World Academy of Science, Engineering and Technology
International Journal of Civil and Environmental Engineering
Vol:12, No:10, 2018

Evaluation of Best-Fit Probability Distribution for Prediction of Extreme Hydrologic Phenomena

Karim Hamidi Machekposhti, Hossein Sedghi

Abstract**—**The probability distributions are the best method for forecasting of extreme hydrologic phenomena such as rainfall and flood flows. In this research, in order to determine suitable probability distribution for estimating of annual extreme rainfall and flood flows (discharge) series with different return periods, precipitation with 40 and discharge with 58 years time period had been collected from Karkheh River at Iran. After homogeneity and adequacy tests, data have been analyzed by Stormwater Management and Design Aid (SMADA) software and residual sum of squares (R.S.S). The best probability distribution was Log Pearson Type III with R.S.S value (145.91) and value (13.67) for peak discharge and Log Pearson Type III with R.S.S values (141.08) and (8.95) for maximum discharge in Jelogir Majin and Pole Zal stations, respectively. The best distribution for maximum precipitation in Jelogir Majin and Pole Zal stations was Log Pearson Type III distribution with R.S.S values (1.74&1.90) and then Pearson Type III distribution with R.S.S values (1.53&1.69). Overall, the Log Pearson Type III distributions are acceptable distribution types for representing statistics of extreme hydrologic phenomena in Karkheh River at Iran with the Pearson Type III distribution as a potential alternative.

Keywords**—**Karkheh river, log pearson type III, probability distribution, residual sum of squares.

I. Introduction

THE extreme hydrological events such as flood (extreme streamflow) can have intense impacts on activity of humans and society. In nature, one of the basic reasons for extreme flood is extreme rainfall and therefore the probability of occurrence of a specific extreme rainfall. Since the extreme rainfalls are stochastic processes, probability theory and frequency analysis are used in hydrology immense and are needed to understand and describe the phenomena such as rainfall and flood (extreme streamflow). Choosing a probability distribution that provides a good fit to precipitation and flood flows has been a topic of interest in hydrology, meteorology, and others.

Concentrating of frequency analyses is on estimation of return periods of a phenomenon which associated with annual rainfall and streamflow (discharge) of diverse values. The estimated value of the random variable is also estimated for a given probability. A random variable such as rainfall or streamflow (discharge) is an amount that depends on chance of the values or range of values can be predicted only with probability, not with certainty. A distribution function provides a probabilistic model of phenomenon represented by a particular random variable [9]. Sheng and Michio showed that Pearson Type III and Log Pearson Type III distributions are acceptabale for annual, seasonal and monthly precipitation in Japan [13]. Khosravi et al. found that Log Pearson Type III distribution is suitable for annual discharges (peak and mean discharges) estimation in Minab River at Iran [10].

Strupczewski et al. [14] studied an application of distributions to modelling of annual peak flows in general and of Polish data sets in particular. Mohammed et al. evaluated the probability distribution models for the prediction of inflows of Kainji reservoir, Niger state, Nigeria and found that the Gumbel Extermal Type I (EVI) model gave the best fit to inflow series [11]. Standard probability distribution functions commonly used in water resources engineering have been identified in the literature, e.g. [1]-[8], [12], [15].

Karkheh river basin has experienced extreme rainfall events during the winter and spring seasons that last for several hours and lead to flash flood. The location of interest of this study is Karkheh river basin in Khuzestan province at Iran (Jelogir Majin and Pole Zal gauging stations) since flash floods occur in this River every year. We selected Jelogir Majin and Pole Zal gauging stations in this basin because they are located above the Karkheh dam reservoir. We want to evaluate the best probability distribution for entrance flow to the reservoir. The period for study is 40 years (1966-2015) for precipitation and 58 years (1958-2015) for flood flows. This series after homogeneity test and statistical adequacy was compared and evaluated by use of SMADA software graphical test and Residual Sum of Squares (R.S.S).

The objectives of this study are:

(i) To perform the frequency analysis of annual precipitations and discharges with six commonly used probability distributions (Normal, 2 Parameter Log Normal, 3 Parameter Log Normal, Pearson Type III, Log Pearson Type III and Gumbel Extermal Type I) to Karkheh river basin.
(ii) To identify the most appropriate probability distribution.
(iii) To estimate the annual precipitations and discharges for selected return periods (T= 2yrs, 5yrs, 10yrs, 25yrs, 50yrs, 100yrs and 200 years).

The Karkheh basin in west of the Iran, located in the central and southern regions of the Zagros mountain range and its area is more than 50000 km^2. In terms of the geographical coordination, this region has been extended between 46˚ 57′ - 49˚ 10′ E longitudes and 31˚ 48′ - 34˚ 56′ N latitudes. There

Karim Hamidi machekposhti is Ph.D. in Water Sciences and Engineering Department, Science and Research Branch, Islamic Azad University, Tehran, Iran (phone: 00989197704935- fax: 00982144865179-82; e-mail: karim.hamidi@srbiau.ac.ir).

Hossein Sedghi is with Water Sciences and Engineering Department, Science and Research Branch, Islamic Azad University, Tehran, Iran (e-mail: h.sedghi1320@gmail.com).

are several hydrometric stations on Karkheh River in this basin. The 900-km long Karkheh River (the Karkheh River is formed from connecting Kashkan and Saymareh rivers and directly connected to the Karkheh dam reservoir is the largest surface reservoir in the region, which has an important role in supplying water to the region) is the third largest river in Iran based on annual average flow. The basin's climate of Karkheh is very similar to mediterranean climate with hot/dry summers and mild/wet winters. The average of annual rainfall in this basin is 750 mm in the northern mountains and 150 mm in the southern arid plains. The study area position is depicted in Fig. 1.

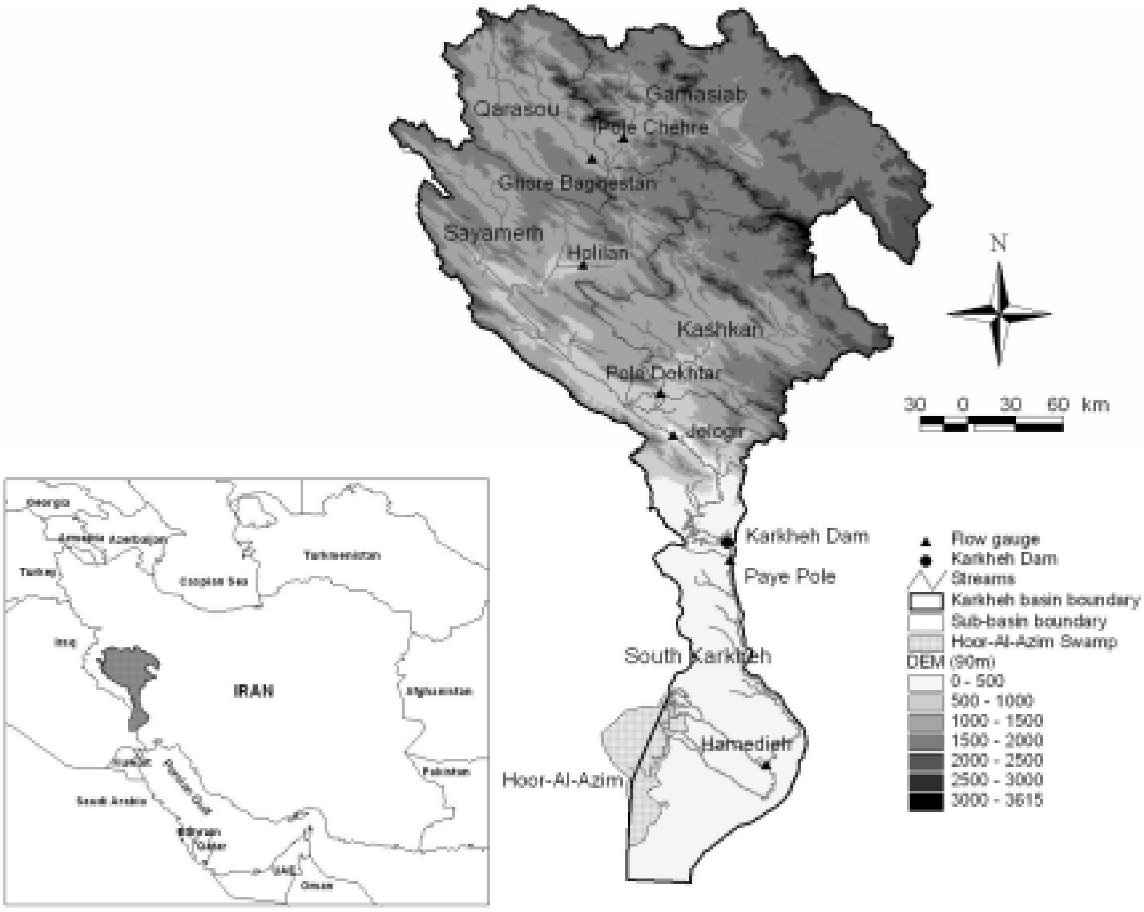

Fig. 1 The Karkheh River Basin

II. Methods

This study respectively was directed to annual rainfall (maximum precipitation) and flood flow (peak and maximum discharges) data recorded at Karkheh River for the years 1966 to 2015 and the years 1958 to 2015, respectively. The data for these years were taken from Iran Water Resources Management Organization (IWRMO). In this study, we wanted to examine the ability of annual data instead daily or monthly or other measures. If we can prove that annual data of hydrologic phenomenon (such as rainfall and discharge) are appropiate for prediction, then we can easily predict flooding and reduce flood damages. In order to control the data in two mentioned stations, we must check three conditions; adequacy, accuracy, and relevance. We use run test method for homogeneity and Water Resources Council (WRC) test for outlier detection of data. Our precipitation and discharge data have all above conditions and there is no outlier data. Therefore, these data are suitable for estimation. For the estimation, precipitation and discharge changes with different return periods were used theoretical probability distributions. It provides to estimate the water potential of Karkheh River with different return periods. The probability that events such as floods, wind storms or tornadoes will occur is often expressed as a return period. The inverse of probability (generally expressed in %) gives the estimated time interval between events of a similar size or intensity.

In this study, SMADA software is used for annual extreme rainfall and flood flows data. SMADA is a collection of tools to assist in analysis and design of stormwater systems. These include tools to perform hydraulic calculations, hydrologic calculations, hydrograph generation, statistical calculations, BMP selection, and pollutant loading. The system is expanded with new tools on a regular basis. The SMADA software facilitates the both the probability and probability distributions analysis in combination. The subprogram DISTRUB 2.0 was used for fitting the disribution.

The best distributions were determined with comparison of observational and predicted data. For this purpose, we apply Residual Sum of Squares (R.S.S) for each distribution. Equation (1) calculates R.S.S

$$R.S.S = \sqrt{\frac{\sum (Q_e - Q_0)^2}{n-m}} \quad (1)$$

In the above equation, Q_e and Q_o are the predicted and observed values for data respectively, n and m are the numbers of data and distribution parameter. The value of m is 2 for Normal, 2 Parameter Log Normal and Gumbel Extermal Type I distributions and also is 3 for Pearson Type III, Log Pearson Type III and 3 Parameter Log Normal distributions.

The objective of the study was to estimate the probable precipitation and discharge for development of decision support system for farm level irrigation scheduling.

III. Result and Discussion

Annual extreme precipitations and discharges for Karkheh River at two stations are checked whether outliers exist in the data series before using them. High and low outlier value had not been seen in the studied series, therefore we could start the modelling and predicted the extreme annual rainfall and streamflow (discharge) in the studied station.

The plot of the predicted precipitations and discharges against return period are presented in Figs. 2-7. In Tables I-VI, the Residual Sum of Squares (R.S.S) values of this series for six common distributions can be seen. In these tables, Q_2, Q_3, etc. show the amount of precipitations or discharges with different return periods.

According to Figs. 2 and 3 and Tables I and II, Pearson Type III and Log Pearson Type III has the maximum fitting for annual maximum precipitations in the two stations studied. In Jelogir Majin and Pole Zal stations, the minimum R.S.S value for annual maximum precipitation are 1.90, 1.74, 1.69 and 1.53, respectively. So they are the best distribution for estimating precipitation data.

For peak and maximum discharges in these two stations according to Figs. 4-7 and Tables III and VI. Log Pearson Type III has the highest fitting and has the best distribution. In Jelogir Majin station, the minimum R.S.S value for annual peak and maximum discharges are 145.91 and 141.08 and in Pole Zal station are 13.67 and 8.95, respectively. Also, we saw in Figs. 4-7 and Tables III and VI that Normal and Log Normal distributions are unsuitable and have the minimum fitting between actual and predicted values because the negative values show in the Normal distribution that it is not rational, and there are not negative values for discharge in the nature. Also, the two above mentioned distributions had the highest R.S.S value, therefore they are unsuitable. 3 Parameter Log Normal, Pearson Type III and Gumbel Extremal Type I distributions have relatively the best fitting in the experimental and estimated curves apparently but the estimated R.S.S values are high, therefore they are unsuitable.

In overall, Log Pearson Type III and then Pearson Type III distributions are recommended for estimation of annual precipitation and discharge for the Karkheh river basin. This is consistent with the findings obtained by Sheng and Michio [13] and other researchers.

IV. Conclusion

A total of six probability distributions are applied to the series of annual extreme precipitations and discharges (peak, maximum values) of two stations for Karkheh river basin. The conclusions obtained from this study are as below.

- Based on the analysis of statistical tests, Log Pearson Type III and then Pearson Type III distributions proves to be the most appropriate distribution for annual maximum precipitations and Log Pearson Type III distribution for annual peak and maximum discharges at two stations under study for Karkheh river basin.
- In overall, Log Pearson Type III distribution is recommended for estimation of annual precipitation and flood flows (discharge) for the Karkheh river basin with the Pearson Type III distribution as a potential alternative.
- The main inputs of rainfall-runoff models and designing of hydrologic structures are estimating the rainfall and streamflow (discharge) with various frequencies and durations.
- Future research can be carried out by using the other stations in Karkheh river basin to verify that Log Pearson Type III and Pearson Type III distributions are the recommended Distributions.

TABLE I
Annual Maximum Precipitation with Different Return Period (mm) in Jelogir Majin Station

Return Period (year)	Probability Distribution					
	Normal	**2 Parameter Log Normal**	**3 parameter Log Normal**	**Pearson Type III**	**Log Pearson Type III**	**Gumbel Extremal Type I**
Q_{200}	89.53	101.96	97.55	99.35	107.31	104.12
Q_{100}	85.87	95.12	91.94	93.37	99.24	96.43
Q_{50}	81.86	88.15	86.10	87.15	91.19	88.72
Q_{25}	77.40	80.99	79.93	80.60	83.08	80.94
Q_{10}	70.50	71.04	71.04	71.22	72.05	70.47
Q_5	64.02	62.82	63.36	63.21	63.18	62.17
Q_3	57.98	56.02	56.75	56.41	55.99	55.59
Q_2	51.64	49.67	50.32	49.92	49.42	49.65
RSS	2.93	1.93	2.07	**1.90**	**1.74**	2.35

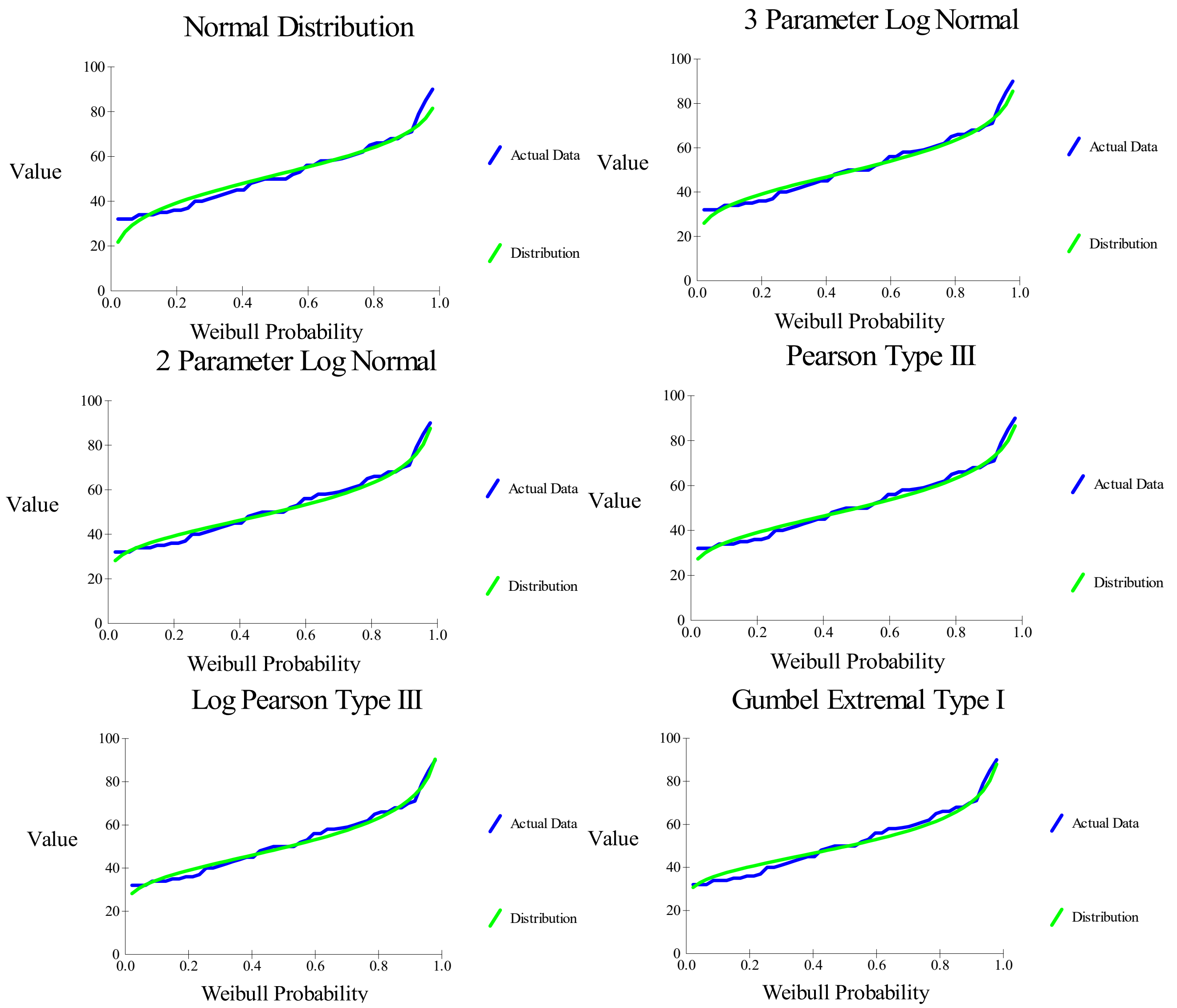

Fig. 2 Observational and predicted annual maximum precipitation values of distributions by SMADA in Jelogir Majin station

TABLE II
ANNUAL MAXIMUM PRECIPITATION WITH DIFFERENT RETURN PERIOD (MM) IN POLE ZAL STATION

Return Period (year)	**Probability Distribution**					
	Normal	**2 Parameter Log Normal**	**3 parameter Log Normal**	**Pearson Type III**	**Log Pearson Type III**	**Gumbel Extremal Type I**
Q_{200}	97.17	110.66	99.84	100.50	101.78	112.99
Q_{100}	93.19	103.22	95.26	95.78	97.27	104.65
Q_{50}	88.84	95.66	90.33	90.71	92.28	96.28
Q_{25}	84.00	87.90	84.94	85.18	86.69	87.85
Q_{10}	76.52	77.11	76.78	76.85	78.02	76.49
Q_{5}	69.49	68.20	69.33	69.29	69.98	67.49
Q_{3}	62.95	60.82	62.55	62.67	62.67	60.35
Q_{2}	56.07	53.92	55.60	55.48	55.26	53.91
RSS	1.81	2.38	1.79	**1.69**	**1.53**	3.07

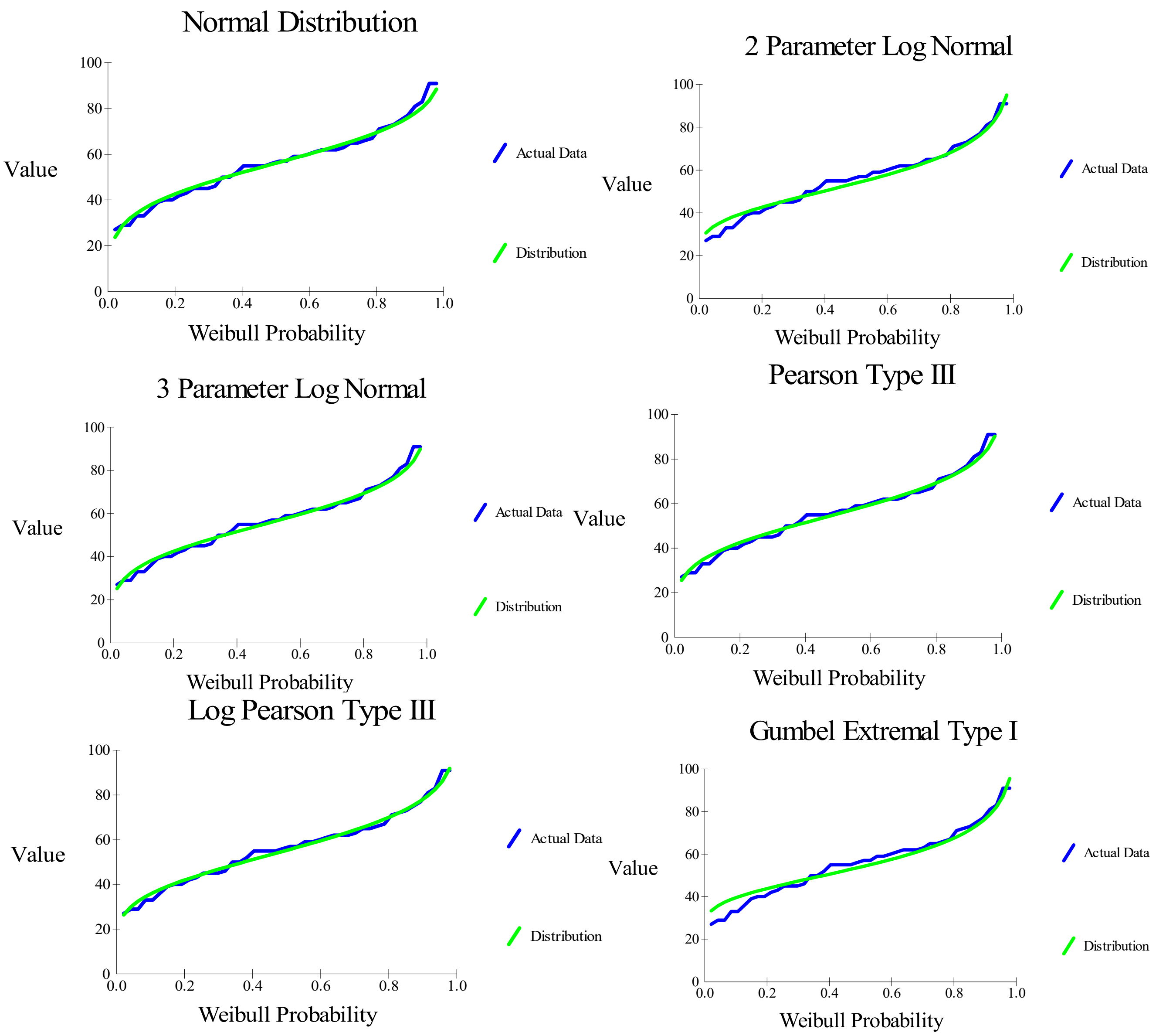

Fig. 3 Observational and predicted annual maximum precipitation values of distributions by SMADA in Pole Zal station

TABLE III
ANNUAL PEAK DISCHARGE WITH DIFFERENT RETURN PERIOD (M^3/S) IN JELOGIR MAJIN STATION

Return Period (year)	Probability Distribution					
	Normal	**2 Parameter Log Normal**	**3 parameter Log Normal**	**Pearson Type III**	**Log Pearson Type III**	**Gumbel Extremal Type I**
Q_{200}	3678.02	5602.78	5027.02	5291.30	5650.23	4532.98
Q_{100}	3443.62	4744.95	4406.03	4623.01	4896.14	4052.79
Q_{50}	3187.47	3956.97	3804.64	3963.88	4163.48	3570.85
Q_{25}	2902.64	3233.45	3219.31	3313.53	3452.76	3085.32
Q_{10}	2461.62	2365.22	2460.59	2465.62	2546.21	2430.84
Q_5	2047.91	1763.97	1884.21	1830.07	1881.76	1912.85
Q_3	1662.41	1342.14	1443.27	1362.03	1398.60	1501.47
Q_2	1257.22	1007.02	1061.81	983.96	1007.66	1130.50
RSS	367.82	187.38	191.53	176.83	**145.91**	234.79

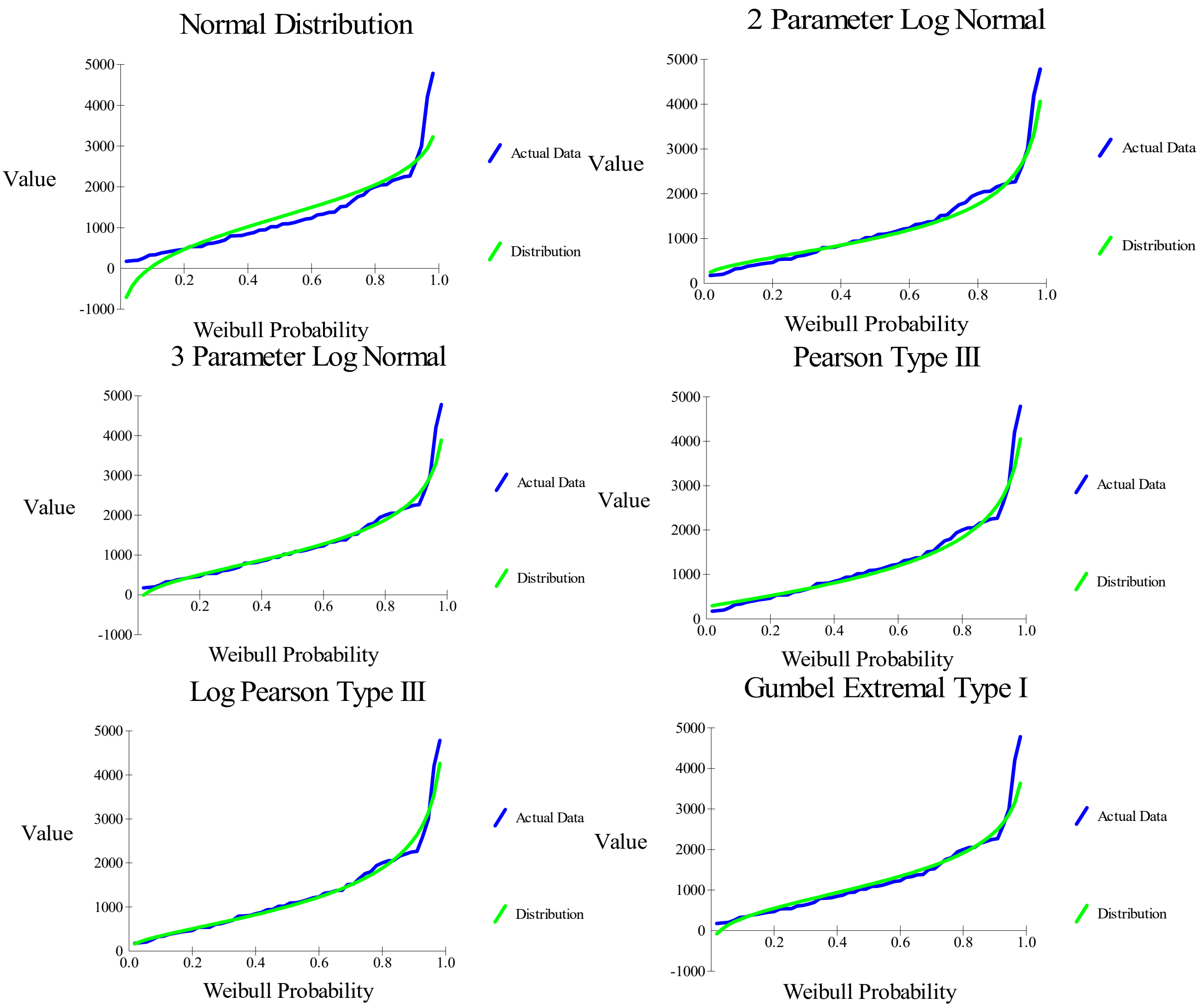

Fig. 4 Observational and predicted annual peak discharge values of distributions by SMADA in Jelogir Majin station

TABLE IV
ANNUAL PEAK DISCHARGE WITH DIFFERENT RETURN PERIOD (M^3/S) IN POLE ZAL STATION

Return Period (year)	Probability Distribution					
	Normal	**2 Parameter Log Normal**	**3 parameter Log Normal**	**Pearson Type III**	**Log Pearson Type III**	**Gumbel Extremal Type I**
Q_{200}	750.59	1051.46	910.36	938.86	1064.21	910.96
Q_{100}	706.62	916.64	824.94	848.35	945.56	820.89
Q_{50}	658.57	788.99	738.57	756.25	827.16	730.49
Q_{25}	605.15	667.82	650.43	661.86	708.73	639.41
Q_{10}	522.42	515.86	528.74	531.56	551.15	516.65
Q_5	444.82	404.90	429.06	425.79	429.34	419.49
Q_3	372.51	323.10	347.25	340.56	336.01	342.32
Q_2	296.51	254.87	271.38	263.833	256.28	272.74
RSS	47.87	26.38	24.16	20.02	**13.67**	28.67

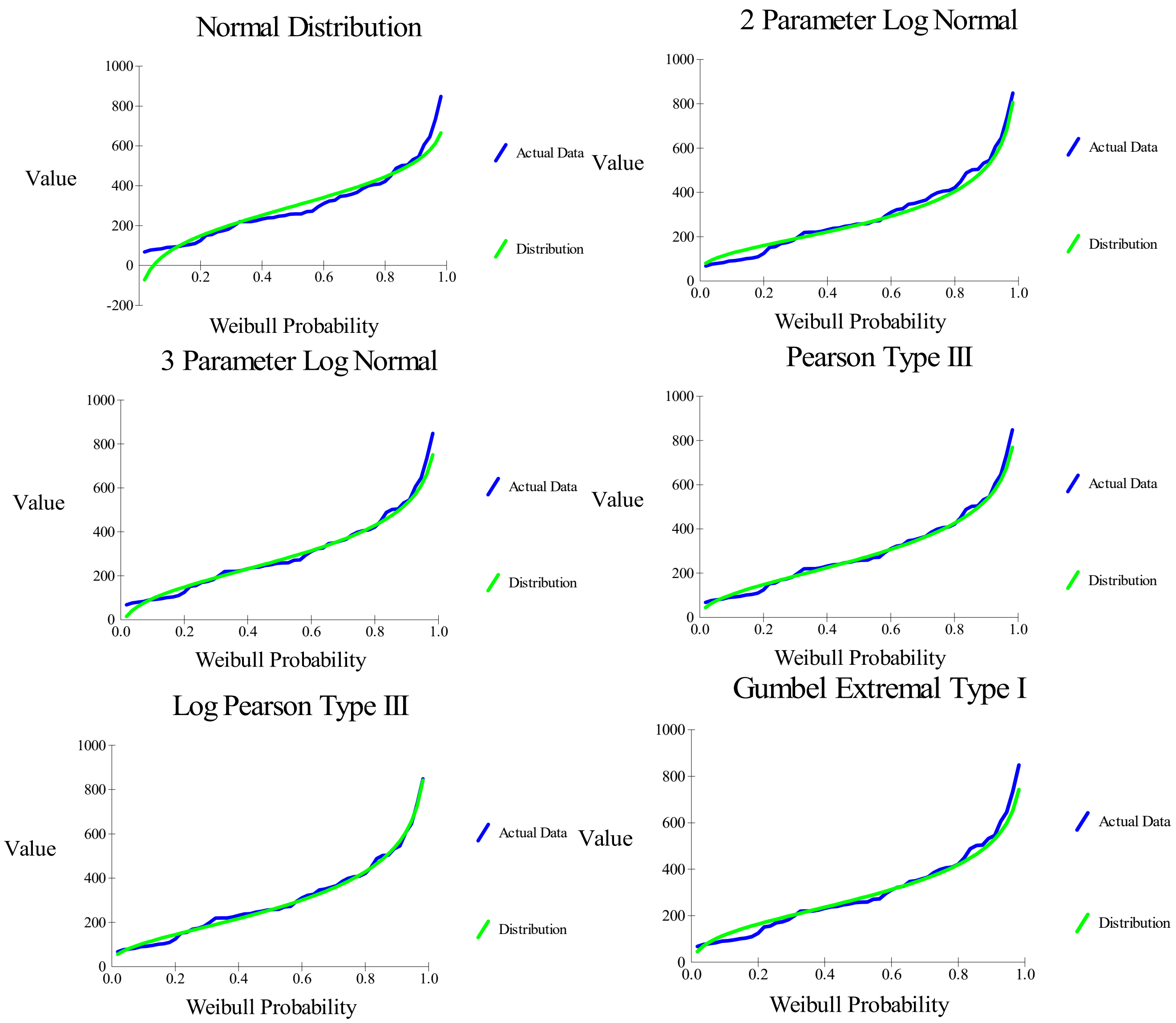

Fig. 5 Observational and predicted annual peak discharge values of distributions by SMADA in Pole Zal station

TABLE V
ANNUAL MAXIMUM DISCHARGE WITH DIFFERENT RETURN PERIOD (M^3/S) IN JELOGIR MAJIN STATION

Return Period (year)	Probability Distribution					
	Normal	2 Parameter Log Normal	3 parameter Log Normal	Pearson Type III	Log Pearson Type III	Gumbel Extremal Type I
Q_{200}	446.58	616.16	536.56	552.60	599.44	540.37
Q_{100}	420.87	539.86	487.65	500.81	538.61	487.70
Q_{50}	392.77	467.24	438.06	447.99	476.82	434.83
Q_{25}	361.52	397.90	387.30	393.73	413.84	381.56
Q_{10}	313.14	310.27	316.93	318.56	327.99	309.76
Q_{5}	267.76	245.70	259.02	257.24	259.81	252.94
Q_{3}	225.47	197.68	211.27	207.58	206.30	207.81
Q_{2}	181.02	157.30	166.80	162.60	159.57	167.11
RSS	27.12	15.66	14.23	12.11	**8.95**	16.85

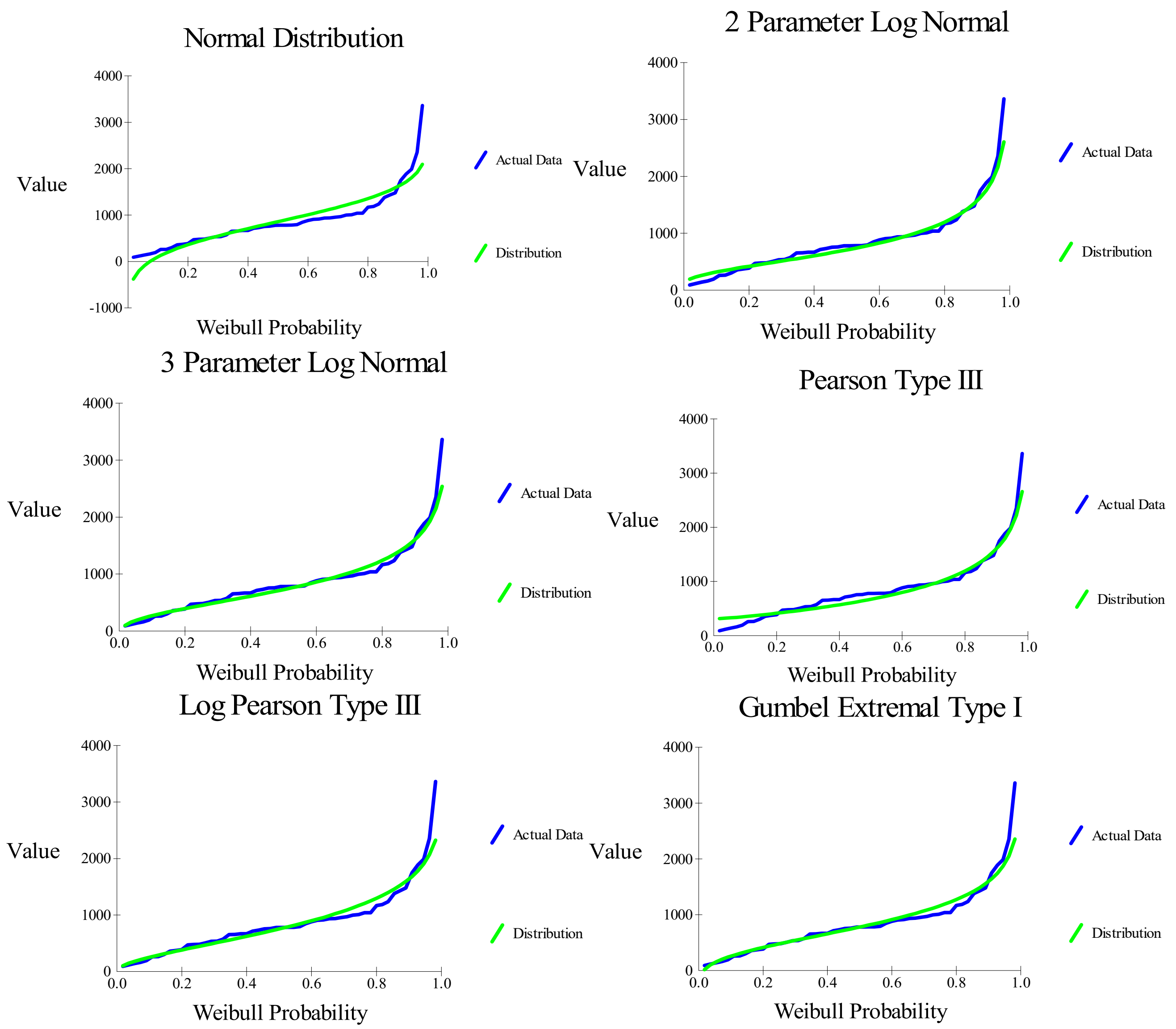

Fig. 6 Observational and predicted annual maximum discharge values of distributions by SMADA in Jelogir Majin station

TABLE VI
ANNUAL MAXIMUM DISCHARGE WITH DIFFERENT RETURN PERIOD (M^3/S) IN POLE ZAL STATION

Return Period (year)	**Probability Distribution**					
	Normal	**2 Parameter Log Normal**	**3 parameter Log Normal**	**Pearson Type III**	**Log Pearson Type III**	**Gumbel Extremal Type I**
Q_{200}	446.58	616.16	536.56	552.60	599.44	540.37
Q_{100}	420.87	539.86	487.65	500.81	538.61	487.70
Q_{50}	392.77	467.24	438.06	447.99	476.82	434.83
Q_{25}	361.52	397.90	387.30	393.73	413.84	381.56
Q_{10}	313.14	310.27	316.93	318.56	327.99	309.76
Q_5	267.76	245.70	259.02	257.24	259.81	252.94
Q_3	225.47	197.68	211.27	207.58	206.30	207.81
Q_2	181.02	157.30	166.80	162.60	159.57	167.11
RSS	241.06	148.20	144.25	147.87	**141.08**	162.58

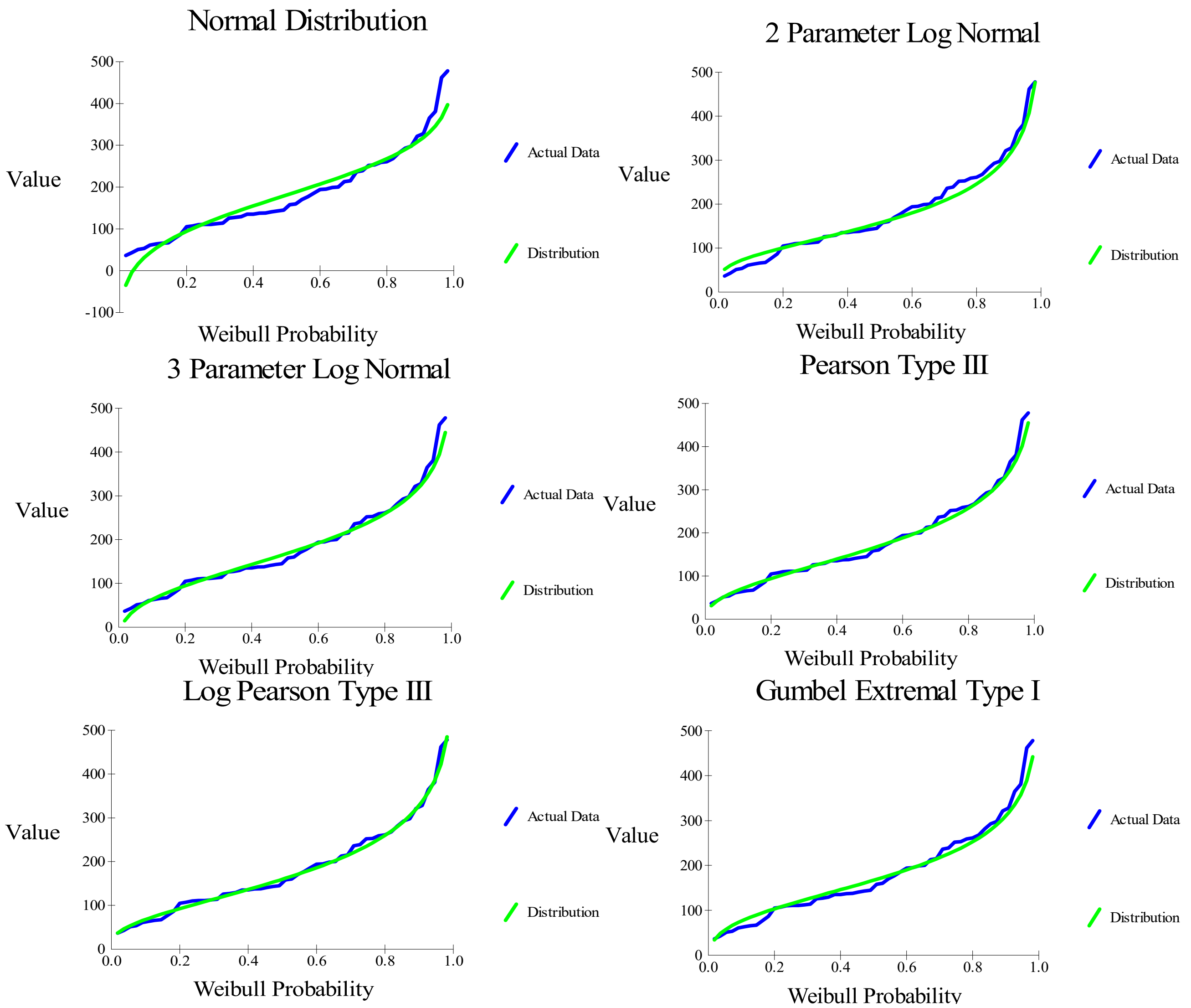

Fig. 7 Observational and predicted annual maximum discharge values of distributions by SMADA in Pole Zal station

References

[1] Abida H, Ellouze M. Probability distribution of flood flows in Tunisia. Hydrology and Earth System Sciences Discussions. 2007; 4(2): 957-981.

[2] Alias NE, Takara K. Probability Distribution for Extreme Hydrological Value-Series in the Yodo River Basin, Japan and Kuala Lumpur, Malaysia. In Proceedings of the 2nd International Conference on Water Resources. Langkawi, Malaysia. 2012.

[3] Atroosh KB, Moustafa AT. An Estimation of the Probability Distribution of Wadi Bana Flow in the Abyan Delta of Yemen. Journal of Agricultural Science. 2012; 4(6): 80-89.

[4] Garba H, Ismail A, Tsoho U. Fitting Probability Distribution Functions to Discharge Variability of Kaduna River. International Journal of Modern Engineering Research (IJMER). 2013; 3(5): 2848-2852.

[5] Haktan Y. Comparison of Various Flood Frequency Distribution using Annual Flood Peaks Data of Rivers in Anatolis. J. Hdrol. 1992; 136: 1-31.

[6] Hamidi machekposhti K, Sedghi H, Telvari A, Babazadeh H. Determination of Suitable Probability Distribution for Annual Discharges Estimation (Case Study: Karkheh River at Iran). International Journal of Probability and Statistics. 2016; 5(3): 73-81.

[7] Hosking JRM, Wallis JR. Some Statistics Useful in Regional Frequency Analysis. Water Resources Research. 1993; 29(2): 271-281.

[8] Ibrahim MH, Isiguzo EA. Flood Frequency Analysis of Guara River Catchment at Jere, Kaduna State, Nigeria. Scientific Research and Essay. 2009; 4(6): 636–646.

[9] Izinyon, OC, Ihimekpen N, Igbinoba GE. Flood Frequency Analysis of Ikpoba River Catchment at Benin City Using Log Pearson Type III Distribution. Journal of Emerging Trends in Engineering and Applied Sciences (JETEAS). 2011; 2(1): 50-55.

[10] Khosravi GH, Majidi A, Nohegar A. Determination of Suitable Probability Distribution for Annual Mean and Peak Discharges Estimation (Case Study: Minab River- Barantin Gage, Iran). International Journal of Probability and Statistics. 2012; 1(5): 160-163.

[11] Mohammed JM, Otache YM, Jibril I, Mohammed ISH. Evaluation of Best-Fit Probability Distribution Models for the Prediction of Inflows of Kainji Reservoir, Niger State, Nigeria. Air, Soil and Water Research. 2017; 10: 1–7.

[12] Ni Lar W, Khin Maung W. The Probability Distributions of Daily Rainfall for Kuantan River Basin in Malaysia. International Journal of Science and Research (IJSR). 2014; 3(8): 977-983.

[13] Sheng Y, Michio H. Probability distribution of annual, seasonal and monthly precipitation in Japan. Hydrological sciences journal. 2007; 52(5): 863-877.

[14] Strupczewski WG, Kochanek K, Markiewicz I, Bogdanowicz E, Weglarczyk S, Singh VP. On the tails of distributions of annual peak flows. Hydrology Research. 2011; 42(2-3): 171-192.

[15] Vogel R, Wilson I. Probability distribution of annual maximum, mean and minimum stream flows in the United States. Journal of Hydrologic Engineering. 1996; 1(2): 69-76.