ISSN: 0067-2904
GIF: 0.851

Statistical Analysis of Extreme Rainfall Data in Baghdad City

Tasnim H.K. Al-Baldawi*, Zinah Zaid Ali Al-Zuabidi
Department of Mathematics, College of Science, Baghdad University, Baghdad, Iraq

Abstract

Studying extreme precipitation is very important in Iraq. In particular, the last decade witnessed an increasing trend in extreme precipitation as the climate change. Some of which caused a disastrous consequences on social and economic environment in many parts of the country. In this paper a statistical analysis of rainfall data is performed. Annual maximum rainfall data obtained from monthly records for a period of 127 years (1887-2013 inclusive) at Baghdad metrology station have been analyzed. The three distributions chosen to fit the data were Gumbel, Fréchet and the generalized Extreme Value (GEV) distribution. Using the maximum likelihood method, results showed that the GEV distribution was the best followed by Fréchet distribution.

Keywords: Extreme value Distributions, Gumbel Distribution, Fréchet Distribution, QQ plots

التحليل الاحصائي لبيانات القيم المتطرفة للامطار في مدينة بغداد

تسنيم حسن كاظم البلداوي*، زينة زيد علي الزبيدي
قسم الرياضيات، كلية العلوم، جامعة بغداد، بغداد، العراق

الخلاصة

تعد دراسة ظاهرة الامطار الغزيرة من الامور المهمة في العراق. فقد شهد العقد الاخير اتجاها متزايدا في غزارة الامطار كمظهر من مظاهر التغير المناخي ادى البعض منها الى حدوث عواقب كارثية على البيئتين الاجتماعية والاقتصادية في عدة اجزاء من البلاد. في هذا البحث اجرينا تحليلا احصائيا لبيانات الامطار حيث قمنا باستخراج القيم العظمى السنوية لكميات الامطار من السجلات الشهرية للسنوات1887–2013 من محطة بغداد للانواء الجوية لغرض تحليلها. تم اختيار ثلاثة توزيعات وهي كل من توزيع كمبل وتوزيع فريشيه وتوزيع القيمة المتطرفة العام. باستخدام طريقة الامكان الاعظم اظهرت النتائج ان توزيع القيمة المتطرفة العام كان الافضل يليه توزيع فريشيه.

1. Introduction

Throughout the last few decades there has been a considerable evidence of a global climate change. Among others, the global warming phenomena may have a serious impact on hydrology. Urban hydrology will probably be perturbed by global change. In particular, existing classical water drainage infrastructure may no longer work properly. Availability of probabilistic rainfall scenarios should then allow an investigation of the potential impacts of climate change on urban water management strategies. Extreme value theory is the branch of statistics that deals with modeling rare events, so that the prediction of future extreme levels might be expected based on a historical series of observations. The application of the extreme value theory (EVT) becomes widespread and popular since 1980's. One of the main application areas of EVT is the analysis of extreme precipitation events. However, in Iraq very few studies presented dealing with this issue. One of which a study presented by Dawood (2009) on a probability analysis of extreme monthly rainfall data in Mosul city, north of Iraq.

*Email: tasnim@scbaghdad.edu.iq

In addition to Gumbel distribution, he selected the normal, Pearson Type III, Lognormal, and the 3-parameter lognormal distributions. He concluded that all the five distributions are suitable to represent the data [1]. Al Aboodi (2014) presented a somewhat similar probability analysis of extreme monthly rainfall in Baghdad city, the middle of Iraq for the period (1887-1958). He applied the same probability distributions of his predecessor and concluded that the lognormal, the 3-parameter lognormal and Gumbel distributions were accepted by the chi-square test as adequate models for the data [2]. AghaKouchak and Nasrollahi (2010) compared various semi-parametric and parametric extreme value index estimators in order to characterize the tail behavior of long-term daily rainfall time series. The analyses showed that the parametric methods are superior to the semi parametric approaches. In Particular, the maximum likelihood estimators are found to be more robust and consistent for practical applications [3].

In this study we focus only on the results about maximum monthly rainfall records in Baghdad city for the period (1940-2013 inclusive).

2. Parametric modeling of Annual Maxima

Three parametric distributions are considered to represent the extreme annual rainfall data in Baghdad. The selected distributions are chosen such that its limiting distribution is unbounded; that is, their upper endpoint tends to infinity.

2.1. Gumbel Distribution

The Gumbel distribution, named after one of the pioneer scientists in practical applications of the Extreme Value Theory (EVT), the German mathematician Emil Julius Gamble (1891-1966), has been extensively used in various fields including hydrology for modeling extreme events [4].

In probability theory and statistics, the Gumbel distribution is commonly used in practice to model the distribution of the maximum (or the minimum) of a number of samples of various distributions. Such a distribution might be used to represent the distribution of maximum level of a river in a particular year if there was a distribution of the maximum values for the past ten years. It is useful in predicting the chance that an extreme earthquake, flood or other natural disaster will occur [5].

The Grumble distribution has a constant positive skewness and is commonly used for hydrological analyses. The maxima from any distribution that converges on an exponential function at the positive tail (normal, chi-square, lognormal etc.) will have a Gumbel distribution.

The cumulative distribution function of the Grumble distribution is [6]:

$$F(x) = \exp[-exp - \frac{(x-\epsilon)}{\alpha}] \quad (1)$$

Where ϵ is the location parameter and α is the scale parameter.

The probability density function is:

$$f(x) = \alpha^{-1} exp - \frac{(x-\epsilon)}{\alpha} \exp[-exp - \frac{(x-\epsilon)}{\alpha}], \text{ for } x \in R \quad (2)$$

2.2. Fréchet distribution

The Fréchet distribution is a special case of the generalized extreme value distribution. It has the cumulative distribution function [4]:

$$F(x) = \exp(-x)^{-\alpha} \quad , \quad x > 0 \quad (3)$$

Where $\alpha > 0$ is a shape parameter. It can be generalized to include a location parameter μ and a scale parameter $\beta > 0$ with the cumulative distribution function:

$$F(x) = \exp\ [-\left(\frac{x-\mu}{\beta}\right)^{-\alpha}] \quad if \quad x > \mu \quad (4)$$

The Fréchet distribution was named after Maurice Fréchet (1878-1973), a French mathematician who devised one possible limiting distribution for a sequence of maxima, provided convenient scale normalization. In applications to finance, the Fréchet distribution has been of great use apropos to the adequate modeling of market-returns which are often heavy-tailed. This is a positively skewed distribution. If the raw data are a Fréchet distribution then their logarithms will follow a Gamble distribution.

The two parameter fréchet distribution has the cumulative distribution function:

$$F(x) = \exp\left[-\left(\frac{x}{\beta}\right)^{-\alpha}\right] \quad (5)$$

The probability density function is:

$$f(x) = \frac{\alpha}{\beta}(\frac{x}{\beta})^{-\alpha-1} \exp[-\left(\frac{x}{\beta}\right)^{-\alpha}] \quad (6)$$

2.3. Generalized Extreme Value (GEV) distribution

Let Mn = max (Y_1, . . . , Yn) where Yi are i.i.d. from a continuous cumulative distribution function F. Suppose we can find normalizing constants $a_n > 0$ and b_n such that

$$P\left(\frac{M_n - b_n}{a_n} \leq y\right) \to G(y)$$

As $n \to \infty$, where G is some proper distribution function. Then G is necessarily one of three possible types of limiting distribution functions which have been called the Gumbel type, Fr´echet type and Weibull type. Nowadays, it is realized that it is more convenient to consider the generalized extreme value (GEV) distribution, which holds the three types as special cases. It has a cumulative distribution function of the form [7]:

$$G(y) = \exp\{-\left[1 + \frac{\xi(y-\mu)}{\sigma}\right]_+^{-\frac{1}{\xi}}\} \quad (7)$$

Where μ is the location parameter, σ is the scale parameter and ξ is the shape parameter, , and $h_+ = \max(h, 0)$.

Differentiation of (7) gives the probability density function of the GEV distribution [7]:

$$g(y) = \frac{1}{\sigma}[1 + \xi\left(\frac{y-\mu}{\sigma}\right)]^{-(\frac{1}{\xi}+1)} \exp\{-\left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\} \quad (8)$$

The (GEV) distribution is very flexible and is the distribution that is recommended for use in many studies concerning extreme events.

3. Maximum Likelihood Estimation

There are many different methods used in estimating the parameters of probabilistic models. Some of these methods are simple and preliminary but is less efficient, others are more complicated but more efficient. We will use the maximum likelihood method in estimating the parameters of the selected models because it is one of the best methods with high efficiency. It is one of the most widely used methods of statistical estimation has become generally accepted as the most robust procedure. Numerical maximum likelihood estimation requires maximization of the log–likelihood function or equivalently minimization of its negation. As there is no analytical solution for obtaining the maximum likelihood estimates of the parameters for the models considered, we need to maximize the log–likelihood numerically. We do this using EasyFit Software.

3.1. Gumbel MLE [8]

The likelihood function of the Gumbel model is given by

$$L(\alpha, \epsilon) = (\frac{1}{\alpha})^n \, exp \, \{-\sum_{i=1}^{n}[\frac{(x_i - \epsilon)}{\alpha} + exp\left\{-\frac{(x_i - \epsilon)}{\alpha}\right\}]\}$$

The negative log-likelihood function is

$$-\ln(L) = n ln(\alpha) + \sum_{i=1}^{n} \frac{x_i - \epsilon}{\alpha} + \sum_{i=1}^{n} exp - [\frac{x_i - \epsilon}{\alpha}] \quad (9)$$

3.2. Fréchet MLE [9]

The likelihood function of the Fréchet model is given by

$$L(\alpha, \beta) = \frac{\alpha^n}{\beta^n} \frac{\beta^{n(\alpha+1)}}{\prod_{i=1}^{n}(x_i)^{\alpha+1}} \, exp - \sum_{i=1}^{n}\left(\frac{\beta}{x_i}\right)^{\alpha}$$

The negative log-likelihood function is

$$-lnL = -n ln\, \alpha - n\alpha ln\, \beta + (\alpha + 1)\sum_{i=1}^{n} \ln x_i + \sum_{i=1}^{n}\left(\frac{\beta}{x_i}\right)^{\alpha} \quad (10)$$

3.3. GEV MLE [10]

The likelihood function of the GEV model is given by

$$L(\mu, \sigma, \xi) = (\frac{1}{\sigma})^n \prod_{i=1}^{n}[1 + \xi\left(\frac{y_i - \mu}{\sigma}\right)]^{-(\frac{1}{\xi}+1)} exp \sum_{i=1}^{n}\{-\left[1 + \xi\left(\frac{y_i - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\}$$

The negative log-likelihood function is

$$-\ln(L) = n log\sigma + \left(1 + \frac{1}{\xi}\right)\sum_{i=1}^{n} \log\left[1 + \xi\left(\frac{y_i - \mu}{\sigma}\right)\right] + \sum_{i=1}^{n}[1 + \xi\left(\frac{y_i - \mu}{\sigma}\right)]^{-1/\xi} \quad (11)$$

4. Empirical Results and Analysis

4.1. Fitting Annual Maxima

The data utilized for the analysis represents maximum annual rainfall totals in Baghdad city obtained from monthly records for a period of 127 years (1887-2013 inclusive). Data shown in appendix are provided by Iraqi Metrological Organization and Seismology. Tables-1,-2 show some summary statistics of the data.

Table 1- Descriptive Statistics

Statistic	Value
Sample Size	114
Range	194.7
Mean	54.677
Variance	1136.4
Std. Deviation	33.71
Coef. of Variation	0.61654
Std. Error	3.1573
Skewness	2.0997
Excess Kurtosis	5.3862

Table 2- Percentile Values

Percentile	Value
Min	13.5
5%	19.6
10%	25.2
25% (Q1)	32.175
50% (Median)	44.95
75% (Q3)	63.825
90%	90.15
95%	148.48
Max	208.2

In practical applications, an appropriate estimator may be selected according to goodness of fit tests rather than on theoretical considerations. Therefore, an important step is to test whether the resulting model fits the observations. Table-3 lists the maximum likelihood estimators along with the goodness of fit statistics for each model. According to the goodness of fit statistics the three models seems to be accepted.

Table 3- Fitting Results and Goodness of Fit Statistics

#	Distribution	Parameters	Kolmogorov-Smirnov		Anderson-Darling		Chi-Squared	
			Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gumbel	σ=26.284 μ=39.506	0.09993	3	2.5094	3	14.059	3
2	Fréchet	α=2.3515 β=36.752	0.08785	2	1.1648	2	1.7425	2
3	GEV	k=0.24305 σ=17.934 μ =38.718	0.03958	1	0.21964	1	1.532	1

4.2. QQ plots

Given an ordered sample of independent observations $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ from a population with estimated distribution function $\hat{F}$, a quantile plot consists of the points

$$\left\{\left(\hat{F}^{-1}\left(\frac{i}{n+1}\right), x_{(i)}\right) : i = 1, \ldots, n\right\}$$

If $\hat{F}$ is a reasonable estimate of F, then the quantile plot should also consist of points close to the unit diagonal.

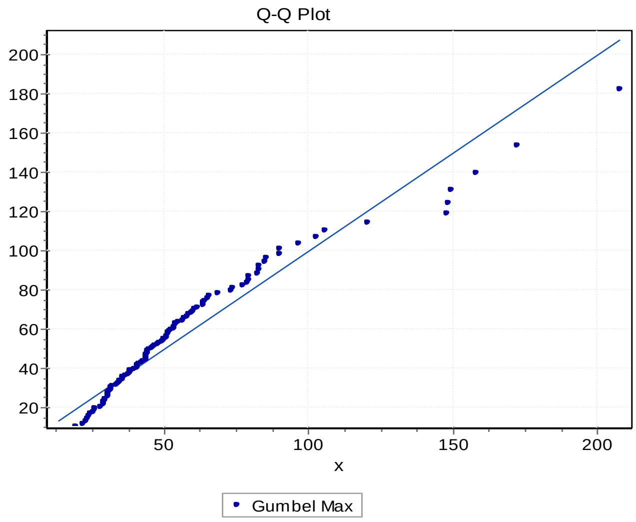

Figure 1- A quantile – quantile plot for the fitted Gumbel model to annual maximum rainfall data in Baghdad

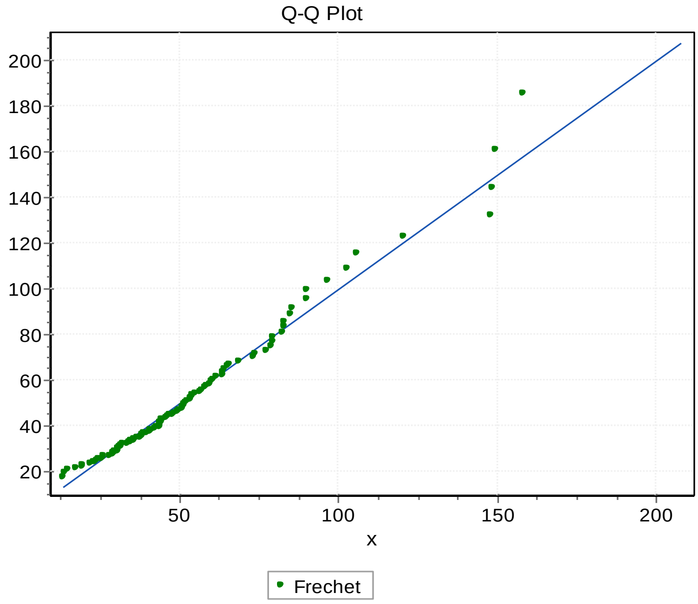

Figure 2- A quantile – quantile plot for the fitted Fréchet model to annual maximum rainfall data in Baghdad

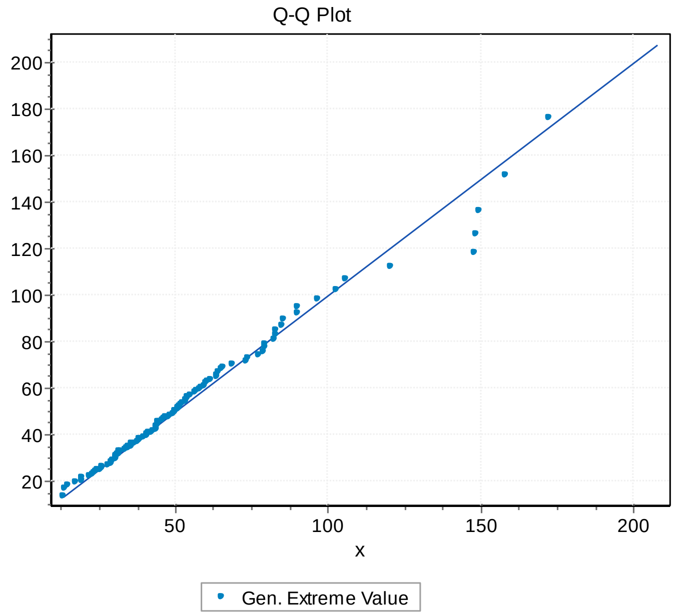

Figure 3- A quantile – quantile plot for the fitted GEV model to annual maximum rainfall data in Baghdad

As shown in Figures-1, -2 and =3 the data approximately lined up on the quantile plots, which indicates it was a good fit.

5. Results and Conclusions

1. According to the QQ plots given in Figures-1, - 2, and -3, it is clear that the GEV distribution is fitted better than the Gumbel and the Fréchet distributions.
2. Conclusions based on the goodness of fit tests (Kolmogorov Smirnov, Anderson Darling, and Chi-Squared) shows that the GEV distribution is fitted better followed by the Fréchet distribution. And hence, we can conclude that the shape parameter ξ is actually different from zero.

References

1. Dawood, A., S. **2009**. Probability Analysis of Extreme Monthly Rainfall in Mosul City, North of Iraq. *Marsh Bulletin*, 4(1), pp:60-74.
2. Ali H. Al Aboodi. **2014**. Probability Analysis of Extreme Monthly Rainfall in Baghdad City, Middle of Iraq. *Basra Journal for Engineering Sciences*, 14 (1).
3. Amir AghaKouchak and Nasrin Nasrollahi **2010**. Semi-parametric and Parametric Inference of Extreme Value Models for Rainfall Data, *Water Resour Manage* 24, pp:1229–1249.
4. Gumbel, E. J. **1958**. *Statistics of Extremes*, Columbia University Press, New York.

5. 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*. ASIN B0007DSHG4.
6. Hosking, J. R. M. **1990**. The theory of probability weighted moments, Research Report RC12210, IBM Thomas J. Watson Research Center, New York.
7. Beirlant, J, Goegebeur, Y., Segers, J., and Teugels, J.**2005**. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons Ltd. p:490.
8. Davison, A. C., and Smith R. L. **1990**. Models for exceedances over high threshold (with discussion), *J. R. Stat. Soc. B* 52(3), pp: 393-442.
9. Marwa, A., A., Hegazy Zaher and Elsherpieny, E.A. **2013**. Optimum Group Limits for Maximum Likelihood Estimation of the exponentiated Fréchet Distribution Based on Grouped Data. *British Journal of Applied Science & Technology* 3(4), pp: 1464-1480.
10. Coles, S. **2001**. *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

Appendix: Annual rainfall totals (mm) in Baghdad city

Year	Max.(mm)	Year	Max.(mm)	year	Max.(mm)
1887	41.7	1930	83.3	1973	53.7
1888	59.9	1931	M	1974	148.7
1889	55.1	1932	M	1975	79.8
1890	149.9	1933	M	1976	29.2
1891	90.2	1934	M	1977	44.2
1892	17.8	1935	M	1978	44.2
1893	65.8	1936	M	1979	38.5
1894	208.2	1937	M	1980	44.5
1895	41.1	1938	97.1	1981	32.1
1896	120.9	1939	36	1982	29.9
1897	58.7	1940	51	1983	14.3
1898	31.8	1941	41.1	1984	49.6
1899	29.5	1942	22.5	1985	34
1900	42.9	1943	31.3	1986	45.1
1901	13.5	1944	47.2	1987	19.6
1902	51.8	1945	54.5	1988	50.2
1903	26.2	1946	51.6	1989	56.7
1904	29.7	1947	52.4	1990	36.1
1905	51.1	1948	24.8	1991	M
1906	43.9	1949	63.8	1992	25.6
1907	105.9	1950	38.4	1993	102.9
1908	35.1	1951	85.5	1994	41.3
1909	19.6	1952	26.3	1995	48
1910	36.8	1953	34.5	1996	40.2
1911	61.7	1954	90.1	1997	44
1912	30.7	1955	54.1	1998	42.4
1913	46.7	1956	44.8	1999	30.8
1914	M	1957	77.8	2000	29.6
1915	M	1958	74.2	2001	23.5
1916	M	1959	73.8	2002	38.4
1917	46	1960	28.2	2003	M
1918	68.8	1961	57.2	2004	M
1919	82.8	1962	48.7	2005	60.6
1920	65.5	1963	44	2006	52.7
1921	85.3	1964	44.4	2007	32.2
1922	50	1965	58.3	2008	23.7
1923	64.5	1966	31.1	2009	15.1
1924	53.8	1967	38.9	2010	32
1925	36.1	1968	148.4	2011	31
1926	158.2	1969	34.9	2012	83.2
1927	24.4	1970	78.9	2013	172.7
1928	61	1971	63.9		
1929	37.8	1972	79.5		