Kufa Journal of Engineering
Vol. 10, No. 4, October 2019, P.P. 29-41
Received 20 June 2018, accepted 13 March 2019

PERFORMANCE OF THE PROBABILITY DISTRIBUTIONS FOR PLOTTING POSITIONS IN ESTIMATING THE MAXIMUM DISCHARGES OF ADHAIM RIVER

Abdulwahab M. Younis[1], Ihsan F. Hasan[2] and Younis N. Saeed[3]

[1] **Asst. Professor, Dams and Water Resources Engineering, University of Mosul, Iraq. Email: amyggv56@yahoo.com**

[2] **Lecturer, Dams and Water Resources Research Center, University of Mosul, Iraq. Email: obaidyhassan@gmail.com**

[3] **Asst. Lecturer, Dams and Water Resources Research Center, University of Mosul, Iraq. Email: younisy@yahoo.com**

HTTP://DX.DOI.ORG/10.30572/2018/KJE/100403

ABSTRACT

In this study, the peak discharges of Adhaim River were plotted against their hydrologic years. Two probability distributions and five plotting positions formulas were fitted to the annual maximum discharges AMD. The performances of the probability distributions with plotting position formulas were evaluated using the coefficients of determination R^2, root mean square errors RMSE, Mean Absolute Percent Error MAPE, and absolute differences between predicted and observed discharges. The annual maximum discharges of Adhaim River vary in magnitude from 111 to 3520 m^3/sec for record periods. The mean of annual maximum discharges 753 m^3/sec. The Lognormal LN and Log-Person type III LP3 distributions have the highest R^2 using Weibull plotting position formula WPP of 0.992 and 0.989, respectively. The LN had minimum RMSE and MAPE of 69.13 and 7.3%, respectively, when conformed with California plotting position CPP. The LP3 had minimum RMSE and MAPE of 93.6 and 6.5% respectively when conformed with Chegadayev plotting position CHPP. The minimum absolute differences at return periods of 25, 50,100, and 200 years were obtained when LN conformed with Hazen plotting position HPP and when LP3 conformed with CPP.

KEYWORDS: Flood, Plotting Positions, Probability distributions, Annual maximum discharges.

1. INTRODUCTION

Floods are natural hazards that cause human and economic losses, destruction of agricultural lands, and major property losses (Fill and Stedinger, 1995). One method of decreasing flood damages and economic losses is to use flood frequency analysis for determining efficient designs of hydraulic structures. In hydrology, estimation of peak discharges for design purposes on catchments with only limited available data is still a continuing problem (Blazkovaa and Bevenb, 1997), therefore; the deriving flood frequency curve considers an elegant method to solve this problem, which needs to availability the observed historical data for the study area such as mean annual flow.

The theoretical probability distributions are either continuous or discrete. Most of the hydrological variables are considered as continuous random variables therefore; the continuous distributions are used widely in hydrology (Viessman and Lewis, 2003). Several studies used different distributions for describing flood data all over the world, such as Generalized Extreme Value GEV and LP3 distributions in USA (Vogel et al., 1993), GEV, Generalized Logistic GLO and Generalized Pareto GPA distributions in India (Bhuyan et al., 2010), GLO, GEV, Pearson type III P3 and GPA in Turkey (Saf, 2009), GEV, P3, GLO, and GPA distributions in Tunisia (Abida and Ellouze, 2008) and Log normal and LP3 in Nigeria (Izinyon and Ajumuka, 2013). The three-parameter log-Pearson type III distribution is the most frequently used distribution in the USA, whereas the lognormal distribution in China (Singh, and Strupczewski, 2002).

The Plotting positions (PP) formulas have been used in estimating magnitudes of hydrological events and their corresponding return periods, detecting outliers, fitting distributions to the data, and in evaluating the adequacy of fit of the alternative parametric floods frequency models (Adeboye and Alatise, 2007). All plotting position formulas give similar values near the center of the distribution but may vary considerably in the tails (Haan, 1994). There are several formulas of plotting positions presented by researchers (Chow et al., 1988; Ewemoje and Ewemooje, 2011; Mehdi and Mehdi, 2011; and Makkonen, 2006).

In the present study two commonly used probability distributions (Log-Normal Distribution, and Log-Pearson Type III Distribution) and five different plotting position (PP) formulas Table 1 were applied and compared to select the best flood frequency distribution which best fits the annual maximum flood flows of Adhaim river basin in Iraq.

2. STUDY AREA

The Adhaim River is one of the tributaries of the Tigris River which travels 230 km from its source in the hills between Lesser Zab and Diyala rivers to the end of its estuary into Tigris River at 15 km south of Balad city Fig. 1. Its catchment area is about 13 000 km^2 totally lies in Iraq (Arslan, 2016). It is an area with practically no snowfall, and even rainfall is limited. Thus, effective flow occurs during the rainy season only, therefore it's considered an intermittent river.

The flood of Adhaim River usually occurs early from the Tigris River flood, so its water flows to the Tigris River earlier than the time of water levels rising in the Tigris River. Therefore, the effect of the Adhaim River's flood on the Tigris River's flood is small except in some years where an early flood occurs in both of the Tigris and Adhaim Rivers, so the flood of Adhaim River will have a clear effect in increasing the water levels of the Tigris River although this increase is small. Annual maximum daily Discharge data of Adhaim River basin for 39 years were used in this study, the measured annual maximum Daily Discharge of Adhaim River during the observation period at Injana gaging station ranged from 3520 m^3/sec to 111 m^3/sec, The mean of annual maximum discharges 753 m^3/sec. The maximum monthly Discharge ranged from 424 m^3/sec to 18 m^3/sec, the catchment area at Injana gaging station (Latitude 34° 30′ 00″ N, Longitude 44° 31′ 00″ E) is about 9840 km^2 at this station (Saleh, 2010).

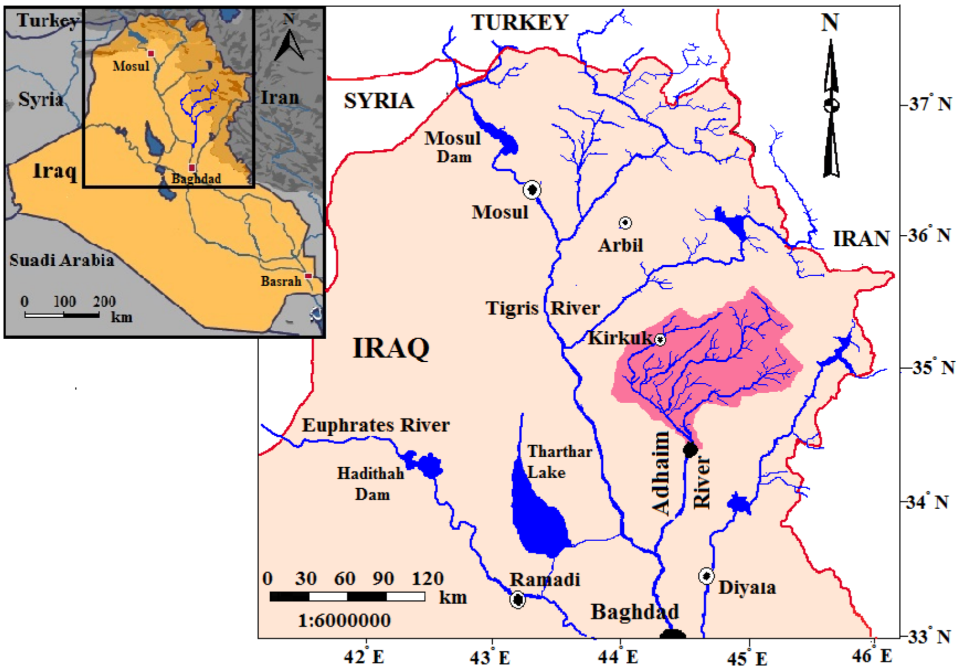


Fig. 1. The locations of Adhaim River.

3. METHODOLOGY

There are many distributions which are used in flood frequency analysis. Distributions used in the present study are:

3.1. Log-Normal Distribution (LN)

A probability density function (PDF) is a continuous mathematical expression that determines the probability of a particular event (Izinyon and Ajumuka, 2013). The PDF of the lognormal distribution with parameters μ and σ (which denotes as location and scale parameters respectively) of such a variable y= *ln x* is given by Eq.1 (Chow et al., 1988).

$$f(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} exp\left(\frac{-y-\mu_y}{2\sigma_y^2}\right) \quad , \quad x > 0 \qquad 1$$

Where $\boldsymbol{x}$ is random variable, the mean $\mu_y = \bar{y}$ and standard deviation $\sigma_y = S_y$ were computed using Eq.2, Eq.3 and Eq.4.

$$y = Log(Q_{max}) \qquad 2$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y \qquad 3$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n}(y-\bar{y})^2}{n-1}} \qquad 4$$

Where, $\bar{y}$ is mean of y; S_y is standard deviation of y; n is number observations; Q_{max} is annual maximum discharges (m^3/sec).

The intermediate variable W and frequency factors Z corresponding to the return periods of the ranked annual maximum discharges were computed by the Eq.5 and Eq.6 (Chow et al., 1988).

$$W = \sqrt{ln\frac{1}{p^2}} \quad , \quad 0 < p \leq 0.5 \qquad 5$$

$$Z = w - \frac{2.515517 + 0.80285W + 0.01032W^2}{1 + 1.432788W + 0.189269W^2 + 0.001308W^3} \qquad 6$$

Where p is the probability of exceedence. The predicted flood discharge Q_T at various return periods T is computed by Eq.7 and Eq.8 (Makkonen, 2006).

$$Y_T = Log\ Q_T = \bar{y} + ZS_y \qquad 7$$

$$Q_T = 10^{(y_T)} \qquad 8$$

3.2. Log Pearson Type III Distribution (LP3)

The LP3 Probability Density Function (PDF) expressed as in Eq.9 (Ewemoje and Ewemooje, 2011).

$$f(x) = \frac{1}{\beta^2 \Gamma(\alpha)} (x-\gamma)^{\alpha-1} e^{-(x-\gamma)/\beta} \qquad , x \geq \gamma \qquad 9$$

Where Γ(α) is the Gamma function. The parameters α, β and γ are related to the first three moments of the random variable X, Eq.10, Eq.11 and Eq.12 (Izinyon and Ajumuka, 2013).

$$\mu_y = \gamma + \frac{\alpha}{\beta} \qquad 10$$

$$s_y = \frac{\alpha}{\beta^2} \qquad 11$$

$$g_s = \frac{2}{\sqrt{\alpha}} \qquad 12$$

In LP3 method, the rearranged annual maximum discharge data is first transformed to logarithms discharges of base 10. Then In addition to the mean and the standard deviation, the coefficient of skewness g_s (which denotes as shape parameter) was computed to determine predicted discharges by the Eq.13 (Haan, 1994).

$$g_x = \frac{n \sum_{i=1}^{n} (y - \bar{y})^3}{(n-1)(n-2) S_y^3} \qquad 13$$

The frequency factor K_T of the LP3 is computed using Eq.14 (Chin, 2006).

$$K_T = \frac{1}{3k} [\{(Z-k)k + 1\}^3 - 1] \qquad 14$$

Where $k = \frac{g_s}{6}$, When $g_s = 0$; $K_T = Z$. The predicted flood discharge Q_T at various return periods T is computed in Eq.15 and Eq.16 (Adeboye and Alatise, 2007).

$$Y_T = Log\, Q_T = \mu_y + K_T S_y \qquad 15$$

$$Q_T = 10^{Y_T} \qquad 16$$

4. PLOTTING POSITIONS

Plotting position refers to the probability value assigned to each piece of data to be plotted (Ewemoje and Ewemooje, 2011). In case flood frequency analysis, it is common to plot both the assumed population and the peak discharges of the sample. To plot the sample values on frequency paper, it is necessary to assign an exceedence probability to each magnitude. A plotting position formula is used for this purpose. Numerous methods have been proposed for the determination of plotting positions; most plotting position formulas are represented in the Eq.17 (Ewemoje and Ewemooje, 2011).

$$P(Q \geq Q_T) = \frac{m-b}{n+1-2b} \qquad 17$$

Where P is the probability of exceedence (probability a given flood magnitude will be equaled or exceeded), Q is predicted discharge (m^3/s), Q_T is flood discharge at return period equaled or exceeded (m^3/s), n is the number of years of record, m is the rank of annual maximum discharges series arranged in descending order of magnitude, b is the plotting position parameter between 0 and 1, depending upon the plotting position formula and on the theoretical distribution (Portela and Delgado, 2009). For example, b=0 for all Weibull formula, 0.44 for Gringorten formula, 0.5 for Hazen formula and 0.3 for Chegodayev formula. The Used plotting position formulas is shown in Table 1.

Table 1. Plotting Position Formulas used in the present study

Plotting Positions	Formula
California (CPP)	$p(Q \geq Q_T) = \frac{m}{n}$
Hazen (HPP)	$p(Q \geq Q_T) = \frac{2m-1}{2n}$
Weibull (WPP)	$p(Q \geq Q_T) = \frac{m}{n+1}$
Chegadayev (CHPP)	$p(Q \geq Q_T) = \frac{m-0{,}3}{n+0.4}$
Gringorten (GPP)	$p(Q \geq Q_T) = \frac{m-0.44}{n+0.12}$

5. PERFORMANCE INDICATORS

Performance indicators were used to evaluate the performance and comparison of the plotting position formulas using Log-Normal and Log Pearson Type III probability Distributions. In this study, three performance evaluations that are the Coefficient of Determination R^2, Eq. (18), Root Mean Squared Error RMSE, Eq. (19), and Mean Absolute Percent Error MAPE, Eq. (20) were used. The Absolute Differences between the observed and the predicted discharges at return periods of 25, 50, 100 and 200 years were also obtained.

$$R^2 = \frac{(\sum(O_i - \bar{O})(P_i - \bar{P}))^2}{\sum(O_i - \bar{O})^2 \sum(P_i - \bar{P})^2} \qquad 18$$

$$RMSE = \sqrt{\frac{\sum(O_i - P_i)^2}{n}} \qquad 19$$

$$MAPE = \frac{\sum\left(\frac{|O_i - P_i|}{O_i}\right)}{n} \quad 20$$

Where O_i, P_i are observed and predicted discharges at time i respectively, $\bar{O}$, $\bar{P}$ are mean value of the observed and predicted discharges respectively. A high R^2 value indicates a good model fit with observed data. RMSE and MAPE with low values are indicating accurate model prediction.

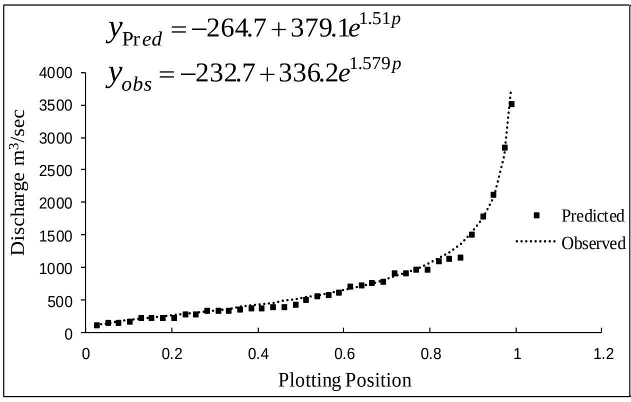
6. RESULTS AND DISCUSSION

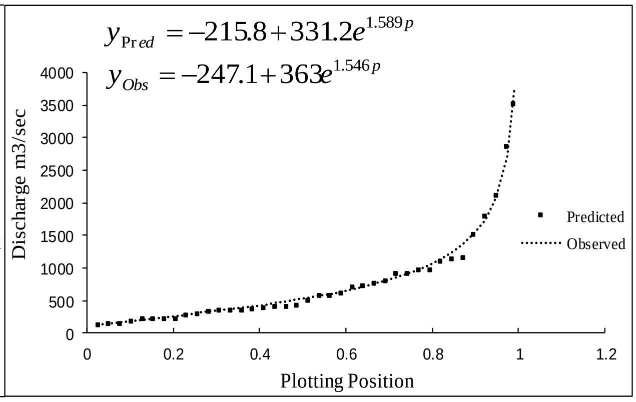
6.1. Maximum discharges using LN distribution:

The relations between values of the observed and predicted discharges using Log-Normal distribution conformed with the five plotting positions formulas for Adhaim River are given in Fig. 2. Regression testing was carried out using SPSS Sigma plot, the best equation represents this relation was Exponential Growth (Double, 3 Parameters) which its general formula as in the Eq.21.

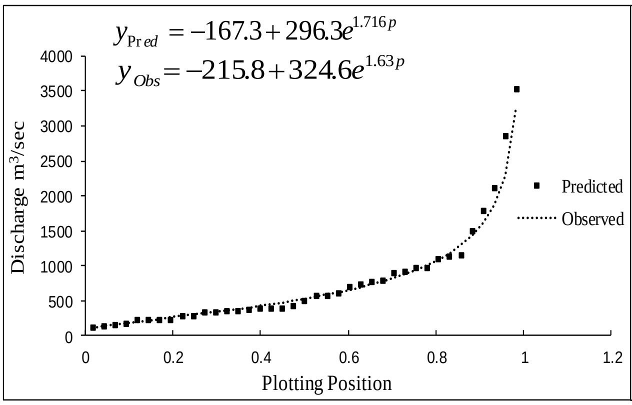
$$y = y_0 + ae^{bp} \quad 21$$

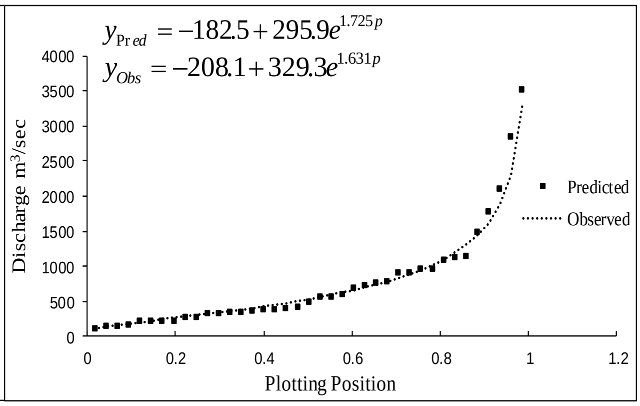
Where y is peak discharge, p is probability and a, b, y_o are equation's constants.





A B





C D

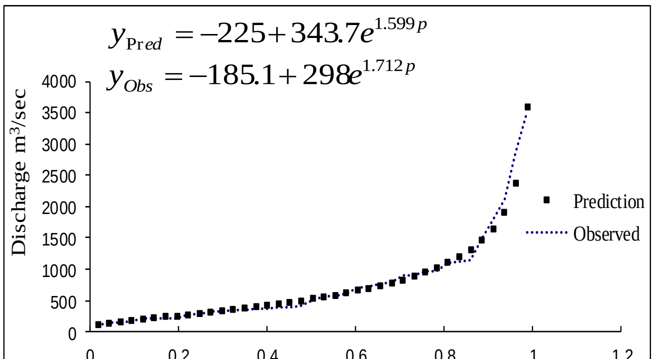


Fig. 2. Observed and predicted annual Maximum discharges using LN with: A: California, B: Hazen, C: Weibull, D: Chegodayev and E: Gringorten plotting positions.

Fig. 2 illustrates the relationship between observed and predicted annual Maximum discharges versus its occurrence probability for the five different plotting position Formulas in case of using LN Distribution, from this figure it can be seen that all plotting position formulas have similar values near the center of the distribution but varied considerably in the tails, this was more clearly in the Weibull plotting position. Discharge values at return periods of 25, 50, 100 and 200 years were estimated for the five different plotting position Formulas and Absolute Difference between the observed and predicted discharges at used return periods were obtained in case of using Log Normal Distribution as shown in Table 2.

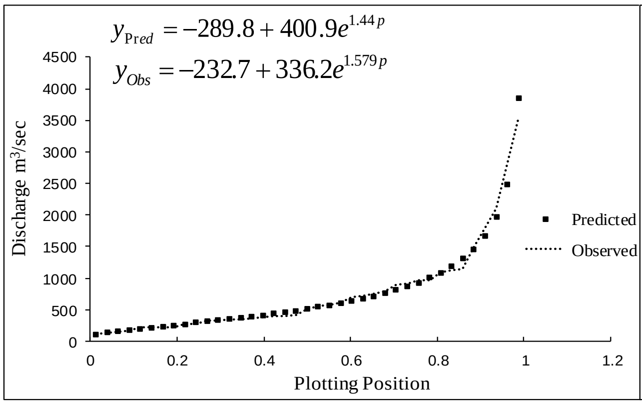
Table 2. Predicted discharge values with Absolute Differences at used Return Periods.

Log – Normal Distribution					
	Plotting Positions	Return Periods (T)			
		25	50	100	200
Predicted Flows (m^3 / sec)	California	139	126.4	120.1	117.20
	Hazen	138.95	127.4	121.35	118.3
	Weibull	150.15	139.5	134	131.15
	Grigorten	140.40	129	123.33	120.7
	Chegodayev	143	132	126.3	123.65
Absolute Differences	California	13.58	12.14	11.38	11.04
	Hazen	1.829	1.34	0.65	0.25
	Weibull	19.48	20	19.8	19.7
	Grigorten	6.38	5.7	5.3	5.24
	Chegodayev	8.44	8.22	7.7	7.8

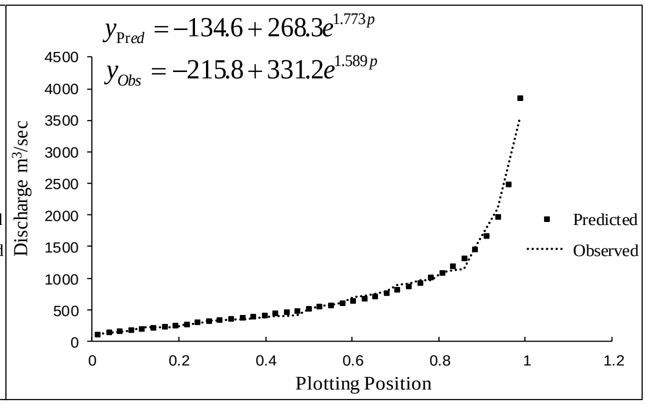
6.2. Maximum discharges using LP3 Distribution:

The relations between values of the observed and predicted discharges using Log Pearson Type III Distribution conformed with the five plotting positions formulas for Adhaim River are given in Fig. 3. Regression testing was carried out using SPSS Sigma plot, the best equation represents this relation was Exponential Growth (Double, 3 Parameters) which its general formula as in the equation 21:

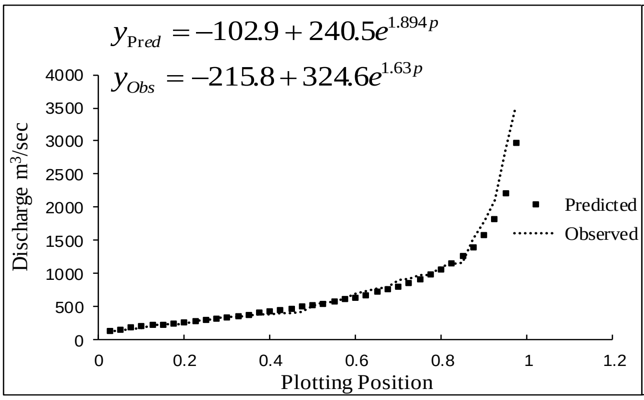
Fig. 3 illustrates the relationship between observed and predicted annual Maximum discharges versus its occurrence probability for the five different plotting position Formulas in case of using LP3 Distribution, from this figure it can be seen that all plotting position formulas have similar values near the center of the distribution but varied considerably in the tails, this was more clearly in the Weibull plotting position. Discharge values at return periods of 25, 50, 100 and 200 years were estimated for the five different plotting position Formulas and Absolute Difference between the observed and the predicted discharges at used return periods were obtained in case of using Log Person Type III Distribution as shown in Table 3.



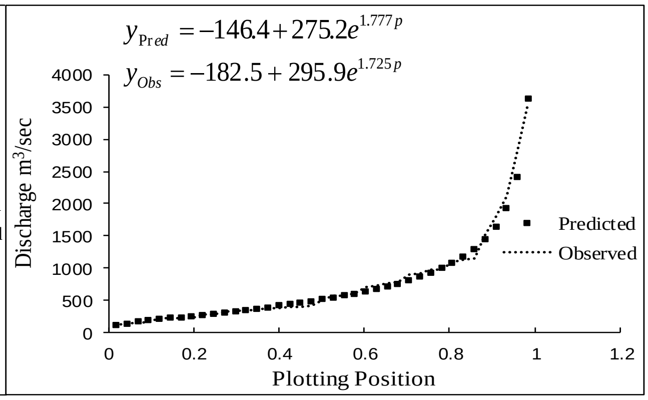
A



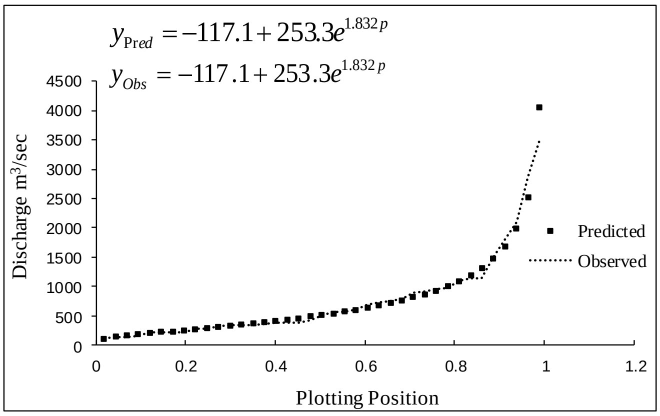
B



C



D



E

Fig. 3. Observed and predicted annual Maximum discharges using LP3 with: A: California, B: Hazen, C: Weibull, D: Chegodayev and E: Gringorten plotting positions.

Table 3. Predicted discharge values with Absolute Differences at used Return Periods.

Log – Person Type III Distribution					
	Plotting Positions	Return Periods (T) 25	50	100	200
Predicted Flows (m^3 / sec)	California	134.66	122.7	116	113.85
	Hazen	153	143.38	138.3	136
	Weibull	156.3	146.8	142.01	139.7
	Grigorten	155.4	145.25	140.6	138.12
	Chegodayev	149	138.55	133.3	131
Absolute Differences	California	9.45	8.4	8.1	7.7
	Hazen	16	17.2	17.6	18
	Weibull	25.7	27.32	27.9	28.3
	Grigorten	21.43	22	22.68	22.8
	Chegodayev	14.5	14.7	14.8	15.04

To check the performance of the models considered in the present study, a comparison is carried out between the observed and predicted discharge values of LN and LP3 using the five different plotting position Formulas. The statistical results of Coefficient of Determination R^2, Root Mean Squared Error RMSE, and Mean Absolute Percent Error MAPE were chosen for this comparison and as shown in Table 4.

Table 4. Comparison of the Statistical results from different Formulas.

Plotting Positions	Probability Distribution					
	Log – Normal			Log – Person Type III		
	R^2	RMSE	MAPE	R^2	RMSE	MAPE
California	0.991	69.13	7.3%	0.986	129.4	8.5%

Hazen	0.989	112.5	8.8%	0.982	97.6	8.9%
Weibull	0.992	201.6	9.1%	0.989	159.2	9.1%
Grigorten	0.990	101	7.7%	0.984	120	7.4%
Chegodayev	0.989	120	8.1%	0.987	93.6	6.5%

Note: shaded: best-fit Formula

Table 4 showed that the minimum errors were 69.13 and 7.3% using Root Mean Squared Error and Absolute Percent Error tests respectively obtained by using CPP Formula with the LN. This means that LN with CPP Formula is the best distribution for using in statistical prediction for Adhaim river basin. After that the minimum Root Mean Squared Error and minimum Absolute Percent Error were 93.6 and 6.5% respectively by using CHPP Formula with the LP3, the highest Coefficient of determinations R^2 were 0.992 and 0.989 by using WPP Formula with the LN and LP3 respectively. Generally on the basis of statistical results shown in Table 4, LN is more suitable than the LP3 for Adhaim river basin using different plotting position Formulas.

The minimum absolute differences were 1.829, 1.34, 0.65, and 0.25 at return periods of 25, 50, 100, and 200 years respectively in case of using the LN when conform with HPP formula while the minimum absolute differences were 9.45, 8.4, 8.1, 7.7 at return periods of 25, 50, 100, 200 years respectively in case of using the LP3 when conform with CPP formula. LN gives small absolute differences in comparison with LP3; this indicates that LN is more suitable than the LP3 distribution for study area at different return periods. The graphical evaluation includes visual comparison of predicted Discharges with 25, 50, 100 and 200 years return period for different plotting position Formulas as illustrated in Fig. 4.

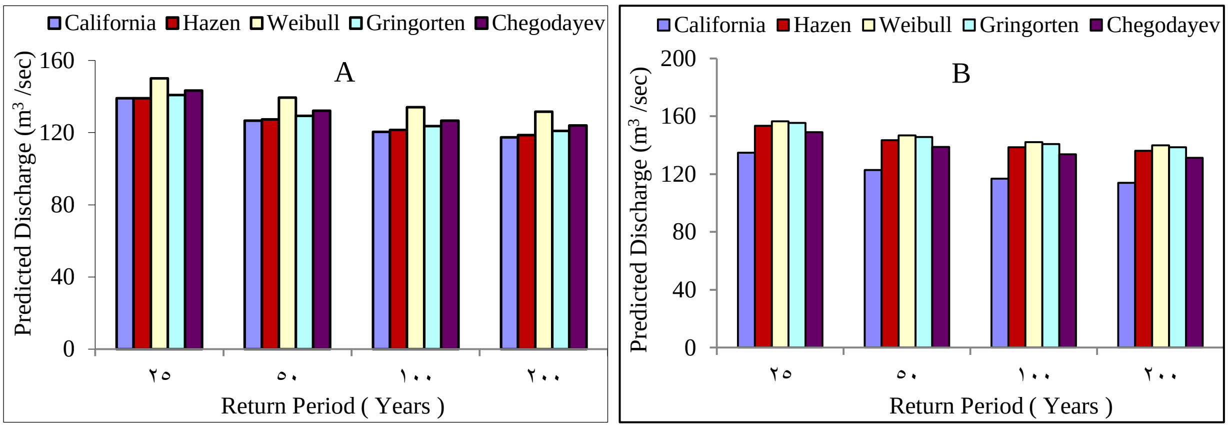


Fig. 4. Graphical Comparison of Different PP at used Return Periods for A: LN and B: LP3.

From Fig. 4 the comparison between predicted annual Maximum discharges for LN appears that CPP, HPP, and GPP Formulas gives similar values between themselves with respect to each return period while WPP, CHPP Formulas gives variant values with respect to each return period. For LP3 the predicted annual Maximum discharge values were similar in case of HPP, WPP, and GPP Formulas with respect to each return period while CPP, CHPP Formulas gives variant values form other Formulas with respect to each return period.

7. CONCLUSIONS

1. All plotting position formulas in particular Weibull formula more clearly give similar values near the center of the distribution but varied considerably in the tails. Furthermore, this formula has the highest Coefficient of determinations for LN and LP3.
2. The CPP Formula with the LN gives minimum RMSE and MAPE, and then CHPP Formula with the LP3 give minimum RMSE and MAPE.
3. LN conformed with HPP formula is the suitable distribution in predicting annual maximum discharges at various return periods which gives minimum absolute differences. While LP3 conformed with CPP formula is the suitable distribution in predicting annual maximum discharges at various return periods which gives minimum absolute differences.
4. Generally LN is the most suitable distribution for Adhaim river basin in comparison with LP3 using different plotting position Formulas, While LN is the best fitted distribution for annual maximum discharges of Adhaim river basin when it is conformed with CPP formula.

8. REFERENCES

Abida, H. and Ellouze, M. (2008) 'Probability distribution of flood flows in Tunisia' Hydrol. Earth Syst. Sci., 12:703–714.

Adeboye, O. B. and Alatise, M. O. (2007) 'Performance of Probability Distributions and Plotting Positions in Estimating the Flood of River Osun at Apoje Sub-basin, Nigeria' Agricultural Engineering International: CIGR journal. Vol. IX. July.

Arslan, C. A. (2016) 'Performance of Artificial Neural Networks, Support Vector machine and Fuzzy logic networks ANFIS in Monthly Streamflow Forecasting for Diyala, Adhaim and Elkhazer Rivers Northern Iraq' International Journal of Scientific and Engineering Research, 7(6), 105-114.

Bhuyan, A. Borah, M. and Kumar, R. (2010) 'Regional Flood Frequency Analysis of North-Bank of the River Brahmaputra Using LH-Moments' Water Resour. Manage. 24:1779–1790.

Blazkovaa, S. and Bevenb, K. (1997) 'Flood frequncey prediction for data limited catchments in the Czech Republic using Astochastic rainfall model and topmodel' Journal of Hydrology 195: 256–278. www. ivsl. Org.

Chin, D. A. (2006) 'Water Resources Engineering' 2nd Edition Person Prentice Hall, Upper Saddle River, New Jersey.

Chow V. T. Maidment, D. R. and Mays, L. W. (1988) 'Applied Hydrology' McGraw-Hill, New York, 572.

Ewemoje, T. A. and Ewemooje, O. S. (2011) 'Best Distribution and Plotting Positions of Daily Maximum Flood Estimation at Ona River in Ogun-Oshun River Basin, Nigeria' Agricultural Engineering International: CIGR Journal 13(3), 1-11.

Fill, H. D. and Stedinger, J. R. (1995) 'Homogeneity test based upon Gumble distribution and critical appraisal of Dalrymple test' Journal of Hydrology 166: 81-105. www. ivsl. Org.

Haan, C. T. (1994) 'Statistical Methods in Hydrology' Iowa State University Press. Ames.

Izinyon, O. C. and Ajumuka, H. N. (2013) 'Probability distribution models for flood prediction in Upper Benue River Basin–Part II' Civil and Environ. Res. Journal, 3(2), 62-74.

Makkonen, L. (2006) 'Plotting Positions in Extreme Value Analysis' Journal of Applied Meteorology and Climatology, 45, 334-340.

Mehdi, F. and Mehdi, J. (2011) 'Determination of Plotting Position Formula for the Normal Log Normal Pearson (III) Log-Pearson(III) and Gumble Distributional Hypotheses Using the Probability Plot Correlation Coefficient Test' World Applied Sciences Journal 15(8):1181-1185.

Portela, M. M. and Delgado, J. M. (2009) 'A new plotting position concept to evaluate peak flood discharges based on short samples' Water Resour. Manage. Vol 125, 415-427.

Saf, B. (2009) 'Regional Flood Frequency Analysis Using L-Moments for the West Mediterranean Region of Turkey' Water Resour. Manage. 23:531–551.

Saleh, D.K. (2010) 'Stream gage descriptions and streamflow statistics for sites in the Tigris River and Euphrates River Basins, Iraq' U.S. Geological Survey Data Series 540.

Singh, V. P. and Strupczewski, W. G. (2002) 'On the status of Flood frequency analysis' Hydrol. Process, 16, 3737–3740.

Viessman, W. J. and Lewis, G. L. (2003) 'Introduction to Hydrology' 5th ed., Pearson Education Inc. New York.

Vogel, R. M. Thomas, W. O. and McMahon, T. A. (1993) 'Flood-flow frequency model selection in the south western United States' Water Resour. Plan. Manage. 119(3), 353-366.