ZANCO Journal of Pure and Applied Sciences
The official scientific journal of Salahaddin University-Erbil
https://zancojournals.su.edu.krd/index.php/JPAS

ISSN (print):2218-0230, **ISSN (online):** 2412-3986, **DOI:** **http://dx.doi.org/10.21271/zjpas**

RESEARCH PAPER

Fitting of Rainfall Data in Erbil City Using Statistical Distribution Techniques

Khalid A. ABBAS[1], Hassan Sadi Ibrahim[1]

[1] Department of Physics, College of Science, Salahaddin University-Erbil, Kurdistan Region, Iraq.

A B S T R A C T:

Rainfall Statistical distribution fitting is essential for the design of water management and water related infrastructure. Future prediction of rainfall events can be made such as floods and drought for a given area of study if the statistical distribution was known. In this paper the rainfall data of 48 years in Erbil meteorological station in Erbil city was analyzed and fitted to several types of statistical distributions to find the best distribution.

The Normal, Gamma and Weibull distributions model were used for annual and monthly rainfall and to the goodness of fit was found based on the p-value and the Anderson-Darling tests. The result shows that Weibull and Gamma functions are successful for all cases, while Normal function failed in a number of months but did very well in others.

KEY WORDS: Rainfall, Statistical distribution, Normal, Gamma, Weibull, p-value, Anderson-Darling test, Erbil.
DOI: http://dx.doi.org/10.21271/ZJPAS.35.6.2
ZJPAS (2023) , 35(6);9-15.

1. INTRODUCTION:

Rainfall data analysis strongly depends on the statistical distribution. Therefore, it is essential to know the best distribution that the data follows to make better understanding of rainfall events specially when studying the extreme rainfall events such as drought and flood. This topic has long been studied and many researchers used many probability distributions that give a good fit to rainfall data, among them normal, Weibull, Gamma, lognormal, and Pearson type distribution were used.

Longe ago (Fisher, 1924) showed in his study that the rainfall distribution during a season has a greater influence crop yield rather than the total amount of rainfall. In his research (Ozturk, 1981) reviewed a mixture of Gamma and Poisson distribution models for precipitation totals and their applications, and made an application for monthly precipitation totals.

* **Corresponding Author:**
Khalid A. ABBAS
E-mail: khalid.abbas@su.edu.krd
Article History:
Received: 24/04/2023
Accepted: 04/06/2023
Published: 15/12/2023

(Al-Mansory, 2005) selected six distributions for the data set of Basrah station, south of Iraq for 75-given years, and used six distributions (Normal, Log-Normal, Log-Normal type III, Pearson type III, Log-Person type III, and Gumbel). Results showed that Gumbel and Person type III distributions were the better for describing maximum monthly rainfall in Basrah station.

For arid and semi-arid regions (Maliva and Missimer, 2012) the Rainfall data are best fit using probability distribution functions (Normal, Log-Normal, Gamma, Weibull, and Gumbel). Our region of study is considered semi-arid region (Saeed, and Abbas 2014), and we expect such distribution will have best fit for our data.

(Husak et al.2006) used Gamma distribution applied on monthly rainfall in Africa for drought monitoring applications. To find the goodness of fit they determined the Kolmogorov–Smirnov test, and compared these results against Weibull distribution. Their result showed that Gamma distribution was suitable for nearly 98% over all months.

(Alghazali and Alawadi, 2014), used Normal, Gamma and Weibull distributions to fit to rainfall data from thirteen Iraqi stations, and used two goodness of fit tests (Chi-Square and Kolmogorov-Smirnov) tests, they found that Gamma distribution was suitable for five stations, while, Normal and Weibull distributions were not suitable for any station.

2.Used probability distributions

The following three probability distributions are used in this paper: Normal, Weibull and Gamma.

2.1. Normal

Normal (or Gaussian) probability distribution is widely used. Its corresponding probability distribution function (PDF) is (C. Walck, 2007):

$$F(X) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2} \quad \text{......} \; (1)$$

where:

μ: the mean

σ: the standard deviation

The distribution can be standardized by using the standardized variable z = (x − μ)/σ, which has a mean of zero and standard deviation of 1. Standard variable Z is referred to as the X. By allowing μ = 0 and σ =1, the density function for Z can be determined from the definition of a normal distribution. The relevant distribution function is provided by:

$$F(Z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2} \quad \text{......} \; (2)$$

2.2. Weibull

Based on the value of the shape parameter β, the Weibull distribution is a flexible distribution that can adopt the traits of other types of distributions, and it is widely used in meteorology.

The two-parameter Weibull distribution's probability density function f(x) and cumulative distribution function F(x) are provided by (Reddy, 1997):

$$f(x) = \frac{\alpha}{\beta}\, x^{\alpha-1}\, e^{-(x/\beta)^{\alpha}} \quad \text{......} \; (3)$$

$$F(X) = 1 - e^{-(x/\beta)^{\alpha}} \quad \text{......} \; (4)$$

where:

β: scale parameter

α: shape parameter

2.3. Gamma

A versatile life distribution model is the gamma distribution. The probability distribution function (PDF) of Gamma is (C. Walck, 2007):

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}\, X^{\alpha-1}\, e^{X/\beta} \quad \text{......} \; (5)$$

Where:

α: shape parameter

β: scale parameter

The Cumulative Distribution Function for Gamma distribution is:

$$F(x) = \frac{\Gamma_x \beta(\alpha)}{\Gamma(x)} \quad \text{......} \; (6)$$

3.Goodness of Fit (GOF)

The effectiveness of a random sample in fitting a theoretical probability distribution function is determined by the goodness of fit test. To put it another way, these tests demonstrate how well the chosen distribution matches the data. The general process entails defining a test statistic, which is a function of the data measuring the gap between the hypothesis and the data, and then calculating the likelihood of obtaining data that have an even larger value of this test statistic than the value observed, assuming the hypothesis is true. We refer to this likelihood as the confidence level.

3.1. Anderson-Darling test

The particular distribution is used by the Anderson-Darling test to determine critical values. This offers the benefit of enabling a more sensitive test but the drawback of requiring the calculation of critical values for each distribution. The following criteria define the Anderson-Darling test statistic (D'Agostino and Stephens, 1986):

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln F(X_i) + \ln(1 - F(X_{n-i+1}))] \quad \text{......} \; (7)$$

The exact distribution that is being examined determines the critical values for the Anderson-Darling test. If the test statistic, A, is higher than the critical value, the test is a one-sided test, and the hypothesis that the distribution is of a particular form is rejected. Keep in mind that the Anderson-Darling statistic for a particular distribution may be multiplied by a constant (which often depends on the sample size, n). The critical values should be measured against this.

3.2. Kolmogorov-Smirnov Test

Given a finite data set, the Kolmogorov-Smirnov (K-S) test is a goodness-of-fit test that is used to detect if an underlying probability distribution deviates from a proposed distribution. Following is a step-by-step process for doing the K-S test given a series of sample values ($x_1,x_2,\ldots,x_i$) obtained from a population X, is (Soong, 2004):

- The sample values are listed in descending order of magnitude and are indicated by the symbol (x_i).
- The relationship $S(x_i)=i/N$, where N is the total number of observations, yields the observed distribution functions $S(x_i)$.
- Equation yields the distribution function $F(x_i)$ at each x_i using the predicted distribution, and the deviations D_2 are calculated from: $D_2=S(x_i)-F(x_i)$
- The critical value determined in statistical tables is contrasted with the highest absolute value of D_2, which was acquired from the previous Equation. The tested distribution is appropriate for representing the observed data if D_2 is less than the crucial value; otherwise, it is ineffective.

4.Results and Discussion

The Statistical distribution fitting was performed using (Normal, Gamma and Weibull) distribution models on a series of 48 years monthly and annual rainfall data in Erbil city.

(Table 1) shows the Statistics of monthly and annual rainfall data in Erbil for 48-years, the high standard deviation of the rainfall can be noticed from the table which is a normal characteristic of rainfall in the region.

Table (1) Statistics of monthly and annual rainfall data (mm) in Erbil for 48-years.

Month	Mean	SD.	Median	Min.	Max.	Skewness	Kurtosis
Jan	74.0	40.1	66.0	1.6	177.3	0.40	-0.40
Feb	73.1	39.7	69.9	5.7	189.0	0.65	0.39
Mar	71.3	46.7	61.1	6.4	237.9	1.22	2.31
Apr	49.0	36.3	42.4	1.0	170.5	1.16	1.57
May	15.5	19.2	10.4	0.1	95.2	2.28	6.28
Oct	14.4	20.1	7.9	0.1	88.1	2.24	5.06
Nov	45.8	46.6	28.0	1.0	200.7	1.79	2.82
Dec	68.3	44.4	60.1	0.1	181.8	0.82	0.08
Annual	412.5	117.3	406.3	227.8	737.7	0.45	-0.05

Fig.1 shows the best fitting Probability plot for monthly rainfall data (the unsuitable ones are not shown), and results shows that Jan and Feb months are best fitted by Normal distribution. The Weibull distribution is best fitted for Mar, Apr, and Dec months, While the rest of months May, Oct, and Nov are best fitted with Weibull distribution function.

These results are concluded using the p-value at confidence level of 0.05, shown on Fig.1 for each monthly rainfall and on Fig.3 for annual rainfall, also supported by Anderson-Darling test at same confidence level shown in (Table 2).

For the annual rainfall the best fitted distribution is the Normal function, with p-value of 0.755. From (Table 2), It can be seen that more than one function is suitable for fitting, such as the case of Jan month (and other months) where all three functions fitted well, but only one was selected based on the highest p-value. In case when the p-vales were equal the best was chosen depending on the Anderson-Darling test result (which is 0.752 at confidence level of 0.05). The best fitting Histogram plots are shown on Fig.2 for monthly rainfall data also on Fig.3 for annual rainfall data with shape and scale parameters for each best fitting distribution.

Also, it has been observed from (Table 2), Weibull and Gamma distributions could be fitted to all months, but the Normal distribution failed in months (Mar, Apr, May, Oct, Nov, and Dec), non the less Normal distribution was best fit in other cases.

The high variability of rainfall data seen in (Table 1) which results a large standard deviation had its effect on the shape of Probability plot seen in Fig.1 especially on the starting month of rainfall season in Oct and the ending month in May, in which the rainfall amount is small, effect seen in one tail of the plot.

5.Conclusion

Rainfall analysis result revealed that there is no one function that best fits all cases. The best fitting probability distribution for Erbil city from (Normal, Gamma and Weibull) was as follows:

Normal: Jan, Feb, Annual.
Weibull: Mar, Apr, Dec.
Gamma: May, Oct, Nov.

The obtained fitting results were compared and best fit was selected based on the p-value at confidence level of 0.05, and also Anderson-Darling test was used in the procedure.

In most cases, more than one function was suitable for the fitting but only one was chosen as the best. Other cases (Mar, Apr, May, Oct, Nov, and Dec), were found where Normal function failed to fit the data, but others did well, accordingly cautious is advised using the Normal function.

Table (2) The results of A-D tests and p-value for Normal, Weibull and Gamma distribution functions.

Month	Distribution	Anderson Darling	p-value
Jan	Normal	0.381	0.381
	Weibull	0.232	>0.250
	Gamma	0.515	0.215
Feb	Normal	0.357	0.442
	Weibull	0.13	>0.250
	Gamma	0.3	>0.250
Mar	Normal	0.823	0.031
	Weibull	0.18	>0.250
	Gamma	0.213	>0.250
Apr	Normal	1.011	0.01
	Weibull	0.166	>0.250
	Gamma	0.24	>0.250
May	Normal	3.429	<0.005
	Weibull	0.534	0.18
	Gamma	0.388	>0.250
Oct	Normal	4.494	<0.005
	Weibull	0.49	0.223
	Gamma	0.407	>0.250
Nov	Normal	3.581	<0.005
	Weibull	0.582	0.135
	Gamma	0.577	0.169
Dec	Normal	0.896	0.02
	Weibull	0.239	>0.250
	Gamma	0.485	0.245
Annual	Normal	0.243	0.755
	Weibull	0.286	>0.250
	Gamma	0.205	>0.250

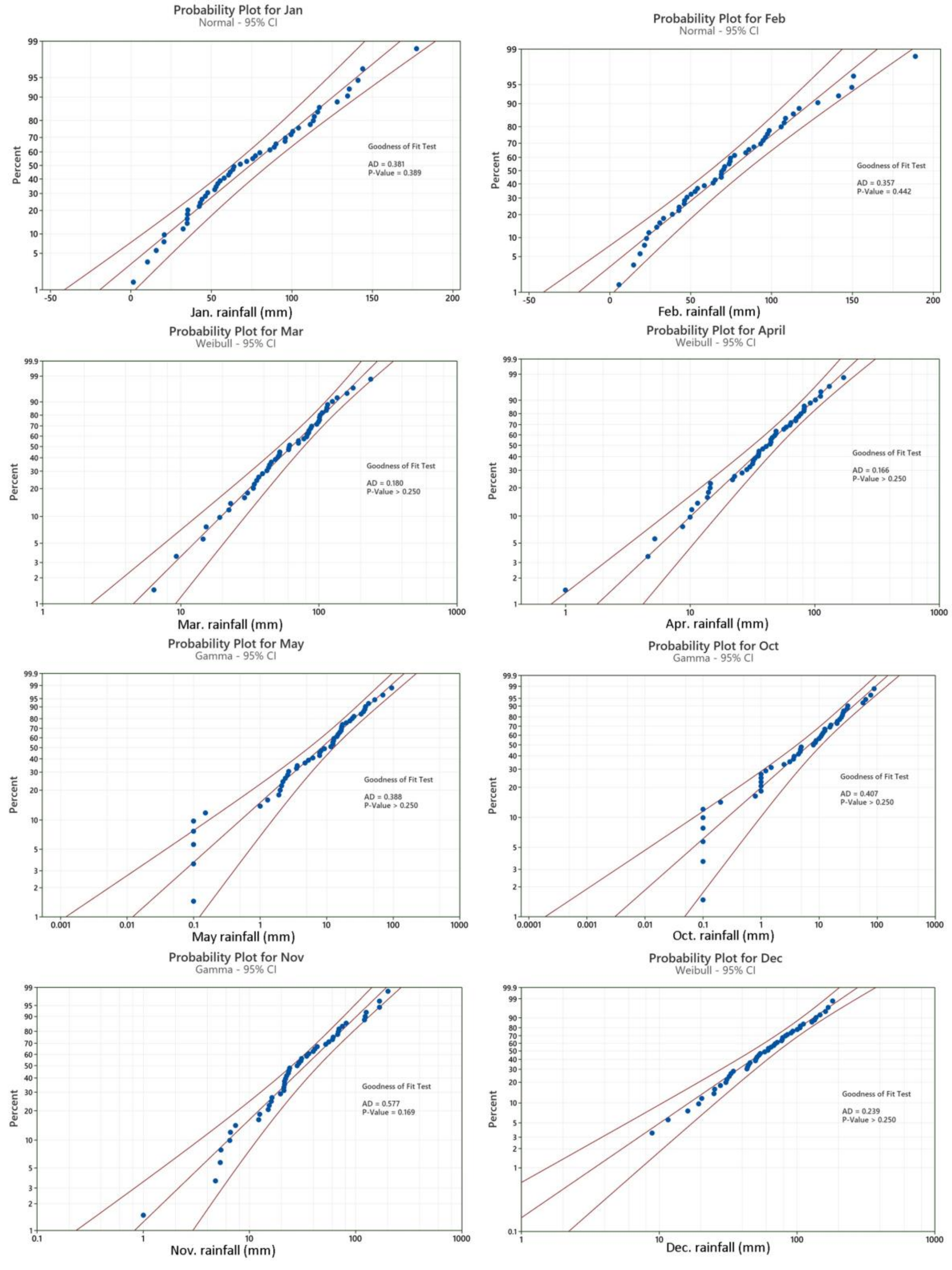

Figure 1: The best fitting Probability plot for monthly rainfall data.

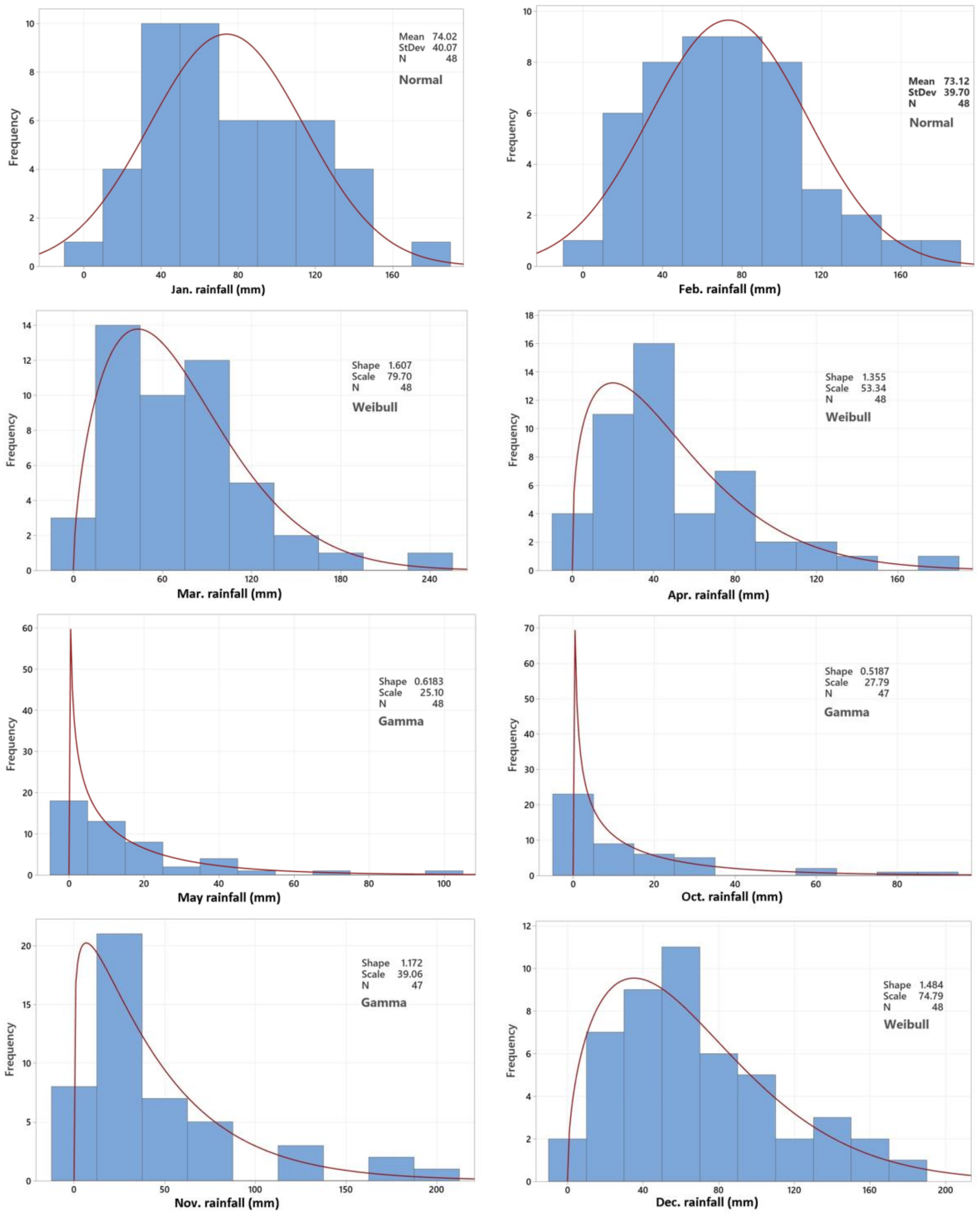

Figure 2: The best fitting Histogram plots for monthly rainfall data.

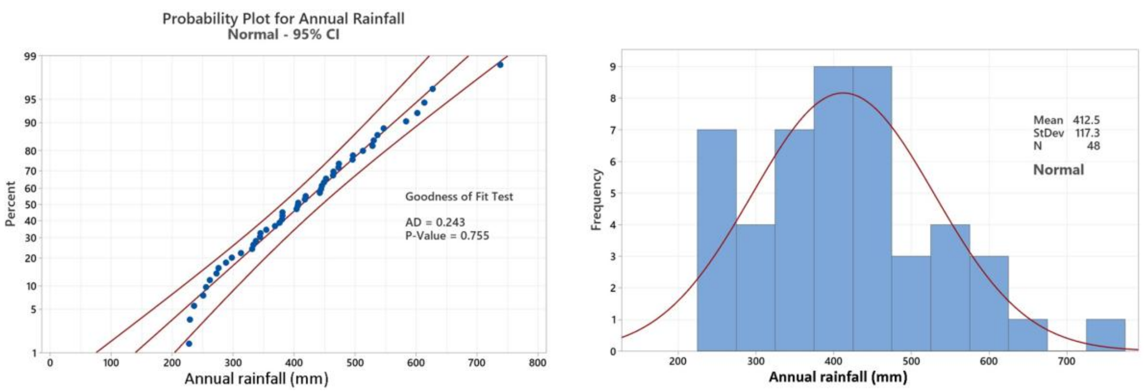

Figure 3: The best fitting Probability plot and Histogram plot for Annual rainfall data.

References

Fisher R.A. The influence of the rainfall on the yield of wheat at Rothamsted. Philosophical transaction of the Royal Society of London. 1924; Series B, Vol. 213.

Ozturk A. (1981), "On the Study of a Probability Distribution for Precipitation Totals", Department of Statistics, Division of Mathematical Science, Mimeograph Series (81-3).

Al-Mansory H. B. (2005), "Statistical Analysis of Extreme Monthly Rainfall in Basrah City, South of Iraq", Marina Mesopotamica, Basrah, Iraq, Vol. 20, 2, PP. (283-296).

Maliva R. and Missimer T. (2012), "Arid Lands and Water Evaluation and Management", Springer-Verlag Berlin Heidelberg, ISBN 978-3-642-29104-3.

Saeed, Mohammed & Abbas, Khalid. (2014). ANALYSIS OF CLIMATE AND DROUGHT CONDITIONS IN THE FEDRAL REGION OF KURDISTAN. International Scientific Journal Environmental Science.

Husak G. J., Michaelsen J. and Funk C. (2006), "Use of the Gamma Distribution to Represent Monthly Rainfall in Africa for Drought Monitoring Applications", International Journal of Climatology, Published online in Wiley Inter. Science, Royal Meteorological Society, Santa Barbara, U. S. A., DOI: 10.1002/joc.

Najm Obaid Salim Alghazali, Dhelal Adnan Hussein Alawadi (2014) "Fitting Statistical Distributions of Monthly Rainfall for Some Iraqi Stations", Civil and Environmental Research, ISSN 2224-5790, Vol.6, No.6, 2014.

C. Walck, ''Hand-Book on Statistical Distribution for Experimentalists'', University of Stokholm,2007.

Reddy J. R. (1997), "Stochastic Hydrology", Laxmi Publications, 2nd Edition, New Delhi, India, ISBN 10:817008086X.

D'Agostino, Ralph B, Stephens, Michael A., "GOODNESS-OF-FIT TECHNIQUES", MARCEL DEKKER, INC.,1986, ISBN: 0-8247-8705-6.

T. T. Soong, Fundamentals of Probability and Statistics for Engineers, John Wiley and Sons, New York, USA, 2004.