PROBABILITY ANALYSIS OF EXTREME MONTHLY RAINFALL IN BAGHDAD CITY, MIDDLE OF IRAQ

Ali H. Al Aboodi
Civil Eng. Department, Eng. College, Basrah University, Basrah, Iraq

Abstract

Data sets of highest monthly rainfall for the period (1887-1958) are used for evaluating the proper theoretical statistical distribution of extreme monthly rainfall in Baghdad city. The frequency analyses and most statistical test were done using a commercial version of HYFRAN. Five distributions are used in this research, which are: - Normal, Pearson Type III, Lognormal, 3-parameter lognormal and Gumbel. Estimation of theoretical distribution is achieved by using maximum likelihood method and adequacy test is carried out using chi-square test. Lognormal, 3-parameter lognormal, and Gumbel distributions seem to be suitable for representing of maximum monthly rainfall in the study area.

Keywords: Probability, Hyfran, Extreme, Rainfall, Baghdad City

التحليل الاحتمالي لقيم الأمطار الشهرية العظمى في مدينة بغداد وسط العراق

علي حسن العبودي
قسم الهندسة المدنية, كلية الهندسة, جامعة البصرة, البصرة, العراق

الخلاصة

استخدمت قيم الأمطار الشهرية العظمى للفترة (1887-1958) لتقييم التوزيع الاحتمالي النظري الموائم في مدينة بغداد. انجزت التحليلات الترددية وجميع الاختبارات الإحصائية باستخدام النسخة التجارية لبرنامج (HYFRAN). اختبرت خمس توزيعات في هذا البحث هي الطبيعي, وبيرسون النوع الثالث, اللوغارتيمي الطبيعي, اللوغارتيمي الطبيعي ذي الثلاث معالم, وكامبل. تم تخمين التوزيع النظري باستعمال طريقة الاحتمال الأكبر إضافة إلى اختبار الموائمة باستخدام مربع كاي. التوزيع اللوغارتيمي الطبيعي, اللوغارتيمي الطبيعي ذي الثلاث معالم, وكامبل, هم التوزيعات المثلى لقيم الأمطار الشهرية العظمى في منطقة الدراسة.

Introduction

In many applications, the extreme values of a process are more importance. Such as, rainfall, runoff, material strength, pollutant concentrations and insurance claim sizes are all examples of processes of this type. The probability analyses of extreme value have played an important role in engineering practice for water resources design and management. The recent developments in the statistical theory of extreme values can be applied to improve the rigor of hydrologic applications and to make such analyses more physically meaningful. Hydrologic processes in nature are governed by the laws of chance, the use of probability theory and mathematical statistics is unavoidable in the extraction of information from hydrologic data and for the best mathematical description of these processes [1].

Rainfall frequency analyses are often used to aid the design of many hydraulic structures, such as design of storm sewer network [2]. These studies provide the necessary information for the development of a design storm, which represent the probability of occurrence of heavy rainfall, and are used in the design of hydraulic structures [3].

Although the fundamental probabilistic theory of extreme values has been well developed for a long time [4], but the statistical modeling of extremes remains a subject of active research [5]. Several researchers have provided useful applications of extreme value distributions to rainfall data from different regions of the world, such as Bahkar et al., (2006) observed the frequency analysis of consecutive days peaked rainfall at Banswara, Rajasthan, India, and found gamma distribution as the best fit as compared by other distribution and tested by Chi-square value [6]. Hanson et al., (2008) analysis indicated that Pearson type III distribution fits the full record of daily precipitation data and Kappa distribution best describes the observed distribution of wet-day daily rainfall [7]. Olofintoye et al., (2009) examined that 50% of the total station number in Nigeria follows log-Pearson type III distribution for peak daily rainfall, while 40% and 10% of the total station follows Pearson type III and log-Gumbel distribution respectively [8].

Data has been adopted in this research for the long-term from 1887 until 1958 in order to estimate all the climatic changes taking place during this period; this reflects the reality of the climate in Baghdad city. This research focuses on the extreme values of rainfall distribution and taking a long time gives the perception for the nature of the distributions of extreme values. In the present paper, a statistical modeling approach is achieved based on Maximum Likelihood (ML) to estimate the probability analysis of extreme monthly rainfall in Baghdad city. This study is useful for providing the basic information of storm sewer network design and hydraulic structures.

Location of Study Area

Baghdad is situated in the middle of Iraq; this city also lies on the Tigris River at its closest point to the Euphrates, (40 km) to the west. The Diyala River joins the Tigris just southeast of the city and borders its eastern suburbs. (see Fig.1). The terrain surrounding Baghdad is a flat alluvial plain (34 meters) above sea level. Historically, the city has been inundated by periodic floods.The city is located on a vast plain bisected by the Tigris River. The Tigris splits Baghdad in half, with the Eastern half being called 'Risafa' and the Western half known as 'Karkh'. The land on which the city is built is almost entirely flat and low-lying, being of alluvial origin due to the periodic large floods which have occurred on the river.

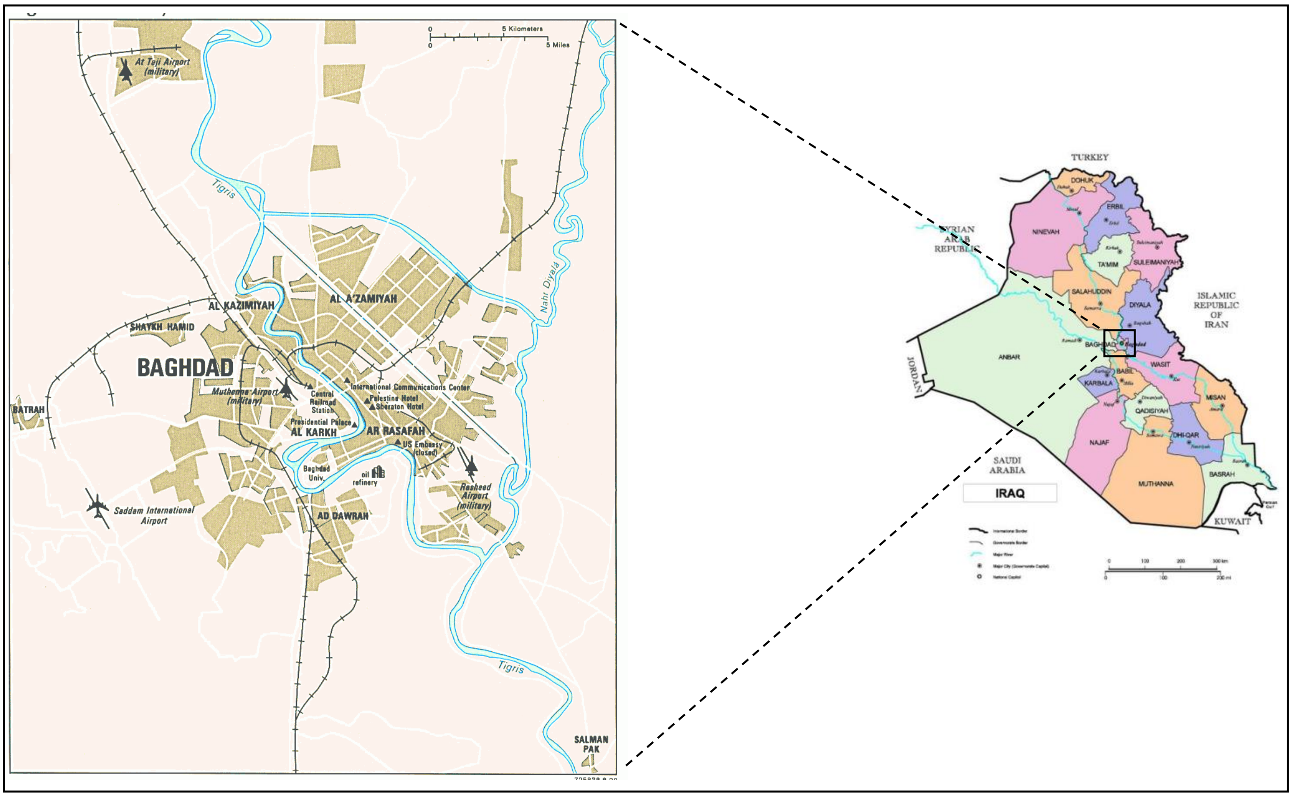

Figure (1) Location of study area in reference to map of Iraq.

Climate of Study Area

Baghdad has a hot arid climate and according to the maximum temperatures, one of the hottest cities in the world. In the summer season from June to August, the average maximum temperature is as high as 44 °C accompanied by blazing sunshine, and even at night temperatures in summer are seldom below 24 °C. Though the humidity is very low (usually under 10%) due to Baghdad has long distance from the indicated by a dash (-) as the period (1931-1935), while, no rainfall which is indicated by a zero (0.0). A "trace" is indicated by (T) and is defined as rainfall of less than measurable quantity. If the record for any month is incomplete, it is indicated as a missing data. In general, rainfall observations have not been made during the summer months, when normally there is no rainfall. Rainfall during the months of May and October is usually such a very small percentage of annual rainfall. The metrological station is located on altitude line (33^0 $20'$) and longitude line (44^0 $24'$), and it has elevation (34.1 m) above sea level.

Extreme Value Technique

The basic theory of extreme value is extreme value theory. If the random variables (X1, X_2 …) whose common cumulative distribution function is F, i.e.

$F(x) = \Pr \{X_i \le x$ ...(1)

Also let $M_n = \max (x_1, \ldots, x_n)$ denote the n^{th} sample maximum of the process. Then

$P_r \{M_n \le x\} = F(x)^n$ ...(2)

Equation (1) is of no immediate interest, since it simply says that for any fixed (x) for which F(x) <1, we have $P_r \{M_n \le x\} \rightarrow 0$

For non-trivial limit results we must renormalize: find $a_n > 0$, b_n
Such that

$$\Pr\left\{\frac{M_n - b_n}{a_n} \le x\right\} = F\left(a_n x + b_n\right)^n \rightarrow H(x)$$
…(3)

Arabian Gulf, dust storms from the deserts to the west are a normal occurrence during the summer. In the winter season, from December to February, by contrast, Baghdad has maximum temperatures averaging 15 to 16 °C. Minima can indeed be very cold, the minimum average occur at January is around 4 °C, but temperatures below 0 °C are not uncommon during this season [9].

Annual rainfall, almost entirely confined to the period from November to March, averages around 140 millimeters, but has been as high as 575 millimeters and as low as 23 millimeters. In the summer season rainfall is almost completely unavailable at this time of year [9].

History and Description of Precipitation Data Collection

All the data of precipitation was obtained from the report of hydrological survey of Iraq [10]. Rain gages installed by or for the Directorate of Meteorology are British Meteorological office standard non-automatic type with either eight or five inch diameter rims. These gages consist of a funnel which drains into a collector bottle inside a cylindrical over flow can. The precipitation data as shown in appendix (A) is presented as monthly totals in millimeters. A missing recorded data is derived rigorously by Gnedenko (1943) [12], asserts that if

$\xi > 0$ to the Fre'chet distribution with $\alpha = 1/\xi$, $\xi < 0$ to the Weibull distribution with $\alpha = -1/\xi$

$$\xi > 0 : \text{"}long-tailed\text{"}\ case,$$
$$1 - F(x) \propto x^{-1/\xi}$$
$$\xi = 0 : \text{"}exponential-tail\text{"}$$
$$\xi < 0 : \text{"}short-tailed\text{"}\ case,$$
$$finite\ endpoint\ at\ \mu - \xi/\varphi$$

Method of Research

HYFRAN (HYdrological FRequency ANalyses) program (commercial version) [13] is used in this research. The frequency analyses and most statistical test were done by HYFRAN; this program is developed by the chair in Stastical Hydrology, INRS-EAU. The data entered into spreadsheet of the program, HYFRAN automatically calculates the values of the basic statistics as shown in table (1), and plots the non-exccedance probability and histogram of observation data (see figures. 2 &3).

Empirical non-exccedance probability is calculated using Weibul formula:

$$F[x[k]] = [k-a]/[n-2a+1]$$
$$0 < a \le 0.5$$

Where, k=rank and n= number of observations, and for Weibul a=0, therefore

The three type theorem originally stated without detailed mathematical proof by Fisher and Tippett (1928) [11], and later a nondegenerate (H) exist (i.e., a distribution function which does not put all its mass at a single point), it must be one of three types:-

$$H(x) = \exp(-e^{-x}), \qquad \text{all } x, \ldots (4)$$

$$H(x) = \begin{cases} 0 & x < 0 \\ \exp(-x^{-\alpha}), & x \ge 0 \end{cases} \quad \ldots (5)$$

$$H(x) = \begin{cases} \exp(-/x/^{\alpha}), & x < 0 \\ 1, & x \ge 0 \end{cases} \quad \ldots (6)$$

Here two distribution functions H_1 and H_2 are said to be of the same type if one can be derived from the other through a simple location-scale transformation,

$$H_1(x) = H_2(Ax + B), \quad A > 0 \ \ldots(7)$$

Very often, (4) is called the Gumbel type, (5) the Fre'chet type and (6) the Weibull type. In (5) and (6), $\alpha > 0$

The three types may be combined into a single Generalized Extreme Value (GEV) distribution:-

$$H(x) = \exp\left\{-\left[1 + \xi \frac{x-\mu}{\varphi})\right]_+^{-1/\xi}\right\}, \ \ldots(8)$$

(y_+ = max(y, 0) where μ is a location parameter, $\varphi > 0$ is a scale parameter and ξ is a shape parameter. The limit $\xi \to 0$ corresponds to the Gumbel distribution,

$$F[x[k]] = P_a = k/n+1$$

to 8). Maximum likelihood method is applied to estimate theoretical distribution parameters (table 2). Adequacy test is achieved by using Chi-square test, these values is compared with tabulated values of (0.01) significant level (table 3). According to this test, Lognormal, 3-parameter lognormal, and Gumbel distributions are the proper for describing of extreme monthly rainfall in the study area.

The return period T (in years) is the reciprocal of P_a, or in mathematical notation

$$T= 1/ P_a$$

Five distributions are selected for representing the probability analyses of extreme rainfall in Baghdad city, which are: Normal, Pearson Type III, Lognormal, 3-parameter lognormal and Gumbel and using (99%) confidence interval (figures from 4

Table (1) Basic statistics of data set

Basic statistics	Value	Number of data
Minimum	14	630
Maximum	210	
Average	58	
Standard deviation	35	
Median	51	
Coefficient of variation (Cv)	0.6	
Skewness coefficient (Cs)	2.0	
Kurtosis coefficient (Ck)	8.0	

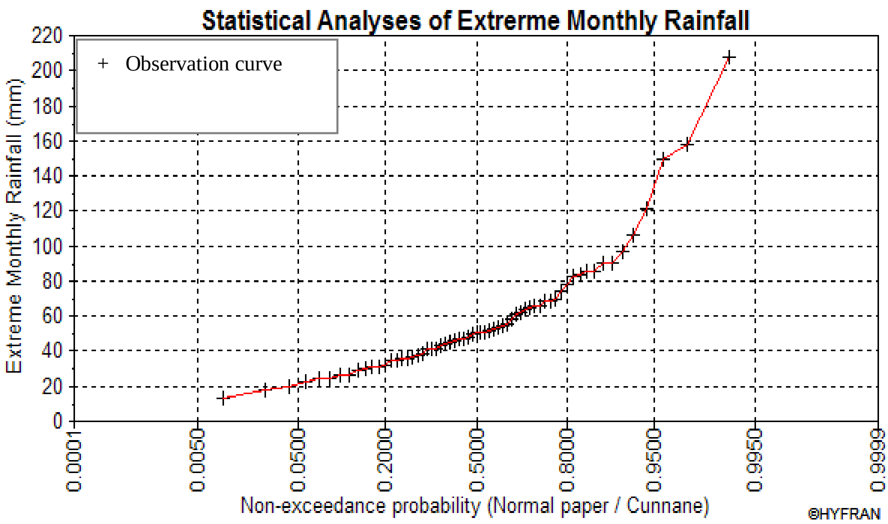

Fig.(2) Observation of dataset on probability paper

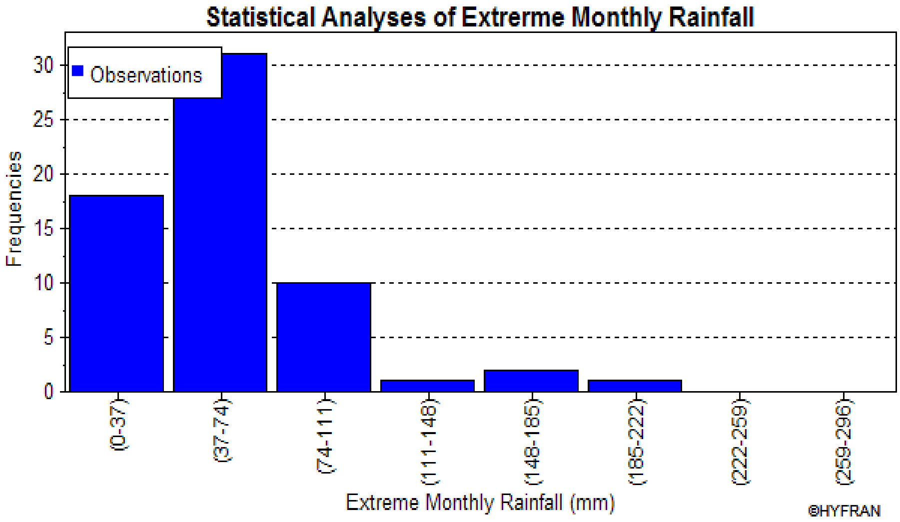

Fig.(3) Histogram of observation dataset

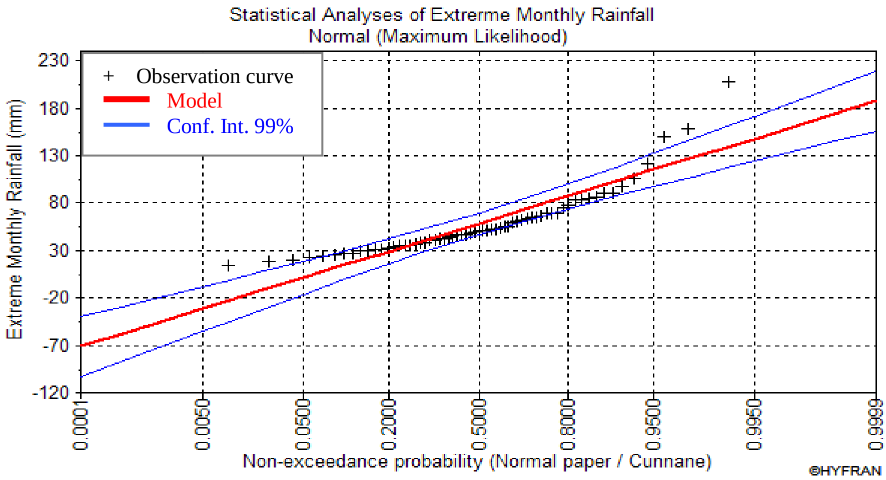

Fig.(4) Empirical, theoretical probabilities and confidence interval against extreme monthly rainfall (normal distribution)

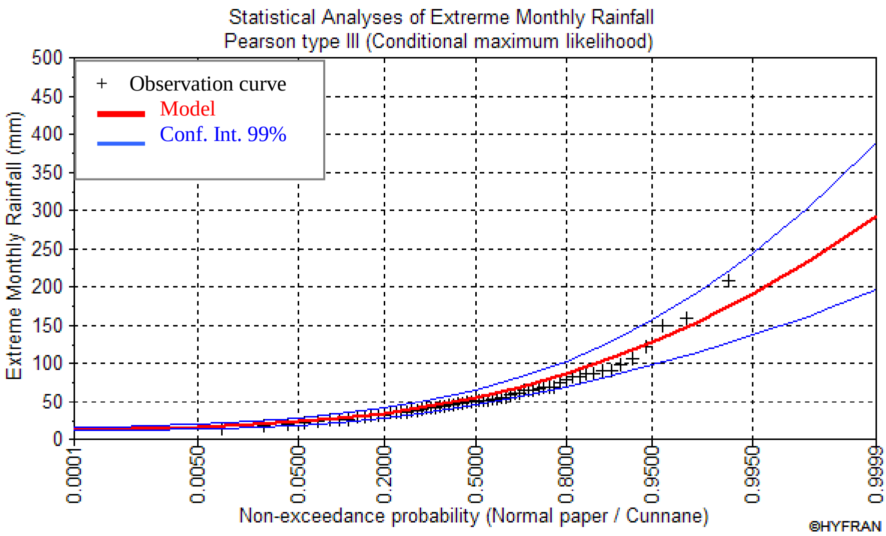

Fig.(5) Empirical, theoretical probabilities and confidence interval against extreme

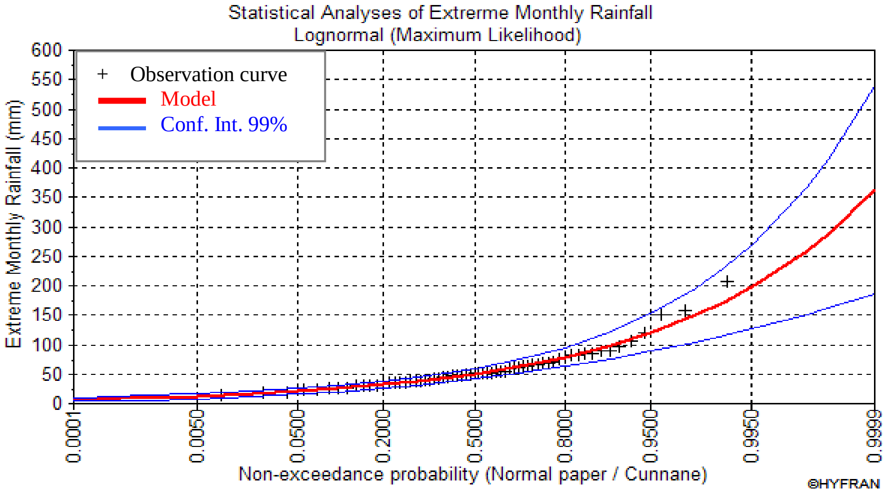

Fig.(6) Empirical, theoretical probabilities and confidence interval against extreme monthly rainfall (lognormal distribution)

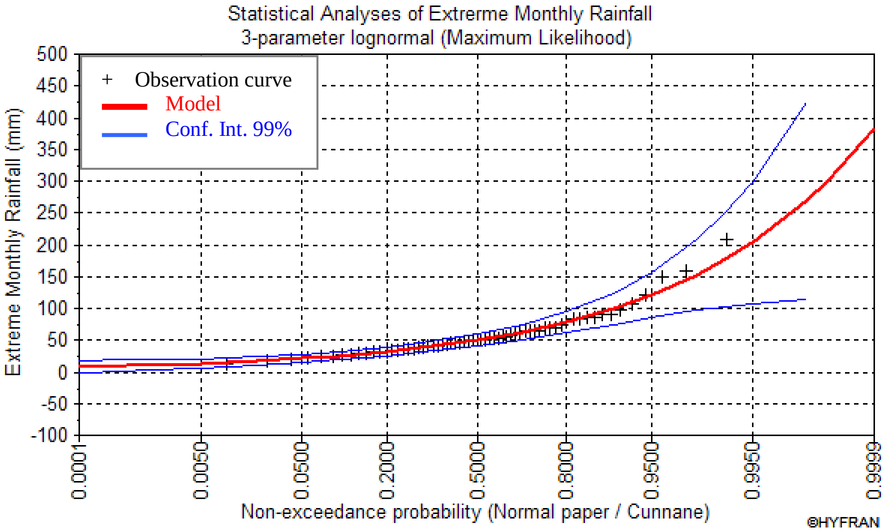

Fig.(7) Empirical, theoretical probabilities and confidence interval against extreme monthly rainfall (3-parameters lognormal distribution)

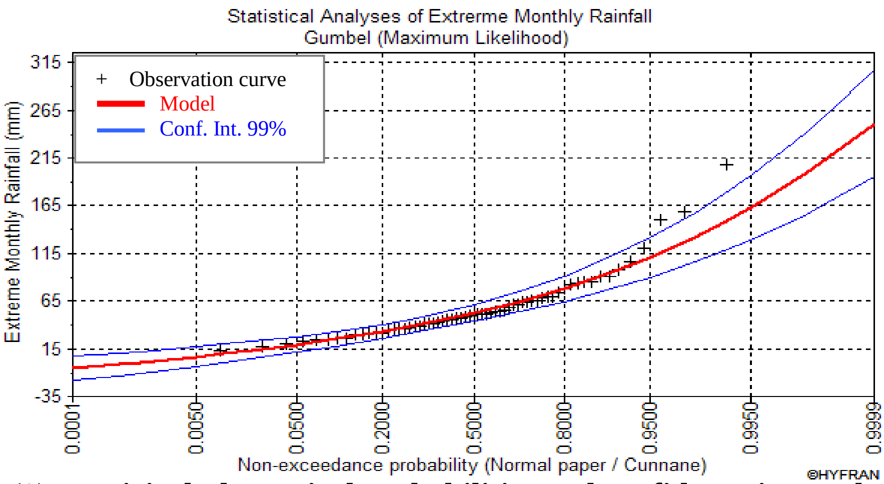

Fig.(8) Empirical, theoretical probabilities and confidence interval against extreme monthly rainfall (Gumbel distribution)

Table (2) Fitting models with estimated parameters

No.	Distribution	Mathematical formula	Estimated parameters
1.	Normal	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$	$\mu=57.9952$ $\sigma=34.727$
2.	Pearson type III	$f(x)=\frac{\alpha^{\lambda}}{\Gamma(\lambda)}(x-m)^{\lambda-1}e^{-\alpha(x-m)}$	$\alpha=0.0429606$ $\lambda=2.08366$ $m=13.5$
3.	Lognormal	$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{-\frac{[\ln x-\mu]^2}{2\sigma^2}\right\}$	$\mu=3.91756$ $\sigma=0.53062$
4.	3-parameters lognormal	$f(x)=\frac{1}{(x-m)\sigma}\exp\left\{-\frac{[\ln(x-m)-\mu]^2}{2\sigma^2}\right\}$	$\mu=3.85364$ $\sigma=0.560561$ $m=2.69114$
5.	Gumbal	$f(x)=\frac{1}{\alpha}\exp\left\{-\frac{x-u}{\alpha}\exp\left(-\frac{x-u}{\alpha}\right)\right\}$	$\alpha=22.4302$ $u=43.9002$

Table (3) Adequacy test using Chi-square

No.	Distribution	Chi-Square test	
		Tabulated value (0.01) significant level	Calculated value
1.	Normal	13.277	<u>18.43</u>
2.	Pearson type III	11.344	<u>35.89</u>
3.	Lognormal	13.277	1.29
4.	3-parameters lognormal	11.344	2.24
5.	Gumbal	13.277	2.56

Conclusions

For representing the probability analyses of extreme rainfall in Baghdad city, five distributions are selected which are: Normal, Pearson Type III, Lognormal, 3-parameter lognormal and Gumbal. Maximum likelihood method is applied to estimate theoretical distribution parameters. According to the comparison between the values that obtained from adequacy test (Chi-square values) with tabulated values of (0.01) significant level, Lognormal, 3-parameter lognormal, and Gumbel distributions seem to be suitable for representing of extreme monthly rainfall in Baghdad city.

References

[1]Huff, F.A., and J.R. Angel, (1992), Rainfall frequency atlas of the Midwest. Champaign, Illinois: Illinois Water Survey.

[2]Faiers, G.E., B.D. Keim, and K.K. Hirschboeck, (1999), a synoptic evaluation of frequencies and intensities of extreme three- and 24-hour rainfall in Louisiana. Professional Geographer 46(2):156-163.

[3]Nguyen, V.T., Nguyen T.D., and Ashkar F., (2002), Regional frequency analysis of extreme rainfalls. Water Science and Technology 45(2):75-81.

[4]Lead better M., Lindren G, Rootz H. Extremes and related properties of random sequences and processes. New York: Springer; 1983.

[5]Coles S. An introduction to statistical modeling of extreme values.
London: Springer; 2001.

[6]Bhakar, S.R., Bansal, A.N., Chhajed, N., and Purohit, R.C. Frequency analysis of consecutive day’s maximum rainfall at Banswara, Rajasthan, India. ARPN Journal of Engineering and Applied Sciences. 2006;1(3): 64-67.

[7]Hanson, L.S., and Vogel, R. The probability distribution of daily rainfall in the United States. Proc. In World Environment and Water Resources Congress Conference. 2008.

[7]Olofintoye, O.O., Sule, B.F., and Salami, A.W. Best-fit Probability distribution model for peak daily rainfall of selected Cities in Nigeria. New York Science Journal. 2009;2(3).

[8]Funding seismographic & meteorological commission.

[9]Summary of monthly precipitation at stations in Iraq for period (1887-1958), (1964), Republic of Iraq, Developing Board, Baghdad, Iraq, 62 p.

[10]Fisher, R.A. and Tippett, L.H.C. (1928), Limiting forms of the frequency distributions of the largest or smallest member of a sample. Proc. Camb. Phil. Soc. 24, 180-190.

[11]Gnedenko, B.V. (1943), Surla distribution limite du terme maximum d'une serie aleatoire. Ann. Math. 44, 423-453.

[12]HYFRAN Hydrological frequency analysis, primarily developed by Bernard Bobe´e

Appendix (A) Monthly meteorological data of Baghdad station for period (1887-1958)

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1887	-	-	-	-	-	-	-	-	-	0.0	0.3	41.7
1888	1.8	59.9	20.8	68.6	16.8	1.0	0.0	0.0	0.0	0.0	29.7	13.0
1889	-	-	-	-	-	0.0	0.0	0.0	0.0	0.0	0.0	55.1
1890	19.0	149.9	128.8	72.9	0.0	0.0	0.0	26.9	0.0	0.0	2.5	114.6
1891	41.1	53.8	15.2	12.7	19.6	0.0	0.0	0.0	0.0	1.3	41.1	90.2
1892	17.8	11.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.1	0.0
1893	41.9	19.0	19.0	5.1	0.0	0.0	0.0	0.0	0.0	1.5	3.3	65.8
1894	37.8	208.2	111.5	63.0	0.8	0.3	0.0	0.0	0.0	3.1	122.9	26.2
1895	41.1	15.0	1.8	6.6	7.6	0.0	0.0	0.0	0.0	3.3	33.9	37.3
1896	120.9	7.9	92.5	12.5	1.3	0.0	0.0	0.0	0.0	0.0	2.5	4.8
1897	19.0	58.7	16.5	11.4	6.9	0.0	0.0	0.0	0.0	0.0	18.8	37.1
1898	31.8	31.0	30.0	7.9	13.0	0.0	0.0	0.0	0.0	0.0	17.8	28.7
1899	6.6	3.6	14.0	8.6	1.3	0.0	0.0	0.0	0.0	2.5	27.4	29.5
1900	10.6	29.5	26.2	0.0	0.0	0.5	0.0	0.0	0.0	1.3	42.9	35.6
1901	13.5	0.0	7.1	5.6	4.3	0.0	0.0	0.0	0.5	0.0	1.3	5.1
1902	4.8	9.6	45.0	47.5	0.0	0.0	0.0	0.0	0.0	12.2	51.8	12.5
1903	18.3	26.2	11.7	11.4	0.0	0.0	0.0	0.0	0.0	0.0	5.1	5.3
1904	24.9	3.8	20.8	19.8	24.9	0.0	0.0	0.0	0.0	12.7	0.8	29.7
1905	12.2	6.6	51.1	3.6	0.8	0.0	0.0	0.0	0.5	0.0	1.3	5.8
1906	22.6	7.9	9.7	14.0	0.8	0.0	0.0	0.0	0.0	0.0	43.9	18.8
1907	29.0	24.4	105.9	60.5	20.1	0.0	0.0	0.0	0.0	6.6	4.1	4.3
1908	35.1	16.3	16.5	6.4	2.8	0.0	0.0	0.0	0.0	0.0	0.0	12.2
1909	1.5	17.8	7.1	8.4	3.6	0.0	0.0	0.0	0.0	6.4	6.4	19.6
1910	33.5	13.0	36.8	5.6	5.1	0.0	0.0	0.0	0.0	0.0	20.1	25.1
1911	61.7	14.5	60.2	24.1	5.3	0.0	0.0	0.0	0.0	0.0	18.3	36.8
1912	26.4	13.2	18.8	2.0	-	0.0	0.0	0.0	0.0	3.6	0.8	30.7
1913	46.7	23.6	13.5	3.3	1.5	0.0	0.0	0.0	0.0	0.0	16.5	27.4
1914	34.8	48.3	18.5	53.1	3.8	0.0	0.0	0.0	-	-	-	-
1915	-	-	-	-	-	-	-	-	-	-	-	-
1916	-	-	-	-	-	-	-	-	-	-	-	-
1917	-	0.0	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	46.0
1918	23.1	36.6	21.8	68.8	4.6	0.0	0.0	0.0	-	0.8	62.0	33.5
1919	82.8	29.0	3.1	65.5	1.8	2.3	0.0	0.0	0.0	0.0	0.0	29.7
1920	1.5	38.1	65.5	1.8	2.3	0.0	0.0	0.0	4.3	9.1	10.0	51.6
1921	43.2	5.8	3.3	5.3	29.2	0.0	0.0	0.0	0.0	0.0	85.3	46.0
1922	21.6	38.9	50.0	2.5	4.6	0.0	0.0	0.0	0.0	0.0	1.8	33.3
1923	8.6	64.5	4.1	13.7	0.3	0.5	0.0	0.0	0.0	0.8	0.5	23.4
1924	53.8	15.5	24.6	0.3	5.6	0.0	0.0	0.0	0.0	5.1	0.3	45.2
1925	5.3	36.1	12.2	16.0	0.3	0.0	0.0	0.0	0.0	17.8	16.5	4.3
1926	66.3	53.8	41.7	15.2	0.0	1.0	0.0	0.0	0.0	0.5	158.2	23.9
1927	1.8	12.2	0.0	7.4	17.0	0.0	0.0	0.0	0.0	0.0	24.4	6.9
1928	22.6	45.7	1.3	2.3	0.0	0.0	0.0	0.0	0.0	0.0	61.0	1.8
1929	1.5	14.2	0.5	17.8	11.4	0.0	0.0	0.0	0.0	5.3	3.3	37.8
1930	83.3	20.3	0.0	7.6	2.0	0.0	0.0	0.0	0.0	0.0	12.5	41.4
1931	-	-	-	-	-	-	-	-	-	-	-	-
1932	-	-	-	-	-	-	-	-	-	-	-	-
1933	-	-	-	-	-	-	-	-	-	-	-	-
1934	-	-	-	-	-	-	-	-	-	-	-	-
1935	-	-	-	-	-	-	-	-	-	-	-	-
1936	-	-	-	-	-	-	-	-	-	-	-	-
1937	-	-	-	18.2	6.5	-	-	-	-	21.2	35.4	T

1938	71.0	37.9	69.6	6.1	0.6	-	T	-	T	T	34.3	97.1
1939	36.0	28.1	16.3	25.0	T	T	-	-	T	0.2	30.3	8.9
Continued Appendix (A)												
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1940	51.0	34.4	9.2	10.7	0.2	-	-	T	-	9.0	0.9	11.4
1941	0.6	41.1	12.7	32.0	-	T	-	-	-	4.6	T	38.6
1942	4.2	11.8	16.4	T	T	T	-	-	-	8.0	22.5	19.6
1943	15.1	23.0	31.3	7.5	T	-	-	-	-	3.2	0.3	4.7
1944	13.6	2.0	23.5	11.3	3.3	-	-	-	-	T	47.2	19.5
1945	54.5	14.1	1.1	2.7	0.9	-	-	-	-	0.9	23.2	40.3
1946	30.2	22.1	50.9	8.2	2.4	-	-	-	-	0.1	9.2	24.0
1947	16.2	11.2	52.4	0.7	8.4	-	-	-	T	0.1	33.4	23.8
1948	5.6	20.8	20.6	12.4	0.1	-	-	-	-	T	1.4	24.8
1949	12.7	31.9	54.9	8.2	0.2	-	-	-	-	-	0.1	63.8
1950	8.9	18.8	13.1	0.3	29.2	-	T	-	-	1.9	38.4	10.2
1951	37.2	85.5	45.5	1.1	0.2	-	-	-	1.3	T	33.8	20.9
1952	7.9	10.6	5.0	26.3	0.8	-	-	-	-	-	0.8	20.9
1953	5.4	34.5	14.8	2.6	T	-	-	-	-	1.9	29.0	8.9
1954	13.8	44.4	69.5	21.6	T	-	-	-	-	T	10.8	27.0
1955	90.1	54.1	11.3	15.1	31.7	13.0	2.5	-	-	-	12.5	27.5
1956	7.5	8.5	26.1	44.8	0.0	0.0	0.0	0.0	T	0.0	T	4.6
1957	12.1	51.5	58.9	72.9	28.5	T	0.0	0.0	0.0	2.8	77.8	31.5
1958	74.2	6.8	3.2	T	2.0	0.0	0.0	0.0	0.0	T	20.3	19.2