International Journal of Scientific Research in Civil Engineering
© 2020IJSRCE | Volume 4 | Issue 4 | ISSN: 2456-6667

Evaluation of Estimators of Probability Distributions for Frequency Analysis of Rainfall and River Flow Data

N. Vivekanandan
Central Water and Power Research Station, Pune, Maharashtra, India
E-mail: anandaan@rediffmail.com

Article Info
Volume 4, Issue 4
Page Number: 67-75
Publication Issue:
July-August, 2020

ABSTRACT

Assessment of extreme rainfall and peak flood for a given return period is of utmost importance for planning and design of hydraulic structures. This can be achieved through Extreme Value Analysis (EVA) of rainfall and Flood Frequency Analysis (FFA) of river flow data by fitting 2-parameter Log Normal, Extreme Value Type-1, Generalized Extreme Value and Log Pearson Type-3 (LP3) distributions to the annual maximum series of observed data. Based on the intended applications and the variate under consideration, method of moments and Maximum Likelihood Method (MLM) are used for determination of parameters of the distributions. The adequacy of fitting probability distributions applied in frequency analysis of rainfall and river flow data was evaluated by quantitative assessment using Goodness-of-Fit (viz., Chi-square and Kolmogorov-Smirnov) and diagnostic (viz., Correlation Coefficient and Root Mean Squared Error) tests, and qualitative assessment by the fitted curves of the estimated values. Based on quantitative and qualitative assessments, the study shows the LP3 (MLM) is better suited for estimation of extreme rainfall and peak flood amongst four distributions adopted in EVA and FFA.

Article History
Accepted : 01 Aug 2020
Published: 06 Aug 2020

Keywords: Chi-Square, Correlation Coefficient, Kolmogorov-Smirnov, Log Pearson Type-3, Maximum Likelihood Method, Rainfall, Root Mean Squared Error

I. INTRODUCTION

Assessment of Extreme Rainfall (ER) and Peak Flood (PF) is considered as the important parameters for planning and design of hydraulic structures, river protection works and development of integrated water resources management projects. For this purpose, Extreme Value Analysis (EVA) of rainfall and Flood Frequency Analysis (FFA) of river flow data are generally carried out [1]. This can be achieved by fitting probability distributions to the series of observed Annual 1-day Maximum Rainfall (AMR) and Annual Peak Flood (APF) data.

A number of probability distributions viz., 2-paramater Log-Normal (LN2), Extreme Value Type-1 (EV1), Generalised Extreme Value (GEV) and Log Pearson Type-3 (LP3) distributions are widely applied in EVA and FFA [2]. Generally, Method of Moments (MoM) is used in determining the parameters of the probability distributions. Sometimes, it is difficult to assess the exact information about the shape of a distribution that is conveyed by its third and higher order moments. Also, when the sample size is small, the numerical values of sample moments can be very different from those of the probability distribution from which the sample was drawn. It is also reported that the estimated parameters of the distributions fitted by MoM are often less accurate than those obtained by other parameter estimation procedures viz., Maximum Likelihood Method (MLM), method of least squares and probability weighted moments [3]. AlHassoun [4] carried out a study on developing empirical formula to estimate rainfall intensity in Riyadh region using EV1 (also known as Gumbel), LN2 and LP3. He concluded that the LP3 distribution gives better accuracy amongst three distributions studied in estimation of rainfall intensity. Mohammed and Azhar [5] derived hydrometeorological approach to estimate the design flood at Kol Dam in the Satluj River Basin using Snyder's probable maximum flood hydrograph and standard project hydrograph with Central Water Commission of India recommendations.

Suhartano et al. [6] applied the Normal, LN2, LP3 and EV1 distributions to analyse the design flood by FFA in Lesti sub watershed. Ul Hassan et al. [7] adopted the GEV, Pearson Type-3, EV1, 3-parameter Log Normal and Generalized Logistic distributions in estimating the flood at five gauging sites of Torne River. Moreover, when different distributional models are applied for EVA and FFA, a common problem that arises is how to determine which model fits best for a given set of data. This can be answered by formal statistical procedures involving Goodness-of-Fit (GoF) and diagnostic tests; and the results are quantifiable and reliable. Qualitative assessment is made from the fitted curves of the estimated ER and PF. For quantitative assessment on rainfall and river flow data within the observed range, Chi-square (χ^2) and Kolmogorov-Smirnov (KS) tests are applied [8]. A diagnostic test viz., Correlation Coefficient (CC) and Root Mean Squared Error (RMSE) is used for the selection of best fit probability distribution for estimation of ER and PF.

This paper presents a study on evaluation of estimators of probability distributions adopted in EVA and FFA with illustrative example and the results obtained from the study.

II. METHODOLOGY

The procedures involved in EVA and FFA are: (i) prepare the observed AMR and APF series from the daily data series; (ii) determination of parameters of LN2, EV1, GEV and LP3 by MoM and MLM; and estimate the extreme values (i.e., 1-day maximum rainfall or peak flood) for different return periods; (iii) check the adequacy of fitting probability distributions using GoF and diagnostic tests; (iv) conduct quantitative and qualitative assessments; and (v) analyse the results and suggestions made thereof. Table 1 presents the Cumulative Distribution Function (CDF) and quantile estimator (x(T)) of probability distributions adopted in EVA and FFA. Procedures for determination of parameters of the distributions by MoM and MLM are available in the text book titled 'Flood Frequency Analysis' by Rao and Hamed [9].

TABLE 1

CDF, QUANTILE ESTIMATORS OF LN2, EV1, GEV AND LP3 DISTRIBUTIONS

Distribution	CDF (F(x))	Quantile estimator (x(T))
LN2	$F(x)=\phi\left(\frac{\ln(x)-\mu_y}{\sigma_y}\right)$, σ_y>0, x>0	$x(p)=\exp(\mu_y+\sigma_y\phi^{-1}(P))$
EV1	$F(x)=\exp\left[-\exp\left(-\frac{x-\xi}{\alpha}\right)\right]$, -∞<x<∞, α>0	$x(T)=\xi-\alpha\ln[-\ln(1-(1/T))]$
GEV	$F(x)=\exp\left(-\left[1-\frac{k(x-\xi)}{\alpha}\right]^{1/k}\right)$, α>0	$x(T)=\xi+\frac{\alpha}{k}\left(1-[-\ln(1-(1/T))]^{k}\right)$
LP3	$F(x)=\begin{cases}G\left(k,\frac{\ln(x)-\xi}{\alpha}\right) & ,\alpha>0\\ 1-G\left(k,\frac{\ln(x)-\xi}{\alpha}\right) & ,\alpha<0\end{cases}$	No explicit expression of the quantile function is available.

In Table 1, ξ is the location parameter, α is the scale parameter, k is the shape parameter, x is the variable, μ_y and σ_y are the average and standard deviation of the logarithmic transformed series of x (i.e., y=ln(x)), P is the probability of exceedance, ϕ and $\phi^{-1}(\ldots)$ are the CDF and quantile function of the standard normal distribution, G(…) is the incomplete Gamma integral, F(x) is the CDF of x and x(T) is estimated value of the variable for a return period (T).

A) Goodness-of-Fit Tests

GoF tests are essential for checking the adequacy of probability distributions to the Annual Maximum Series (AMS) of observed data used in EVA and FFA. Out of a number GoF tests available, the widely accepted GoF tests are χ^2 and KS [10], which are used in the study. The theoretical descriptions of GoF tests statistic are given as below:

$$\chi^2=\sum_{j=1}^{NC}\frac{\left(O_j(x)-E_j(x)\right)^2}{E_j(x)} \quad \ldots (1)$$

where, $O_j(x)$ is the observed frequency value of x for j[th] class, $E_j(x)$ is the expected frequency value of x for j[th] class and NC is the number of frequency classes. The rejection region of χ^2 statistic at the desired significance level (η) is given by $\chi^2_C \geq \chi^2_{1-\eta,NC-m-1}$. Here, m denotes the number of parameters of the distribution and χ^2_C is the computed value of χ^2 statistic by probability distributions.

$$KS=\overset{N}{\underset{i=1}{Max}}\left|F_e(x(i))-F_D(x(i))\right| \quad \ldots (2)$$

where, $F_e(x(i))=i/(N+1)$ is the empirical CDF of x(i), $F_D(x(i))$ is the computed CDF of x(i), x(i) is the observed data for i[th] sample and N is the number of sample values [11]. If the computed values of GoF tests statistic given by the distribution are less than its theoretical values at the desired significance level then the distribution is considered to be acceptable for EVA or FFA at that level.

B) Diagnostic Test

Sometimes the GoF test results would not offer a conclusive inference thus posing a problem for the user in selecting a suitable probability distribution (with parameter estimation method) for their application. In such cases, a diagnostic test in adoption to GoF is applied for making inference. The selection of best fit probability distribution for estimation of ER and PF can be performed through CC and RMSE, which is defined as below:

$$CC=\frac{\sum_{i=1}^{N}(x(i)-\bar{x})(y(i)-\bar{y})}{\sqrt{\left(\sum_{i=1}^{N}(x(i)-\bar{x})^2\right)\left(\sum_{i=1}^{N}(y(i)-\bar{y})^2\right)}} \quad \ldots (3)$$

$$RMSE = \left((1/N) \sum_{i=1}^{N} (x(i) - y(i))^2 \right)^{1/2} \quad \ldots (4)$$

where, y(i) is the estimated value of x(i) for i^{th} sample, $\bar{x}$ is the average of observed values and $\bar{y}$ is the average of estimated values.The distribution has high CC and minimum RMSE is considered as better suited for estimation of ER and PF [12].

III. APPLICATION

In this paper, a study on evaluation of probability distributions viz., LN2, EV1, GEV and LP3 adopted in EVA and FFA was carried out. The parameters of distributions were determined by MoM and MLM; and are used for estimation of ER and PF. Daily rainfall data observed at Agartala rain-gauge station for the period 1901 to 2014 and daily APF data observed at Haora gauging site for the period 1990 to 2009 was used. The AMR and APF series was extracted from the daily data series and used in EVA and FFA. From the scrutiny of the daily rainfall data, it was observed that the data for four years (1952, 1954, 2004 and 2005) are missing. However, the data for the missing years were not considered in EVA. From the scrutiny of river flow data, it was observed that there are no missing values in the daily data series. Table 2 gives the descriptive statistics of AMR and APF data considered in the study.

TABLE 2
DESCRIPTIVE STATISTICS OF AMR AND APF

Data series	Average	SD	CS	CK	Minimum	Maximum
AMR	150.2 mm	90.1 mm	5.442	40.283	61.4 mm	880.0 mm
APF	177.8 m^3/s	79.0 m^3/s	1.285	1.871	78.0 m^3/s	394.9 m^3/s
SD: Standard Deviation; CS: Coefficient of Skewness; CK: Coefficient of Kurtosis						

IV. RESULTS AND DISCUSSIONS

By applying the procedures of EVA and FFA, as described above, parameters of the LN2, EV1, GEV and LP3 distributions were determined by MoM and MLM, and are used for estimation of ER (i.e., 1-day maximum) and PF. The EVA and FFA results are presented in Tables 3 and 4 while the plots are shown in Figures 1 and 2. For FFA of river flow data, as the observed AMS is available only for 20-years, the PFs for return period beyond 100-year are not estimated and therefore not presented in Table 4. From the results, it is noted that the estimated ER and PF obtained from LP3 (MLM) is relatively higher than those values of LN2, EV1 and GEV distributions. From Figure 1, it can be seen that the plots of ER estimates using LN2, GEV and LP3 distributions are in the form of exponential curve while the pattern of EV1 plots are in linear. Likewise, from Figure 2, it is noted that the plots of PF estimates using LN2, GEV, EV1 and LP3 distributions are in the form of linear curve.

A) Analysis Based on GoF Tests

The GoF test values for the AMR and APF series were computed by LN2, EV1, GEV and LP3 distributions, and the results are presented in Table 5. For the present study, number of frequency class is considered as 11 for AMR series while 5 for APF series while computing χ^2 test statistic. Also, for AMR series, the theoretical value of χ^2 at 5% significance level was determined based on the degrees of freedom viz., seven for 3-parameter distributions (viz., GEV and LP3) and eight for 2-parameter distributions (viz., EV1 and LN2). Likewise, for APF series, the theoretical value of χ^2 at 5% significance level was determined based on the degrees of freedom, viz., one for GEV and LP3 distributions while two for

EV1 and LN2. Based on GoF tests results, the following observations were drawn from the study:

i) χ^2 and KS tests results indicated that LN2, EV1 and GEV distributions (using MoM and MLM) are not acceptable for EVA of rainfall for Agartala.

ii) χ^2 and KS test results confirmed the applicability of LP3 (using MoM and MLM) distribution for EVA of rainfall.

iii) χ^2 and KS test results supported the use of LN2, EV1, GEV and LP3 distributions for FFA.

TABLE 3

EXTREME RAINFALL (mm) ESTIMATES USING LN2, EV1, GEV AND LP3 DISTRIBUTIONS

Return period (year)	LN2		EV1		GEV		LP3	
	MoM	MLM	MoM	MLM	MoM	MLM	MoM	MLM
2	128.8	137.4	135.4	137.6	127.8	130.8	125.4	145.5
5	205.4	188.5	215.0	183.7	190.3	180.0	178.1	208.7
10	262.1	222.4	267.8	214.1	242.4	220.5	227.2	276.3
20	320.6	254.9	318.4	243.4	302.4	266.8	287.1	366.9
25	340.0	265.2	334.5	252.7	323.7	283.2	309.2	402.2
50	402.3	297.2	383.9	281.3	397.7	339.5	387.9	535.7
100	467.9	329.3	433.0	309.6	485.0	405.5	484.9	714.2
200	537.4	361.6	481.9	337.9	588.3	483.0	604.4	953.3
500	635.5	405.1	546.4	375.2	754.8	606.7	806.1	1398.0
1000	714.9	438.7	595.2	403.4	908.2	719.7	1000.3	1868.9

TABLE 4

PEAK FLOOD (m^3/s) ESTIMATES USING LN2, EV1, GEV AND LP3 DISTRIBUTIONS

Return period (year)	LN2		EV1		GEV		LP3	
	MoM	MLM	MoM	MLM	MoM	MLM	MoM	MLM
2	162.5	163.4	164.8	164.3	163.9	162.4	160.5	159.5
5	232.2	229.9	234.6	227.1	233.0	226.1	230.6	228.1
10	279.9	274.9	280.9	268.6	279.8	270.2	281.5	279.0
20	326.6	318.6	325.2	308.5	325.5	314.0	333.6	331.9
25	341.6	332.6	339.3	321.1	340.1	328.2	350.8	349.5
50	388.4	376.1	382.6	360.1	385.7	372.9	406.2	406.8
100	436.1	420.2	425.6	398.8	431.7	418.7	464.9	468.2

TABLE 5

COMPUTED VALUES OF GOF TESTS STATISTIC USING LN2, EV1, GEV AND LP3 DISTRIBUTIONS

Data series	GoF tests	LN2		EV1		GEV		LP3	
		MoM	MLM	MoM	MLM	MoM	MLM	MoM	MLM
AMR	χ^2	57.400	56.500	66.200	61.450	23.800	23.450	4.600	4.752
	KS	0.205	0.208	0.216	0.219	0.130	0.135	0.057	0.070
APF	χ^2	1.000	1.000	2.000	2.000	1.000	1.000	1.000	1.000
	KS	0.051	0.052	0.050	0.051	0.050	0.051	0.062	0.061
Theoretical values: $\chi^2_{0.5,7}=14.067$; $\chi^2_{0.5,8}=15.057$; $\chi^2_{0.5,1}=3.841$; $\chi^2_{0.5,2}=5.991$; $KS_{0.5,110}$=0.125; $KS_{0.5,20}$=0.269									

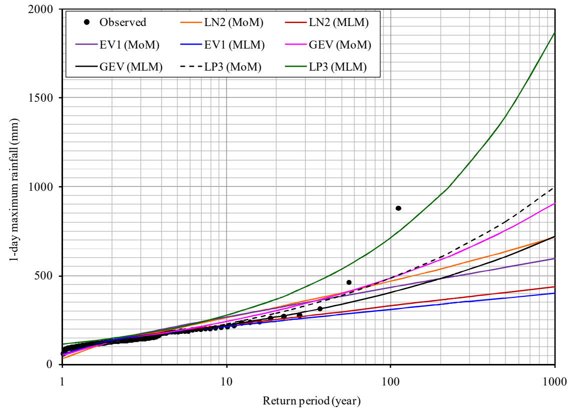

Figure 1: Plots 1-day maximum rainfall using LN2, EV1, GEV and LP3 distributions with observed data for Agartala

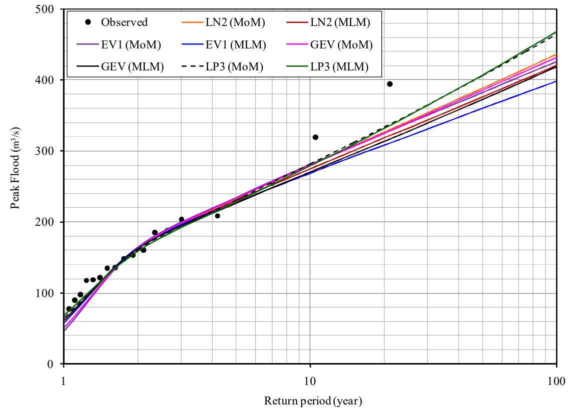

Figure 2: Plots of peak flood using LN2, EV1, GEV and LP3 distributions with observed data for Haora

B) Analysis Based on Diagnostic Test

In addition to GoF tests, for identifying the best suitable probability distribution amongst four distributions adopted in EVA and FFA, second line of action, i.e., CC and RMSE was applied and these values were computed for LN2, GEV, EV1 and LP3 distributions, and the results are presented in Table 6. From the diagnostic test results, it is noted that the RMSE values given by LP3 (MLM) for AMR series while LP3 (MoM) for APF series were found as minimum when compared with the corresponding values of LN2, EV1 and GEV. Also, from Table 6, it is noted that the CC values given by LN2, EV1, GEV and LP3 distributions vary from 0.823 to 0.927 for Agartala while 0.983 to 0.992 for Haora.

TABLE 6

VALUES OF INDICATORS OF DIAGNOSTIC TEST USING LN2, EV1, GEV AND LP3 DISTRIBUTIONS

Data series	Indicators	LN2		EV1		GEV		LP3	
		MoM	MLM	MoM	MLM	MoM	MLM	MoM	MLM
AMR	CC	0.861	0.827	0.823	0.823	0.900	0.897	0.921	0.927
	RMSE (mm)	20.222	26.507	22.350	27.686	18.499	22.169	18.110	15.798
APF	CC	0.987	0.986	0.983	0.983	0.985	0.987	0.991	0.992
	RMSE (m^3/s)	17.477	19.713	17.811	22.533	17.750	21.275	16.072	16.864

C) Selection of Probability Distribution

Based on EVA and FFA results obtained from quantitative assessment by using GoF and diagnostic tests, it was observed that the analysis offered diverging inferences and thus called for qualitative assessment. Hence, the best fit for estimation of rainfall and PF was re-assessed through fitted curves of the estimated values together with values of the indicators used in diagnostic test; and accordingly final selection was made.

i) RMSE values indicated that LP3 (MLM) for Agartala while LP3 (MoM) for Haora could be used for frequency analysis of rainfall and river flow data respectively.
ii) However, the estimated parameters of distributions fitted using MoM are often less accurate than MLM. Hence, for Haora, the RMSE value obtained from LP3 (MoM) is not considered for the selection of best fit for estimation of PF.
iii) In light of the above, after eliminating the RMSE value of LP3 (MoM), it is identified

that RMSE value of LP3 (MLM) is the second minimum next to LP3 (MoM).

iv) CC values on estimation of rainfall and flood using LP3 (MLM) distribution are noted to be 0.927 and 0.992 respectively.

Hence, qualitative assessment (plots of EVA results) of the outcomes was weighed with values of the indicators (viz., CC and RMSE) used in diagnostic test and accordingly LP3 (MLM) was found to be best fit for estimation of ER and PF. The estimated values (i.e., 1-day maximum rainfall for Agartala and peak flood for Haora) with 95% confidence limits using LP3 (MLM) distribution are presented in Table 7 while the plots are shown in Figures 3 and 4. From these figures, it can be seen that 90% of observed AMR and 100% of observed APF are within the confidence limits of the estimated values.

TABLE 7

ER AND PF ESTIMATES WITH 95% CONFIDENCE LIMITS USING LP3 (MLM) DISTRIBUTION

Return period (year)	1-day maximum rainfall (mm)	Confidence limit		Peak flood (m^3/s)	Confidence limit	
		LCL (mm)	UCL (mm)		LCL (m^3/s)	UCL (m^3/s)
2	145.5	120.1	170.9	159.5	130.1	188.9
5	208.7	160.0	257.4	228.1	180.0	276.3
10	276.3	200.3	352.2	279.0	215.2	342.8
20	240.2	240.2	493.7	331.9	250.3	413.5
25	402.2	254.5	550.0	349.5	262.4	436.7
50	535.7	300.4	770.9	406.8	300.0	513.6
100	714.2	355.0	1073.5	468.2	342.1	594.4
200	953.3	425.3	1481.7			
500	1398.0	575.0	2221.0			
1000	1868.9	750.2	2987.7			

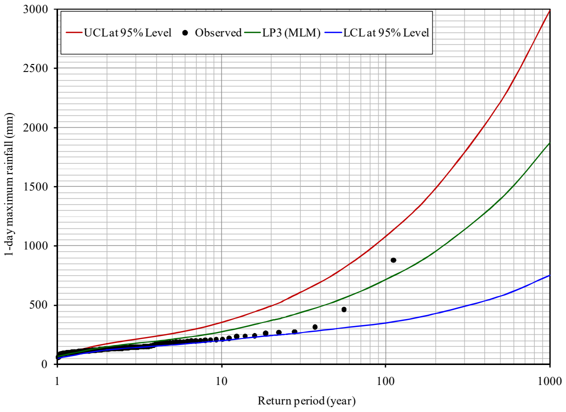

Figure 3: Plots of estimated extreme rainfall using LP3 (MLM) with confidence limits and observed data

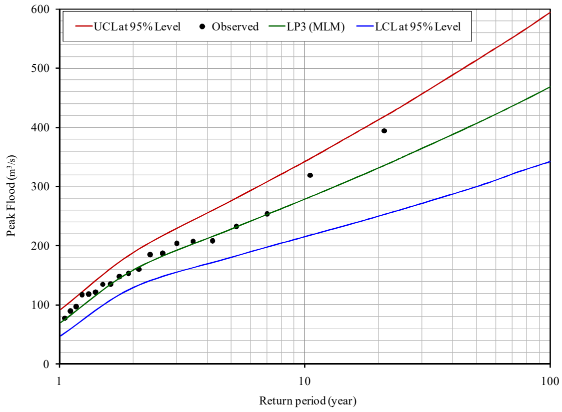

Figure 4: Plots of estimated peak flood using LP3 (MLM) with confidence limits and observed data

V. CONCLUSIONS

Frequency analysis of hydrometeorological parameters is essential in design consideration of establishment of hydraulic and civil structures. An effort is made to conduct a study on evaluation of estimators of LN2, EV1, GEV and LP3 distributions adopted in EVA for Agartala and FFA for Haora with a specific objective to identify best suitable distribution amongst four distributions for estimation of ER and PF. The adequacy of fitting probability distributions was checked by quantitative (viz., GoF and diagnostic tests) and qualitative (viz., fitted curves of the estimated values) assessments. Based on the results of the data analysis, the following conclusions were drawn from the study:

i) χ^2 and KS test results supported the use of LP3 (using MoM and MLM) for EVA of rainfall for Agartala.

ii) χ^2 and KS tests results didn't confirm the applicability of LN2, EV1 and GEV distributions for EVA of rainfall for Agartala while MoM and MLM is applied for determination of the parameters of the distributions.

iii) χ^2 and KS tests results indicated that LN2, EV1, GEV and LP3 distributions are acceptable for FFA for Haora.

iv) CC values on rainfall and flood estimation obtained from LP3 (MLM) distribution are noted to be 0.927 and 0.992 respectively.

v) Qualitative assessment (plots of estimated values) of the outcomes was weighed with the values of the indicators used in diagnostic test and accordingly LP3 (MLM) was found to be better suited for estimation of ER and PF.

By considering the data length (i.e., 110 years for Agartala and 20 years of Haora) of AMR and APF series available for the study, the study suggested that the estimated values of rainfall for return period beyond 250-year while peak flood for return period beyond 100-year may be cautiously used due to uncertainty in the higher order return periods while designing the hydraulic structures in the respective sites.

VI. ACKNOWLEDGEMENTS

The author is grateful to the Director In-charge, Central Water and Power Research Station, Pune, for providing the research facilities to carry out the study. The author is thankful to the Chief Engineer, Central Water Commission, Silchar and India Meteorological Department, Pune for the supply of river flow and rainfall data used in the study.

VII. REFERENCES

[1] AERB, Extreme values of meteorological parameters, AERB Safety Guide No. NF/SG/ S-3, Atomic Energy Regulatory Board (AERB), 2008.

[2] G. Di Balldassarre, A. Castellarin and A. Brath. 2006. Relationships between statistics of rainfall extremes and mean annual precipitation: an application for design storm estimation in northern central Italy. Hydrology and Earth System Sciences, 10 (2): 589–601.

[3] B. Naghavi, F.X. Yu and V.P. Singh. 1993. Comparative evaluation of frequency distributions for Louisiana extreme rainfall. Water Resources Bulletin, 29 (2): 211-219.

[4] S.A. AlHassoun. 2011. Developing empirical formulae to estimate rainfall intensity in Riyadh region. Journal of King Saud University-Engineering Sciences, 23 (1): 81–88.

[5] S. Mohammed and H. Azhar. 2017. Estimation of design flood at Kol dam using hydrometeorological approach. International Journal of Environmental Sciences & Natural Resources, 4 (1): 1-6.

[6] E. Suhartanto, M.L. Lily, N. Dina, I.H. Febri and A.K. Dwi. 2018. Estimation of Design Flood with Four Frequency Analysis Distributions. Asian Journal of Applied Science and Technology, 2 (1): 13-27.

[7] M. Ul Hassan, O. Hayat and Z. Noreen. 2019. Selecting the best probability distribution for at-site flood frequency analysis; a study of Torne River. SN Applied Science 1 (12): Article ID: 1629.

[8] USWRC. 1982. Guidelines for determining flood flow frequency. United States Water Resources Council (USWRC), Bulletin No. 17B (Revised), Washington, DC, New York.

[9] A.R. Rao and K.H. Hamed, 2000. Flood frequency analysis. CRC Press, Florida, USA,

[10] J. Zhang. 2002. Powerful goodness-of-fit tests based on the likelihood ratio. Journal of Royal Statistical Society, 64 (2): 281-294.

[11] P.E. Charles Annis. 2009. Goodness-of-Fit tests for statistical distributions. http://www.statisticalengineering.com.

[12] J. Chen and B.J. Adams. 2006, Integration of artificial neural networks with conceptual models in rainfall-runoff modelling, Journal of Hydrology, 318 (1-4)): 232-249.

Cite this article as:

N. Vivekanandan, "Evaluation of Estimators of Probability Distributions for Frequency Analysis of Rainfall and River Flow Data", International Journal of Scientific Research in Civil Engineering (IJSRCE), ISSN: 2456-6667, Volume 4, Issue 4, pp. 67-75, July-August, 2020. URL: http://ijsrce.com/IJSRCE204412.