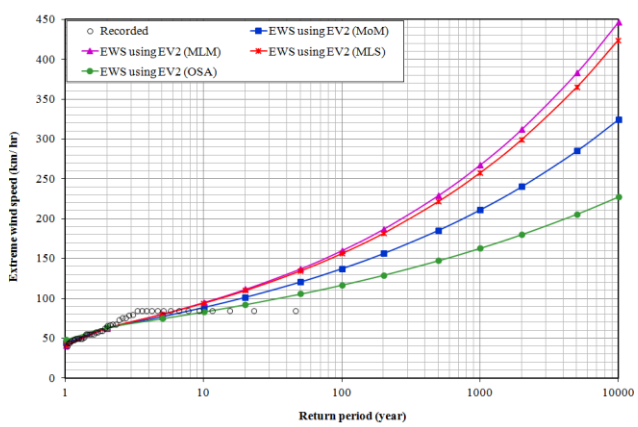

FIGURE 2: PLOTS OF RECORDED AND ESTIMATED EWS BY EV2 DISTRIBUTION (USING MOM, MLM, MLS AND OSA) FOR DELHI

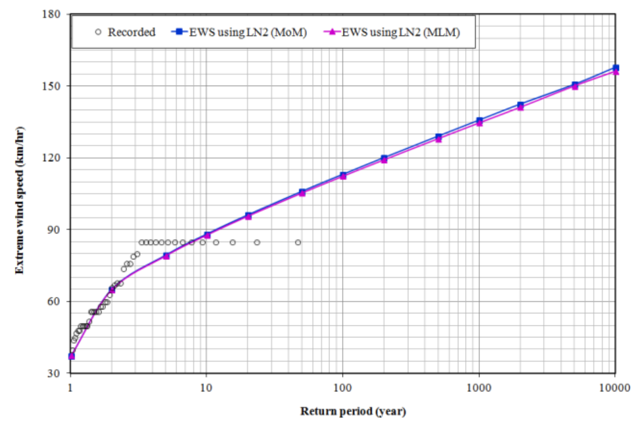

FIGURE 3: PLOTS OF RECORDED AND ESTIMATED EWS BY LN2 DISTRIBUTION (USING MOM AND MLM) FOR DELHI

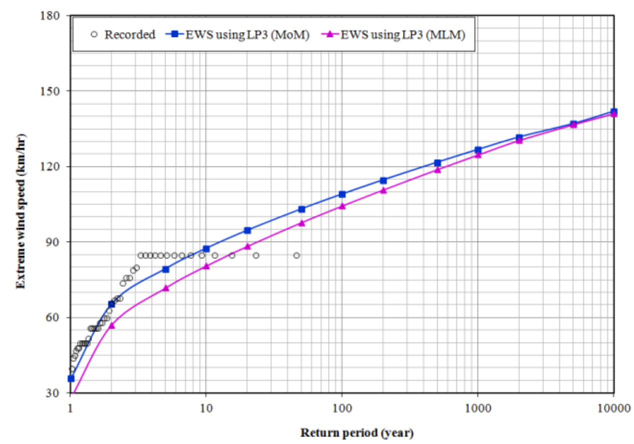

FIGURE 4: PLOTS OF RECORDED AND ESTIMATED EWS BY LP3 DISTRIBUTION (USING MOM AND MLM) FOR DELHI

B) Analysis Based on Diagnostic Test

In addition to GoF test results, D-index is used for the selection of suitable probability distributions for EVA of wind speed for Delhi. The D-index values for EV1, EV2, LN2 and LP3 distributions are computed from Eq. (3) and the results are presented in Table 7.

TABLE 7: COMPUTED VALUES OF D-INDEX BY EV1, EV2, LN2 AND LP3 DISTRIBUTIONS

Distribution	D-index			
	MoM	MLM	MLS	OSA
EV1	0.877	0.853	1.283	0.510
EV2	1.433	2.358	2.199	0.805
LN2	0.749	0.688	-	-
LP3	0.544	0.551	-	-

From Table 7, it may be noted that the D-index value of 0.510 computed by EV1 (OSA) distribution is found to be minimum when compared with the corresponding values of other three distributions when MoM and MLM is applied for determination of parameters in addition to MLS and OSA for EV1 and EV2. Based on GoF and diagnostic test results, it is observed that the EV1 (using OSA) is better suited probability distribution for estimation of EWS for Delhi.

V. CONCLUSIONS

The paper presents the computer aided procedure for determination of parameters of EV1, EV2, LN2 and LP3 distributions (using MoM and MLM) in addition to MLS and OSA for EV1 and EV2 for EVA of wind speed for Delhi. The study shows the selection of a suitable distribution is evaluated by GoF (using A^2 and KS) and diagnostic (using D-index) tests. The A^2 test results do not support the use of EV1, EV2, LN2 and LP3 distributions whereas KS test results confirm the use of these probability distributions for modelling AMWS recorded at Delhi. Based on the evidence of GoF and diagnostic test results, the study identifies the EV1 (using OSA) is better suited probability distribution for estimation of EWS for Delhi. The study suggests the 10000-year return period Mean+1σ value of 155.6 km/hr computed by EV1 (OSA) distribution could be considered for design purposes while planning and design of civil structures in the vicinity of Delhi.

ACKNOWLEDGEMENTS

The author is grateful to the Director, Central Water and Power Research Station, Pune, for providing the research facilities to carry out the study. The author is thankful to M/s Nuclear Power Corporation of India Limited, Mumbai and India Meteorological Department, Pune, for providing the wind speed data.

REFERENCES

[1] Atomic Energy Regulatory Board (AERB), “Extreme values of meteorological parameters”, AERB/ NF/ SG/ S-3, 2008.

[2] B. Bobee and F. Askhar, “The Gamma family and derived distributions applied in hydrology”, Water Resources Publications, 1991.

[3] E. Simiu, N.A. Heckert, J.J. Filliben and S.K. Johnson, “Extreme wind load estimates based on the Gumbel distribution of dynamic pressures: an assessment”, Structural Safety, vol. 23, pp. 221-229, 2001.

[4] S. Bivona, R. Burlon and C. Leone, “Hourly wind speed analysis in Sicily”, Renewable Energy, vol. 28, pp. 1371-1385, 2003.

[5] A.N. Celik, “On the distributional parameters used in assessment of the suitability of wind speed probability density functions”, Energy Conversion and Management, vol. 45, pp. 1735-1747, 2004.

[6] J.P. Palutikof, B.B. Brabson, D.H. Lister and S.T. Adcock, “A review of methods to calculate extreme wind speeds”, Meteorological Applications, vol. 6, pp. 119-132, 1999.

[7] M. Kunz, S. Mohr, M. Rauthe, R. Lux and Ch. Kottmeier, “Assessment of extreme wind speeds from regional climate models – Part 1: Estimation of return values and their evaluation”, Natural Hazards Earth System Sciences, vol. 10, pp. 907-922, 2010.

[8] C.Morgan, L. Matthew, R.M. Vogel, L.G. Baise and Eugene, “Probability distributions for offshore wind speeds”, Energy Conversion and Management, vol. 52, pp. 15–26, 2011.

[9] Gh.I. El-Shanshoury and A.A. Ramadan, “Estimation of extreme value analysis of wind speed in the North-Western Coast of Egypt”, Arab Journal of Nuclear Science and Applications, vol. 45, pp. 265-274, 2012.

[10] B.H. Lee, D.J. Ahn, H.G. Kim and Y.C. Ha, “An estimation of the extreme wind speed using the Korea wind map”, Renewable Energy, vol. 42, pp. 4-10, 2012.

[11] R. Daneshfaraz, S. Nemati, H. Asadi and M. Menazadeh, “Comparison of four distributions for frequency analysis of wind speed: A case study”, Journal of Civil Engineering and Urbanism, vol. 3, pp. 6-11, 2013.

[12] S.M. Lawan, W.A.W.Z. Abidin, W.Y. Chai, A. Baharun and T. Masri, Statistical modelling of long-term wind speed data, American Journal of Computer Science and Information Technology, vol. 3, pp. 1-6, 2015.

[13] A.R. Rao and K.H. Hamed, “Flood frequency analysis”, CRC Publications, Washington D.C., New York, 2000.

[14] N. Cook, “Comments on plotting positions in extreme value analysis”, Journal of Applied Meteorology and Climatology, vol. 50, pp. 255-266, 2011.

[15] P.E. Charles Annis, “Goodness-of-Fit tests for statistical distributions”, [http://www.statisticalengineering.com/goodness.html], 2009.

[16] United States Water Resources Council (USWRC), “Guidelines for determining flood flow frequency”, Bulletin No. 17B, 1981.