Marsh Bulletin 4(1)(2009) 60-74

MARSH BULLETIN

Amaricf_Basra office@yahoo.com
abdulalwan@yahoo.com
.marshbulletin@yahoo.com

Probability Analysis of Extreme Monthly Rainfall In Mosul City, North of Iraq

A.S. Dawood
Civil Eng. Dept., Eng. Coll., Basrah Univ., Basrah, Iraq

Abstract

This study is an attempt at evaluating the proper theoretical statistical distribution of extreme monthly rainfall in Mosul. That is why data for the period (1923-1958) have been made use of to have all relevant information about sets of the highest monthly rainfall then. The frequency analyses and all needed statistical tests were done using the final version 1.1 of HYFRAN Software that operate under windows environment. The following Five distributions are used: Normal, Pearson Type III, Lognormal, 3-parameter lognormal and Gumbal. We arrived at the estimation of the theoretical distribution by using maximum likelihood method. The adequacy test is carried out by applying the chi-square test. All five distributions have been found to be suitable for representing of extreme monthly rainfall in the area under investigation.

1-Introduction

Rainfall analysis for any region is essential in planning and design of irrigation and drainage systems and overall programme of command area development. As the distribution of rainfall varies over space and time, it is required to analyze the data covering long periods and recorded at various locations to obtain reliable information. Further, these data needs to be analyzed in different ways depending on the problem under consideration. For example, analysis of consecutive days rainfall is more relevant for drainage design of agricultural lands, whereas analysis of weekly rainfall data is more useful for planning cropping pattern as well as water management practices than that of monthly, seasonal and annual data. The analysis of rainfall data for computation of expected rainfall of a given frequency is commonly done by utilizing different probability distributions. Different distribution functions commonly used are Gumbel, lognormal, Pearson type III, log Pearson type III, and Gamma distribution. Since no single distribution can be described as the best, it is required to compare the suitability of different probability distributions.

With the possibility of including additional rainfall data using computer programs, a statistical analysis is possible for many locations. The output of a hydrologic system is treated as stochastic, space-independent, and time-independent when there is no correlation between adjacent observations. Random variables are statistically described by a probability distribution. The probability of an event is the chance that it will occur based on an observation of the random variable. The larger the sample size of random variables the better the estimate of the probability of the event. To determine the probability of events occurring, probability distribution functions are fit to the data to determine the appropriate function to use for estimation of the events. There are two methods for fitting distributions to data: method of moments and method of maximum likelihood. Karl Pearson developed the method of moments in 1902. Pearson's method of moments selects probability function parameters such that the moments are equal to those of the sample data. The first moment for each observation is the sample mean, the second is the variance, and the third is the coefficient of skewness. R.A. Fisher developed the method of maximum likelihood in 1922. Fisher's method for determining the parameter of a probability distribution is finding the parameter values which maximizes the likelihood of occurrence. The method of maximum likelihood is considered the most theoretically correct for fitting probability distributions.

Frequency analyses of hydrologic data use probability distributions to relate the magnitude of extreme events to their frequency of occurrence. The distribution functions most often used when estimating hydrologic events were: Normal, 2 Parameter Log Normal, 3 Parameter Log Normal, Pearson Type III, Log Pearson Type III, and Gumbel Type I. Monthly precipitation events tend to follow the normal distribution, and distribution varies over a continuous range and is symmetric about the mean but allows negative values. However, hydrologic variables tend to be skewed and all are non-negative. The log normal distribution eliminates the problem of non-negative variables as the data are greater than zero, permits the skewness of the data, and does require the data to be symmetric about the logarithm of the mean. The Pearson Type III methods transform the mean, standard deviation, and the coefficient of skewness into the three parameters of the distribution function. When the data are greatly skewed, the log transformation of the Pearson Type III is used to reduce the skewness. Next, the Gumbel Type I distribution, also known as the Extreme Value Type I distribution, is a two-parameter distribution. One parameter is the most probable value of the distribution and the second is a measure of dispersion. Extreme value distributions have been widely used in hydrology (Chow *et al.,*: 1988). Analyzing data for the largest or smallest observations from sets of data became the basis for using the Gumbel (Extreme Value) Type I distribution. In the early 1980s, a comparison of the various distribution functions available for analysis of rainfall frequencies showed that the Log Pearson Type III distribution became method of choice. As a result, in 1981 the U.S. Water Resources Council (now called Interagency Advisory Committee on Water Data) recomme-nded the Log Pearson Type III distribution be used in an effort to promote consistency for flood flow analysis. An independent study showed that this distribution was the most appropriate method of estimation of rainfall data (Naghavi *et al.*, 1991)**.**

Rainfall frequency analyses are often used to aid the design of hydraulic structures, such as the design of storm sewer network (Huff and Angel :1992), and Faires *et al.* :1997). These studies provide the information necessary for the development of a design storm, to represent the probability of occurrence of heavy rainfall, and are used in the design of hydraulic structures (Nguyen et al. :2002). Such studies, however, primarily focus on point rainfall totals, rather than areal totals. When a drainage system is designed, it must handle rainfall within a given area, not just rain from a single point.

2- Location of study area

Mosul is the second largest city in Iraq. It is located along the Tigris River in Ninawa. The Tigris River runs through the center of Mosul, bisecting the city into eastern and western halves. This city is 396 km (250 miles) northwest of Baghdad. The original city stands on the west bank of the Tigris River, opposite the ancient city of Nineveh on the east bank, but the metropolitan area has now grown to encompass substantial areas on both banks, with five bridges linking the two sides (see figure 1).

Mosul city is located topographically in a depression surrounded by hilly lands from the east and west slopping toward city centre and Tigris River. This characteristic brought this city to deliver big quantities of surface runoff during rainfall, flowing through number of main Wadies to Tigris River. Mosul located on Latitude (36^0 $20^'$) North and Longitude (43^0 $08^'$) East, and it has elevation (222 m) above sea level.

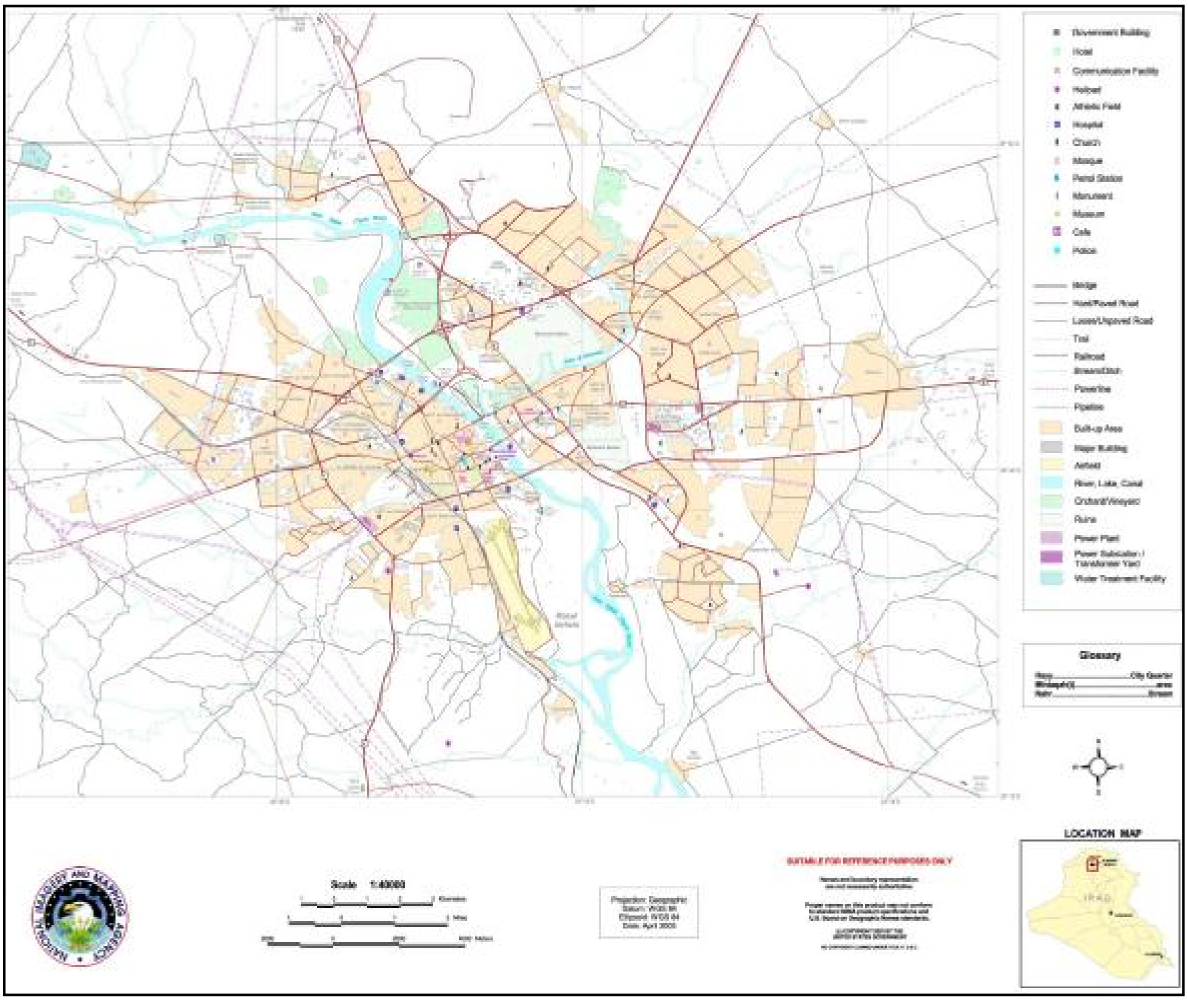

Fig.(1) map of study area.

Climate of study area

The Mosul is of a temperate-warm climate. Mosul is called Um Al-Rabi'ain (The City of Two Springs), because autumn and spring are very much alike there. The city enjoys the Mediterranean climate. That is, it is hot in Summer which starts in May and ends at the end of September. The temperature varies a lot and it may approach 50 °C. In Winter which starts in December and ends at the end of February. Mosul may witness the fall some frost. This indicates this city is really cold in Winter. The climate of the governorate of Mosul varies according to topographical differences. The temperature degrees are (3 °C – 6 °C) in Winter and (30 °C – 40 °C) in Summer.

Mosul (Al Mawsil) shares the severe alternations of temperature experienced by upper Mesopotamia . The summer heat is extreme, and in winter frost is not unknown . Nevertheless the climate is considered healthy and agreeable; copious rains fall in general in winter.

Rainfalls due to the effect of winds blowing from both the Mediterranean Sea Area and the Red Sea Area. The Red Sea winds are less frequent. When the two fronts of winds meet over the Mosul, the city has heavy rain that leads to the flooding of the streets of the city. This flood took place in 1982.

History and description of precipitation data collection

Rainfall is generally measured as an accumulated depth over time. Measurements represent the amount caught by the gage opening and are valid only for the gage location. The amount collected may be affected by gage location and physical factors near the gage. Application over large areas requires a study of adjacent gages and determinations of a weighted rainfall amount. The data of precipitation was obtained from the report of hydrological survey of Iraq (summary of monthly precipitation at stations in Iraq for the period 1887-1958) . 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. 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. Data of rainfall is totaled in calendar years and water years, the latter extending from October of the pervious calendar year through September of the calendar year concerned. 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 Latitude (36^0 $19^{'}$) and longitude (43^0 $09^{'}$), and it has elevation (222 m) above sea level.

Extreme value theory (EVT)

Extreme value theory (EVT) is a separate branch of statistics that deals with extreme events. This theory is based on the extremal types theorem, also called the three types theorem, stating that there are only three types of distributions that are needed to model the maximum or minimum of the collection of random observations from the same distribution.

Broadly speaking, there are two principal kinds of model for extreme values.

The oldest group of models are the block maxima models; these are models for the largest observations collected from large samples of identically distributed observations.

Let $(X1, X_2 \ldots)$ be identically distributed random variables with unknown underlying distribution function. Then common cumulative distribution function is F,

$$F(x) = \Pr\{X_i \le x\}$$

Also let $M_n = \max(x_1, \ldots, x_n)$ denote the nth sample maximum of the process. Then

$$P_r\{M_n \le x\} = F(x)^n \quad \ldots(1)$$

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

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

Such that

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

The three type theorem originally stated without detailed mathematical proof by Fisher and Tippett (1928), and later derived rigorously by Gnedenko (1943), asserts that if a non degenerate (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:

$$H(x) = \exp(-e^{-x}), \text{ all } x, \quad \ldots(2)$$

$$H(x) = \begin{cases} 0 & x<0 \\ \exp(-x^{-a}), & x>0 \end{cases} \quad \ldots(3)$$

$$H(x) = \begin{cases} \exp(-/x/^{a}), & x<0 \\ 1, & x>0 \end{cases} \quad \ldots(4)$$

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,

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

Very often, (2) is called the Gumbel type, (3) the Fréchet type and (4) the Weibull type. In (3) and (4), $\alpha > 0$

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

$$H(x) = \exp\left\{-(1+\xi\frac{x-\mu}{j})_+^{-1/\xi}\right\}, \quad \ldots(5)$$

($y_+ = \max(y, 0)$) where μ is a location parameter, $j > 0$ is a scale parameter and ξ is a shape parameter. The limit $\xi \to 0$ corresponds to the Gumbel distribution, $\xi > 0$ to the Fréchet distribution with $\alpha = 1/\xi$, $\xi < 0$ to the Weibull distribution with $\alpha = -1/\xi$

.

Methodology

The program used in this research to do most of the statistical tests is commonly known as HYFRAN; which is a name composed of the first pairs of letters of the words: HYdrological FRequency Analyses. This program was developed at the National Institute of Scientific Research of the University of Québec, with funding from Hydro-Québec and the Natural Sciences and Engineering Research Council (NSERC).

HYFRAN includes a number of powerful, flexible, user-friendly mathematical tools that can be used for the statistical analysis of extreme events. It can also perform more basic analysis of data time series. HYFRAN can be used in any study that requires fitting of statistical distribution to an independent and identically distributed data series. Applications are found in various technical areas such as engineering, environment, meteorology, medical sciences. As soon as the data entered into the spreadsheet, this software automatically calculates the basic statistics as shown in table (1), and plots the non-exccedance probability and histogram of observation data (see figures. 2 and 3).

Empirical non-exccedance probability is calculated using Weibul formula:

$$F[x[k]] = [k - a]/[n - 2a + 1]$$

$$0 < a \leq 0.5$$

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

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

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

$$T = 1/P_a$$

Normal, Pearson Type III, Lognormal, 3-parameter lognormal and Gumbal are the five distributions selected for representing the probability analyses of Extreme rainfall in this study. The (99%) confidence interval is also used here (figures from 4-8). Maximum likelihood method is applied to estimate theoretical distribution parameters (table 2). The adequacy test is achieved by using Chi-square test, these values are compared with tabulated values of (0.01) significant level (table 3). According to this test, , all the distributions are found to be appropriate for describing of extreme monthly rainfall in the study area.

Table (1) Basic statistics of data sets

Basic statistics	Value	Number of data
Minimum	63.5	36
Maximum	182	
Average	114	
Standard deviation	31.1	
Median	111	
Coefficient of variation (Cv)	0.273	
Skewness coefficient (Cs)	0.388	
Kurtosis coefficient (Ck)	2.14	

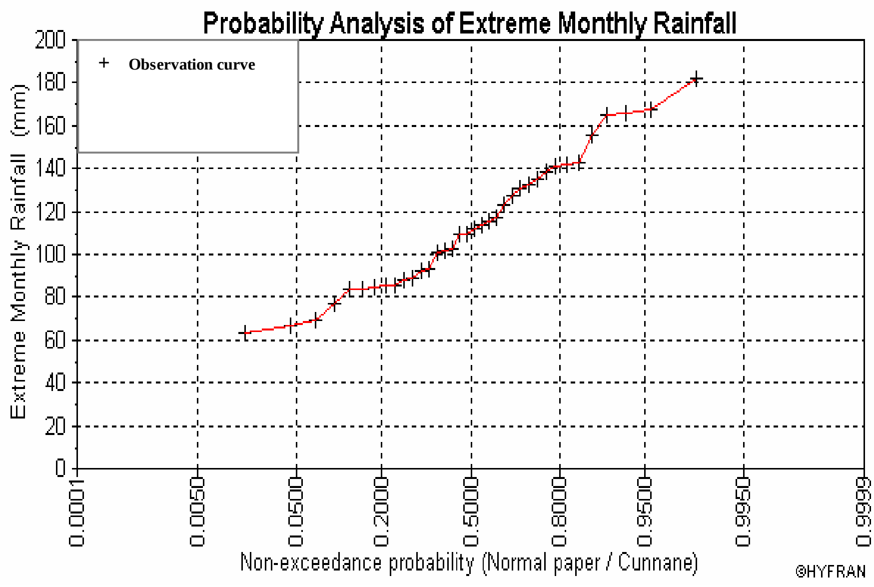

Fig.(2) Observation of dataset on probability paper

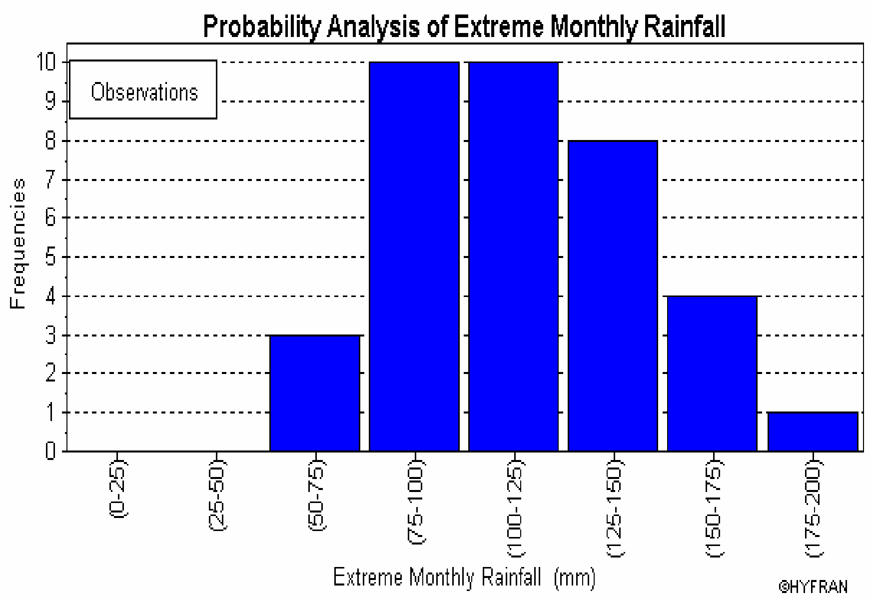

Fig.(3) Histogram of observation dataset

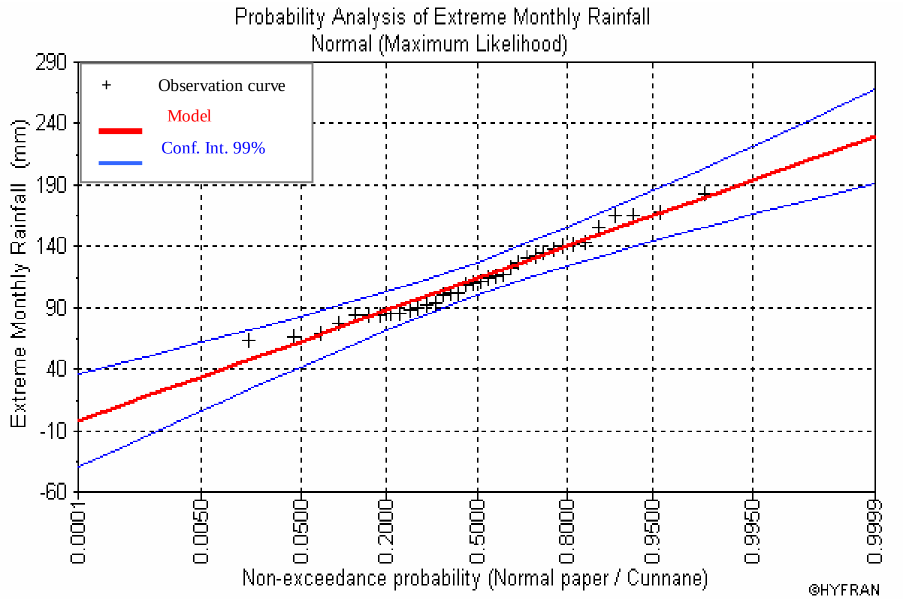

Fig.(4)Empirical, theoretical probabilities and confidence interval against extreme monthly rainfall (normal distribution)

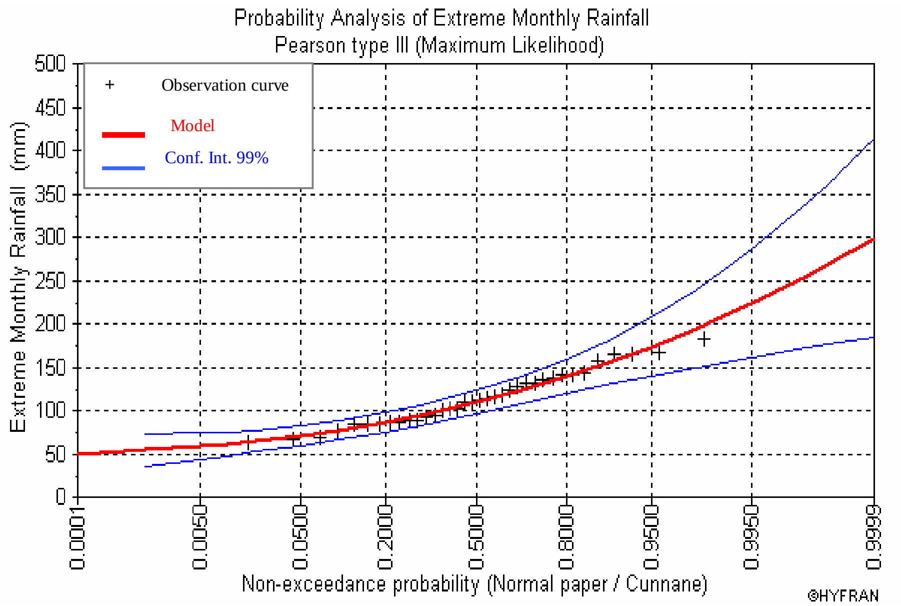

Fig.(5) Empirical, theoretical probabilities and confidence interval against extreme monthly rainfall (Pearson type III distribution)

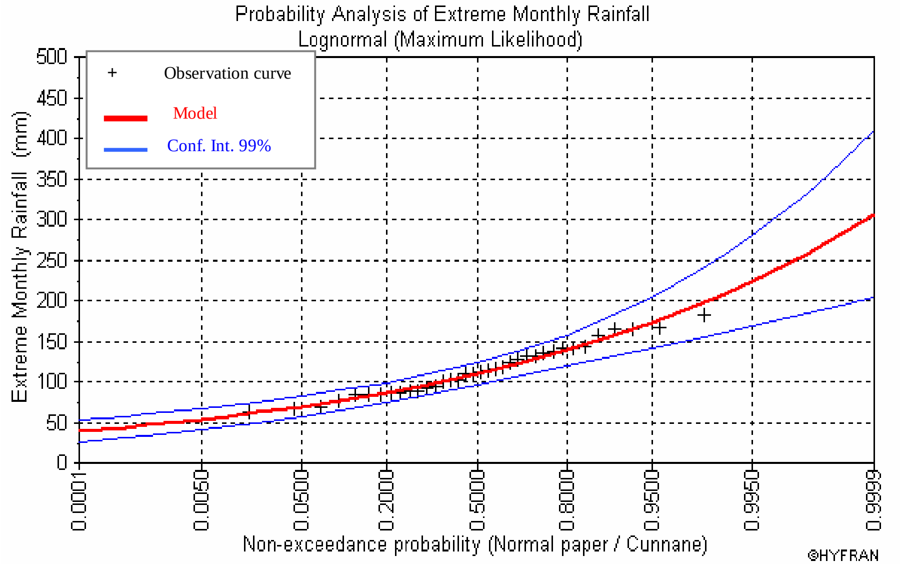

Fig.(6)Empirical, theoretical probabilities and confidence interval against extreme monthly rainfall (lognormal distribution)

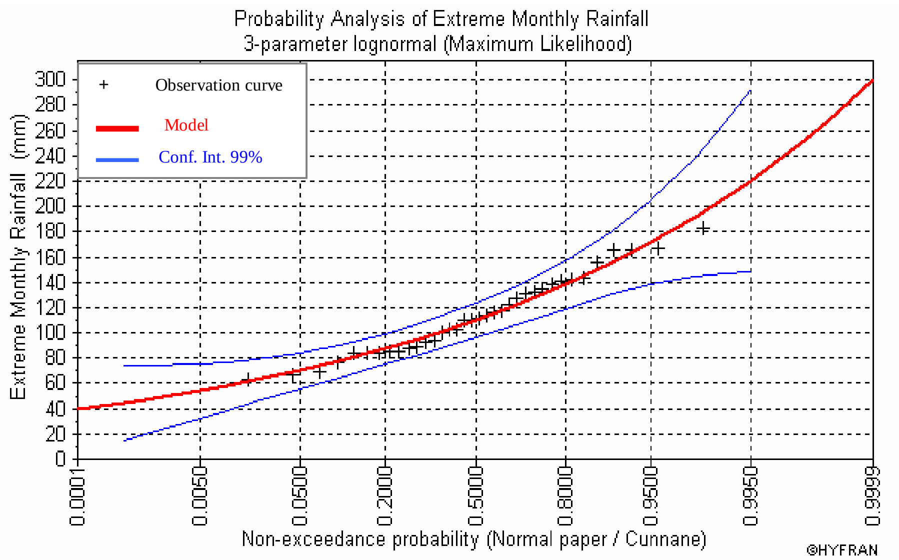

Fig.(7)Empirical, theoretical probabilities and confidence interval against extreme monthly rainfall (3-parameters lognormal distribution)

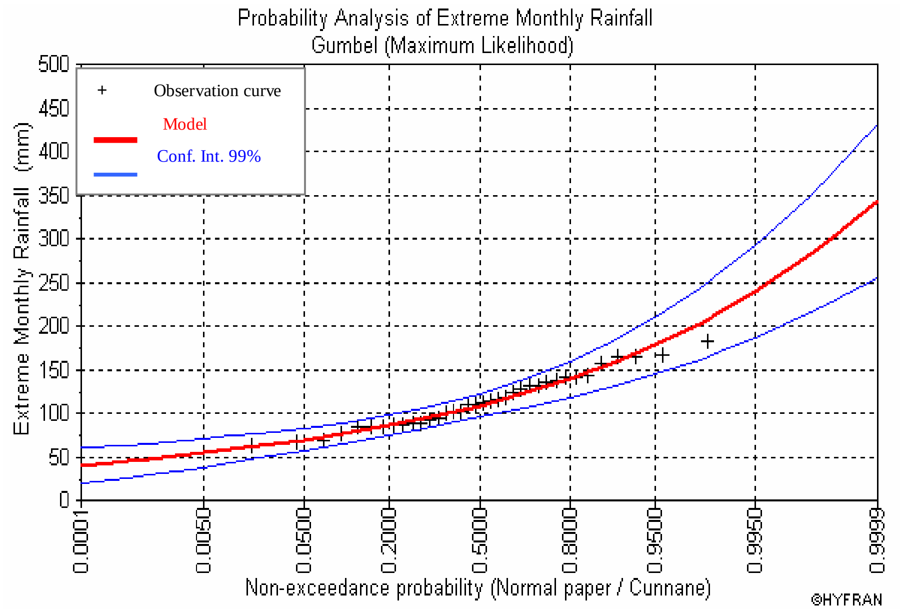

Fig.(8) Empirical, theoretical probabilities and confidence interval against extreme monthly rainfall (Gumbal distribution)

Table (2) Fitting the models with the estimated parameters

No.	Distribution	Mathematical formula	Estimated parameters
1.	Normal	$f(x)=\frac{1}{s\sqrt{2p}}\exp\left\{-\frac{(x-m)^2}{2s^2}\right\}$	$m = 113.739$ $s = 31.1066$
2.	Pearson type III	$f(x)=\frac{a^l}{\Gamma(a)}(x-m)^{l-1}e^{-a(x-m)}$	$a = 0.0675167$ $l = 4.67117$ $m = 44.5535$
3.	Lognormal	$f(x)=\frac{1}{xs\sqrt{2p}}\exp\left\{-\frac{[\ln x-m]^2}{2s^2}\right\}$	$m = 4.69717$ $s = 0.276707$
4.	3-parameters lognormal	$f(x)=\frac{1}{(x-m)s}\exp\left\{-\frac{[\ln(x-m)-m]^2}{2s^2}\right\}$	$m = 4.71862$ $s = 0.267044$ $m = -2.29191$
5.	Gumbal	$f(x)=\frac{1}{a}\exp\left\{-\frac{x-u}{a}\exp\left(-\frac{x-u}{a}\right)\right\}$	$a = 26.5816$ $u = 98.5719$

Table (3) Adequacy test using Chi-square

No.	Distribution	Chi-Square test	
		Tabulated value (0.01) significant level	Calculated value
1.	Normal	13.277	4.00
2.	Pearson type III	11.344	4.89
3.	Lognormal	13.277	4.00
4.	3-parameters lognormal	11.344	4.00
5.	Gumbal	13.277	4.89

4- Conclusions

The five distributions: Normal, Pearson Type III, Lognormal, 3-parameter lognormal and Gumbal are selected to represent the probability analyses of Extreme rainfall in city of Mosul. For the purpose of estimating the theoretical distribution parameters, the Maximum Likelihood Method was applied to the data in this research. estimate. As a result of to the comparison between the values that obtained from the adequacy test (Chi-square values) with their counter-part tabulated values of (0.01) significant level, All the five distributions were found to be suitable for representing of extreme monthly rainfall in Mosul city.

4- References

Chow, Ven Te., David R. Maidment, and Larry W. Mays. Applied Hydrology McGraw-Hill, New York, 1988.

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.

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.

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

HYFRAN Hydrological frequency analysis, primarily developed by Bernard Bobe´e.A demo version is available via www.wrpllc.com/books/hyfran.html.

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

Naghavi, Babak, Vijay P. Singh, Fang Xin Yu. "LADOTD 24-Hour Rainfall Frequency Maps and I-D-F- Curves (Summary Report)". Louisiana State University, Baton Rouge, LA., August 1991.

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

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

Appendix (A) Monthly meteorological data of Mosul station

for period (1923-1958)

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.	Calendar Year	Water Year
1923	66.6	85.8	30.4	41.9	7.4	T	T	--	T	13.8	7.8	31.4	285.1	--
1924	93.4	65.6	47.4	19.2	14.6	2.6	--	--	T	29.9	29.3	63.5	365.5	[illegible]
1925	19.7	21.1	69.5	15.6	11.3	1.1	--	--	T	23.1	19.7	69.3	250.4	261.0
1926	79.2	163.6	96.5	25.5	15.6	T	0.3	--	--	1.6	87.1	109.8	579.2	492.8
1927	8.0	85.6	9.7	71.3	53.2	T	--	--	T	3.7	46.3	52.7	330.5	426.3
1928	43.5	95.4	28.3	2.1	0.1	0.3	1.1	--	1.3	0.3	109.4	13.4	295.2	274.8
1929	34.3	73.3	27.1	42.9	4.8	T	--	--	--	T	141.0	15.9	339.3	305.5
1930	61.7	61.0	2.6	102.1	1.6	2.0	T	T	T	T	20.2	76.3	327.5	387.9
1931	39.6	92.7	22.0	114.1	5.1	1.4	--	T	--	4.5	27.0	13.6	320.0	371.4
1932	46.4	66.5	45.4	12.0	9.2	--	--	--	0.1	--	13.7	16.4	209.7	224.7
1933	132.3	66.9	76.5	52.3	1.0	0.3	--	--	T	T	4.7	40.8	374.8	359.4
1934	40.9	59.6	48.3	49.2	50.8	--	--	--	--	2.2	6.9	63.5	321.4	294.3
1935	22.9	37.9	11.6	41.5	1.9	--	--	--	T	6.5	84.0	33.7	240.0	188.4
1936	52.0	84.6	19.9	43.2	61.6	T	T	--	--	T	79.4	60.1	400.8	385.5
1937	84.4	13.1	10.7	44.2	7.5	0.0	T	T	0.0	54.6	127.0	13.0	354.5	299.4
1938	155.9	70.3	42.4	41.0	27.0	T	T	0.0	1.5	T	110.4	72.2	520.7	532.7
1939	73.3	59.5	138.1	108.4	5.8	0.0	0.0	--	T	11.3	53.1	135.3	584.8	567.7
1940	122.8	59.5	36.2	66.2	5.3	T	--	T	T	30.7	33.8	52.2	406.7	489.7
1941	38.3	93.2	80.7	28.7	T	--	--	--	0.5	--	20.4	165.5	427.3	358.1
1942	62.4	76.4	51.2	7.6	0.2	T	--	T	0.6	8.3	89.2	9.9	305.8	384.3
1943	51.5	134.8	103.6	42.3	8.5	T	2.3	--	--	14.1	1.7	48.7	407.5	450.4
1944	141.9	5.7	42.6	40.7	21.2	T	T	0.0	0.0	3.9	126.1	60.4	442.5	316.6
1945	105.0	15.7	18.5	12.3	3.3	T	T	0.0	T	5.2	80.4	115.8	356.2	345.2
1946	84.0	102.6	63.5	74.5	79.2	3.3	T	--	T	13.3	10.6	71.2	502.2	608.5
1947	88.0	29.8	36.8	8.3	9.6	1.1	T	0.0	--	0.1	42.9	36.7	253.3	268.7
1948	30.1	26.9	28.6	142.9	22.0	T	0.0	T	--	T	9.3	113.3	373.1	330.2
1949	42.6	129.8	167.4	74.6	10.9	--	--	0.0	0.0	T	T	116.6	541.9	547.9
1950	117.0	86.9	105.0	11.6	69.8	0.3	T	0.0	T	8.1	14.2	29.4	442.3	507.2
1951	[illegible]	76.9	47.2	45.2	22.5	T	--	T	T	22.9	24.6	72.5	369.9	301.6
1952	57.3	182.3	75.9	42.6	6.2	T	T	--	T	T	5.8	55.7	405.8	464.3
1953	76.4	84.1	131.1	75.7	9.5	3.8	T	--	0.0	9.3	45.1	102.6	537.6	442.1
1954	48.9	137.1	165.0	128.6	6.4	T	T	0.2	T	3.8	19.1	76.1	585.2	643.2
1955	41.5	59.7	39.7	73.7	6.1	0.2	0.0	0.0	1.8	0.0	41.7	100.7	365.1	321.7
1956	92.6	31.9	50.3	52.0	1.5	T	0.0	T	0.9	1.5	15.1	50.9	296.7	371.6
1957	55.3	45.2	111.6	85.8	67.7	4.5	0.0	T	7.1	3.5	44.1	35.1	459.9	444.7
1958	83.9	5.0	39.9	7.1	2.6	0.1	0.0	0.0	0.0	0.4	11.5	57.7	208.2	221.3

التحليل الاحتمالي لقيم الأمطار الشهرية العظمى في مدينة الموصل شمال العراق

عمار سلمان داود

قسم الهندسة المدنية, كلية الهندسة, جامعة البصرة, البصرة, العراق

الملخص

بالإمكان اعتبار هذه الدراسة محاولة لتقييم التوزيع الاحتمالي النظري الخاص بقيم الأمطار الشهرية العظمى في مدينة الموصل. ولهذا، فقد تم استخدام القيم المسجلة للأمطار طيلة الفترة (1923-1958). و لتحليل البيانات المتوفرة لدينا قمنا بإجراء التحليلات الترددية وجميع الاختبارات الإحصائية وذلك باستعمال النسخة 1,1 لبرنامج هايفران (**HYFRAN**) الذي يعمل تحت نظام ويندوز(**WINDOWS**): وتم اختبار خمس توزيعات هنا ألا وهي, الطبيعي, وبيرسون النوع الثالث, اللوغارتيمي الطبيعي, اللوغارتيمي الطبيعي ذي الثلاث معالم, وكامبل. وقد تم التوصل إلى تخمين التوزيع النظري باستعمال طريقة الاحتمال الأكبر أما اختبار الموائمة فكان عن طريق استخدام مربع كاي. وثبتت النتائج أن جميع التوزيعات هي المثلى لقيم الأمطار الشهرية العظمى لمنطقة الدراسة.