IOSR Journal of Applied Physics (IOSR-JAP)
e-ISSN: 2278-4861.Volume 8, Issue 5 Ver. III (Sep - Oct. 2016), PP 19-22
www.iosrjournals.org

Goodness of Fitting of Annual Rainfall Data of Four Stations of Iraq

Firas S. Basheer
(Department of Atmospheric Sciences, College of Science, AL-Mustansiriyah University, Baghdad, Iraq)

Abstract: *Rainfall is very important to humankind, because it represents a pure source of water aside from the Earth's. Therefor many researchers are interested in Rainfall. In this study, the monthly data of the amount of rainfall of four stations (Baghdad, Mousl, Rutba, and Basra) of Iraq were collected by means of algorithm, standerd deviation, maximum and minimum were computed. To understand the behavior of rainfall, we tested the data according to Kolmogorov-Smirnov and Chi-Squared tests. The 16 probability distributions were applied to know which distribution is suitable for these data. We found out that the Mosul station has biggest amount of rainfall from other stations and these data follow Weibull distribution which is considered the best distribution according to the tests whereas the Baghdad station take Log-normal distribution, likewise, Rutba station, in spite of maximum and minimum monthly rainfall, Baghdad has bigger statistics than Rutba. Exceptionally, Basra station takes Logistic distribution. All processes use Microsoft Excel and MATLAB programming.*

Keywords: *agricultural, atmospheric sciences, climate, rainfall, urban.*

I. Introduction

The rainfall factor is very important for human because it has a big impact on life such as agriculture, urban and the management of water resources. Therefore, there are a lot of researchers' interest circumvent around the selection of which of the distributions could be convenient to rainfall date. Tao et al reported that several probability models have been developed to describe the distribution of annual extreme rainfalls at a single site. The selection of an appropriate model depends mainly on the characteristics of available rainfall data at the particular site [1].Suhaila J. and A. Jemain classified daily rainfall data according to four rain types of sequence of wet days (Type 1, 2, 3 and 4) [2].They examined four distributions to fit daily rainfall amount in Peninsular Malaysia. Also, they estimated the parameters for each distribution by using the maximum likelihood method. Based on these goodness-of-fit test, the Mixed Exponential is found to be the most appropriate distribution for explaining the daily rainfall amount in Peninsular Malaysia. Olofintoye et al used some distributions annual rainfall data for 20 stations to state the peak daily rainfall distribution characteristics in Nigeria [3].They found that the best distribution of 50% of the total station number that log-Pearson type III, while Pearson type III distribution is a second best at 40% of the total stations, and lastly was log-Gumball at 10% of the total stations. Mohita A. S. and J. B. Singh processed rainfall data to identify the maximum rainfall on any one day, of any week and month, in a monsoon season and in a year. The rainfall data were analyzed to find the best fit probability distribution for each period of study and the trend has been presented in this study. The lognormal and gamma distribution were found as the best fit probability distribution for the annual and monsoon season period of study, respectively, generalized extreme value distribution was observed in most of the weekly period as best fit probability distribution. The best fit probability distribution of monthly data was found to be different for each month [4]. Oseni B. A. and F. J. Ayoola presented several types of statistical distributions to explain rainfall distribution in Ibadan metropolis of 30 years [5]. According to recorded data at rain gauge station, they used exponential, gamma, normal and Poisson distributions compared to identifying the optimal model for daily rainfall amount. They found that exponential distribution is the best model and the second best model is normal also, Poisson model that has the same estimated rainfall amount for explaining the daily rainfall in Ibadan metropolis models.

In this paper, we fit the annual rainfall data of four stations of Iraq on distributions and select which one is best fit according to Kolmogorov-Smirnov and Chi-square tests.

II. Methodology

We obtained the annual rainfall data of four satiations of Iraq (Baghdad, Mousl, Basra, Rutba) from Iraqi Meteorological Organization And Seismology and calculated the means, stander deviation, and maximum and minimum of data that showed in table 1 [6].

Table 1: shows statistical description of four stations.

stations	maximum	minimum	Std.deviation	Mean	Observation
Baghdad	336.000	49.900	58.436	135.269	73
Rutba	263.800	23.301	53.095	115.644	82
Mosul	633.000	165.100	111.765	377.420	74
Basra	61.955	31.900	61.955	141.989	73

III. Climate of Iraq

The climate in Iraq is mainly of the continental, subtropical semi-arid type, with the north and north-eastern mountainous regions having a Mediterranean climate. Rainfall is seasonal and occurs in the winter from December to February, except in the north and northeast of the country, where the rainy season is from November to April. Average annual rainfall is estimated at 216 mm, but ranges from 1 200 mm in the northeast to less than 100 mm over 60% of the country in the south. Winters are cool to cold, with a daily temperature of about 16 °C dropping at night to 2 °C with a possibility of frost. Summers are dry and hot to extremely hot, with a shade temperature of over 43 °C during July and August, yet dropping at night to 26 °C. Iraq can be divided into four agro-ecological zones [7] :

- Arid and semi-arid zones with a Mediterranean climate. A growing season of about nine months, over 400 mm of annual winter rainfall, and mild/warm summers prevail. This zone covers mainly the northern governorates of Iraq. Major crops include wheat, barley, rice and chickpea. Other field crops are also produced in smaller quantities. There is some irrigation, mainly from springs, streams and bores.

- Steppes with winter rainfall of 200–400 mm annually. Summers are extremely hot and winters are cold. This zone is located between the Mediterranean zone and the desert zone. It includes the feed barley production areas, limited wheat production, and it has limited irrigation.

- The desert zone with extreme summer temperatures and less than 200 mm of rainfall annually. It extends from just north of Baghdad to the Saudi Arabian and Jordanian borders. It is sparsely populated and cultivated with just a few crops in some irrigated spots [8].

- The irrigated area which extends between the Tigris and Euphrates rivers from the north of Baghdad to Basra in the south. Serious hazards of this area are of poor drainage and salinity. The majority of the country's vegetables, sunflower and rice are produced in this zone [9].

IV. Steps of mothed

Step I: Fitting the probability distribution

The probability distributions viz. normal, lognormal, gamma, weibull and pearson are generalized where extreme value were identified to evaluate the best fit probability distribution for rainfall. In addition, the different forms of these distributions were also tested and thus the total 16 probability distributions viz. normal, lognormal (2P, 3P), gamma (2P, 3P), generalized gamma (3P, 4P), log-gamma, weibull (2P, 3P), pearson 5 (2P, 3P), pearson 6 (3P, 4P), log-pearson 3, were generalized under extreme value and applied to find out the best fit of probability distribution. The description of various probability distribution functions viz. density and testing the goodness of fit function, range, and the parameter involved are presented in table 1.

The goodness of fit test measures the compatibility of random sample with the theoretical probability distribution. The goodness of fit tests is applied for testing the following null hypothesis:

H0: the maximum monthly rainfall data follow the specified distribution

AH: the maximum monthly rainfall data does not follow the specified distribution.

The following goodness-of-fit tests viz. Kolmogorov-Smirnov test and the chi-square test at α (0.01) level of significance for the selection of the best fit probability distribution.

(i) Kolmogorov-Smirnov Test:

The Kolmogorov-Smirnov statistic (D) is defined as the largest vertical difference between the theoretical and the empirical cumulative distribution function (ECDF):

$$D = \max_{1 \le i \le n} \left(F_i(X_i) - \frac{i-1}{n}, \frac{i}{n} - F(X_i) \right) \tag{1}$$

Where, Xi = random sample, i =1, 2… n.

$$CDF = F_n(X) - \frac{1}{N}.[\text{Number of observations } X_n \le (2)] \tag{2}$$

This test is used to decide if a sample comes from a hypothesized continuous distribution.

(ii) Chi-Squared Test:

The Chi-Squared statistic is defined as

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{3}$$

Where
Oi = observed frequency
Ei = expected frequency
'i'= number of observations (1, 2, …….k) Calculated by :

$$E_i = F(X_2) - F(X_1) \tag{4}$$

F = the CDF of the probability distribution being tested
The observed number of observation (k) in interval 'i' is computed from equation given below:

$$K = 1 + \log_2 n \tag{5}$$

n = sample size
This test is for continuous sample data only and is used to determine if a sample comes from a population with a specific distribution.

Step II: Identification of best fit probability distribution

The two goodness of fit tests mentioned above were fitted to the total rainfall data treating different data set. The test statistic of each test were computed and tested at (α=0.05) level of significance. Accordingly, the ranking of different probability distributions were marked from 1 to 18 based on minimum test statistic value. The distribution holding the first rank was selected for the two tests independently. The assessments of all the probability distribution were made on the bases of total test score obtained by combining the entire two tests. Maximum score 18 was awarded to rank first probability distribution based on the test statistic and lesser scores were awarded to the distribution having rank more than 1, that is 2 to 18. Thus, the total score of the entire two tests were summarized to identify the best fit distribution on the bases of highest score obtained.
The probability distribution having the maximum score was included as a fourth probability distribution in addition to three probability distributions which were previously identified.

V. Results and discussion

We found - according to monthly rainfall data - that Mosul station has biggest amount of rainfall from other stations; it was 633mm, and these data follow Weibull distribution which is the best distribution according to the tests also, the minimum value was 165 mm, whereas the Baghdad station take Log-normal distribution. Likewise, Rutba station, in spite of maximum and minimum rainfall, Baghdad has bigger rates than Rutba. Exceptionally, Basra station takes Logistic distribution and the maximum and minimum rainfall were 61.955 mm and 34.900 mm respectively. Figure 1 show the density function of distributions for four stations and Tables 2, 3 show the results of tests and the parameters of best distributions.

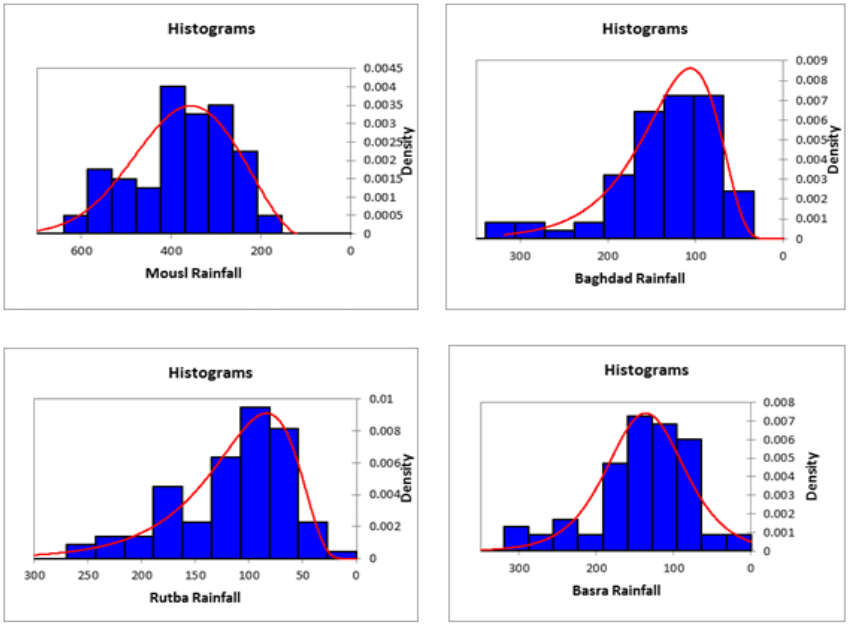

Figure 1: Explain the Histograms density function of distributions of Rainfall data

Table 2: test ranking first position of four stations

Stations	Test ranking first position			
	Kolmogorov Smirnov		Chi-square	
	Distribution	Statistic	Distribution	Statistic
Baghdad	Log-normal	0.992	Log-normal	0.299
Rutba	Log-normal	0.897	Log-normal	0.134
Mosul	Weibull (3)	0.933	Weibull(3)	0.264
Basra	Logistic	0.906	Logistic	0.005

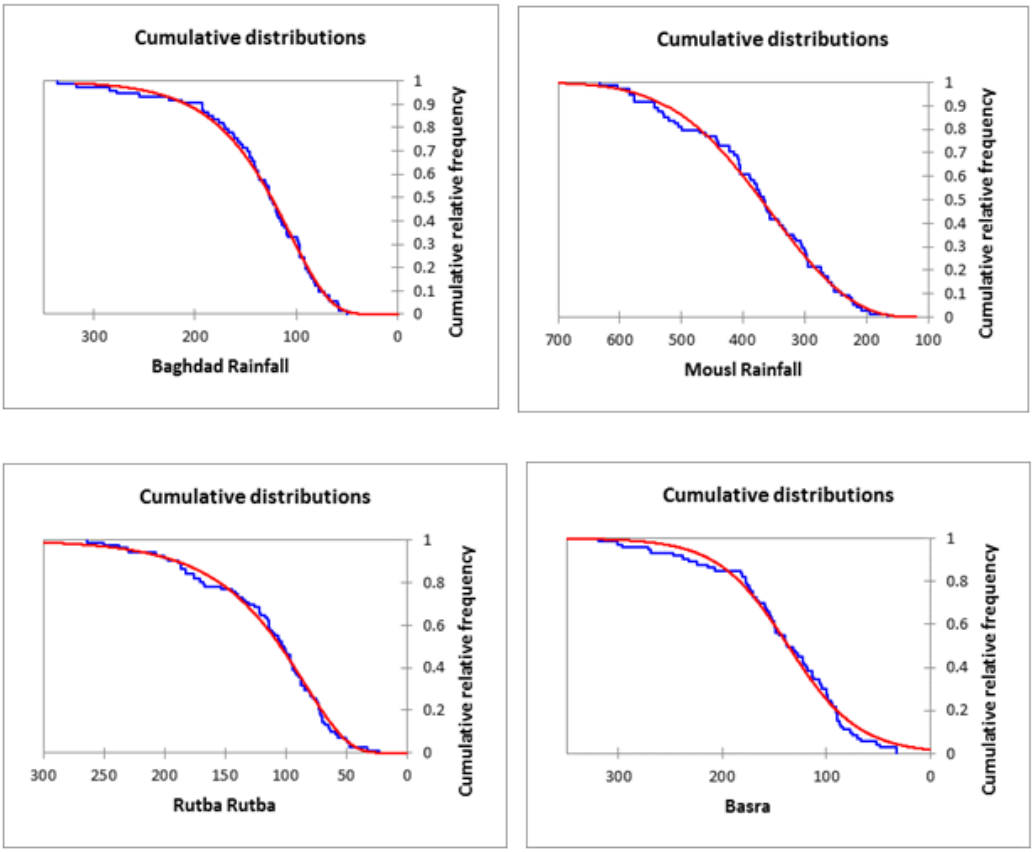

Figure 2: Explain the cumulative distribution of Rainfall data

Table 3: the parameters of distributions

stations	distributions	Parameters		
Baghdad	**Log-normal**	μ=4.826	Sigma =0.403	
Rutba	**Log-normal**	μ=4.647	Sigma =0.469	
Mosul	**Weibull (3)**	Beta= 2.501	Gamma=2890893	μ=120.162
Basra	**Logistic**	μ=136.583	Sigma=33.712	

References

[1]. Tao, D.Q, Nguyen, V.T and Bourque, A (2002). On selection of probability distributions for representing extreme precipitations in Southern Quebec. Annual Conference of the Canadian Society for Civil Engineering. 5th -8th June 2002. pp 1-8.

[2]. Suhaila Jamaludin, Abdul Aziz Jemain, Fitting the Statistical Distributions to the Daily Rainfall Amount in Peninsular Malaysia, Vol 46 (June 2007)

[3]. Olofintoye, O.O et al, Best–fit Probability distribution model for peak daily rainfall of selected Cities in Nigeria, New York Science Journal, 2009, 2(3), ISSN 1554-0200

[4]. Mohita A. S. and J. B. Singh, Use of Probability Distribution in Rainfall Analysis, School of Basic and Applied Science, Shobhit University, Meerut, Uttar Pradesh, 250110, India 2010;3(9)

[5]. Oseni, B. A. and F. J. Ayoola., Fitting The Statistical Distributions For Daily Rainfall In IBADAN, Based On Chi-Square And Kolmogorov- Smirnov Goodness-Of-Fit Tests., European Journal of Business and Management, Vol 4, No.17, 2012

[6]. IOMS, Iraqi Meteorological Organization and Seismology, http://meteoseism.gov.iq

[7]. Azhar Hadi Mahdi, Statistical Distributions of Wind Speed for Baghdad, Basra and Mosul Governorates, Eng. and tech. Vol: 31, NO: 4, 2013.

[8]. Sabereh Darbandi, Comparison of Four Distributions for Frequency Analysis of Wind Speed, Environment and Natural Resources Research, Vol. 2, No. 1; March 2012.

[9]. FAO, towards sustainable agricultural development in Iraq: The Transition from Relief, Rehabilitation and Reconstruction to Development 2003, p222.