Jordan Journal of Civil Engineering, Volume **14**, No. **4, 2020**

Flood Frequency Analysis of Annual Maximum Streamflows at Selected Rivers in Iraq

Ihsan F. Hasan [1)]

[1)] Lecturer, Dams and Water Resources Research Center, University of Mosul, Mosul, Iraq.
E-Mails: ihsanfh@uomosul.edu.iq; obaidyhassan@gmail.com

ABSTRACT

In this study, thirteen different probability distributions were used to analyze flood frequencies based on the annual maximum daily flood discharge data for AMF from 10 stream gauging stations in the Tigris River basin in the northeast of Iraq. Methods of moments, maximum likelihood and L-moments were used to estimate the parameters of these distributions. L-moment ratio diagram and goodness-of-fit tests were used in order to determine the best fitting probability distribution for AMF. The results showed that the Log-Pearson type-3 with method of moments, Log-Logistic type-3 with the maximum likelihood method and Generalized Extreme value with L-moments method were identified as the best-fit distributions for the selected gauging stations rather than the other distributions. A comparison was performed between the above three best-fitted distributions by using Root Mean Square Error. It has been observed that Log-Pearson type-3 with method of moments for parameter estimation was the most suitable distribution that fitted the annual maximum daily flood data for the study area.

KEYWORDS: Flood frequency analysis, Log-Pearson 3, Probability distributions, Tigris river.

INTRODUCTION

Flood frequency analysis is a technique used to estimate the flood magnitude corresponding to any required return period (Bhattarai et al., 2005). It is often used by engineers and hydrologists in the planning, design and management of various water resource projects and hydraulic structures, such as dams, spillways, weirs, barrages, irrigation projects, flood control measures,… etc. (Al-Qudah, 2011; Kordrostami et al., 2020; Mishra et al., 2010; Saf, 2009; Arystambekova and Davletgaliev, 2018).

Flood frequency analysis is generally carried out by fitting peak flow observations to a suitable probability distribution. Two approaches are mainly applied; using an annual maximum flow (AMF) and peaks-over-threshold method (POT) observations (Bezak, 2014). The AMF approach is generally favored for recurrence intervals greater than 10 years, because it is straightforward to ensure statistical independence between annual flood data and allows the analysis to better focus on low-exceedence probability events (Saf, 2009).

The probability distribution function plots the event magnitude against the proportion of events greater than or equal to that event (Selaman et al., 2007). Numerous flood frequency studies have focused on identifying appropriate probability distributions for AMF data. Commonly used probability distributions for hydrologic AMF data are: Normal, Lognormal, Exponential, Gamma, Pearson Type III, Log-Pearson Type III, generalized extreme value, Pareto and Gumbel distributions (Hu et al., 2019; Seckin et al., 2014; Haberlandt and Radtke, 2014; Ganamala and Kumar, 2017).

There are many ways of fitting frequency distributions to sample data; e.g. graphical method, method of moments, L-moments, maximum likelihood, least squares, probability weighted moments, … etc. All methods quantify parameters of statistical distributions using the sample data (Pearson, 1991). The good estimator of the parameters may be obtained by selecting the proper method of estimation (Bhuyan et al., 2010).

Received on 20/4/2020.
Accepted for Publication on 30/8/2020.

Several studies estimate the probability distribution parameters from the hydrologic data by using method of moments, maximum likelihood (ML) method and L-moment method. L-moment and L-moment ratio diagram based on commonly used frequency distributions have been used by Noto and Loggia (2009) for analysis of annual maximum peak discharge data recorded from gauging sites in Sicily, Italy, in order to derive regional flood frequency curves, where generalized extreme value distribution was identified as the robust distribution for the study area. L-moments method was also used by Kumar and Chatterjee (2005) in regional flood frequency analysis for gauged and ungauged basins.

Baratti et al. (2012) proposed an original approach to infer the flood frequency distribution at seasonal and annual time scales for the Blue Nile river at the Sudan–Ethiopia Border; the parameters of the frequency distributions are fitted by maximum likelihood of both seasonal and annual peaks. Ghorbani et al. (2010) analyzed flood frequencies by fitting eighteen different distributions to the maximum annual discharges from 6 gauging stations in the Aji river basin in Iran and the parameters of these distributions were estimated using the method of maximum likelihood and the method of moments. Engeland et al. (2018) used the maximum likelihood method to estimate the applied distribution parameters at selected catchments in Norway.

The aim of this study is to compare and evaluate different methods of flood frequency analysis for the annual maximum flood discharge of selected rivers' gauging stations in Iraq, specifically looking at the comparison of the application of the methods of moments, maximum likelihood and L-moments to find the parameters of a particular probability distribution. The study also seeks to determine the best fitting probability distributions in order to suggest the most suitable distribution based on different goodness-of-fit statistics, such as Anderson-Darling test chi-square, Kolmogorov–Smirnov, L-moment ratio diagram and Root Mean Square Error (RMSE).

STUDY AREA

Ten gauging stations from a set of seven river catchments located between latitudes 34° 06′ 01″N and 37° 08′ 00″N and longitudes 42° 41′ 00″E and 45° 45′ 00″E in the northeastern region of Iraq have been selected for this study. The selected region extends over an area of 140 000 km^2, which forms nearly 5.9% of the total geographical area of the country. The catchment areas range from 1060 km^2 for the smallest to 46700 km^2 for the largest, with a mean of 509.5 km^2. The mean elevation varies from 403 m a.s.l to 934 m a.s.l with large variations within each catchment. As a consequence, the mean annual rainfall varies between 813 mm and 3018 mm over the catchments. The location of the catchments is shown in Figure 1.

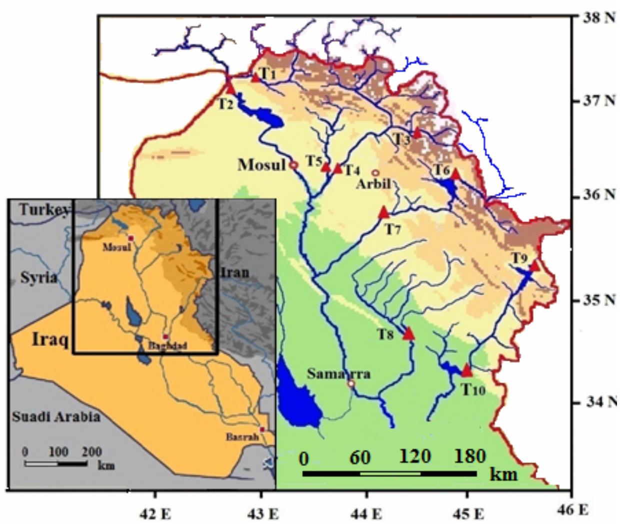

Figure (1): The location of the selected stream gauging stations

1: Zakho Station. **T2:** Tusan Station. **T3:** Balikian Station. **T4:** Eski Kalak Station. **T5:** Manquba Station. **T6:** Dokan Station. **T7:** Altun Kupri Station. **T8:** Injana Station. **T9:** Derbendikhan Station. **T10:** Diyala Station.

ANNUAL MAXIMUM FLOOD DATA ANALYSIS

Analysis of gauged data permits an estimate of the future peak discharge in terms of its probability or frequency of exceedance at a given site. This is done by statistical methods when the available data is sufficient at the site to make a meaningful statistical analysis. Bulletin 17B suggests that at least 10 years of recorded data are necessary to warrant a statistical analysis by methods which are presented herein.

In this research, from the 10 hydrometric gauging stations existing at the selected seven rivers, the available average daily instantaneous discharges (in m^3/s) were obtained. Stream flows varied over 18-56 water years in record lengths at these stations, with a mean of about 36 years. Most of the data was recorded between 1957 and 2013. The water year in Iraq begins from the 1st of October to the 30th of September. The use of recorded data from stations is limited to only those prior to the construction of a dam upstream of the site. Table 1 summarizes the main physiographic characteristics of the study Area (Younis and Hasan, 2014).

Table 1. Main characteristics of selected stream gauging stations

River	Gauging station	Latitude (N)	Longitude (E)	Elevation (m)	Catchment Area (km^2)	Record length (year)
Khabur	Zakho	37° 08′ 00″	42° 41′ 00″	440	3500	27
Tigris	Tusan	37° 04′ 00″	42° 23′ 00″	325	46700	28
Balikian	Balikian	36° 39′ 00″	44° 30′ 00″	535	1060	18
Khazir	Manquba	36° 18′ 00″	43° 33′ 00″	255	2900	37
Greater Zab	Eski-Kalak	36° 16′ 00″	43° 39′ 00″	299	20500	55
Lesser Zab	Dokan	35° 57′ 14″	44° 57′ 10″	418	7000	56
Lesser Zab	Altun-Kupri	35° 45′ 41″	44° 08′ 52″	256	13800	20
Adhaim	Injana	34° 30′ 00″	44° 31′ 00″	86	9840	35
Diyala	Derbendikhan	35° 08′ 00″	45° 45′ 00″	413	17800	47
Diyala	Diyala	34° 06′ 01″	45° 42′ 02″	119	29700	42

The annual maximum daily flood discharge data for AMF from each of the ten streamflow gauging stations has been used in the frequency analysis for the study. The annual maximum daily flood discharges *versus* time of occurrence within the hydrological year (1st of October to 30th of September) have been presented in Figure 2. One can see that the catchments have most of their annual maxima in late winter or early spring, during snowmelt and heavy rain. Controlling flood generation depends on various factors, such as catchment size, elevation distribution, precipitation climatology, … etc.

FLOOD FREQUENCY ANALYSIS METHODOLOGY

Flood frequency analysis refers to procedures that use recorded flood data to select and fit a probability distribution model to predict the magnitude of flood by using return periods at a particular location in the catchment area. All the calculations of flood frequency analysis were carried out by using Excel program and EasyFit software, version 5.5.

Probability Distributions

The proper choice of a statistical distribution which best fits the AMF data is necessary to analyze flood frequency for a particular area. In the study area, there is no fixed distribution used for flood frequency analysis of the annual maximum daily discharge series; however, some countries have adopted a standard distribution. Therefore, in this study, thirteen commonly used probability distributions in the analysis of hydrologic extreme variables; namely, Exponential 2P, Gamma, Gamma 3P, Generalized Gamma, Generalized Extreme Value, Generalized Pareto, Gumbel, Logistic, Log-Logistic (3p), Log-Pearson3, Normal, Lognormal Type III and Weibull (3P), are considered as potential candidates to find the best fitting distribution. The parameters of each distribution are estimated by using methods of moments, L-moments and maximum likelihood. Then, these thirteen distributions are compared based on goodness -of- fit tests to suggest the most suitable distribution for flood frequency analysis in each gauging station at the selected rivers. A Probability Density Function (PDF) is a continuous mathematical expression that determines the probability of a particular event (Izinyon, 2013). For more details, the essential properties and probability density functions of these distributions can be obtained from references, such as Hosking and Wallis (1997) and Bhuyan and Borah (2009).

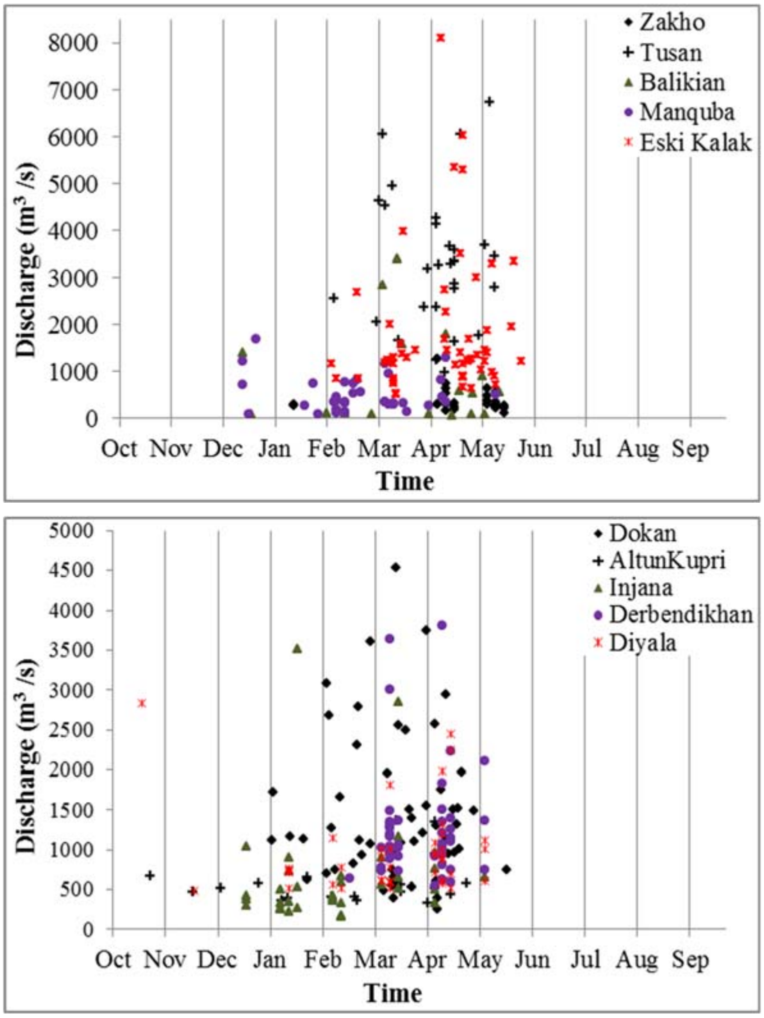

Figure (2): Annual maximum daily flood *versus* time of occurrence for the selected stream gauging stations

METHODS OF PARAMETER ESTIMATION

In this study, the methods of moments (MOM), maximum likelihood (ML) and L-moments (LMOM) are the basic methods used for distribution parameter estimation in flood frequency analysis and are explained in the following sub-sections. Table (2) shows the probability distributions and their parameter estimation methods used in this study.

Table 2. Probability distributions and their parameter estimation methods

Distribution	Symbol	Parameter Estimation Method
Exponential (2P)	EXP	Maximum likelihood
		L-moments
Gamma	GAM	Method of moments
Gamma (3P)	GAM3	Maximum likelihood
Generalized Gamma	GGAM	Maximum likelihood
		L-moments
Generalized Extreme Value	GEV	L-moments
Generalized Pareto	GPA	L-moments
Gumbel	GUM	Method of moments
		L-moments
Logistic	LOG	Method of moments
		L-moments
Log-Logistic (3p)	LLOG	Maximum likelihood
Log-Pearson Type III	LP3	Method of moments
Normal	NOR	Maximum likelihood
		L-moments
Log-Normal (3P)	LN3	Maximum likelihood
Weibull (3P)	W3	Maximum likelihood

Method of Moments

The method of moments is based on observation, noting that the parameters of probability destitution can usually be expressed in terms of the first few moments of the distribution (Chin, 2006). These moments can be estimated using sample statistics. Then, the parameters of the distribution can be calculated using the relationship between the parameters and moments. The three moments most often used in hydrology are mean, standard deviation and skewness, as shown in the following equations (Chin, 2006; Karamouz et al., 2003):

$$\hat{\mu}_x = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

$$\hat{\sigma}_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2 \tag{2}$$

$$\hat{g}_x = \frac{N}{(N-1)(N-2)} * \frac{\sum_{i=1}^{N}(x_i-\mu)^3}{\sigma^3} \tag{3}$$

where $\hat{\mu}_x$, $\hat{\sigma}_x^2$ and $\hat{g}_x$ are unbiased estimates of the mean, variance and skewness of the population. A hat (^) above a parameter indicates an estimated value.

Maximum Likelihood Method

Consider that there are n random observations $x_1, x_2, \ldots, x_n$, which have a joint probability distribution given by $fx(x_1, x_2, \ldots, x_n|\theta_1, \theta_2, \ldots, \theta_m)$, where $\theta_1, \theta_2, \ldots, \theta_m$ are the parameters and m is the number of unknown parameters. Maximum likelihood method selects the parameters that maximize the likelihood of the random observations, where:

$$L(\theta_1, \ldots, \theta_m) = \prod_{i=1}^{n} fx(x_i|\theta_1, \theta_2, \ldots, \theta_m) \tag{4}$$

Commonly, a logarithmic transformation of the likelihood function is performed; it is more convenient to maximize the natural logarithm of the likelihood function than the likelihood function itself (Chin, 2006). The logarithm of the likelihood function can also be maximized as (Karamouz et al., 2003):

$$\log(L(x)) = \log\prod_{i=1}^{n} fx(x_i|\theta_1, \ldots, \theta_m = \sum_{i=1}^{n}\log[fx(x_i|\theta_1, \ldots, \theta_m)] \tag{5}$$

For maximizing the above sum, partial derivatives of $L(x)$ with respect to each of the parameters $(\theta_1, \ldots, \theta_m)$ should be equated to zero:

$$\frac{\partial log(L(x))}{\partial \theta_1} = 0, \frac{\partial log(L(x))}{\partial \theta_2} = 0, \ldots, \frac{\partial log(L(x))}{\partial \theta_m} = 0 \quad (6)$$

Then, the above equations can be solved simultaneously to obtain the m maximum likelihood parameters $\theta_1, \theta_2, \ldots, \theta_m$.

L- Moments Method

Hosking (1990) introduced the concept of Linear-moments concept as the parameter estimation method for various probability distributions. The r^{th} Probability Weighted Moment (PWM) of the random variable X is defined by Greenwood et al. (1979) as in Equation (7):

$$\beta_r = E[X\{F_x(X)^r\}] = \int_{-\infty}^{+\infty} xF(x)^r f(x)dx \quad (7)$$

where β_r is the r^{th}- order PWM, $F(x)$ and $f(x)$ are the cumulative distribution function and probability density function of X. When $r = 0$, β_0 is equal to the mean of the distribution $\mu = E[x]$. consider a random sample size of N. First, rank the measured values as $x_1 \leq x_2 \leq \ldots \leq x_N$. Unbiased sample estimates of the PWMs, b_r, of β_r for any distribution can be computed as follows (Hosking and Wallis, 1997):

$$b_0 = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$b_1 = \frac{1}{N(N-1)}\sum_{i=2}^{N}(i-1)x_i$$

$$b_2 = \frac{1}{N(N-1)(N-2)}\sum_{i=3}^{N}(i-1)(i-2)x_i$$

$$b_3 = \frac{1}{N(N-1)(N-2)(N-3)}\sum_{i=4}^{N}(i-1)(i-2)(i-3)x_i \quad (8)$$

where N is the sample size, x is the arranged data value and i is the rank of the value in ascending order. Hosking (1990) defined r^{th} L-moments, λ_r of a random variable as a linear combination of probability-weighted moments, β_r. For any distribution, the first four L-moments are easily computed from PWMs using the following equations:

$$\lambda_1 = \beta_0$$
$$\lambda_2 = 2\beta_1 - \beta_0$$
$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$$
$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \quad (9)$$

The sample estimates of the first four L-moments are calculated by substituting b_0 to b_3 for β_0 to β_3, respectively in the above equations. The L-moments of various probability destitutions are given in terms of the parameters of the distributions by Hosking (1990) for more details. The method of L-moments obtains parameter estimates by equating the first four L-moments to the corresponding destitution parameters and then solving for destitution parameters.

The L-moment ratios; L-coefficient of Variation L-Cv (τ_2), L-coefficient of Skewness (L-Skewness) (τ_3) and L-coefficient of Kurtosis (L-Kurtosis) (τ_4) are defined by:

$$\text{L-Cv } (\tau_2) = \lambda_2/\lambda_1, \quad \text{L- Skewness } (\tau_3) = \lambda_3/\lambda_2, \quad \text{L-Kurtosis } (\tau_4) = \lambda_4/\lambda_2 \quad (10)$$

Selection of Appropriate Distributions

L-moment ratio diagram is a graph between L-kurtosis (τ_4) *versus* L-skewness (τ_3), which is useful as a guide in the selection of an appropriate probability distribution for a particular site. This diagram gives a visual indication of which distributions may be expected to give a good fit to a data sample. The distribution nearest to the coordinate of the sample LMOM ratio is taken as the best distribution to fit the actual data and the furthest is the least-fit distribution to represent the data. In this study, the L-moment ratio diagram has been employed to select the suitable probability distribution only in case of using L-moment method for parameter estimation.

Goodness -of- fit tests are used to assist in finding the distribution that best fits the given data. These tests describe the differences between the observed data values and the expected values calculated from the specific distribution in case of parameter estimation by using the methods of moments and maximum likelihood. In this study, Anderson-Darling (A-D), Kolmogorov-Smirnov (K-S) and Chi-Squared (x^2) tests were used for the goodness-of-fit tests, as described below.

Anderson-Darling test compares the fit of an observed cumulative distribution function to an expected cumulative distribution function. The A-D test is preferred, because it is a modification of the Kolmogorov-Smirnov (K-S) test and is the most powerful test available at present (Millington *et al*., 2011).

The A-D test statistic (A^2) is:

$$A^2 = n - \frac{1}{n}\sum_{i=1}^{N}\left[(2i-1)\left(lnF(x_i) + ln\left(1 - F(x_i)\right)\right)\right] \quad (11)$$

Kolmogorov-Smirnov test is based on the maximum difference between observed and expected frequencies. The test statistic (D) is:

$$D = max\ |F_o(x) - F_e(x)| \quad (12)$$

where $F_o(x)$ is the observed cumulative relative frequency and $F_e(x)$ is the expected cumulative relative frequency. The value of D must be less than a tabulated value of D at the specified confidence level for the distribution to be accepted.

Chi-squared test statistic can be estimated as follows:

$$X^2 = \sum_{i=1}^{k}\frac{(O_i - E_i)^2}{E_i} \quad (13)$$

k is determined by: $k = 1 + log_2 N$.

$\boldsymbol{O_i}$ is the observed frequency, $\boldsymbol{E_i}$ is the expected frequency and ***N*** is the sample size.

In the above **A-D, K-S** and **Chi-sq** tests, the test rejects the hypothesis regarding the distribution if the test statistic obtained is greater than the critical value at a given significance level (α). The significance level most commonly used is α=0.05 and the critical value is then compared with the test distribution statistic to determine whether it can be rejected or not.

RESULTS AND DISCUSSION

A comparison of flood frequency analyses was generally performed based on different probability distributions and the parameters were estimated using either the methods of moments (MOM), maximum likelihood (ML) or L-moments (LMOM). The agreement between the results of theoretical probability distributions and observations was measured by using L-moment ratio diagram and goodness-of-fit tests.

Figure 3 shows the L-moment ratio diagram to select the suitable probability distribution in case of parameter estimation by L-moment method. Based on the graph, the average value of the samples is closer to GEV distribution. Therefore, the GEV distribution was appropriate for the annual maximum streamflow data in the selected 10 gauging stations rather than the other distributions.

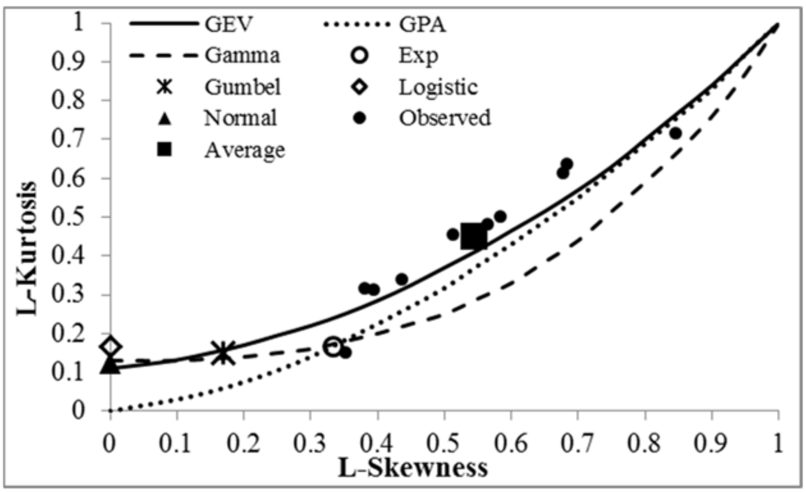

Figure (3): L-moment ratio diagrams for the 7 distributions using L-moment for parameter estimation

In case of parameter estimation by using the methods of moments and maximum likelihood, the goodness-of-fit tests based on **A-D**, **K-S** and **Chi-sq**. test statistics were used for evaluating and comparing the fitted distributions to select the most valid one. Tables 3 and 4 show the computed values of **A-D**, **K-S** and **Chi-sq.** test statistics for each distribution used in each gauging station.

Table 3. Computed values of goodness-of-fit tests using method of moments

Test	Gauging Station	Distribution			
		GAM	GUM	LOG	LP3
Anderson-Darling	Zakho	1.2831	1.3059	2.0214	0.5962
	Tusan	0.1719	0.2316	0.3104	0.16181
	Balikian	0.9858	0.9645	1.3651	0.9798
	Manquba	0.4719	0.7187	1.6612	0.4485
	Eski Kalak	3.5928	4.0325	5.5574	0.5228
	Dokan	0.4562	0.6751	2.1537	0.2039
	Altun Kupri	1.1861	1.0905	1.8258	0.2437
	Injana	2.6740	2.9189	3.7656	0.2554
	Derbendikhan	2.4058	2.2385	3.5890	0.5861
	Diyala	2.1214	2.0430	3.3676	0.4111
Kolmogorov-Smirnov	Zakho	0.1736	0.1924	0.2633	0.1485
	Tusan	0.0846	0.0954	0.1091	0.0798
	Balikian	0.2509	0.2406	0.2376	0.2085
	Manquba	0.1379	0.1695	0.2358	0.1260
	Eski Kalak	0.2018	0.2055	0.2785	0.112
	Dokan	0.0853	0.0973	0.1612	0.0586
	Altun Kupri	0.1965	0.1949	0.2668	0.1067
	Injana	0.2258	0.2210	0.2376	0.0759
	Derbendikhan	0.1766	0.1765	0.2331	0.1203
	Diyala	0.1942	0.1842	0.2161	0.0919
Chi-Squared	Zakho	1.752	4.2654	10.898	1.7433
	Tusan	0.0418	0.14775	1.2715	0.1295
	Balikian	1.9616	0.5025	0.6959	2.8196
	Manquba	2.4425	4.4477	8.9561	1.6480
	Eski Kalak	21.504	30.028	26.809	4.1233
	Dokan	4.0405	1.7612	10.184	6.0918
	Altun Kupri	2.9302	2.8426	5.6503	0.6547
	Injana	10.527	6.6165	6.7695	1.3053
	Derbendikhan	11.513	11.358	17.654	1.6271
	Diyala	4.8415	3.3799	6.5599	1.2963

Note: Shaded: best-fit distribution.

Table 4. Computed values of goodness-of-fit tests using ML method

Test	Gauging Station	Distribution						
		EXP	**GAM3**	**GGAM**	**LLOG**	**LN3**	**NOR**	**W3**
Anderson-Darling	Zakho	3.292	0.54	1.182	0.537	0.657	2.146	0.986
	Tusan	3.600	0.166	0.170	0.155	0.158	0.328	0.179
	Balikian	6.081	1.396	0.861	2.012	0.743	1.353	1.462
	Manquba	2.460	4.661	0.505	0.503	0.479	1.639	0.575
	Eski Kalak	3.121	3.302	3.169	0.469	0.632	5.894	5.312
	Dokan	3.540	0.269	0.485	0.219	0.204	2.195	0.362
	Altun Kupri	1.964	4.433	1.211	0.209	0.229	1.988	4.154
	Injana	2.432	2.960	1.985	0.172	0.281	4.348	2.789
	Derbendikhan	3.177	0.898	2.057	0.403	0.577	4.124	1.046
	Diyala	2.368	4.341	2.093	0.308	0.333	3.683	4.225
Kolmogorov-Smirnov	Zakho	0.223	0.185	0.219	0.151	0.170	0.248	0.173
	Tusan	0.228	0.083	0.085	0.077	0.078	0.121	0.079
	Balikian	0.378	0.222	0.249	0.215	0.216	0.235	0.219
	Manquba	0.183	0.200	0.153	0.118	0.126	0.218	0.158
	Eski Kalak	0.163	0.135	0.231	0.105	0.125	0.266	0.140
	Dokan	0.145	0.072	0.093	0.058	0.060	0.168	0.083
	Altun Kupri	0.133	0.148	0.218	0.094	0.109	0.265	0.112
	Injana	0.108	0.159	0.179	0.065	0.067	0.245	0.134
	Derbendikhan	0.161	0.132	0.184	0.096	0.112	0.244	0.135
	Diyala	0.105	0.094	0.163	0.086	0.087	0.222	0.088
Chi-Squared	Zakho	10.76	3.810	5.484	1.752	2.072	17.78	5.417
	Tusan	4.733	0.131	0.041	0.136	0.126	2.332	0.051
	Balikian	4.439	4.050	1.936	0.485	0.550	0.837	2.906
	Manquba	1.497	**NO**	0.824	0.335	1.614	13.31	1.505
	Eski Kalak	11.42	8.399	12.27	4.150	5.122	36.85	NO.
	Dokan	5.966	5.805	4.206	1.291	5.010	11.85	5.208
	Altun Kupri	0.395	**NO**	1.458	0.504	0.486	5.327	NO.
	Injana	1.712	4.453	7.540	1.316	0.488	6.568	4.007
	Derbendikhan	5.712	8.769	6.765	1.538	0.905	22.88	7.188
	Diyala	0.657	**NO**	3.983	2.604	4.954	7.761	NO

Note: Shaded: best-fit distribution.

Since the goodness-of-fit test statistics indicate the distance between the data and the fitted distributions, the distribution with the lowest statistic value is the very best fitting model. Except that which is marked by NO, all distributions are acceptable, because they satisfy the criteria. The hypothesis is rejected if the test statistic obtained is greater than the critical value at a significance level (α= 0.05), while the best distribution is that which has the minimum statistic value among the acceptable distributions. Based on Table 5, the LP3 distribution is best fitted for all gauging stations in case of using the method of moments. Similarly, in case of using the maximum likelihood method, the LLOG distribution is best fitted for all gauging stations, as illustrated in Table 6.

Table 5. Selection of best-fit distribution in each station using MOM based on Table 3

Station	Anderson-Darling	Kolmogorov-Smirnov	Chi-Squared	Best fit distribution
Zakho	LP3	LP3	LP3	LP3
Tusan	LP3	LP3	GAM	LP3
Balikian	LP3	LP3	GUM	LP3
Manquba	LP3	LP3	LP3	LP3
Eski Kalak	LP3	LP3	LP3	LP3
Dokan	LP3	LP3	GUM	LP3
Altun Kupri	LP3	LP3	LP3	LP3
Injana	LP3	LP3	LP3	LP3
Derbendikhan	LP3	LP3	LP3	LP3
Diyala	LP3	LP3	LP3	LP3

Table 6. Selection of best-fit distribution in each station using ML method based on Table 4

Station	Anderson-Darling	Kolmogorov-Smirnov	Chi-Squared	Best fit distribution
Zakho	LLOG	LLOG	LLOG	LLOG
Tusan	LLOG	LLOG	GGAM	LLOG
Balikian	LN3	LLOG	LLOG	LLOG
Manquba	LN3	LLOG	LLOG	LLOG
Eski Kalak	LLOG	LLOG	LLOG	LLOG
Dokan	LN3	LLOG	LLOG	LLOG
Altun Kupri	LLOG	LLOG	EXP	LLOG
Injana	LLOG	LLOG	LLOG	LLOG
Derbendikhan	LLOG	LLOG	LN3	LLOG
Diyala	LLOG	LLOG	EXP	LLOG

Table 7 contains the parameter magnitudes of the best-fitted LP3, GEV and LLOG distributions estimated by the method of moments, the maximum likelihood method and the L-moment method, respectively.

Table 7. Parameter values of best-fit distributions for the selected gauging stations

Station	GEV			LP3			LLOG		
	α	β	γ	α	β	γ	α	β	γ
Zakho	104.1	0.349	264.2	12.07	0.146	4.030	2.858	246.8	67.48
Tusan	1274	0.082	2720	14.83	-0.12	9.820	6.092	4833	-1653
Balikian	485.3	0.353	329.9	470.6	0.065	-24.8	0.717	301.3	65
Manquba	235.8	0.187	324.6	53.11	-0.11	11.56	2.227	392.2	10.05
Eski Kalak	539.4	0.452	1113	4.517	0.294	5.967	1.897	879.8	454.6
Dokan	589.8	0.171	984.7	1081	-0.02	27.66	2.521	1086	100.1
Altun Kupri	126.5	0.378	430.4	2.895	0.242	5.568	1.736	180.6	297.3
Injana	212.8	0.451	385.7	5.937	0.278	4.596	1.982	364.9	117.7
Derbendikhan	304.5	0.331	919.8	3.884	0.219	6.168	2.442	599.2	453.7
Diyala	231.5	0.372	691.9	3.429	0.242	5.923	1.449	274.8	481.9

α, β and γ are the scale, shape and location parameters.

In order to compare the best fitted distributions visually, the observed annual maximum flood data was plotted with those results from the best fitted LP3, GEV and LLOG theoretical probability distributions based on methods of moments, maximum likelihood and L-moment, as illustrated for some chosen stations in Figure 4.

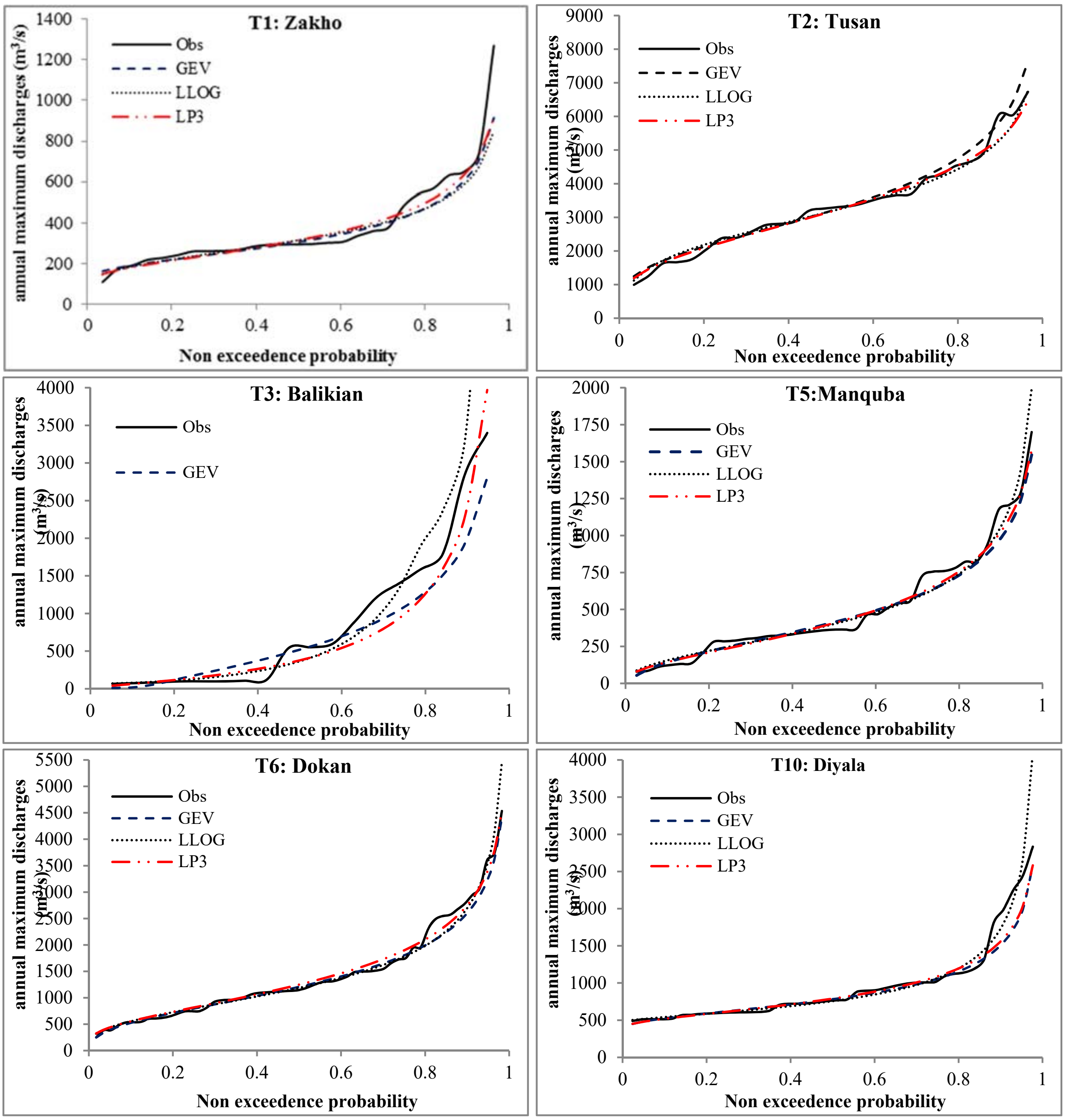

Figure (4): Observed and empirical probability distributions of annual maximum flood

Also, to evaluate the performance of the best modules, the values of Root Mean Square Error (RMSE) obtained from observed AMF data and predicted from probability distributions were used as the comparison basis among the three best-fitted distributions LP3, GEV and LLOG resulting from using methods of moments, maximum likelihood and L-moments, respectively. As presented in Table 8, lower RMSE values give a better accuracy.

Table 8. Statistical values of RMSE for the best-fitted distributions

Stations	GEV	LP3	LLOG
Zakho	**77.8**	75.60	90.7
Tusan	261.6	193.2	**194.0**
Balikian	**319.3**	285.9	331.7
Manquba	**67.2**	55.41	74.3
Eski Kalak	463.5	399.5	**403.3**
Dokan	**125.4**	97.4	168.6
Altun Kupri	114.4	90.5	**104.5**
Injana	298.0	278.3	**292.2**
Derbendikhan	208.3	200.2	**205.9**
Diyala	**148.0**	134.8	209.9

Note: Shaded: best-fit distribution. Bold: alternative candidate.

The results in Figure 4 and Table 8 showed that the LP3 distribution using the method of moments for parameter estimation provided good estimates for annual maximum flood in all the selected river gauging stations in comparison with GEV distribution method and LLOG distribution method. Therefore, the LP3 distribution is identified as the best fitting distribution for the study area.

This is in agreement with the results of previous studies conducted in the study area; for example, the study presented by Hussein and Hameed (2019), where they found that LP3 gave a good fit to the annual maximum discharge for the Adhaim river in Iraq. Sadiq and Safaa (2010) also concluded that LP3 is the best-fit distribution for the observed annual maximum flows for six streams of Tigris river in Iraq.

CONCLUSIONS

This study employed the methods of moments, maximum likelihood and L-moments to identify the probability distributions by using the annual maximum daily flood for flood frequency analysis at 10 selected gauging stations in the Tigris river basin. L-moment ratio diagram and goodness-of-fit test criteria were used to select the suitable probability distribution. In case of using L-moment method for parameter estimation, the L-moment ratio diagram illustrated in Figure 3 presents that the GEV distribution provides the best fit to the observations for all the selected stations. The computed values of A-D, K-S and Chi-sq. test statistics shown in Tables 3, 4, 5 and 6 indicate that the LLOG distribution provides the best fit to the observations for all the selected stations in case of using maximum likelihood method, while the LP3 distribution provides the best fit to the observations for all the selected stations in case of using the method of moments.

Overall, according to the comparison results based on the RMSE values presented in Table 8, LP3 is the most appropriate probability distribution for the observations of the annual maximum daily flood for all the selected stations in the first rank, with the GEV distribution for (T1, T3, T4, T6 and T10) stations and LN3 distribution for (T2, T5, T7, T8 and T9) stations as alternative candidates.

According to the Chi-sq. test statistics shown in Table 4, using GAM3 distribution for **T5**, **T7** and **T10** stations and w3 distribution for **T4**, **T7** and **T10** stations is rejected, because the test statistic values obtained are greater than the critical values at 5% significance level.

Probability distributions identified in the present study will be useful for Iraqi hydrological engineers and practitioners to choose the distribution type of annual maximum flood in their future water resource engineering practices. Differences in flood frequency estimation may result from the use of different parameter estimation methods and different goodness-of-fit test statistics.

REFERENCES

Al-Qudah, K.A. (2011). "Floods as water resource and as a hazard in arid regions: A case study in southern Jordan". Jordan Journal of Civil Engineering, 5 (1), 148-161.

Arystambekova, D., and Davletgaliev, S. (2018). "Impact of reservoir construction on flood characteristics of the rivers of Kazakhstan during spring season". Jordan Journal of Civil Engineering, 12 (2), 314-322.

Baratti, E., Montanari, A., Castellarin, A., Salinas, J.L., Viglione, A., and Bezzi, A. (2012). "Estimating the flood frequency distribution at seasonal and annual time scales." Hydrol. Earth Syst. Sci., 16, 4651-4660.

Bezak, N., Brilly, M., and Sraj, M. (2014). "Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis." Hydrological Sciences Journal, 59 (5), 959-977.

Bhattarai, K.P. (2005). "Flood frequency analysis of Irish river flow data using variants of L-moments". National Hydrology Seminar.

Bhuyan, A., and Borah, M. (2009). "Best fitting probability distributions for annual maximum discharge data of the river Kopili, Assam". Journal of Applied and Natural Science, 1 (1), 50-52.

Bhuyan, A., Borah, M., and Kumar, R. (2010). "Regional flood frequency analysis of north bank of the river Brahmaputra by using LH-moments". Water Resour. Manage., 24, 1779-1790.

Chin, D.A. (2006). "Water-resource engineering". 2nd Edition. Pearson Prentice Hall, Upper Saddle River, New Jersey.

Engeland, F., Wilson, D., Borsányi, P., Roald, L., and Holmqvist, E. (2018). "Use of historical data in flood frequency analysis: A case study for four catchments in Norway". Hydrology Research, 49 (2), 466-486; doi: 10.2166/nh.2017.069.

Ganamala, K., and Kumar, P.S. (2017). "A case study on flood frequency analysis". International Journal of Civil Engineering and Technology, 8 (4), 1762-1767.

Ghorbani, M.A., Ruskeep, H., Singh, V.P., and Sivakumar, B. (2010). "Flood frequency analysis using Mathematica". Turkish J. Eng. Env. Sci., 34, 171-188.

Greenwood, J.A., Landwehr, J.M., Matlas, N.C., and Wallis, J.R. (1979). "Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form". Water Resour. Res., 15 (5), 1049-1054.

Haberlandt, U., and Radtke, I. (2014). "Hydrological model calibration for derived flood frequency analysis using stochastic rainfall and probability distributions of peak flows". Hydrol. Earth Syst. Sci., 18, 353-365.

Hosking, J.R.M. (1990). "L-moments: analysis and estimation of distributions using linear combinations of order statistics". J. Royal Stat. Society B., 52, 105-124.

Hosking, J.R.M., and Wallis, J.R. (1997). "Regional frequency analysis: An approach based on L-moments". New York: Cambridge University Press.

Hu, L., Nikolopoulos, E.I., Marra, F., and Anagnostou, E.N. (2020). "Sensitivity of flood frequency analysis to data record, statistical model and parameter estimation methods: An evaluation over the contiguous United States". J. Flood Risk Management, 13, e12580. https://doi.org/10. 1111/jfr3.12580

Hussein, M.H., and Hameed, I.A. (2019). "An analysis of characteristic discharges for the Adhaim river (Iraq) in multi-year period". River Research and Applications, 35 (4), 375-385.

Izinyon, O.C., and Ajumuka, H.N. (2013). "Probability distribution models for flood prediction in Upper Benue river Basin-Part II". Civil and Environmental Research, 3 (2), 62-74.

Karamouz, M., Szidarovszky, F., and Zahraie, B. (2003). "Water resource system analysis". Lewis Publishers, CRC Press, LLC. ISBN 1-56670-642-4.

Kordrostami, S., Alim, M.A., Karim, F., and Rahman, A. (2020). "Regional flood frequency analysis using an artificial neural network model". Geosciences, 10, 127; doi:10.3390/ geosciences10040127

Kumar, R., and Chatterjee, C. (2005). "Regional flood frequency analysis using L-moments for North Brahmaputra region of India". J. Hydrol. Eng., 4 (3), 240-244.

Millington, N., Das, S., and Simonovic, S.P. (2011). "The comparison of GEV, Log-Pearson type 3 and Gumbel distributions in the Upper Thames river watershed under global climate models". Water Resources Research Report No. 077, 1913-3200.

Mishra, B.K., Takara, K., Yamashiki, Y., and Tachikawa, Y. (2010). “An assessment of predictive accuracy for two regional flood frequency estimation methods”. Annual Journal of Hydraulic Engineering, 54, 7-12.

Noto, L.V., and Loggia, G.L. (2009). “Use of L-moments approach for regional flood frequency analysis in Sicily, Italy”. Water Resour. Manage., 23, 2207-2229.

Pearson, C.P. (1991). “New Zealand regional flood frequency analysis using L-moments”. Journal of Hydrology, NZ, 30 (2), 53-64.

Sadiq, S.M., and Safaa, N.H. (2010). “Comparing of standard and recent extreme probable distributions corresponding to Iraqi stream flow”. Journal of Engineering and Development, 14 (4), 93-108.

Saf, B. (2009). “Regional flood frequency analysis using L-moments for the west Mediterranean region of Turkey”. Water Resource Manage., 23, 531-551.

Seckin, N., Yurtal, R., and Haktanir, T. (2014). “Regional flood frequency analysis for gauged and ungauged catchments of Seyhan river basin in Turkey”. Journal of Engineering Research, 2 (1), 47-71.

Selaman, O.S., Said, S., and Putuhena, F.J. (2007). “Flood frequency analysis for Sarawak using Weibull, Gringorten and L-moments formulae”. Journal of the Institution of Engineers, Malaysia, 68 (1), 43-52.

Younis, A.M., and Hasan, F.I. (2014). "Prediction of flow duration curve for seasonal rivers in Iraq". Jordan Journal of Civil Engineering, 8 (1), 29-42.