Statistical Distributions of Wind Speed for Baghdad, Basra and Mosul Governorates

Azhar Hadi Mahdi
Building and Construction Engineering Department, University of Technology/Baghdad
Email:azharhadi@yahoo.com

Received on: 15/1/2012 & Accepted on: 6/9/2012

ABSTRACT

The goal of this research is to calculate two basic parameters in Gumbel distribution which is known as prevailing model for quantifying risk associated with extreme wind speed, earthquakes and rainfall. This distribution depends on two basic parameters which are location and scale. To estimate these parameters four methods are applied; method of moments , maximum likelihood method , generalized extreme values and generalized Pareto method , for more than fifty years data of maximum wind speeds of three governorates of Iraq (Baghdad , Basra and Mosul) .A simple mathematical method is used in addition of using new software . Results shows similarity of these parameters (location and scale) as shown in tables and graphs for these methods indicated above and we use more tests depends on tabulated values which is improved using Gumbel distribution for extreme values .Fortran program was applied to calculate the equations and derive the parameters and goodness of fit test for the distribution that mentioned in the text.

التوزيعات الآحصائية لسرعة الرياح في المحافظات بغداد،البصرة والموصل

الخلاصة

يهدف هذا البحث الى حساب عاملين متغيرين اساسيين في توزيع كامبل وكما هو معروف هو موديل سائد لحساب المخاطر التي تصاحب حدوث القيم المتطرفة مثلاً سرعة الرياح ،الزلازل وتساقط المطر. ان هذا التوزيع يعتمد على عاملين متغيرين اساسيين هما المركز والميزان. ولتقدير هذه العوامل تم تطبيق اربعة طرق وهي طريقة العزوم ،طريقة اقصى احتمال، طريقة توليد القيم المتطرفة وطريقة توليد باريتو على بيانات اكثر من خمسين سنة لسرعة الرياح السنوية القصوى في العراق ولثلاث محافظات (بغداد ، البصرة والموصل) . وبأستخدام التحليل الرياضي البسيط بالاضافة الى استخدام البرامج الحديثة . اظهرت النتائج تقارب هذه العومل المتغيرة (المركز والميزان) كما هو موضح في الجداول والرسوم للطرق المذكورة اعلاه كما استخدم اختبارات عدة وبالاعتماد على القيم الجدولية لهذه الدوال الاختبارية والتي اثبتت صحة تطبيق توزيع كامبل للقيم المتطرفة . استخدم برنامج فورتران لحساب المعادلات واشتقاق المعالم واختبار جودة المطابقة للتوزيع المذكور في متن البحث .

https://doi.org/10.30684/etj.31.4A13
2412-0758/University of Technology-Iraq, Baghdad, Iraq

INTRODUCTION

The extreme value distributions have many engineering applications; they are used in forecasting events such as wind speed, flood discharges, earthquake magnitudes…etc. The distributions are usually characterized by scale and location parameters. Various traditional methods such as method of moments and maximum likelihood method are used to determine the parameters from available data recording. The software(13) Data plot was used in this paper supports maximum likelihood estimates for the Gumbel, Frechet (maximum case) and generalized Pareto, also the output maximum likelihood includes the method of moment's estimates for the Gumbel distribution. Table (8) Appendix represents Iraqi wind speed data for three stations Baghdad, Basra and Mosul and Table (1) shows the summary for these data. In probability theory and statistics, the Gumbel distribution (named after Emil Julius Gumbel (3) (1891–1966)) is used to model the distribution of the maximum (or the minimum) of a number of samples of various distributions. It is useful in predicting the chance that an extreme earthquake, flood or other natural disaster will occur. The potential applicability of the Gumbel distribution to represent the distribution of maximum relates to extreme value theory which indicates that it is likely to be useful if the distribution of the underlying sample data is of the normal or exponential type. The Gumbel distribution is a particular case of generalized extreme value distribution (also known as the Fisher-Tippett distribution) and the distribution is also known as the log-Weibull distribution and the double exponential distribution(which is sometimes used to refer to the Laplace distribution)(1).

Type-1 Gumbel density function is:

$$f(x/a,b) = abe^{-(be^{-ax}+ax)} \quad ..(1)$$

Where (a) is scale parameter and (b) is location parameter,

And $-\infty < x < \infty$

The distribution is mainly used in the analysis of extreme values and in survival analysis (also known as duration analysis or event-history modeling).

And Type-2 Gumbel probability density function is

$$f(x/a,b) = abx^{-a-1}\, e^{-bx^{-a}} \quad ...(2)$$

For $0 < x < \infty$

This implies that it is similar to the Weibull distributions, substituting b = λ^{-k} and a = - k. Note however that a positive k (as in the Weibull distribution) would yield a negative a, which is not allowed here as it would yield a negative probability density. For $0 < a \le 1$ the mean is infinite. For $0 < a \le 2$ the variance is infinite.The special case b = 1 yields the Fréchet which is Gumbel type II distribution.

PARAMETERS ESTIMATION

There are various methods for estimating the parameters of a probability distribution, both numerical and graphical. Four methods are applied to the data of wind and the results are demonstrated in Table (2):

1. Method of moments.
2. Maximum likelihood.
3. Generalized extreme values.
4. Generalized Pareto PPCC plot.

This involves the following two steps:

- Determination of the best fitting distribution.
- Estimation of the parameters (shape, location and scales) for that distribution.

Method of moments

In statistics, the method of moments is a method of estimation population parameters such as mean, variance, median, etc. (which need not be moments), by equating sample moments with unobservable population moments and then solving those equations for the quantities to be estimated . It is the most natural method, because a large sample is (hoped to be) a faithful image of the unknown distribution (12). Estimating the parameters (θ_1... θ_k) by the methods of moment**s** consists in equating the (known) moments of the sample with the corresponding (unknown) moments of the distribution:

$m_i = \mu_i$ for $i = 1,2,...,k$ …(3)

This is a set of k equations. If this system can be solved:

$\theta_i^* = g_i(m_1, m_2, ..., m_k)$, $i = 1, 2, ..., k$

Then every θ_i^* is a convergent estimator of the corresponding parameter θ_i.

$E[X^k]$ exist for $K < a$ The moments ….(4)

Where a is the scale parameter.

The cumulative distribution function of the Gumbel distribution is

$$F(x;\mu,\beta) = e^{-e^{(\mu-x)/\beta}} \quad \ldots(5)$$

Where the parameters μ is location (real) and $\beta > 0$ is scale (real)

The mean is $\mu + \gamma\beta$ and γ = Euler–Mascheroni constant ≈ 0.57721566

The standard deviation is $\pi/\sqrt{6} \approx 1.28254983$

Therefore, in order to estimate the parameters mentioned above we can apply the next formulas:
The Scale parameter = Sample Standard Deviation * $\sqrt{6\,/\,\pi}$
And the Location parameter = Sample Mean — 0.57721566 * Scale parameter

Maximum Likelihood Method [2], [3]

There are a number of approaches to estimating the parameters of a statistical distribution from a set of data. Maximum likelihood estimates are popular because they have good statistical properties. The primary drawback is that likelihood equations have to be derived for each specific distribution. Maximum Likelihood Estimator would accomplish this by taking the mean and variance as parameters and finding the specific values for these parameters that produces the distribution most likely to have produced the observed results. In general, for a fixed set of data and underlying probability model, the method of maximum likelihood selects values of the model parameters that produce the distribution most likely to have resulted in the observed data (i.e. the parameters that maximize the likelihood function). Maximum likelihood is the extremism estimator based upon the objective function.

$$\ell(\theta) = E[\ln f(x_i|\theta)] \qquad \text{... (6)}$$

To use the method of maximum likelihood, one first specifie the joint density function for all observations. Let the observed values x_1, x_2… x_n be fixed "parameters" of this function, whereas θ will be the function's variable and allowed to vary freely. For an iid (independent and identically distributed) sample this joint density function will be:

$$f(x_1, x_2, \dots\dots x_n|\theta) = f(x_1|\theta).f(x_2|\theta)\dots\dots f(x_n|\theta) \qquad \text{... (7)}$$

From this point of view this distribution function will be called the likelihood:

$$\mathcal{L}(\theta|x_1, \dots\dots., x_n) = f(x_1, x_2, \dots\dots x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \qquad \text{…(8)}$$

In practice it is often more convenient to work with the logarithm of the likelihood function, called the log-likelihood, or its scaled version, called the average log-likelihood:

$$\ln \mathcal{L}\,(\theta|x_1, \dots\dots., x_n) = \sum_{i=1}^{n} \ln f(x_i\,|\theta) \quad , \qquad \hat{\ell} = \frac{1}{n} \ln \mathcal{L} \qquad \text{….(9)}$$

The hat over ℓ indicates that it is akin to some estimator. Indeed, $\hat{\ell}$ estimates the expected log-likelihood of a single observation in the model. The method of maximum likelihood estimates θ_0 by finding a value of θ that maximizes $\hat{\ell}(\theta|x)$. This method of estimation is a maximum likelihood estimator (MLE) of θ_0:

$$\hat{\theta}_{mle} = arg_{\theta \epsilon \ominus} max(\theta|x_1, \dots\dots., x_n) \qquad \text{…(10)}$$

For many models, a maximum likelihood estimator (MLE) can be found as an explicit function of the observed data $x_1 \ldots x_n$. For many other models, however, no closed-form solution to the maximization problem is known or available, and a MLE has to be found numerically using optimization methods. For some problems, there may be multiple estimates that maximize the likelihood. For other problems, no maximum likelihood estimate exists (meaning that the log-likelihood function increases without attaining the supreme value).

GENERALIZED EXTREME VALUES

In probability theory and statistics, the generalized extreme value (GEV) distribution is a family of continuous probability developed within extreme value theory to combine the Gumbel, Fréchet and Weibull families also known as type I, II and III extreme value distributions(6),(7),(8). By the extreme value theorem the GEV distribution is the limit distribution of properly normalized maxima of a sequence of independent and identically distributed random variables. Because of this, the GEV distribution is used as an approximation to model the maxima of long (finite) sequences of random variables. The generalized extreme value distribution has cumulative distribution function:

$$F(x;\mu,\sigma,\xi) = exp\left\{-1\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad \ldots (11)$$

For $1 + \xi(x - \mu) / \sigma > 0$, where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ the scale parameter and $\xi \in \mathbb{R}$ the shape parameter.
The density function is, consequently

$$f(x;\mu,\sigma,\xi) = \frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{(-1/\xi)-1} exp\left\{-1\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \ldots (12)$$

Again, for $1 + \xi(x - \mu) / \sigma > 0$

A. Gumbel or type I extreme value distribution:

$$F(x;\mu,\sigma) = e^{-e^{(\mu-x)/\sigma}} \qquad for \quad x\epsilon\mathbb{R} \quad \ldots (13)$$

B. Fréchet or type II extreme value distribution:

$$F(x;\mu,\sigma,\alpha) = \begin{cases} 0 & x \le \mu \\ e^{-((x-\mu)/\sigma)^{-\alpha}} & x > \mu \end{cases} \quad ..(14)$$

C. Reversed Weibull or type III extreme value distribution:

$$F(x;\mu,\sigma,\alpha) = \begin{cases} e^{-(-(x-\mu)/\sigma)^{\alpha}} & x < \mu \\ 1 & x \ge \mu \end{cases} \quad \ldots (\mathbf{15})$$

Where $\sigma > 0$ and $\alpha > 0$

Generalized Pareto PPCC Plot (15)

The probability plot correlation Coefficient (PPCC) test (Table (5), Table (6)) has been known as powerful and easy test among the goodness of fit tests. In this study, the derivation of the PPCC test statistics for the generalized Pareto distribution was performed by considering sample sizes, significance levels, and shape parameters. In addition, the correlation coefficients between orderly generated data sets and fitted quintiles were computed by using various plotting position formulas. Among the goodness of fit tests, the Probability Plot Correlation Coefficient (PPCC) test has been known as powerful and easy test. Originally, the PPCC test was developed for normality test (Filliben, 1975). Then, the PPCC test has been applied to various probability distributions. Vogel (1986) derived the PPCC test statistics for the Gumbel distribution, and Vogel and Kroll (1989) applied the PPCC test to the 2-parameters Weibull and uniform distributions for low flow frequency analysis**.** Vogel and McMartin (1991) computed the PPCC test statistics of 5% significance level for gamma distribution, and the PPCC test statistics for the GEV distribution are calculated by Chowdhury et al. (1991).The generalized Pareto distribution is a special case of the Wake by distribution and is defined by the following equation:

$$F(x) = 1 - \frac{[1-(\beta/\alpha)(x-x_0)]^1}{\beta} \qquad \text{... (16)}$$

Where x_0 is a location parameter, α is a scale parameter and ß is a shape parameter.

Then, $x_0 \le x < \infty$ for $\beta \le 0$ and $x_0 < x \le x_0 + \alpha/\beta$ for $\beta > 0$.

The PPCC test statistics are influenced by probability distributions, significance levels, sample sizes, plotting position formulas, and shape parameters in case that a given distribution includes a shape parameter. Recommended plotting position formula (Gringorten (1963)) for Gumbel and weibull distributions is:

$$P_i = (i - 0.44) / (n + 0.12) \qquad \text{…(17)}$$

The PPCC plot can be used to estimate the shape parameter of a distribution with single shape parameter. After finding the best value of the shape parameter, the probability plot can be used to estimate the location and scale parameters of a probability distribution generates a PPCC plot (that is, a probability plot correlation coefficient).Table (3) illustrates the results of shape, location, scale and the PPCC value for the three stations.

Table (1) Summary of Wind speed Data.

Stations	Number of records	Maximum value	Mean	Variance	Standard Deviation

Baghdad	57	43.5	24.92	30.207	5.4961
Basra	57	40	22.175	52.076	7.2164
Mosul	54	32	17.481	46.5	6.8191

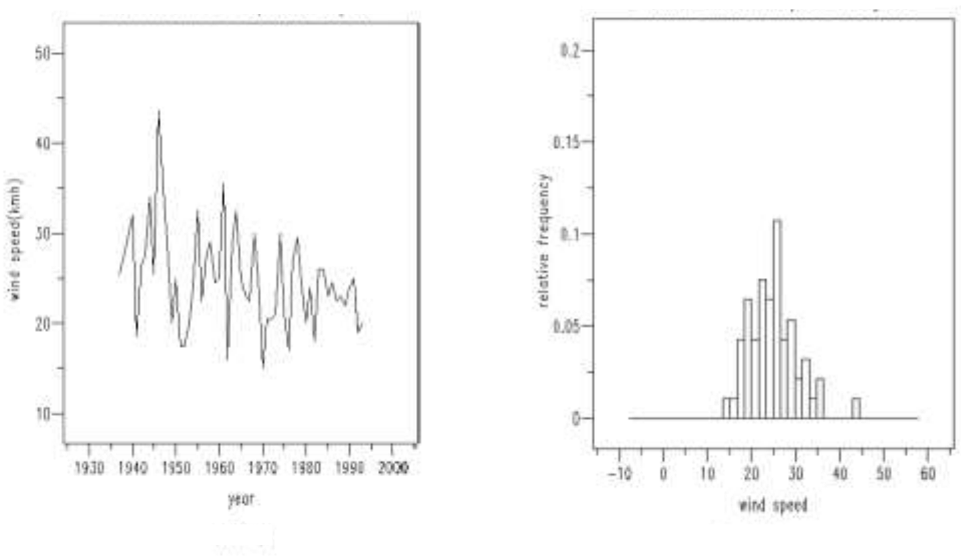


Figure (1) Maximum Annual Wind Speed and relative Histogram for Baghdad Station.

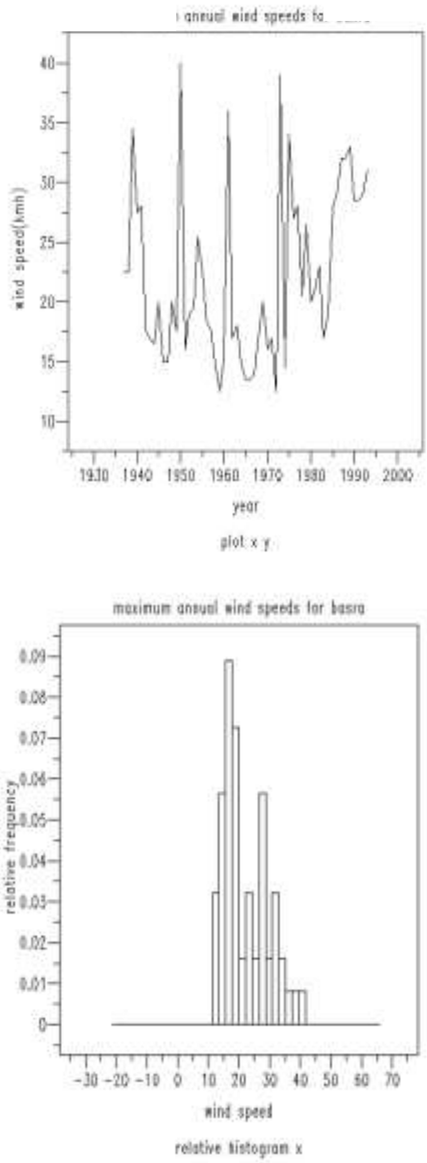


Figure (2) Maximum Annual Wind Speeds and relative histogram for Basra Station.

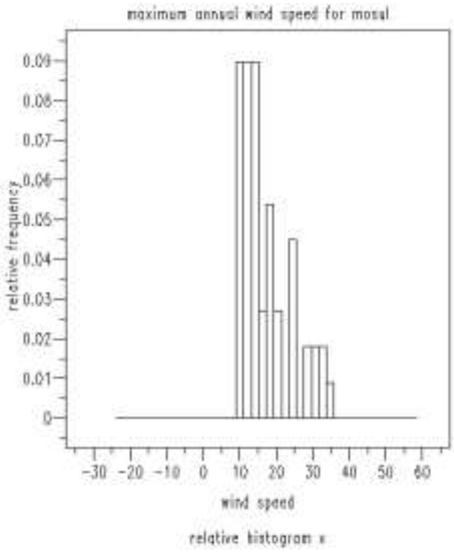


Figure (3) Maximum Annual Wind Speeds and relative histogram for Mosul Station.

KERNEL DENSITY PLOT

Kernel density estimation is a non-parametric way of estimating the probability density function of a random variable. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample. In some fields such as signal processing and econometrics it is also known as the Parzen–Rosenblatt window method, after Emanuel Parzen and Murray

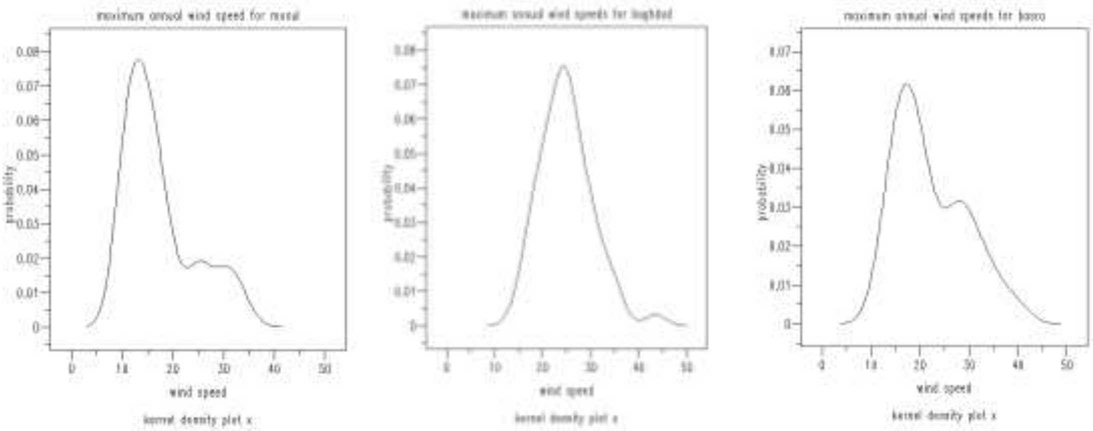
Rosenblatt, who are usually credited with independently creating it in its current form . The kernel density estimate, f(n), of a set of n points from a density f is defined as:

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h\,(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) \qquad \ldots(18)$$

Where K is the kernel function h is the smoothing parameter or window width. Data plot uses a Gaussian kernel function. This down weights points smoothly as the distance from x increases. The width parameter can be set by the user, although Data plot will provide a default width that should produce reasonable results for most data sets. A kernel density plot can be considered a refinement of a histogram or frequency plot. In the exposition above, it is assumed that the data are distributed. Figures (4) show the kernel density plots for the probability of the three set of data.

Table (2) Estimated Parameters

Stations	Method of Maximum Likelihood		Method of Moments for Gumbel I		Method of Moments for Gumbel(II) Frechet		Method of Generalized Extreme Values	
	Location	Scale	Location	Scale	Location	Scale	Location	Scale
Baghdad	22.44	4.40	22.47	4.24	21.86	5.64	22.6	4.75
Basra	18.93	5.79	18.93	5.64	18.03	3.76	18.70	5.66
Mosul	14.23	5.45	14.21	5.36	13.47	3.28	13.82	4.43



Figures (4) Kernel Density plot.

Table (3) Generalized Pareto probability plot.

stations	Shape parameter	Location parameter	Scale parameter	ppcc value
Baghdad	-1.04082	15.84886	18.51229	0.999912
Basra	-0.42857	12.36415	14.09844	0.994061
Mosul	-0.30612	8.635083	11.39913	0.98581

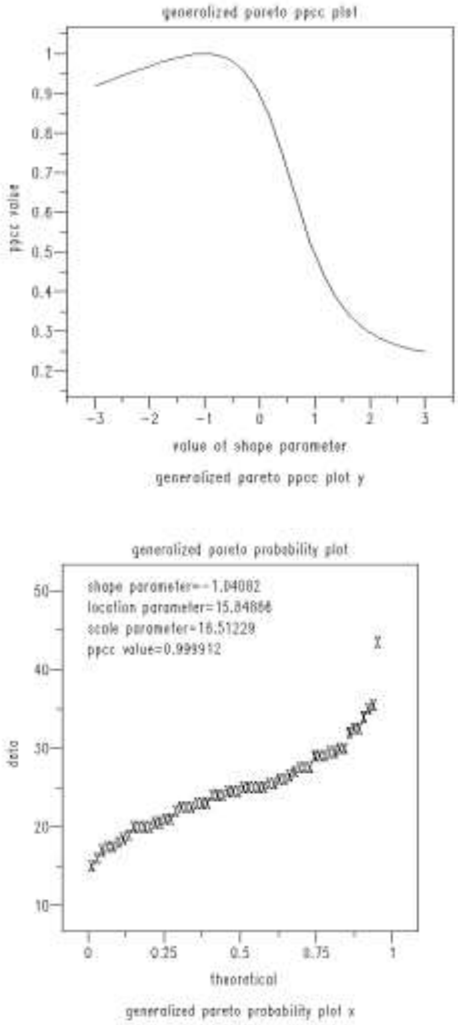
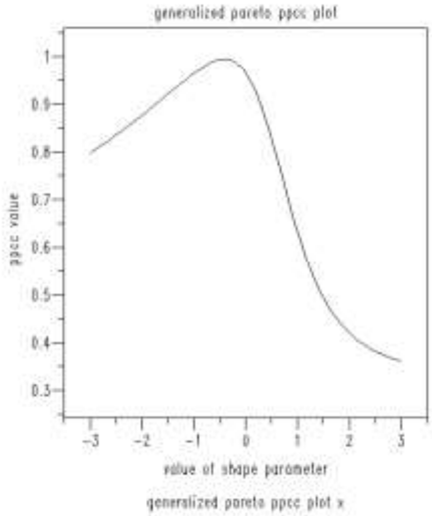


Figure (5) Generalized Pareto PPCC and probability plot for Baghdad Station.



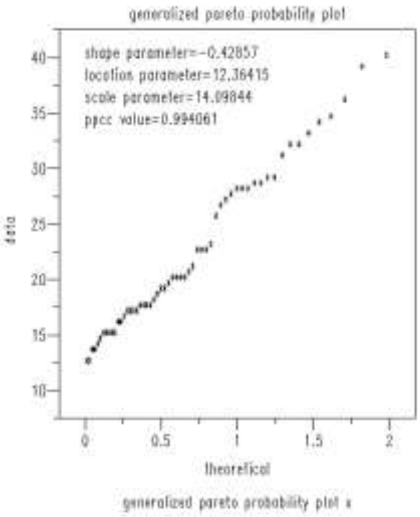


Figure (6) Generalized Pareto PPCC and probability plot for Basra Station.

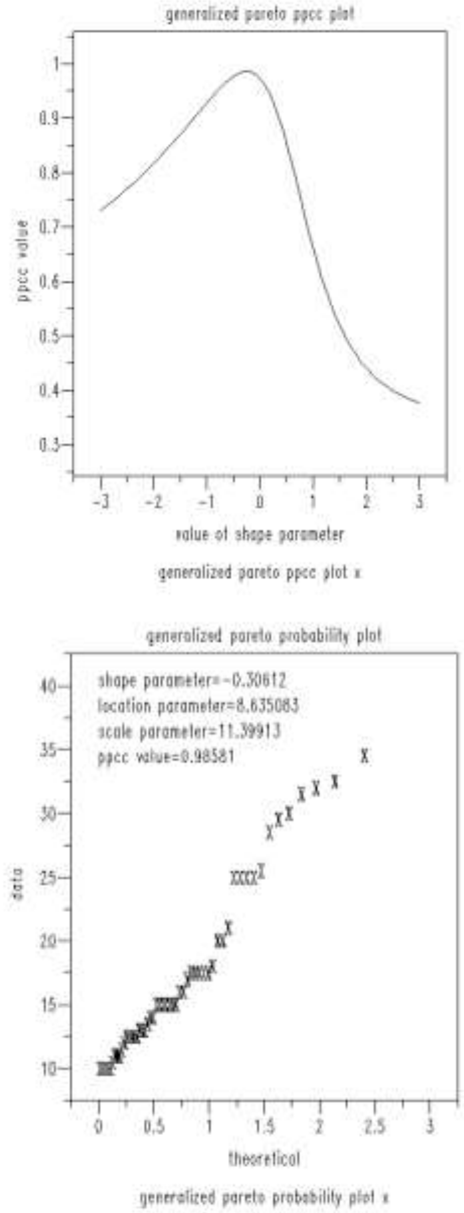
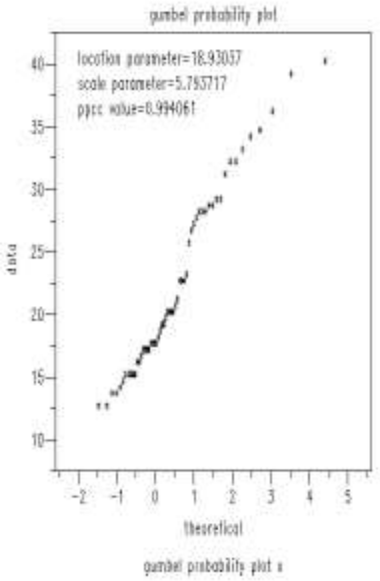
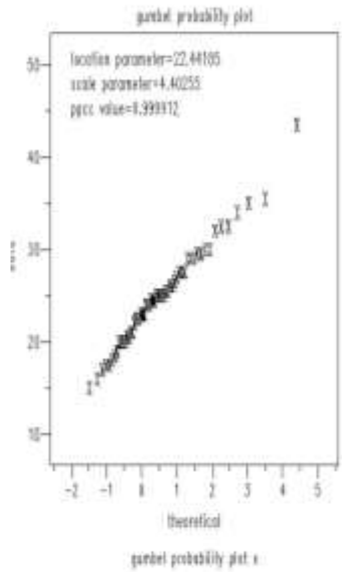


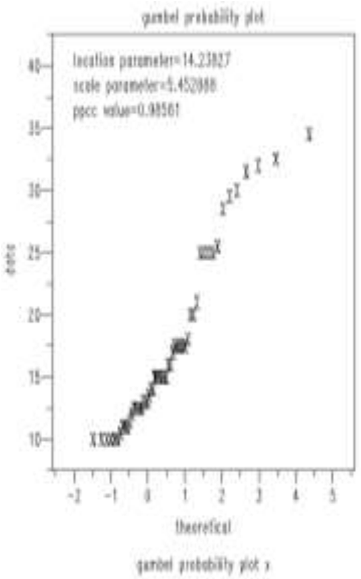
Figure (7) Generalized Pareto PPCC and probability plot for Mosul Station.



Basra



Baghdad



Mosul

Figure (8) Maximum Likelihood Estimation for Gumbel probability plot.

GOODNESS OF FIT TESTS

A statistical hypothesis test is a method of making decisions using data, whether from a controlled experiment or an observational study. This is one of the most useful aspects of statistical inference, since many types of decision-making problems, tests, or experiments in the engineering world can be formulated as hypothesis-testing problems. Three types of test was adapted the analysis to improve that extreme values distribution does fit the data. The Kolmogorov–Smirnov, Anderson-Darling and Chi-square tests (15) are used to test if a sample of data came from a population with a specific distribution. The null and alternative hypothesis is defined as follows:

Zero hypothesis (H_0): The data follows a specified distribution.
Alternative hypothesis (H_1): The data do not follow the specified distribution.

1. The Kolmogorov–Smirnov statistic for a given cumulative distribution function is

$$D_n = sup_x|F_n(x) - F(x)| \quad ..(19)$$

Where $sup_x$the set of distances, the Kolmogorov–Smirnov test is is constructed by using the critical values of the Kolmogorov distribution. The null hypothesis is rejected at level α if $\sqrt{n}D_n > K_\alpha$, Where K_α is found from

$$Pr(K \leq K_\alpha) = 1 - \alpha \quad ....(20)$$

2. The Anderson-Darling test is a modification of the Kolmogorov-Smirnov (K-S) test and gives more weight to the tails than does the K-S test. The Anderson-Darling test statistic is defined as: $A^2 = -N - S$, Where

$$S = \sum_{i=1}^{N} \frac{(2i-1)}{N}[log\, F(Y_i) + log(1 - F(Y_{N+1-i}))] \quad ...(21)$$

Where F is cumulative distribution function of interest, and it is a one-sided test and the hypothesis is rejected if the test statistic, A, is greater than the critical value.

3. The Chi-square test commonly used to compare observed data with data we would expect to obtain according to a specific hypothesis**,** the chi-square test is always testing what scientists call the null hypothesis**,** which states that there is no significant difference between the expected and observed result. For the chi-square goodness-of-fit computation, the data are divided into *k* bins and the test statistic is defined:

$$\chi^2 = \sum_{i=1}^{n} \frac{(Oi - Ei)^2}{Ei} \quad ...(22)$$

Where Oi: Observed Values and Ei: Expected Values.
The expected frequency is calculated by:

$$E_i = N(F(Y_v) - F(Y_l))$$

Where F is the cumulative Distribution function for the distribution being tested, Y_v is the upper limit for class i, Y_l is the lower limit for class **i**, and N is the sample size. The test results for wind speed are demonstrated in Table (4).

Table (4) Testing Statistic

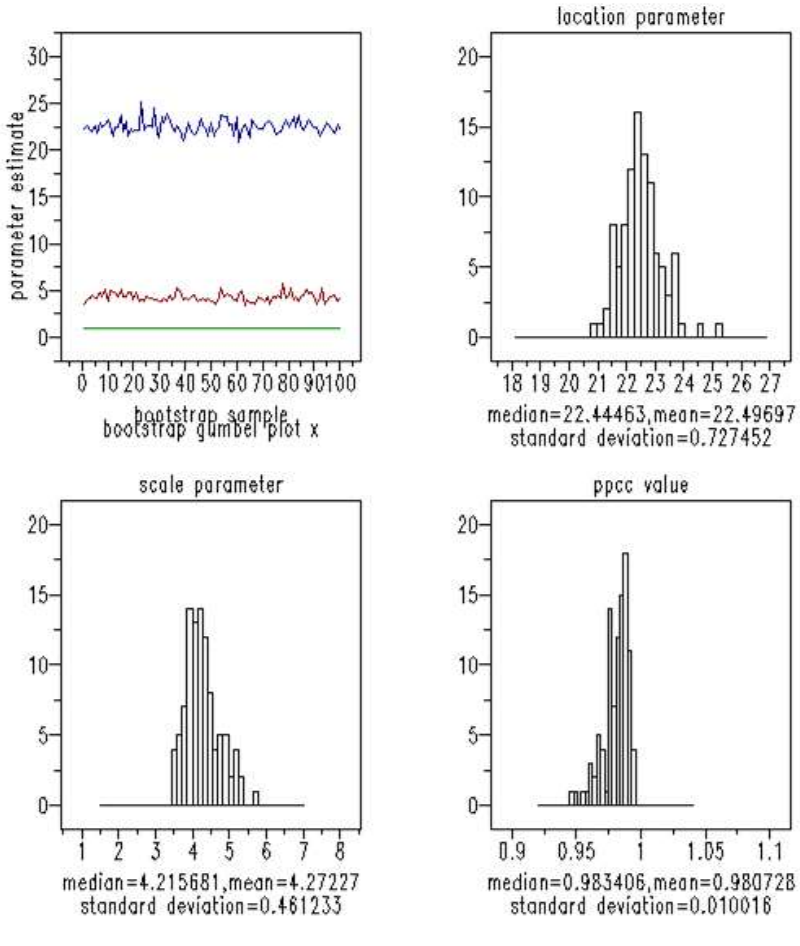
Tests		Kolmogorov Smirnov Statistic	Anderson Darling Statistic	Chi-Squared Statistic
Baghdad Station	Gumbel I	0.09435	0.39098	3.514
	Frechet	0.14993	1.2303	2.0569
	Generalized Extreme	0.7133	0.19545	2.8232
Basra Station	Gumbel I	0.11056	0.89798	4.0685
	Frechet	0.1238	0.65754	7.6352
	Generalized Extreme	0.10339	0.74276	3.4578
Mosul Station	Gumbel I	0.13366	1.1834	1.6034
	Frechet	0.09895	0.51959	1.0019
	Generalized Extreme	0.10555	0.67719	0.80737

Table (5) Kolmogorov-Simirov test for Generalized Pareto PPCC plot

Generalized Pareto	Baghdad station	Basra station	Mosul station
Number of observations	57	57	54
Sample minimum	15	12.5	10
Sample maximum	43	40	34.5
Sample mean	24.921	22.1929	17.3055
Sample standard deviation	5.4468	7.2387	6.8786
test statistic value	0.1140	0.0838	0.1151
CDF value	0.5809	0.2153	0.5583
P-value	0.4191	0.7846	0.4417

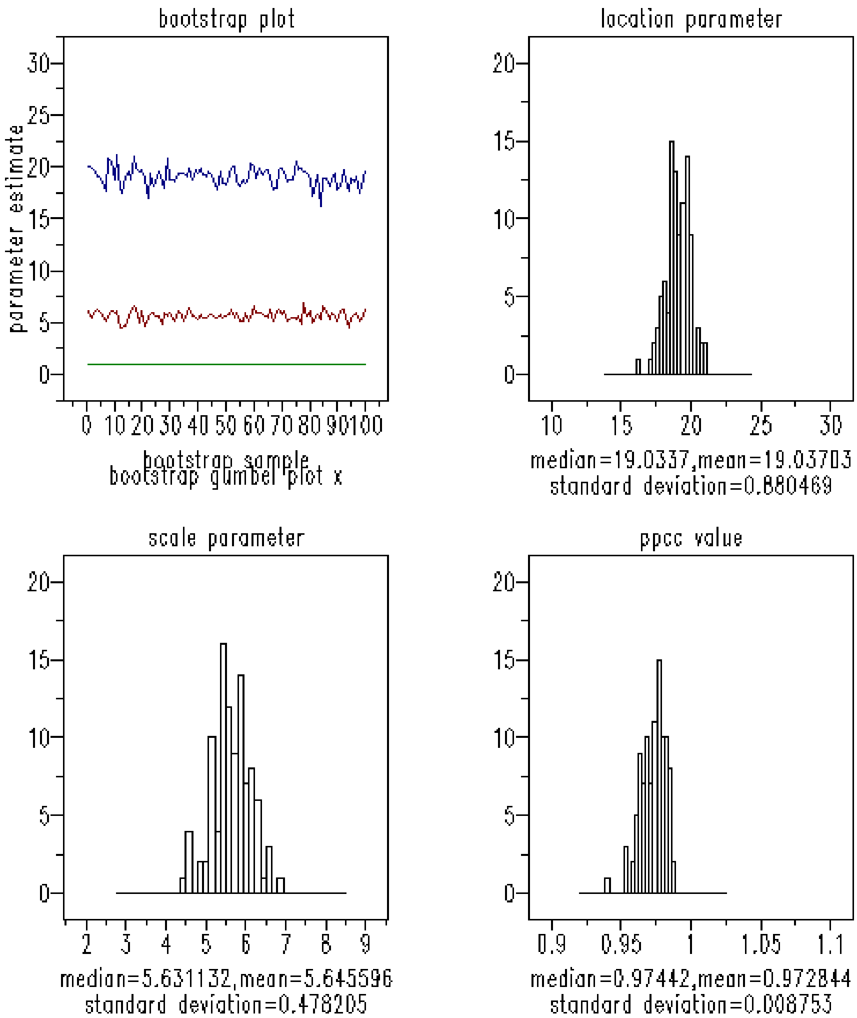
Table (6) fully specified model for Generalized Pareto of the Reference Distribution

Percent Points	Values of Baghdad, Basra stations	Values of Mosul station
0.0	0.000	0.000
50.0	0.107	0.110
75.0	0.133	0.136
90.0	0.159	0.164
95.0	0.177	0.183
97.5	0.192	0.190
99.0	0.213	0.221
99.5	0.230	0.230



Figures (9) Baghdad Station

Bootstrap Gumbel – based Maximum Likelihood. The blue line indicates location parameter; red line indicates scale parameter and the green once indicate PPCC value.
Confidence Interval for Location = (20.5 – 25.0).
Confidence Interval for Scale= (3.5 – 5.9).
Confidence Interval for PPCC value = (0.95 – 1.0). When $\alpha = 0.05$.



Figures (10) Basra station.

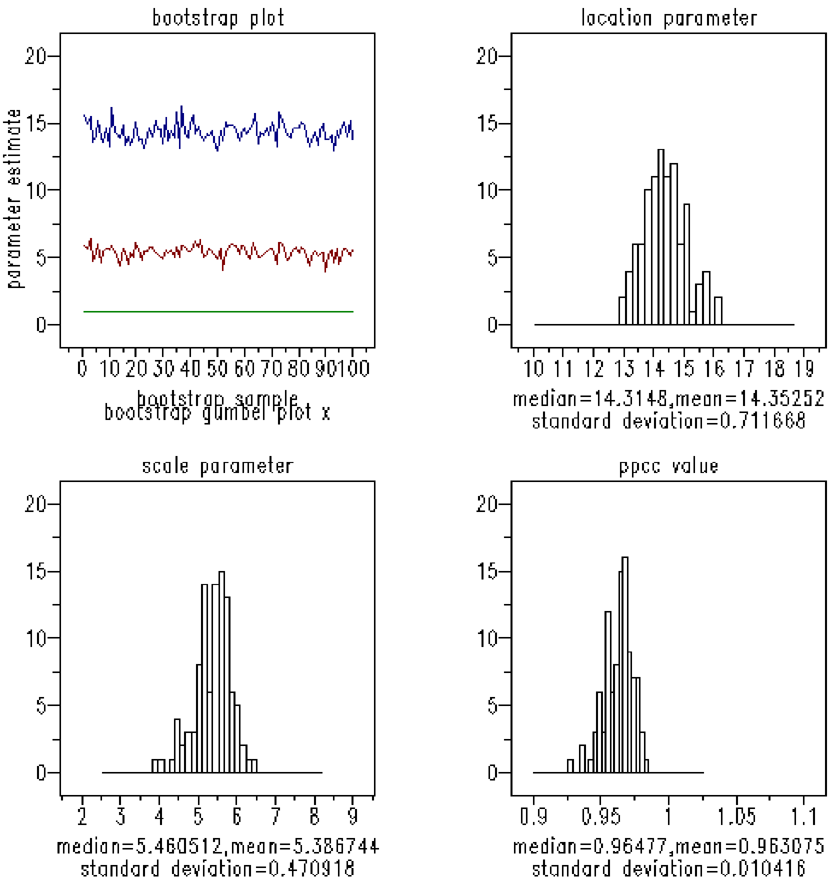
Bootstrap Gumbel – based Maximum Likelihood Confidence Interval for Location and Scale. The blue line indicates location parameter; red line indicates scale parameter and the green once indicate PPCC value.

Confidence Interval for location = (17.0 – 22.0).

Confidence Interval for scale = (4.5 – 7.0).

Confidence Interval for PPCC value = (0.93 – 0.99).

When α = 0.05



Figures (11) Mosul Station

Bootstrap Gumbel – based Maximum Likelihood Confidence Interval for Location and Scale. The blue line indicates location parameter; red line indicates scale parameter and the green once indicate PPCC value.

Confidence Interval for location = (12.7– 16.5).

Confidence Interval for scale = (3.8 – 6.5).

Confidence Interval for PPCC value = (0.92 – 0.97). When α = 0.05.

Table (7) Kolmogorov-Sminov test statistic Conclusions (upper one-tailed test)

α %	CDF %	Baghdad critical values	Basra critical values	Mosul critical values	conclusion
10	90	0.159	0.185	0.164	Accept Ho
5	95	0.177	0.177	0.183	Accept Ho
1	99	0.213	0.213	0.221	Accept Ho

CONCLUSIONS

1. The conclusions based on the plots in figures (1) and (2) shows that the bulk of the annual maximums lie between (15 – 35) (Km/hr.) with probability (0.075) for the mode of Baghdad station, (10 – 25) (Km/hr.) with probability (0.06) for the mode of Basra station and (10 – 20) (Km/hr.) with probability (0.08) for the mode of Mosul station. It seems that all data exhibit slightly skewing to the right.
2. Table (2) shows the convergence in the results of estimated parameters (location, shape) for two distributions (Gumbel I and Gumbel II) by using three methods (moments, maximum likelihood and generalized extreme values) for Baghdad, Basra and Mosul which confirm the validity of the applications.
3. The estimation of the shape parameter in Table (3) is negative values for all stations (-1.04082, -0.42857,-0.30612) which indicates that a reverse Weibull is the appropriate model while the 0 value for the shape parameter indicates that the Gumbel distribution is the appropriate model.
4. The probability plots and the Kolmogorove-Smirnov goodness of fit test indicate that the generalized Pareto provides an adequate distributional model.
5. Conclusion regarding to the distributional model when comparing the values of testing statistic (Kolmogorove-Smirnov, Anderson Darling) in table (4) with the critical values in table (7) there is no evidence from this data to reject the null hypothesis at any level of significance means that data follows Gumbel distribution, also observe occurrence of the test of Chi-Squared Statistic with critical values from statistical tables of Chi-Square (82.27 ,73.29 ,68.78) at levels of significance (1% ,5% ,10%) in the acceptance region .
6. Bootstrap Gumbel – based Maximum Likelihood in figures (9, 10, and 11) estimated values for location and scale parameters occur within the range of the confidence intervals for Baghdad, Basra, and Mosul stations and the Probability Plot Correlation Coefficient PPCC values also occur within the range of confidence intervals.
7. Conclusion based on Probability plot correlation coefficient analyses of sets of maximum yearly wind speed can be also performed to fit reverse Weibull distribution, that in many instances the difference between the respective values of the PPCC will be very small, using the typical steps in developing a reverse Weibull distributional model of the PPCC probability plot approach to estimate the shape, location, and scale parameters. We could obtain confidence intervals for the parameters and for select quintiles using the bootstrap in a similar manner as we did for the Gumbel distribution. Sometimes using the reverse Weibull does not increase the Probability plot correlation coefficient PPCC value. Based on this, we might be inclined to use the simpler Gumbel model.

Appendix

Table (8) (Data of wind speed)

Baghdad station		Basra station		Mosul station	
Year	wind speed	Year	wind speed	Year	wind speed
1937	25.5	1937	22.5	1941	21
1938	27.5	1938	22.5	1942	16
1939	29.5	1939	34.5	1943	11
1940	32	1940	27.5	1944	11
1941	18.5	1941	28	1945	11
1942	26.5	1942	17.5	1946	11.5
1943	27.5	1943	17	1947	10
1944	34	1944	16.5	1948	10.5
1945	25.5	1945	20	1949	10
1946	43.5	1946	15	1950	12.5
1947	35	1947	15	1951	12
1948	29	1948	20	1952	10
1949	20	1949	17.5	1953	15
1950	25	1950	40	1954	17.5
1951	17.5	1951	16	1955	17
1952	17.5	1952	19	1956	17.5
1953	20	1953	19.5	1957	16
1954	24	1954	25.5	1958	14
1955	32.5	1955	22.5	1959	13.5
1956	22.5	1956	18.5	1960	25
1957	27.5	1957	17.5	1961	20
1958	29	1958	15	1962	20
1959	24.5	1959	12.5	1963	25
1960	25	1960	15	1964	17.5
1961	35.5	1961	36	1965	17.5
1962	16	1962	17	1966	17.5
1963	29	1963	18	1967	15
1964	32.5	1964	15	1968	13
1965	25	1965	13.5	1969	12.5
1966	23	1966	13.5	1970	13
1967	22.5	1967	14	1971	12.5
1968	30	1968	17.5	1972	12.5
1969	25	1969	20	1973	15
1970	15	1970	16	1974	13
1971	20.5	1971	17	1975	10
1972	20.5	1972	12.5	1976	15
1973	21	1973	39	1977	14
1974	30	1974	14.5	1978	12.5
1975	21	1975	34	1979	10
1976	17	1976	27	1980	10
1977	27	1977	28	1981	15
1978	29.5	1978	20.5	1982	18
1979	24.5	1979	26.5	1983	15
1980	20	1980	20	1984	15
1981	24	1981	21	1985	25
1982	18	1982	23	1986	34.5
1983	26	1983	17	1987	28.5
1984	26	1984	19	1988	31.5
1985	23	1985	28	1989	30
1986	24.5	1986	29	1990	32.5
1987	22.5	1987	32	1991	25.5
1988	23	1988	32	1992	29.5
1989	22	1989	33	1993	32
1990	24	1990	28.5	1994	25
1991	25	1991	28.5		
1992	19	1992	29		
1993	20	1993	31		

REFERENCES

[1].Willemse, W. J. and Kaas, R., "Rational reconstruction of frailty-based mortality models by a generalization of Gompertz' law of mortality", Insurance: Mathematics and Economics, 40 (3) (2007), 468–484.

[2].Gumbel, E.J. 1954. Statistical theory of extreme values and some practical applications. Applied mathematics series 33. U.S. Department of Commerce, National Bureau of Standards.

[3]. Gumbel, Emil J. (1958), Statistics of Extremes, Columbia University Press, ISBN 0-483-43604-7.

[4].E.Simiua,,I, N.A. Heckertb, J.J. Fillibenb, S.K. Johnson , Extreme wind load estimates based on Gumbel distribution of Dynamic pressures, MD 20899 USA Structural Safety 23 (2001) 221-229.

[5].Demetris K. (2004), Statistics of extremes and estimation of extreme rainfall. National Technical University of Athens, Heroon Polytechneiou 5, GR-157 80 Zographou.

[6].Embrechts, P., Klüppelberg,C, and Mikosch ,T. (1997) Modeling extremely events for insurance and finance. Berlin: Spring Verlag.

[7]. Leadbetter, M.R., Lindgreen, G. and Rootzén, H. (1983). Extremes and related properties random sequences and processes. Springer- Verlag. ISBN 0-387- 90731.

[8].Resnick, S.I. (1987). Extreme values, regular variation and point processes. Springer-Verlag. ISBN 0-387-96481-9.

[9].Rosenblatt, M. (1956). "Remarks on some nonparametric estimates of a density function". Annals of Mathematical Statistics **27**: 832–837.

[10]. Parzen, E. (1962). "On Estimation of a Probability Density Function and Mode". Annals of Mathematical Statistics **33**: 1065–1076.

[11].Stuart. A, Steven F. (1999). Classical Inference and the Linear Model. Kendall's Advanced Theory of Statistics. (Sixth Ed.). London: Arnold. pp. 25.37–25.43. MR1687411. ISBN 0-340-66230-1.

[12].Jack R.Benjamn & Allin C. Cornell (1970) McGraw – Hill Inc. U.S.A ISBN 07-00459-6 (Probability, Statistics, and Decision for Civil Engineers).

[13].James.J.Filliben and Allan Heckert (Data plot version 11/2010) Interactive graphics and Data Analysis Language, Information Technology Laboratory.NIST.

[14].Sooyoung K., Younwoo K., Hongjoon S., Jun-Haeng H. (2008). Ph.D. Candidates, School of Civil and Environmental Engineering, , Derivation of the Probability Plot Correlation Coefficient Test Statistics for the Generalized Logistic and the Generalized Pareto Distributions World Environmental and Water Resources Congress 2008 .

[15].NIST/Sematech-Handbookof Statistical Methods, www.itl.nist.gov/div898/handbook/