Generation Rainfall Intensity Equations for Intensity Duration Frequency (IDF) Curves (Case Study: Salah Al-Din, Iraq)

Asmaa Abdul Jabbar Jamel[ID], Zainab Thair Dawood*[ID]

Civil Engineering Department, College of Engineering, Tikrit University, Salah Al-Din 34001, Iraq

Corresponding Author Email: ZT230014en@st.tu.edu.iq

ABSTRACT

The intensity of rainfall can be considered an important influence in designing and operating hydraulic structures. The intensity-duration-frequency (IDF) curves are very important for planning, managing, operating, and designing all water resource projects. The current study aimed to derive the curves of IDF and equations for the stations (Tikrit, Samraa, Baiji, and Tuz) in Salah Al-Din/Iraq. Using the maximum daily rainfall during the period between 1990 and 2022, developed the empirical equations for estimating rainfall intensity with the various rainfall durations and different return periods (IDF equations), by using three methods of frequency distribution techniques (Gumbel, Log Pearson III, and Log-Normal). By finding all the missing rainfall data using the homogeneity curve expectation maximization (EM) algorithm adopted in SPSS, and checking the consistency of the data using the double mass curve technique. The results showed that the rainfall intensity reduced as the duration increased, while rainfall of any duration showed a higher intensity if the return period of the rainfall was large. A comparison among the three distributions was made using the methods of testing goodness (Chi-Square, Anderson-Darling, and Kolomokorov-Simornov) using Easy Fit software 5.6. The test results proved that the Log Pearson III distribution was the best method for the study area having correlation coefficients 0.92, 1, 0.96, and 0.92 for Tikrit, Samraa, Baiji, and Tuz stations, respectively. Also, Bernard's equation with an error ratio of ($\alpha < 0.05$), can be adopted as a general empirical equation for all hydraulic projects in the study area.

1. INTRODUCTION

Over the past few decades, a lot of research has been done on the relationship between rainfall intensity, duration, and frequency. Engineers utilize rainfall intensity-duration-frequency (IDF) curves as their primary data source to forecast rainfall intensity. IDF curves are a probabilistic tool that may be used in planning and design studies. They make it possible to evaluate the extreme features of rainfall and offer a straightforward way to convey information about local extremes. IDF curve values are frequently the basis for urban drainage design. Urban drainage design is often based on the values provided through IDF curves. This is what was observed in the studies adopted in the current paper, examples of which include the studies [1, 2]. The large increase in population that occurs as a result of urbanization and the development of infrastructure has made many areas throughout Iraq vulnerable to the risks of severe floods that may occur in exceptional circumstances. Statistics and evaluation of heavy rainfall data are very essential in planning and managing water resources projects to design drainage systems and determine the necessary drainage capacity of the channels. It is therefore important to prevent flooding and thus reduce losses and risk assessment in various weather conditions for the successful implementation of infrastructure

projects associated with highways, bridges, airports, city water supply systems, railway lines, small irrigation projects, etc. Studies based on the intensity, duration, and frequency of rainfall have received a lot of attention in previous years [3]. It has become necessary to analyze and understand rainfall behavior due to the increase in urban areas during the last several decades. This requires information on both the volume of rainfall on the surface and the distribution of rain. For the construction of hydraulic structures, return period analysis of short-duration rains is typically employed. On the other hand, urbanization, industry, a shift in lifestyle, and the green revolution are all contributing to the daily rise in water demand [4]. Therefore, it has become necessary to design economical and, at the same time, safe facilities to control these floods using intensity, duration, and frequency (IDF) curves. Which represents the mathematical relationship related to the period of return of rainfall, its intensity, and duration. Zeri et al. [3] in Iraq, Dang [5] in Vietnam, Sangüesa et al. [6] in central Chile, Alramlawi and Fıstıkoğlu [7] in Turkey, and Refaey et al. [8] in Egypt generated IDF curves, and these study results were: Zeri et al. [3] established IDF curves using the Sherman equation for the major cities in Iraq based on observed rainfall data from 2000 to 2022; Dang [5] adopted 0.25 to 8 h with return periods from 2 to 100 years; Ca Mau City in Vietnam derived IDF curves using GEV distribution. Sangüesa et al. [6]

and Alramlawi and Fıstıkoğlu [7] used the annual maximum rainfall series of the study area, which was sampled from the daily downscaled rainfall series. The sampled daily maximum rainfalls were then bias-corrected. The Gamma distribution was best suited to represent the rainfall data for the Al Qusir weather station in Egypt. Ogbozige [9] in Port Harcourt/Nigeria developed the IDF curves using Gumbel, Pearson type III, and LP III. The results of this study showed that the Gumbel distribution is the best distribution for the catchment with the highest (R^2= 0.9865) compared to the distributions Pearson type (III) and LP (III), which have R^2 values of 0.9766 and 0.982, respectively. There was no significant difference between the rainfall intensities predicted from all the IDF equations and those observed in the field. The study was based on t-test analysis and the result showed that ($p < 0.01$), Sherman's equation was adopted in this research. Basumatary and Sil [10] in Barak River Basin/ India generated the rainfall intensity duration frequency curves, this study estimated the maximum rainfall focused on Gumble, LN, and LP III distributions. The goodness fit tests indicated that the LP III method was suitable for the study area with R^2 above 90%. The study used the Bernard equation to estimate the empirical equation and Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Chi-Squared (X^2) tests for the goodness testing by using Easy Fit software. Ewea et al. [11] in the Kingdom of Saudi Arabia derived IDF curves using Gumbel distribution. The goodness of fit shows strong correlations ranging between 0.99 and 0.98 for one of the parameters. Hamaamin [12] developed IDF curves for rainfall in Sulaimani City- Iraq based on the Bernard equation, with a determination coefficient ($R^2 = 1$) and a maximum value of chi-square = 0.744 for a return period of 2 years. Haji [13] in Nablus/ Palestine, the results of this paper proved that Gumbel distribution fits the data and can be adopted for future estimations. Also, the maps for the catchment were developed from the IDF curves. Hussain [14] derived the IDF equation for rainfall at Baiji station in Iraq using Gumbel and LP III distributions. The Weible approach was used to assess and test the maximum rainfall intensity data for the station. Results showed that the optimum distribution was LP III for rainfall intensity with durations of 15-60 min, while the Gumble distribution was the optimum type for rainfall intensity with a duration of 30 min. Since estimations and determinations of rainfall density are necessary for the planning and management of water resources worldwide, the assessment of the relationships is necessary for such intensity. So, the present research aims to develop frequency distribution curves of rainfall data for four stations in Salah Al-Din governorate for years (1990-2022) and to develop a suitable formula for estimating rainfall intensity, taking into account the various rainfall durations and different return periods, as well as to find some important variables for water resource designs and find the best curve fitting that is necessary and that will facilitate research for subsequent researchers.

2. STUDY AREA AND CITY CLIMATE

Salah al-Din is one of the Iraqi governorates, located north of the capital, Baghdad, far from it within limits of 165 km, and its center is the city of Tikrit. It has coordinates of latitude 34° 32′ 1.51" N and longitude 43° 29′ 1.46" E. It has an area of 25,807 km^2, see Figure 1. Therefore, it represents a percentage of (5.6%) of the total area of Iraq. The governorate has a population of 1,237,059 people, according to the 2003

United Nations census. This study area is generally characterized by rainy winters and dry summers, but in terms of the amount of rainfall, it varies according to its different terrain, growing northwards and less southwards. The governorate includes eight districts; the current study area covers the stations (Tikrit, Samarra, Baiji, and Tuz). Table 1 shows the coordinates of the stations of the study area. Figure 2 shows the location of the study area.
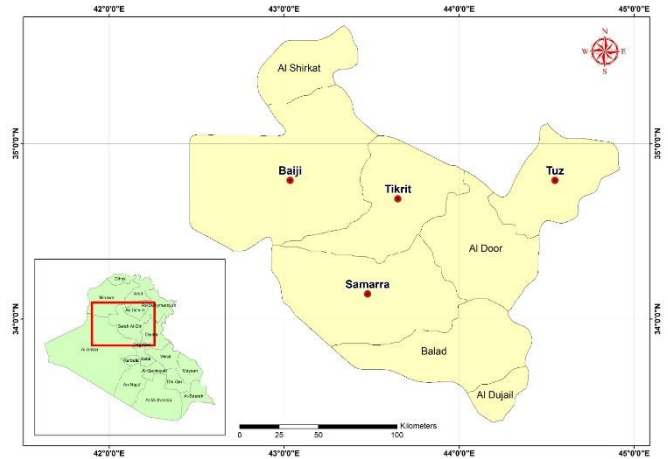

Figure 1. The location of the study area

Table 1. The stations coordinate

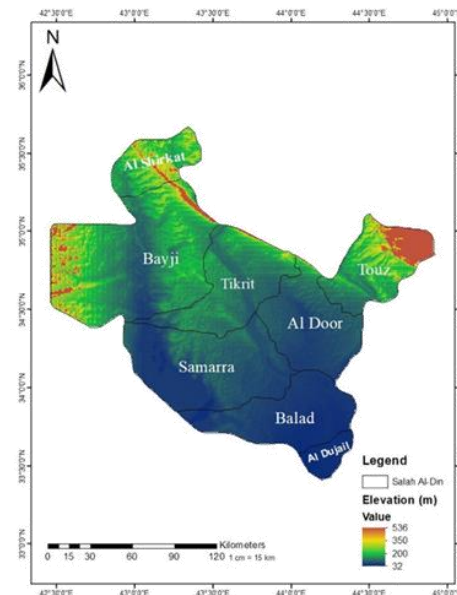| Station Name	Longitude	Latitude
Tikrit	43° 39' 0.27"	34° 41' 3.07"
Samarra	43° 28' 36.65"	34° 8' 30.16"
Baiji	43° 2' 4.78"	34° 47' 20.53"
Tuz	44° 32' 53.25"	34° 47' 20.53"

Figure 2. Elevation map for the study area

3. METHODOLOGY

3.1 Data collection and analysis

By obtaining data from different climate stations for the study area from the General Authority of Meteorology and

Seismology. Stations represent the most extreme climate compared to the other districts. Tikrit and Baiji stations had complete field rain data for a period of 33 years (1990-2022) with very few missing data. As for the Tuz district station, rain data was obtained for 20 years for the period 2003-2022, Samarra station had the period 2003-2012, and the station was permanently closed in 2012. Available rainfall data were analyzed to determine the maximum daily rainfall for each year.

To analyze the IDF curves of rainfall sums at continual duration, it is necessary to get the best-fit probability among the hypothetical distributions. There are essential steps that must be applied to develop the IDF curves for the current study. Figure 3 shows the flowchart, which explains the steps that were followed in the current research to draw IDF curves and then find the appropriate equation for the study area.

3.2 Rainfall duration reduction formula

Through the maximum selected rainfall data, the intensity of rainfall for different return periods can be derived using the Indian meteorological equation (IMD), which is an empirical equation used to estimate the intensity of rainfall for short durations of time, as shown in Eq. (1). The IMD formula has been used in previous research papers; they found that this equation has high accuracy in performance and always gives the best estimation for short durations of rainfall [15, 16].

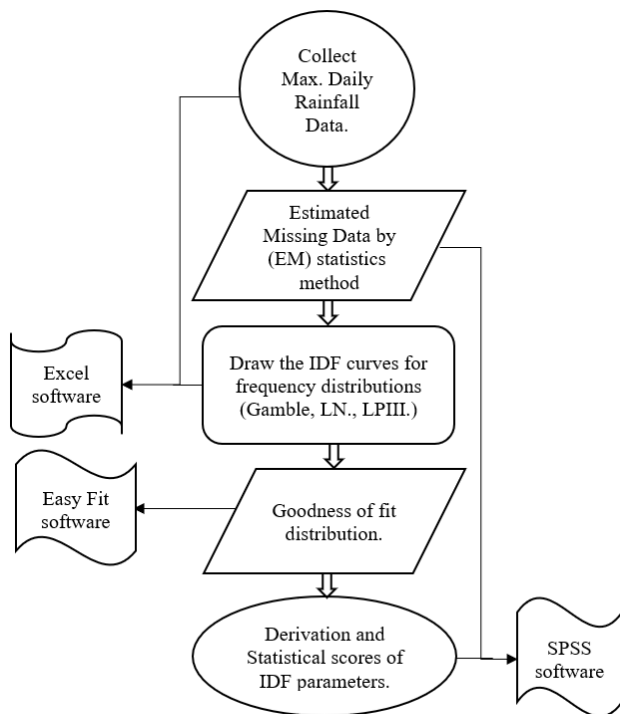$$P_t = P_{(24)} \left(\frac{t}{24}\right)^{\frac{1}{3}} \tag{1}$$

Figure 3. Methodology flowchart for the research

3.3 Missing rainfall data estimation

To compensate for missing data from field stations, it is possible to use one of the statistical methods to find this data. In this study, an iterative method has been used to estimate mean values and covariance matrixes from unavailable data, which is the expectation maximization (EM) algorithm

method. It is a method that uses the probability amplification of the available data, through which probability distribution coefficients are found. This method also depends on the assumption that the missing data are: A set of data is missing at random, meaning it is not affected by certain factors. The data parameters in the mentioned method can be determined using the mean and covariance matrix [16, 17]. Missing data was found for stations by adopting the expectation maximization (EM) method using the SPSS program, and the results were very reasonable and close to the data collected in the field.

3.4 Homogeneity tests

The homogeneity test was conducted to show the extent of homogeneity of the data after the missing data was found. In this test, the monthly rainfall data for the stations (Tikrit, Samarra, Baiji, and Tuz) was used based on the Pettitt test utilizing (XLSTAT) software as shown in Figure 4. In this test, the null hypothesis (H0) was applied when the data were homogeneous, and the alternative hypothesis was applied (Ha) when the data were changed. It means when the calculated value is greater than the significance α level = 0.05, the null hypothesis (H0) will be accepted. According to the Pettitt test, the results showed that Samraa station has the highest p-value while Tuz station has the lowest p-value. Since p-values for all stations are more than 0.05, which means that the null hypothesis cannot be rejected [18-20].
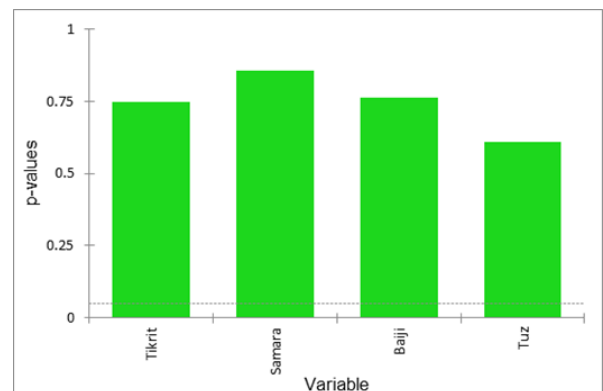

Figure 4. p-values of the tested stations

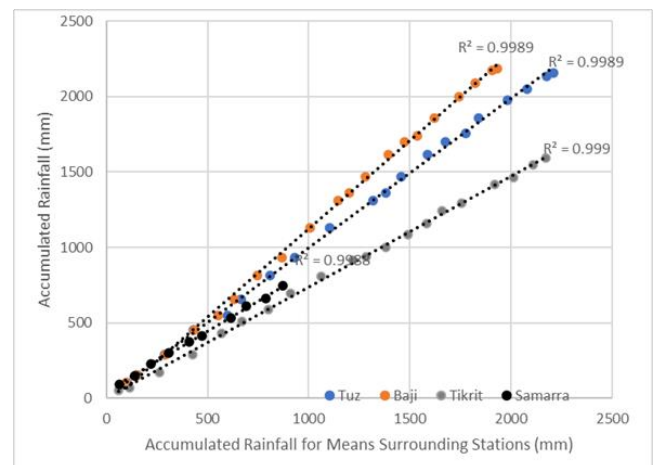### 3.5 The consistency of data

Figure 5. The consistency of data test using double-mass curve technique

A double mass curve is used to check the consistency of the data collected for stations. In this research, rainfall data were used to test the first assessment of consistency using a double mass curve [21, 22], where Figure 5 represents the consistency of the study area with the accumulation of annual rainfall for each station, and depending on the cumulative average annual rainfall for the mean of the surrounding stations, the curves show an approximate straight line. The consistency analysis of the stations adopted in this study confirms that the stations were internally consistent, meaning that the data can be used for statistical analysis.

4. THEORIES OF DISTRIBUTION

The analytical techniques (Gumbel, Log Pearson Type III, and Log-Normal) in this paper were adopted after studying their adoption only among other techniques by the latest research in domains adjacent to the study area and similar in climate to it, and their adequacy and success have been proven, such as [22-25].

4.1 Gumbel distribution

One of the most well-known distributions is the Gumbel distribution, which applies the maximum values of rainfall data for various return periods and at different durations. The Gumbel distribution is the most widely used distribution for IDF analysis due to its eligibility for modeling maximal. It is generally clear and used with ultimate events (peak rainfall) [26]. The method mentioned above is represented by the following Eq. (2) [19, 21, 22]:

$$P_t = P_{ave.} + K_T S \qquad (2)$$

where, $P_{ave.}$ is the average of the maximum precipitation obtained by the Eq. (3):

$$P_{ave.} = \frac{1}{n} \sum_{i=1}^{n} P_i \qquad (3)$$

$$K = \frac{-\sqrt{6}}{\pi} \left(0.5772 + \ln\left(\frac{T}{T-1}\right) \right) \qquad (4)$$

$$S = \left[\frac{1}{n-1} \sum_{i=1}^{n} (P_i - P_{ave.})^2 \right]^{\frac{1}{2}} \qquad (5)$$

Then the rainfall intensity I_T (mm/h) for the return period T can be obtained from Eq. (6):

$$I_t = \frac{P_t}{d} \qquad (6)$$

4.2 Log Pearson type III (LP III)

Similar to the Gumbel model, LP III is used for estimating the frequency intensity. The average and standard deviation of the parameters must be converted to logarithmic data. The LP III probability method is utilized to create different rainfall durations and return periods of rainfall intensity, which produce the IDF curves for the area of study. It is widely applied since its skew parameter allows a better fit to data series where other distributions fail [27].

Eqs. (7)-(10) are the abbreviated formulas for the distribution [19, 22, 27]:

$$P = \log(P_i) \qquad (7)$$

$$P_T = P_{ave.} + K_T S \qquad (8)$$

$$P_{ave.} = \frac{1}{n} \sum_{i=1}^{n} P \qquad (9)$$

$$S = \left[\frac{1}{n-1} \sum_{i=1}^{n} (P_I - P_{ave.})^2 \right]^{\frac{1}{2}} \qquad (10)$$

In this distribution, K_T depends on the return period (T) and the skewness coefficient, which can be determined by the Eq. (11):

$$Cs = \frac{n \sum_{i=1}^{n}(P_i - P_{ave.})^3}{(n-1)(n-2)(S)^3} \qquad (11)$$

P_T, S, and $P_{ave.}$ are, as in the Gumbel method, determined by logarithmically transformed values. K_T represents the frequency factor which can be obtained from tables of K_T values, such as from the table (7.6) of LP III [27]. By expressive Cs and T, K_T for LPT III distribution can be determined. The determination in Eq. (8) will give the maximum value for the exact return period.

4.3 Lognormal distribution theory (LN)

For this method, the frequency factor is computed as in the LP III distribution. The extreme value of intensity must be converted to logarithmic values. So, Eq. (2) is used to obtain the value of extreme intensity. K_T can be obtained from the table (7.6) in the study [20, 27]. The lognormal distribution assumes that the hydrologic quantity distribution forms a lognormal distribution. It applied this method to California frequency analysis, and this method was evaluated to a high degree. This method transforms the peak flow data using a logarithm [12, 26, 28, 29].

5. RESULTS AND DISCUSSION

5.1 Generation IDF curves

In this study, IDF curves were generated for rainfall durations (10, 20, 30, 60, 120, 180, 360, 720, and 1440 min) for return periods (2, 5, 10, 25, 50, and 100 years) for the study area by three distributions. Figures 6-17 show the IDF curves for the four stations.

5.2 The goodness of fit test

The purpose of the goodness-of-fit test is to assess how well the observed frequency in a sample matches the expected frequency derived from the hypothesized distribution. The software tool Easy Fit 5.6 is used to behavior goodness tests, including Chi-squared, Anderson-Darling, and Kolmogorov-Smirnov tests. The minimum value gained from these tests is that the predicted frequencies match the observed frequencies [19]. Table 2 shows the goodness tests used in the current study. It observed frequencies are so close to corresponding expected values, then it is a good fit; otherwise, it is a bad fit. Table 3 shows the results based on the above statistical three methods (Kolmogorov-Smirnov, Anderson-Darling, and Chi-Squared) by using Easy Fit software to perform goodness-of-fit tests.
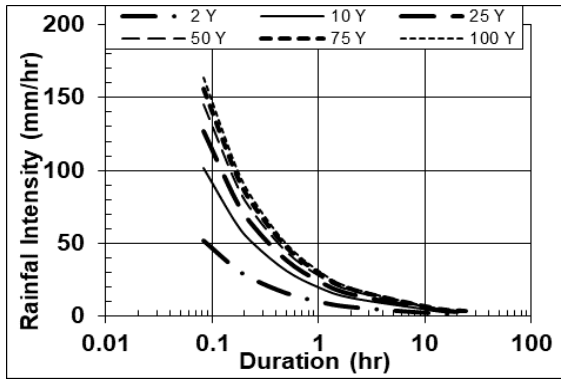
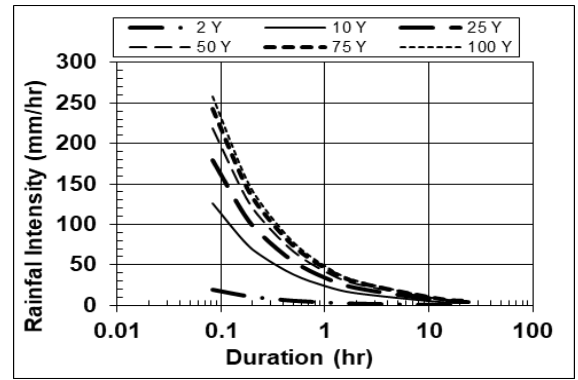**Figure 6.** IDF curves for Tikrit station using Gumbel distribution

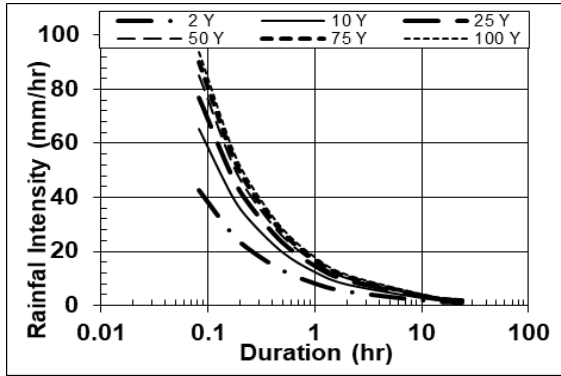**Figure 10.** IDF curves for Tikrit station using LP III distribution

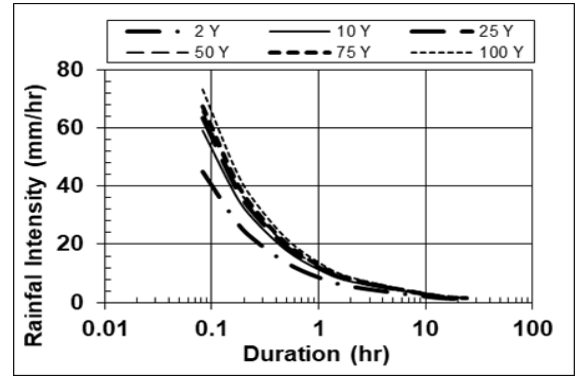**Figure 7.** IDF curves for Samraa station using Gumbel distribution

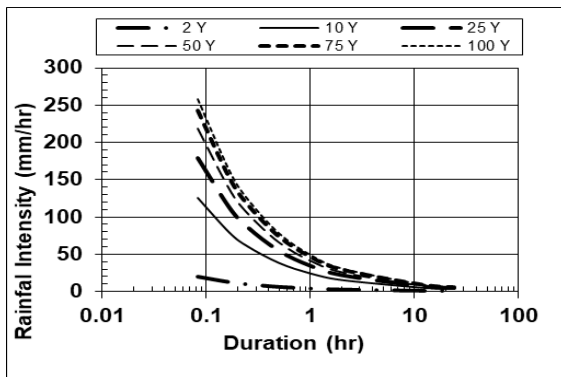**Figure 11.** IDF curves for Samraa station using LP III distribution

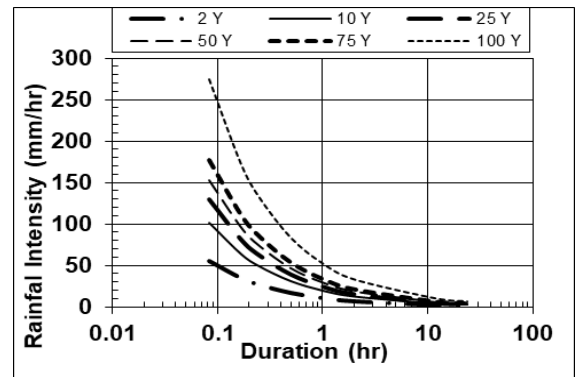**Figure 8.** IDF curves for Baiji station using Gumbel distribution

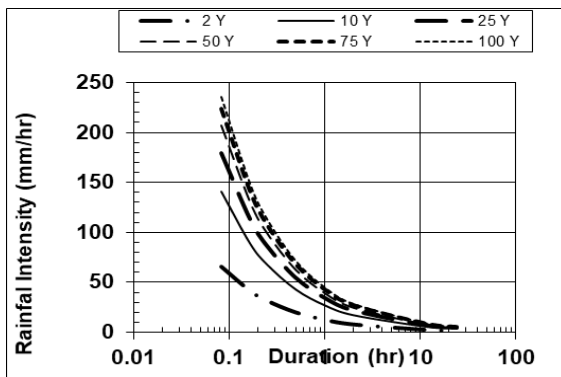**Figure 12.** IDF curves for Baiji station using LP III distribution

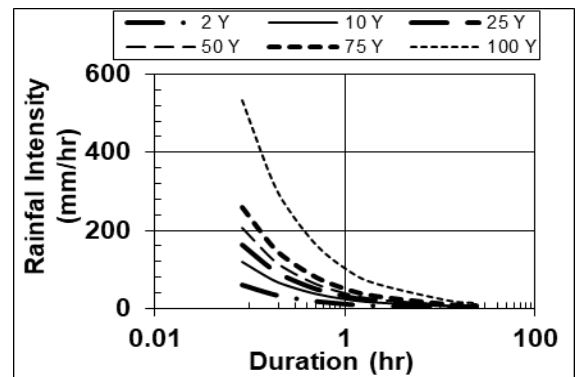**Figure 9.** IDF curves for Tuz station using Gumbel distribution

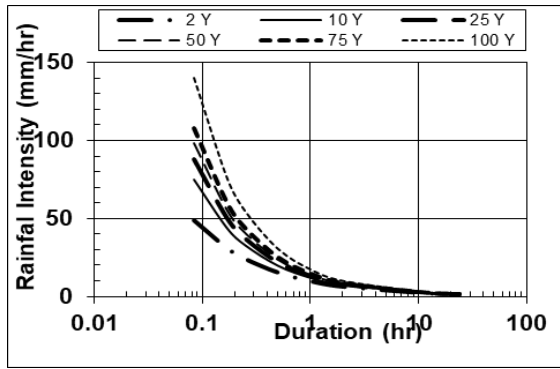**Figure 13.** IDF curves for Tuz station using LP III distribution

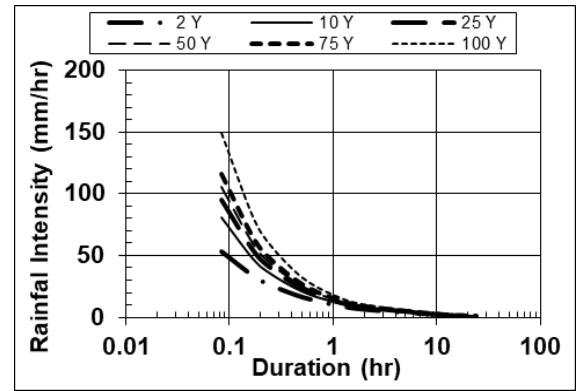**Figure 14.** IDF curves for Tikrit station using LN distribution

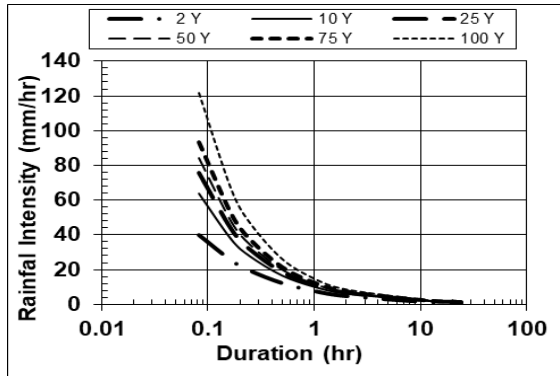**Figure 16.** IDF curves for Baiji station using LN distribution

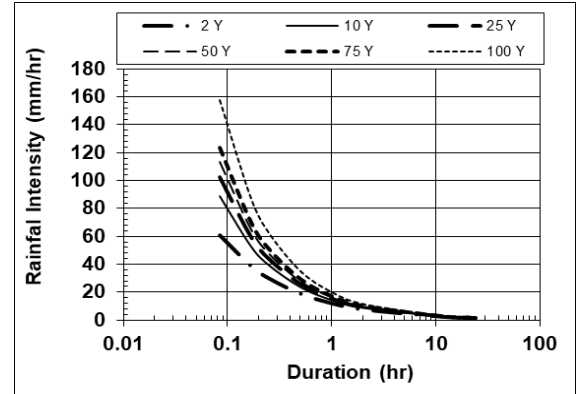**Figure 15.** IDF curves for Samraa station using LN distribution

Figure 17. IDF curves for Tuz station using LN distribution

Table 2. Summary of the goodness of fit tests

The Goodness of Fit Test	Equations	Definitions
Chi-squared	$x^2 = \frac{\sum_{i=1}^{k}(Oi-Ei)^2}{E_i}$ (12)	This test is combined to compare experimental and observed values. Observed values are the rainfall intensities obtained from the distributions, and experimental values represent the rainfall intensities calculated from the empirical formula. The value of chi-squared will be small if the experiential frequencies are close to the equivalent expected frequencies; it is considered a good fit, which leads to acceptance; otherwise, it is a bad fit, which leads to rejection [30].
Kolmogorov-Smirnov (K-S)	$P_{xi} = (\frac{m^*}{n}) + 1$ (13) $\Delta = P_{xi} - F$ (14)	This test is based on a statistic that measures the deviation of the observed cumulative histogram from the hypothesized cumulative distribution function. Relying on the values saved in the Easy Fit 5.6 tables to obtain the tabular value (Δo) of the Kolmogorov-Smirnov statistics for a given degree of probability. We conclude that if the value of the statistic (Δ) is less than the value of (Δo), this means that the distribution is accepted as fitting the assumed probability level [30].
Anderson - Darling (AD)	$A^2 = -n - s$ (15) $S = \sum_{k=0}^{n} \frac{2k-1}{n} [lnF(Y_K) + ln\{(Y_{n+1-k})\}]$ (16)	At the adopted level of significance (α), the theory related to the distributional method is rejected if the test statistic (A^2) is > the critical value [30].

Table 3. Description of the goodness of Fit Tests for stations

Station	Distribution	Kolmogorov-Smirnov		Anderson- Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
Tikrit	LP III	0.07	1	0.18	1	0.50	1
	LN	0.14	2	0.66	2	2.71	2
	Gumbel	0.18	3	1.36	3	3.79	3
Samraa	LP III	0.17	1	0.22	1	N/A*	1
	LN	0.24	2	0.45	2	N/A*	2
	Gumbel	0.26	3	0.73	3	N/A*	3
Baiji	LP III	0.07	1	0.20	1	1.63	1
	LN	0.08	2	0.29	2	1.88	2
	Gumbel	0.10	3	0.45	3	3.52	3

6. RESULTS

As shown in Table 3 the LP III distribution had the first rank on all of the three goodness fit tests, so it is the best distribution for the study area.

Figure 18 shows a radar chart used to indicate parameters R^2 (Coefficient of Determination), Ratio (Bias Ratio), P_{corr} (Pearson Correlation Coefficient), KGE (Kling-Gupta Efficiency) and NSE (Nash-Sutcliffe Efficiency). According to the optimal values for each of these parameters shown in Table A1 (Appendix), each of them can be explained as follows: The chart shows that the value of R^2 for Samarra station = 1, which is the optimal value for this measure, Likewise, for the other stations, it is very close to the optimal value as shown Baiji station and then Tikrit and Tuz stations, and this means that the large proportion of variance in the observed data is explained by the LP III. whereas R^2 indicates how well the observed results are replicated by the model.

Ratio value for Samraa station is very close to the optimal value (1), likewise Baiji and then Tikrit and Tuz stations. This means that there is minimal bias since the ratio between the mean simulated and mean observed values is a ratio close to 1. So, it shows an unbiased model. P_{corr} measures the linear relationship between observed and simulated values, with values ranging from -1 to 1. Values closer to 1 indicate a strong positive correlation, through the chart, it can be interpreted that there is a strong correlation for the Samarra station, being equal to the optimal value, for this metric and the other stations are also very close to 1. As in the previous two measures, Baiji station and then the other two stations.

Since the value 1 represents the perfect fit of the model used according to the KGE metric, so according to what is clear in the radar chart, Samarra, Baiji, Tikrit, and Tuz stations Because it is close to the optimum and according to the sequence suggest the LP III distribution performs well in terms of correlation, bias, and variability. The lowest RE value was for Samraa station.

It is clear from the radar chart that the NSE value for the stations is very close to optimum, and this indicates good predictive performance because NSE measures how well the model predictions match the observed data. A balanced, large shape indicates a well-performing model across all metrics.

The error rate for using the LP III model is very small and within the range (≤ 0.05).

The IDF equation is a mathematical relationship between the rainfall intensity (I) the duration (d), and the return period (T). This empirical equation was practically used to derive the IDF equation [30]. The Bernard equation has been widely used in hydrology applications for maximum intensity, duration, and frequency and represents the relationship between the parameters of duration, which was adopted in the current study Eq. (17).

$$I = \frac{cT^m}{d^e} \qquad (17)$$

Maximum rainfall is a dependent variable in the above equations, while frequencies are an independent variable. c, m, and e are constant parameters associated with the metrological conditions. The SPSS program was utilized to derive the parameters utilized in the empirical equations. It can be described in Table 4 by Eqs. (18)-(29).
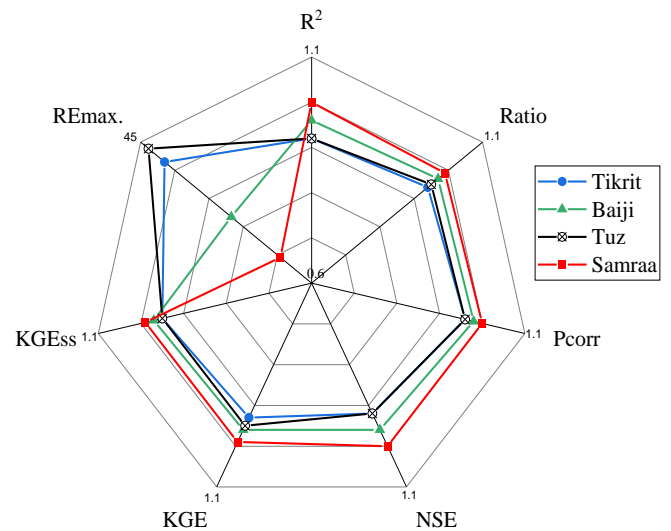

Figure 18. The spider plot of each statistical metric for LP III distribution

Table 4. Summary of factors included in the three methods using Bernard's equation

Station	Gumbel		LP III		LN	
Tikrit	$I = \dfrac{159.16T^{0.27}}{d^{0.66}}$	(18)	$I = \dfrac{31.90\,T^{0.71}}{d^{0.66}}$	(19)	$I = \dfrac{152.73\,T^{0.23}}{d^{0.80}}$	(20)
Samraa	$I = \dfrac{117.21T^{0.19}}{d^{0.66}}$	(21)	$I = \dfrac{125.15T^{0.11}}{d^{0.66}}$	(22)	$I = \dfrac{126.16T^{0.25}}{d^{0.80}}$	(23)
Baiji	$I = \dfrac{131.68T^{0.39}}{d^{0.66}}$	(24)	$I = \dfrac{93.21T^{0.43}}{d^{0.66}}$	(25)	$I = \dfrac{166.77T^{0.23}}{d^{0.79}}$	(26)
Tuz	$I = \dfrac{200.62T^{0.27}}{d^{0.66}}$	(27)	$I = \dfrac{22.19T^{0.89}}{d^{0.66}}$	(28)	$I = \dfrac{188.19T^{0.21}}{d^{0.78}}$	(29)

As for the Baiji station, Hussain in 2006 [14] derived the IDF equation for rainfall. This study is based on field data for a period of ten years (1989-1999) using Gumbel and LP III distributions. Then, according to the Weible method adopted by the study, and Bernard's Eq. (25), Figure 19 shows a comparison between them. The difference between the current equation and the researcher's equation at 15 minutes has a deviation in values greater than $\mp5\%$. As for periods 30 and 60, the convergence of the two equations was less than $\mp5\%$. The difference in sample size and the comprehensiveness of

the current equation for recent years make this difference normal.

Comparing the results of the IDF equation in terms of the value of R^2, Pearson's factor, and confidently 95%, for the current study with the results of these equations for studies of other regions close to the study area in terms of topography and climate, a scattering matrix diagram was used to illustrate the extent of convergence and divergence with previous studies Nasiriyah [28], Najaf [12], Baghdad [31], Basrah [32], Dohuk [33], and Mosul [21] cities, shown in Figure 20. All

results are located within the 95% confidence zone. also, Baghdad and Nasiriyah Cities show the optimum closer relationship with Tikrit city (current study) for correlation factors. From this, it can be said that there is a great convergence in the behavior of nearby areas and the closeness to the topography and weather of the study area, which favored the method of analyzing data using the LP III method.

This study used the graphical representation Taylor diagram, Figure 21 compares multiple models/ datasets against a reference by depicting their correlation, standard deviation, and root mean square error (RMSE) for the three distributions (LP III, Gumbel, and LN). The correlation Coefficient (R^2) value in the diagram as shown the closest to the optimum value (the highest value) is for LP III distribution, this indicates the strongest agreement with the dataset for Samraa station, and the other stations showed values close to that. As shown the datasets with standard deviations similar to the reference are preferable. In all the stations, LP III has the lowest RMSE value as shown in the diagram and this indicates that the dataset is closer to the reference in terms of overall error. This enhances the advantage of the LP III model for analyzing the data of the study area and increases the reliability of its adoption in finding IDF curves and equations in it.
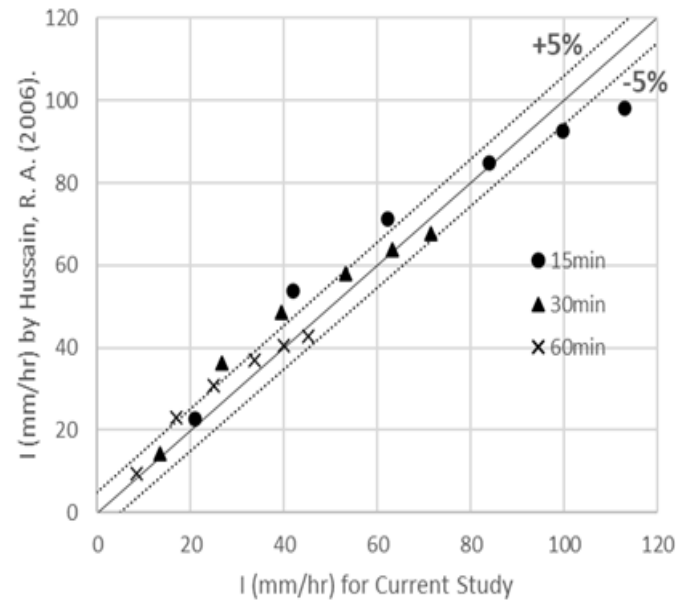

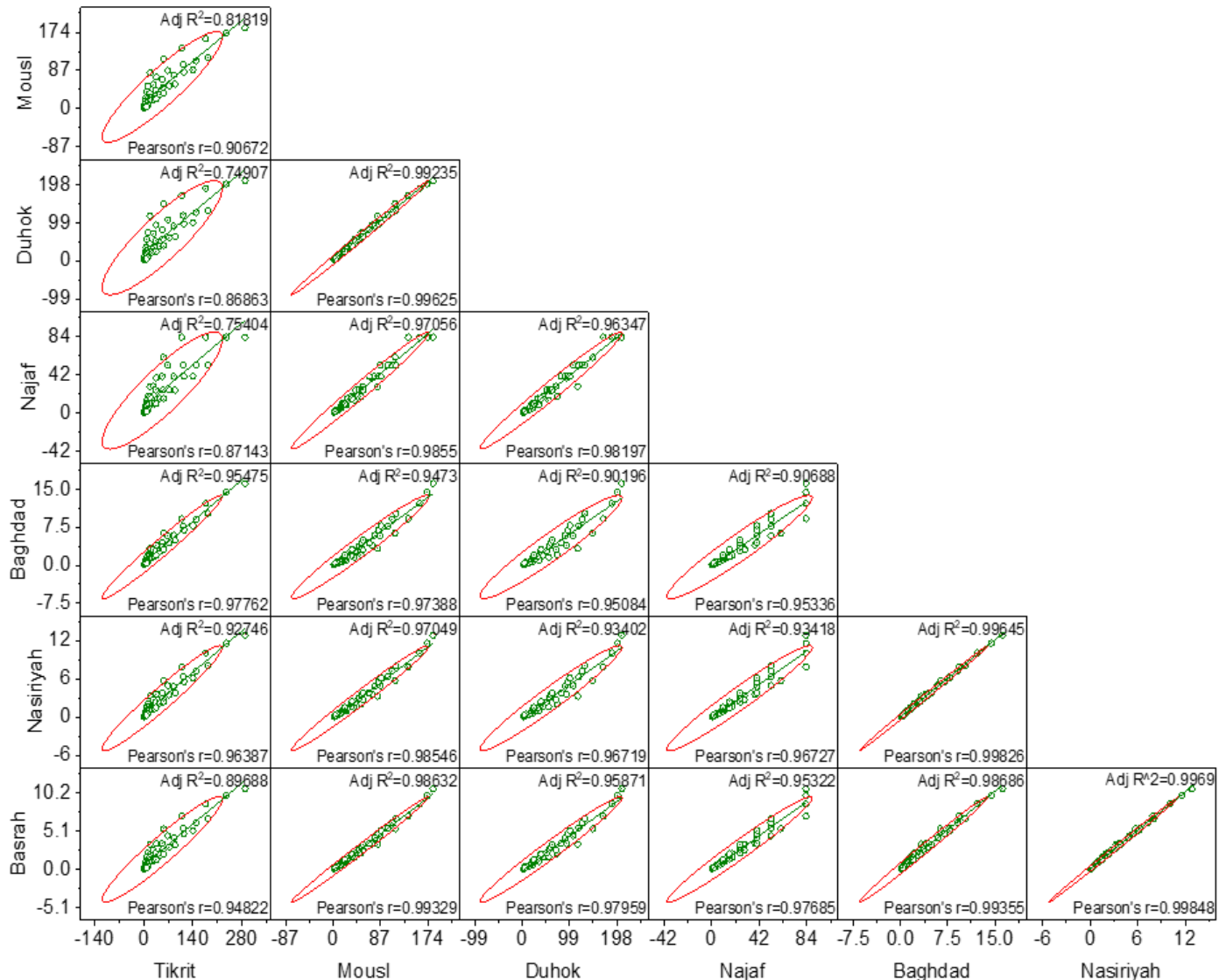**Figure 19.** Comparison between the current study and Hussain [14]

Figure 20. Scatter Matrix for comparison of nearby stations with the current study

(a) Taylor diagram for Samraa station

(b) Taylor diagram for Tikrit station

(c) Taylor diagram for Baiji station
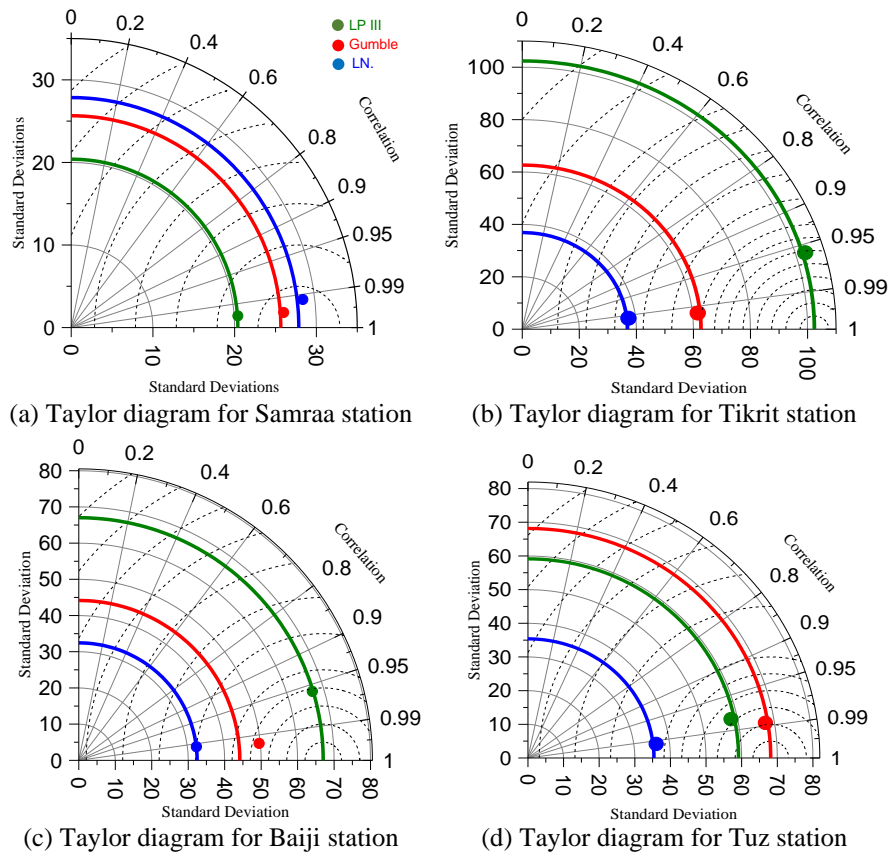
(d) Taylor diagram for Tuz station

Figure 21. Taylor diagram for the stations of the study area

7. CONCLUSIONS

The current study aimed to develop IDF curves and find empirical equations for field data obtained to estimate rainfall intensity in the study area. The obtained IDF curves are usually used when designing all structures for engineering projects. These curves allow for the design of safe and economical flood control structures. Rainfall estimates in this study are in mm and intensity in mm/hr. The different return periods and durations were also analyzed using three techniques (Gumbel, LP III, and LN), and the following was concluded:

1. Based on the cumulative average of the rainfall field data for the study area, it was concluded that all of these field data are internally consistent and are fit for adoption in deducing the IDF curves and equations.
2. Maximum intensity occurs at a return period of 100 years with a duration of 5 minutes, while minimum intensity occurs at a return period of 2 years with a duration of 1440 minutes.
3. By noting the development of IDF curves, which were derived for rainfall durations (10, 20, 30, 60, 120, 180, 360, 720, and 1440 min) for return periods (2, 5, 10, 25, 50, and 100 years) for the study area by three different methods (Gumbel, LP III, and LN), it was indicated that the value of intensities was very high for the LP III distribution compared with the LN and Gumbel methods.
4. The goodness of fit tests showed that all of the statistical metrics values for the empirical Bernard equation were very close to the optimal values, with the highest values for the LP III distribution, with reliability α < 0.05 for each of the three test's variables (Kolmogorov Smirnov, Anderson-Darling, Chi-Squared). Therefore, it presented the best distribution that can be adopted in the Salah-Al-Din

governorate.
5. The most appropriate technique to represent the study area is the LP III, based on all the statistical metrics that are adopted. As a result, it is advised to estimate IDF relationship parameters using this technique. Salah al-Din governorate may adopt the suggested intensity-duration-frequency relationships for design procedures; this distribution is the most appropriate for the approved frequency period compared to the Gumbel and LN distributions, and this behavior in terms of statistical distribution was also recommended by most previous studies for areas similar to Salah al-Din governorate in terms of climate and topography nature, which was adopted in the current study. The validation process confirms that the Bernard equation has the most accuracy in estimating the rainfall intensity. So, it can be used to develop the rainfall intensity in the study area.
6. The study proved the importance of the diagram and used the new statistical approaches in evaluating and comparing the performance of different models or datasets.
7. This study recommends the adoption of the empirical equation for the intensity of rainfall in Salah Al-Din, which will help choose the best pattern for projects using water resources. This equation will serve as a reliable guide to estimate the intensity of the rainfall for any given return period over various durations.

REFERENCES

[1] Kourtis, I.M., Nalbantis, I., Tsakiris, G., Psiloglou, B.E., Tsihrintzis, V.A. (2023). Updating IDF curves under climate change: Impact on rainfall-Induced runoff in urban basins. Water Resources Management, 37(6): 2403-2428. https://doi.org/10.1007/s11269-022-03252-8

[2] Rahman, M.M. (2015). Development of rainfall intensity-duration-frequency relationships from daily rainfall data for the major cities in Bangladesh based on scaling properties. International Journal for Scientific Research & Development, 3(8): 627-631.

[3] Zeri, S.J., Hamed, M.M., Wang, X., Shahid, S. (2023). Utilizing satellite data to establish rainfall intensity-duration-frequency curves for major cities in Iraq. Water, 15(5): 852. https://doi.org/10.3390/w15050852

[4] Yozgatligil, C., Aslan, S., Iyigun, C., Batmaz, I. (2013). Comparison of missing value imputation methods in time series: The case of Turkish meteorological data. Theoretical and applied climatology, 112: 143-167. https://doi.org/10.1007/s00704-012-0723-x

[5] Dang, T.A. (2020). Simulating rainfall IDF curve for flood warnings in the Ca Mau coastal area under the impacts of climate change. International Journal of Climate Change Strategies and Management, 12(5): 705-715. https://doi.org/10.1108/IJCCSM-06-2020-0067

[6] Sangüesa, C., Pizarro, R., Ingram, B., Ibáñez, A., Rivera, D., García-Chevesich, P., Pino, J., Pérez F., Balocchi, F., Peña, F. (2023). Comparing methods for the regionalization of intensity-duration-frequency (IDF) curve parameters in sparsely-gauged and ungauged areas of central Chile. Hydrology, 10(9): 179. https://doi.org/10.3390/hydrology10090179

[7] Alramlawi, K., Fıstıkoğlu, O. (2022). Estimation of intensity-duration-frequency (IDF) curves from large scale atmospheric dataset by statistical downscaling. Teknik Dergi, 33(1): 11591-11615. https://doi.org/10.18400/tekderg.874035

[8] Refaey, M.A., El Naggar, M.H., Mohamad, E.F. (2022). Arid and semi-arid regions IDF curve generation: Case study Al Qusir, Egypt, 13(2): 1-11. https://doi.org/10.17605/OSF.IO/3DHNE

[9] Ogbozige, F.J. (2022). Development of intensity-duration-frequency (IDF) models for manually operated rain gauge catchment: A case study of port harcourt metropolis using 50years rainfall data. Engineering and Technology Journal, 40(5): 627-635. https://doi.org/10.48047/ecb/2023.12.10.534

[10] Basumatary, V., Sil, B.S. (2018). Generation of rainfall intensity-duration-frequency curves for the Barak River Basin. Meteorology Hydrology and Water Management. Research and Operational Applications, 6(1): 47-57. https://doi.org/10.26491/mhwm/79175

[11] Ewea, H.A., Elfeki, A.M., Al-Amri, N.S. (2017). Development of intensity-duration-frequency curves for the Kingdom of Saudi Arabia. Geomatics, Natural Hazards and Risk, 8(2): 570-584. https://doi.org/10.1080/19475705.2016.1250113

[12] Hamaamin, Y.A. (2017). Developing of rainfall intensity-duration-frequency model for Sulaimani city. Journal of Zankoy Sulaimani, 19(3-4): 10634. https://doi.org/10.17656/JZS.10634

[13] Haji, H.F.H. (2017). Developing and mapping rainfall intensity, duration, and frequency curves in al-faria catchment. Doctoral dissertation, An-Najah National University.

[14] Hussain, R.A. (2006). Intensity-duration-frequency analysis for rainfall in Baiji Station. Tikrit Journal of Engineering Sciences, 13(4): 116-135.

[15] Rashid, M., Faruque, S.B., Alam, J.B. (2012). Modeling of short duration rainfall intensity duration frequency (SDRIDF) equation for Sylhet City in Bangladesh. Journal of Science and Technology, 2(2): 92-95.

[16] Schneider, T. (2001). Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. Journal of Climate, 14(5): 853-871. https://doi.org/10.1175/1520-0442(2001)014<0853:AOICDE>2.0.CO;2

[17] Bárdossy, A., Pegram, G. (2014). Infilling missing precipitation records-A comparison of a new copula-based method with other techniques. Journal of Hydrology, 519: 1162-1170. https://doi.org/10.1016/j.jhydrol.2014.08.025

[18] Lumbroso, D.M., Boyce, S., Bast, H., Walmsley, N. (2011). The challenges of developing rainfall intensity-duration-frequency curves and national flood hazard maps for the Caribbean. Journal of Flood Risk Management, 4(1): 42-52. https://doi.org/10.1111/j.1753-318X.2010.01088.x

[19] Ahmed, Z., Rao, D., Reddy, K., Raj, E. (2012). Rainfall intensity variation for observed data and derived data-A case study of Imphal. ARPN Journal of Engineering and Applied Sciences, 7(11): 1506-1513.

[20] Guo, Y. (2006). Updating rainfall IDF relationships to maintain urban drainage design standards. Journal of Hydrologic Engineering, 11(5): 506-509. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:5(506)

[21] Hussein, D.N., AL-Zakar, S.H., Yonis, A.M. (2023). Estimating the intensity equations for rain intensity frequency curves (Mosul/Iraq): Intensity equations for rain. Tikrit Journal of Engineering Sciences, 30(3): 38-48. https://doi.org/10.25130/tjes.30.3.5

[22] Subyani, A.M. (2011). Hydrologic behavior and flood probability for selected arid basins in Makkah area, western Saudi Arabia. Arabian Journal of Geosciences, 4(5): 817. https://doi.org/10.1007/s12517-009-0098-1

[23] Gaznayee, H. (2020). Modeling spatio-temporal pattern of drought severity using meteorological data and geoinformatics techniques for the Kurdistan Region of Iraq. Doctoral Dissertation, Salahaddin University-Erbil.

[24] Lau, A., Behrangi, A. (2022). Understanding intensity-duration-frequency (idf) curves using IMERG sub-hourly precipitation against dense gauge networks. Remote Sensing, 14(19): 5032. https://doi.org/10.3390/rs14195032

[25] Shamkhi, M.S., Azeez, M.K., Obeid, Z.H. (2022). Deriving rainfall intensity-duration-frequency (IDF) curves and testing the best distribution using EasyFit software 5.5 for Kut city, Iraq. Open Engineering, 12(1): 834-843. https://doi.org/10.1515/eng-2022-0330

[26] Al-Awadi, A.T. (2016). Assessment of intensity duration frequency (IDF) models for Baghdad city, Iraq. Journal of Applied Sciences Research, 12(2): 7-11.

[27] Elsebaie, I.H. (2012). Developing rainfall intensity-duration-frequency relationship for two regions in Saudi Arabia. Journal of King Saud University-Engineering Sciences, 24(2): 131-140.

https://doi.org/10.1016/j.jksues.2011.06.001

[28] Dakheel, A.A. (2017). Drawing curves of the rainfall intensity duration frequency (IDF) and assessment equation intensity rainfall for Nasiriyah City, Iraq. University of Thi-Qar Journal, 12(2): 63-82.

[29] Majeed, A.R., Nile, B.K., Al-Baidhani, J.H. (2021). Selection of suitable PDF model and build of IDF curves for rainfall in Najaf city, Iraq. In 3rd International Scientific Conference of Engineering Sciences and Advances Technologies (IICESAT), Babylon, Iraq. https://doi.org/10.1088/1742-6596/1973/1/012184

[30] Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modeling. Journal of Hydrology, 377(1-2): 80-91. https://doi.org/10.1016/j.jhydrol.2009.08.003

[31] Mahdi, E.S., Mohamedmeki, M.Z. (2020). Analysis of rainfall intensity-duration-frequency (IDF) curves of Baghdad city. In 2nd International Conference on Civil and Environmental Engineering Technologies (ICCEET 2020), Najaf, Iraq, 888(1). https://doi.org/10.1088/1757-899X/888/1/012066

[32] Abd Alelah, Z. (2016). Modeling of short duration rainfall intensity duration frequency (SDR-IDF) equation for Basrah City. University of Thi-Qar Journal for Engineering Sciences, 7(2): 56-68. https://doi.org/10.31663/utjes.v7i2.62

[33] Amen, A.R.M., Kareem, D.A., Mirza, A.A., Salih, A.M. (2022). Development of intensity-duration-frequency curves "IDF" for Dohuk City in Kurdistan Region of Iraq. Journal of Duhok University, 25(2): 366-379. https://doi.org/10.26682/sjuod.2022.25.2.34

NOMENCLATURE

AD	Anderson – Darling test.
A	Anderson test value.
C_S	Skewness Coefficient.
c	Constant parameter.
d	Time duration in (h).
e	Constant parameter.
Ei	The expected number.
F	Cumulative Distribution.
Ei	The expected number.
EM	Expectation Maximization algorithm method.
F	Cumulative Distribution.
$F_{(xi)}$	The theoretical cumulative probability.
Ha	Alternative Hypothesis.
H0	Null Hypothesis.
I	Rainfall intensity.
I_T	Rainfall Intensity (mm/h) for each return period (T).
IDF	Intensity – Duration – Frequency.
IMD	Indian Meteorol1ogical Duration.
K_T	The Gumbel frequency factor.
K-S	Kolmogorov-Smirnov.
KGE	Kling-Gupta efficiency score.
KGEss	Kling-Gupta efficiency skill score.
LN	Log Normal.
LP III	Log Pearson type three.
m	Constant parameter.
m^*	The descending order of values for observed data.
NSE	Nash-Sutcliffe efficiency score.
n	The number of historical data points.
Oi	The observed data.
P	The error ratio in t-test.
$P_{ave.}$	Average of annual precipitation data.
P_{corr}	Pearson correlation score.
P_i	Highest daily peak of annual precipitation (mm).
P_t	The required precipitation depth for a duration less than 24 h in (mm).
P_T	Precipitation in (mm) for each return period in (year).
$P_{(24)}$	Daily precipitation in (mm).
p-value	Pettitt test – value.
R^2	The correlation coefficient.
RE	Relative Error score.
S	Standard deviation of precipitation data.
T	Return period in (year).
t	Time (minute).
μ	The mean correlation coefficient.
X^2	Correlation coefficient.
X_i	Estimate value.
Y_i	Gauge-based values.
Δ	Statistic value to fit data.
Δo	Tabular value in Kolmongorov-Simnrov test.

APPENDIX

Table A1. Statistical scores used for deriving IDF formulas

Statistic Metrics	Equations		Range	Optimal Value	Ref.
Mean bias ratio	$Ratio = \frac{\mu_x}{\mu_y}$	(29)	0 to 1	1	[20]
Pearson correlation (P_{corr})	$P_{corr} = \frac{1}{n-1}\sum_{i=1}^{n}(\frac{X_i - \mu_x}{S_t})\left(\frac{y_i - \mu_y}{S_y}\right)$	(30)	-1 to 1	1	[20]
Bias	$Bias = X_i - Y_i$	(31)	$-\infty$ to $+\infty$	0	[21]
Nash-Sutcliffe efficiency (NSE)	$NSE = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - y_i)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \mu_i)^2}$	(32)	0 to 1	1	[20]
Kling-Gupta efficiency (KGE)	$KGE = 1 - \sqrt{\left(1 - \frac{S_x}{S_y}\right)^2 + \left(1 - \frac{\mu_x}{\mu_y}\right)^2 + (1 - P)^2}$	(33)	$-\infty$ to 1	1	[21]
Kling-Gupta efficiency skill score (KGEss)	$KGEss = \frac{KGE - 0.4142}{\sqrt{2}}$	(34)	$-\infty$ to 1	1	[24]
Relative Error (RE)	$RE = \left(\frac{X_i - Y_i}{Y_i}\right) * 100\%$	(35)	$-\infty$ to $+\infty$	0	[26]