Estimate Probability Distribution of Monthly Maximum Daily Rainfall of Iraq

Wedyan G. Nassif1, Osama T. Al-Taai2, Ali J. Mohammed3 and Hasanain K. A. AL-Shamarti4

1,2,3,4Department of Atmospheric Science, College of Science, Mustansiriyah University, Baghdad, Iraq

E-mail: osamaaltaai77@uomustansiriyah.edu.iq

Abstract. Precipitation is very important to humankind, as it represents a pure source of water beside the earth. Therefore, many researchers are interested in rain. In this study, the monthly data for the amount of rainfall for fourteen stations (Baghdad, Basra, Diwania, Kirkuk, Smaw, Tikirit, Rutb, Mousl, Alialgarbi, talafer, Kut, samara, Halî, and Nasiriya) in Iraq was collected from the General Authority for Meteorology and Seismic Monitoring, the maximum understanding of precipitation behavior, was tested the data according to tests Kolmogorov-Smirnov, Anderson, Chi-Squared, The 20 probability distributions were applied to find the appropriate distribution of this data. Was found that the Mosul station has the largest amount of rain from other stations and that this data follows the distribution Log-Gamma Which is considered the best distribution according to the tests, while Basra station takes to distribute Burr (4p), while, Rutba Station takes logistical distribution while Baghdad Log-Pearson 3 Station, despite the maximum monthly precipitation, Baghdad has more statistics than Rutba.

Keywords: Rainfall, frequency, probability, Iraq

1. Introduction

The amount of rain that is received over an area is an important factor in assessing the amount of water available to meet the different requirements of agriculture, industry and other human activities. As the rainfall factor is very important to humans because it has a major impact on life such as agriculture, urban and water resource management. Therefore, there is a lot of interest of researchers about choosing which of the distributions would be appropriate for the date of rainfall [1]. Many researchers have studied the installation of statistical distributions for monthly precipitation for some Iraqi Stations The distributions Gumbel, log Gumbel, Normal and Log- Normal distribution were used in the study. The Normal distribution model was found most appropriate for the prediction of yearly maximum daily-rainfall and the Log-Gumbel distribution model was the most appropriate for the prediction of yearly maximum daily-runoff [2] 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 [3]. In the present study, an attempt was made to determine the statistical parameters and annual one day maximum rainfall (ADMR) at various probability levels using four probability distribution functions, viz., normal, log-normal, log-Pearson type-III and Gumbel distribution (c2) test [4].
Other studies was conducted using the probability of a plot correlation coefficient (PPCC) statistics and L-moment ratio Charts are used to determine the suitability of the monthly regional distribution Rainfall data for 11 stations located in northwestern Iran. Two methods provide Pearson III is the best regional distribution of monthly precipitation data in our study area. Like Regards, PPCC test has been known as one powerful test site among many Fit is fit, but L-Moment is easy and can compare the suitability of many Distributions of many data samples using one graphical tool. [5] Other researchers use monthly precipitation data to demonstrate the relationship between rain and wind speed using the Pearson correlation coefficient [6]. In this paper, we take into account the maximum rainy day data of 14 Iraqi stations on distributions and choose which one is most appropriate according to the tests Kolmogorov-Smirnov, Chi-square and Anderson.

2. The Study Stations
In this paper employed study area is Iraq, where Geographical coordinates it is located between latitudes (29.5-37.5)°N, and between longitudes (38.45-48.45)°E in the Northern Hemisphere. Also, Iraq lies in the south west of Asian continent in the northern part of the Arab homeland, the north border is with turkey, Syria, Jordan from the west, Kuwait and Kingdom Saudi Arabia from south and Iran from the east. And this location determines the closeness or the distance of Iraq from water bodies which have clear impact in the climate and thermal properties of Iraq, where the Mediterranean Sea and the Arabian Gulf are the most influential water bodies in Iraq [7]. The climate of Iraq is a sub-tropical, continental, dry climate with dry hot summers and cool winters, with some precipitation in central and southern of country, and more precipitation in the northern of the country [8]. In this study, the European Center provided rainfall and aerosol data for the period (1980-2018) for four stations as shown in Figure 1.

![Figure 1. The study stations in Iraq](image-url)
3. Materials and Methods

3.1. Datasets
We obtained daily precipitation data for 14 stations of Iraq of period (1980 - 2018) from the Iraqi Meteorological Corporation. Table 1 shows Longitude, Latitude, Mean and standard deviations. The maximum daily precipitation data is sorted of each station which used to analysis the behavior of rainfall over study area.

| Station | Longitude (Degree) | Latitude (Degree) | Station | Longitude (Degree) | Latitude (Degree) |
|---------|--------------------|-------------------|---------|--------------------|-------------------|
| Baghdad | 44.4               | 33.3              | Rutb    | 40.28              | 33.03             |
| Basra   | 47.78              | 30.52             | Mousl   | 43.15              | 36.31             |
| Diwania | 44.95              | 31.95             | Alialgarbi | 46.43             | 32.28             |
| Kirkuk  | 44.35              | 35.74             | talafer | 42.48              | 36.37             |
| Smaw    | 45.27              | 31.27             | Kut     | 45.75              | 32.49             |
| Tikirit | 43.7               | 34.37             | samara  | 43.88              | 34.18             |
| Nasiriya| 46.23              | 31.02             | Hali    | 44.42              | 32.46             |

3.2. Statistical Used
The different Probability distributions are found to describe the variation of maximum Rainfall data which were 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, is generalized under extreme value and applied to figure out the best fit probability distribution. The description of various probability distribution functions by density, range, and the parameter are calculated. The goodness of fit function also is found by the following tests Kolmogorov-Smirnov, Anderson and the chi-square at $\alpha (0.05)$ Significant level.

The Kolmogorov-Smirnov statistic (D) is defined as the largest vertical difference between the theoretical and the empirical cumulative distribution function (ECDF) [10]

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

Where, $X_i =$ random sample, $i = 1, 2, ..., n$.

$$CDF = F_n(X) - \frac{1}{N} [\text{Number of Observations } X_n \leq (2)] \quad (2)$$

This test is used to decide if a sample comes from a hypothesized continuous distribution. Chi-Squared Test: The Chi-Squared statistic is defined as

$$\chi^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i} \quad (3)$$
Where $O_i =$ observed frequency $E_i =$ expected frequency ‘$i$’= number of observations (1, 2, …….k)
Calculated by :

$$E_i = F(X^2) - F(X_i) \quad (4)$$

$F =$ the CDF of the probability distribution being tested

The observed number of observation ($k$) in interval ‘$i$’ is computed from equation given below:

$$k = 1 + \log 2 n \quad (5)$$

$n =$ sample size

**Anderson test:**
The Anderson –Darling test is used to test if a sample of data came from population with a specific distribution [11]. The Anderson-Darling test is defined as the data follow a specified distribution $H_0$: The data do not follow the specified distribution. $H_a$: The Anderson-Darling test statistic is defined as Test Statistic:

$$H_a = -N - S A^2 \quad (6)$$

Where

$$S = \sum_{i=1}^{N} (2i-1) N [\ln F(Y_i) + \ln (1-F(Y_{N+1-i}))] \quad (7)$$

$F$ is the cumulative distribution function of the specified distribution. Note that the $Y_i$ are the ordered data.

The three goodness of fit tests mentioned above were fitted to the total maximum daily rainfall data treating different data set. The test statistic of each test were computed and tested at ($\alpha=0.05$) level of significance. Accordingly the ranking of different probability distributions were marked from 1 to 20 based on minimum test statistic value. The distribution holding the first rank was selected for the three tests independently. The assessments of all the probability distribution were made on the bases of total test score obtained by combining the entire three tests Maximum score 20 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 20. Thus, the total score of the entire three 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.

**4. Results and Discussion**

Table 2, which shows the results of the tests and parameters for the best distribution. Figure 2, 3 and 4 illustrate the density function of the distributions:

The results showed that the Baghdad station in the K.S test takes the probability distribution Log-Gamma. As for the AD and Chi-Squ test it takes the probability distribution Log-Pearson 3 and we notice that the Log-Pearson 3 distribution is the best distribution and the largest amount of precipitation in the most Rainy days is 89.1 mm and its minimum value 10 mm. As for the conductor station in the KS test, it takes the Log-Gamma probability distribution and in the AD test it takes the probability distribution Dagum and the third Chi-Squ test with the probability distribution Pearson 5, the conductor station has the largest amount of rain from the other stations is 96.2 mm and the
minimum value is 12.7 Mm, this data also follows Log-Gamma Distribution, which is the best distribution according to the three tests. The Basra station shows that the KS and AD test takes the probability distribution Burr (4P) and the Chi-Squared test takes the probability distribution Power Function. Basra station data follows that the best Burr (4P) distribution has been found to have the largest amount of rain 73.2 mm, and the lowest amount from rain 6.5 mm. Finally, in Rutba station, it is in the KS test, it takes the probability distribution Gen. Extreme Value. As for the A.D test, it takes the probability distribution Person 5 and in the Chi-Square test. So taking the probability distribution is Log-Logistic. Exceptionally, the Rutba station takes logistical distribution and the maximum precipitation is 48 mm, while the minimum is 7 mm.

Table 2: Test ranking first position parameters of distributions of Iraq stations

| Station   | Pro. Dis                  | test    | Statistic | Critical Value (0.05) | Parameter                  |
|-----------|---------------------------|---------|-----------|-----------------------|----------------------------|
| Baghdad   | Log-Gamma                 | K.S     | 0.07798   | 0.25438               | a=29.904 b=0.10364         |
|           | Log-Pearson 3             | A.D     | 0.199     | 2.5018                | a=9.0076 b=0.18883 g=1.3982 |
|           | Log-Pearson 5             | Chi-Squ.| 0.07146   | 5.9915                | a=9.0076 b=0.18883 g=1.3982 |
| Rutba     | Gen. Extreme Value        | K.S     | 0.08705   | 0.24571               | k=0.17192 s=6.6289 m=15.936 |
|           | Person 5                  | A.D     | 0.29671   | 2.5018                | a=5.3738 b=93.087          |
|           | Log-Logistic              | Chi-Squ.| 0.10906   | 5.9915                | a=3.6669 b=18.448         |
| Basra     | Burr (4P)                 | K.S     | 0.0753    | 0.2499                | a=5.3738 b=93.087         |
|           | Burr (4P)                 | A.D     | 0.21218   | 2.5018                | a=3.6669 b=18.448         |
|           | Power Function             | Chi-Squ.| 0.07671   | 9.4877                | a=3.6669 b=18.448         |
| Mosul     | Log-Gamma                 | K.S     | 0.08617   | 0.24571               | a=29.904 b=0.10364         |
|           | Log-Pearson 3             | A.D     | 0.2813    | 2.5018                | a=9.0076 b=0.18883 g=1.3982 |
|           | Pearson 5                 | Chi-Squ.| 0.23395   | 7.8147                | a=82.676 b=0.0447         |
| Diwaniya  | Log-Gamma                 | K.S     | 0.0702    | 0.24571               | k=1.0085 a=4.4892 b=40.239 |
|           | Log-Pearson 3             | A.D     | 0.1408    | 2.5018                | a=0.19949 a=1.9870E+8     |
|           | Fatigue life(3p)          | Chi-Squ.| 0.13034   | 7.8147                | a=49.024 b=0.06367        |
| Alialghri | Burr(4p)                  | K.S     | 0.07544   | 0.24571               | a=2.1179E+5 b=9.686E-4 g=202.04 |
|           | Burr(4p)                  | A.D     | 0.26148   | 2.5018                | a=0.48187 b=21.06 g=1.4344 |
|           | Burr(4p)                  | Chi-Squ.| 0.5249    | 9.9915                | a=0.19949 a=1.9870E+8     |
| Tikrit    | Gamma (3P)                | K.S     | 0.07304   | 0.24571               | a=29.904 b=0.10364         |
|           | Johnson SB                | A.D     | 0.14666   | 2.5018                | a=9.0076 b=0.18883 g=1.3982 |
|           | Burr(4p)                  | Chi-Squ.| 0.087     | 5.9915                | a=29.904 b=0.10364         |
| Nasiriya  | Log-Logistic (3P)         | K.S     | 0.23607   | 0.24571               | a=29.904 b=0.10364         |
|           | Burr(4p)                  | A.D     | 2.3583    | 2.5018                | a=2.1376E+7 g=2.1376E+7    |
|           | Gen.Gamma                 | Chi-Squ.| 6.6811E-4 | 3.8415                | a=2.9058E+8 b=1.2940E+9 g=1.2940E+9 |
| Kut       | Cauchy                    | K.S     | 0.10673   | 0.24571               | a=1.8271 a=12164.0         |
|           | Burr(4p)                  | A.D     | 0.39458   | 2.5018                | a=5.3738 b=93.087         |
|           | Exponential(2p)           | Chi-Squ.| 0.33769   | 7.8147                | a=3.6669 b=18.448         |
Hali Lognormal (3P) K.S 0.09411 0.24571 k=0.19949 a=1.9870E+8 b=6.2366E+8 g=-6.2366E+8
Fatigue Life (3P) A.D 0.33715 2.5018 k=0.19949 a=1.9870E+8 b=6.2366E+8 g=-6.2366E+8
Fatigue life(3p) Chi-Squ. 0.21613 5.9915 a=0.43822 a=6.5 b=74.346
Kirkuk Pearson 6 (4P) K.S 0.22183 0.24571 k=0.19949 a=1.9870E+8 b=6.2366E+8 g=-6.2366E+8
Inv. Gaussian (3P) A.D 1.6937 2.5018 k=1.0054 a=4.4892 b=40.239
Exponential Chi-Squ. 0.43611 5.9915 a=6.0926 b=225.56
Talafar Dagum (4P) K.S 0.18884 0.24571 k=0.19949 a=1.9870E+8 b=6.2366E+8 g=-6.2366E+8
Log-Logistic (3P) A.D 1.6817 2.5018 k=0.19949 a=1.9870E+8 b=6.2366E+8 g=-6.2366E+8
Exponential Chi-Squ. 0.34041 5.9915 a=49.024 b=0.0637
Smawa Gen. Extreme Value K.S 0.18991 0.24571 k=0.19949 a=1.9870E+8 b=6.2366E+8 g=-6.2366E+8
Burr (4P) A.D 1.2152 2.5018 k=0.19949 a=1.9870E+8 b=6.2366E+8 g=-6.2366E+8
Cauchy Chi-Squ. 0.02784 3.8415 k=0.19949 a=1.9870E+8 b=6.2366E+8 g=-6.2366E+8
Samara Gen. Extreme Value K.S 0.16241 0.24571 k=0.19949 a=1.9870E+8 b=6.2366E+8 g=-6.2366E+8
Dagum A.D 1.1285 2.5018 k=0.19949 a=1.9870E+8 b=6.2366E+8 g=-6.2366E+8
Triangular Chi-Squ. 0.00351 5.9915 k=0.12059 a=1.6545E+7 b=2.1376E+7 g=1.4344

Figure 2: Probability Density Function Kolmogorov Smirnov
Figure 3: Probability Density Function Anderson Darlin
5. Conclusions
A suitable distribution for the 14 studied stations was not found because the Kolmogorov Smirnov test shows that Log-Gamma and Log-Pearson 3 are suitable for the Mosul and Baghdad stations, respectively. While the Chi-Squared test shows that the appropriate distribution for Basra and Rutba stations is Burr (4P) and Log-Logistic, while Anderson test shows that there is no suitable distribution for any station so there is no appropriate distribution for these stations because of one of the appropriate quality tests rejected all tested distributions.

Acknowledgment
An acknowledgment to the European Center for Medium-Range Weather Forecasts (ECMWF) and Giovanni on the data used in this study, and we would also like to thanks Mustansiriyah university for providing scientific support to accomplishing this research.

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] Najm Obaid Salim Alghazali, Dhelal Adnan Hussein Alawadi, Civil Engineering Department, BabylonUniversity, Iraq; Vol.6, No.6, 2014, Civil and Environmental Research.
[3] Firas S. Basheer Journal of Applied Physics, 2278-4861.Volume 8, Issue 5 Ver. III (Sep - Oct. 2016), PP 19-22(3) (Goodness of Fitting of Annual Rainfall Data of Four Stations of Iraq).
[4] Bhim Singh*, Deepak Rajpurohit, Amol Vasishth and Jitendra Singh College of Horticulture and Forestry, MPUAT Campus, Jhalrapatan, Jhalawar - 326 023 (Rajasthan), India Probability analysis for estimation of annual one day maximum rainfall of Jhalrapatan area of rajasthan, India.
[5] Babak Amirataee and Majid Montaseri, Department of Water Engineering, University of Urmia, Urmia, Iran Journal of Urban and Environmental, Engineering, v.7, n.2 p. 247-252, 2013.
[6] Osama Tareq Al-Taai and Wedyan G. Nassif. (2020) Journal Plant Archives Vol. 20, Supplement 2, 2020 pp. 1117-1121 e-ISSN:2581-6063 (online), ISSN:0972-5210 (The Relationship between winds and Rainfall over selected stations in Iraq)
[7] Abbood, Z. M., Al-Taai, O. T. (2018) Calculation of absorption and emission of thermal radiation by clouds cover. ARPN Journal of Engineering and Applied Sciences, 13(9446-9456).

[8] Abbood, Z. M., Al-Taai, O. T. (2018) Study of Absorbance and Emissivity Solar Radiation by Clouds, Aerosols and Some Atmospheric Gases. Journal of Applied and Advanced Research, 3(5), 128-134

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

[10] Chakravarti, Laha, and Roy, (1967). Handbook of Methods of Applied Statistics, Volume I, John Wiley and Sons, pp.392-394.

[11] Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some, Comparisons Journal of the American Statistical Association, 69, pp. 730-737