International Journal of Emerging Engineering Research and Technology
Volume 7, Issue 4, 2019, PP 4-10
ISSN 2349-4395 (Print) & ISSN 2349-4409 (Online)

A Comparative Study on Estimation of Peak Flood Discharge using MOM and LMO Estimators of Probability Distributions

R. S. Bharadwaj[1*], (Mrs.) A. D. Thube[2], N. Vivekanandan[3], C. Srishailam[4]

[1]M.Tech. Scholar, Department of Civil Engineering, College of Engineering, Pune, Maharashtra, India

[2]Associate Professor, Department of Civil Engineering, College of Engineering, Pune, Maharashtra, India

[3]Scientist-B, Central Water and Power Research Station, Pune, Maharashtra, India

[4]Scientist-C, Central Water and Power Research Station, Pune, Maharashtra, India

***Corresponding Author:** R. S. Bharadwaj, M.Tech. Scholar, Department of Civil Engineering, College of Engineering, Pune, Maharashtra, India, E-mail: rohan_bharadwaj90@yahoo.co.in*

ABSTRACT

Estimation of Peak Flood Discharge (PFD) for a particular return period is carried out by fitting Probability Distribution (PD) to the observed discharge data to arrive at a design value for designing civil and hydraulic structures. This paper illustrates the adoption of Exponential, Extreme Value Type-1 (EV1), Extreme Value Type-2, Generalized Extreme Value and Normal PDs for estimation of PFD at Badlapur GD site, Maharashtra. Parameters of the five PDs are determined by method of moments and L-Moments (LMO), and used for estimation of PFD. The adequacy of fitting of PDs is evaluated by Goodness-of-Fit tests viz., Chi-Square and Kolmogorov-Smirnov; and diagnostic tests viz., Mean Absolute Percentage Error and D-Index. The outcomes of the GoF and diagnostic tests results indicated that the EV1 (using LMO) distribution is better suited PD for estimation of PFD at Badlapur. Based on the results obtained from the study, the suggestions are made and presented in the paper.

Keywords: *Chi-Sqaure test, D-index, Peak Flood Discharge, Extreme Value Type-1, Kolmogorov-Smirnov test, L-Moments, Mean Absolute Percentage Error*

INTRODUCTION

Extreme flood events usually cause a lot of damage to life and properties of human society. Determination of the frequencies and magnitudes of those events are important for flood plain management and design of hydraulic structures, civil protection plans, etc. However, length of available records is not enough large to define the risk of flood, extreme rainfall, low-flow, drought, etc. In these cases, Extreme Value Analysis (EVA) involves fitting of samples to a Probability Distribution (PD) is considered as an alternative tool to arrive at a design value [1]. The EVA includes three underlying assumptions such as

- The extremes are random variable, and thus can be described by a PD;
- The data series is independent; and
- The PD does not change from sample to sample (homogeneity).

Number of PDs such as Exponential (EXP), Extreme Value Type-1 (EV1), Extreme Value Type-2 (EV2), Generalized Extreme Value (GEV) and Normal (NOR) are widely adopted in EVA of hydrological variables.

Generally, Method of Moments (MoM) is used in determining the parameters of the PDs.

Sometimes, it is difficult to assess 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 distributions fitted using MoM are often less accurate than those obtained by other parameter estimation procedures such as maximum likelihood method, method of least squares and probability weighted moments [2]. To address these shortcomings, the application of alternative approach, namely L-Moments (LMO) is used for EVA [3]. In the recent past, number of studies has been carried out by different researchers on adoption of PDs for estimation of Peak Flood Discharge (PFD). Topaloglu [4] reported that the frequency analysis of the largest, or the smallest, of a sequence of hydrologic events are being an essential part of the design of hydraulic structures. Guevara [5] carried out hydrological analysis using probabilistic approach to estimate the engineering design parameters of storms in Venezuela. Saf et al. [6] stated that Log-Pearson Type III distribution is more appropriate instead of the widely used Gumbel (also known as EV1) distribution for probability distribution modeling of extreme values. Mujere [7] aimed at analyzing the frequency of Nyanyadzi River floods in Zimbabwe using the Gumbel distribution. Barrati et al. [8] proposed an approach to infer the flood frequency distribution at seasonal and annual time scale to estimate the peak flow that is expected for an assigned return period (T) independently of the season in which it occurs (i.e. annual flood frequency regime) as well as in different selected sub-yearly periods (i.e. seasonal flood frequency regime) for Blue Nile at Sudan-Ethiopia border. Olumide et al. [9] applied NOR and EV1 distributions for prediction of rainfall and runoff at Tagwai dam site in Minna, Nigeria. They have also expressed that the NOR distribution is better suited for rainfall prediction while Log-Gumbel for runoff.

Haberlandt and Radtke [10] carried out model calibration studies in three meoscale catchments in Northern Germany to calibrate a hydrological model directly on PDs of observed peak flows using stochastic rainfall as input if its purpose is the application for derived flood frequency analysis. Mohammed and Azhar [11] 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. [12] applied four probability distributions viz., NOR, Log-NOR, Log Pearson Type-III and Gumbel to analyze the design flood by flood frequency analysis in Lesti sub watershed.

Kolbjørn et al. [13] used annual maximum data from four selected Norwegian catchments, and historical flood information to provide an indication of water levels for the largest floods in the last two to three hundred years. Winter et al. [14] compared 100-year derived flood estimates in 16 catchments in Vorarlberg (Austria) to the flood frequency analysis based on observed discharges and a design storm approach. Moreover, when different distributional models are used for modeling of annual maximum series of either hydrological or hydrometeorological variables (i.e., rainfall, temperature, peak flood, evaporation, etc.), 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 plot of the recorded and estimated PFD. For quantitative assessment on discharge data within the recorded range, Chi-square (χ^2) and Kolmogorov-Smirnov (KS) tests are applied. A diagnostic test of MAPE (%) and D-index is used for the selection of best suitable PD for estimation of PFD. The study compares the five probability distributions used in EVA of discharge data and illustrates the applicability of GoF and diagnostic tests procedures in identifying which distribution is better suited for estimation of PFD.

METHODOLOGY

The procedures involved in EVA of discharge data at Badlapur are:

- Prepare the observed PFD data series from daily stream flow data;
- Determination of parameters of five PDs viz., EXP, EV1, EV2, GEV and NOR using MoM and LMO;
- Estimate the PFD using PDs (using MoM and LMO);
- Check the adequacy of fitting of PDs using GoF tests and diagnostic tests to identify the suitable PD to arrive at a design value;
- Analyze the results and suggestions made thereof.

Table 1 presents the Cumulative Distribution Function (CDF) and quantile estimator (q_T) of PDs considered in the study.

In Table 1, F(q) is the CDF of variable D (i.e., PFD), α is the location parameter, β is the scale parameter, γ is the shape parameter, erf is the error function and T is the return period. For EV1 and EV2 distributions, the reduced variate (Y_T) is defined by Y_T = -ln(-ln(1-(1/T))).

The parameters of the distributions are determined by MoM and LMO, and used to estimate the PFD by the quantile functions of the PDs, as given in Table 1. Theoretical descriptions of the determination of parameters of PDs by MoM and LMO are available in the text book titled 'Flood Frequency Analysis' by Rao and Hamed (2000).

Table1. *CDF and quantile estimator of PDs*

Distribution	CDF	Quantile estimator (qT)
EXP (Scale, Shape)	$F(q;\beta,\gamma)=(1-\exp(-\beta q))^{\gamma}$, q>0, β>0	$q_T=\frac{1}{\beta}\left[-\ln\left(1-\left(1-\left(\frac{1}{T}\right)\right)\right)^{(1/\gamma)}\right]$
EV1 (Location, Scale)	$F(q;\alpha,\beta)=\exp\left(-\exp\left(-\frac{(q-\alpha)}{\beta}\right)\right)$, q>0, β>0	$q_T=\alpha+Y_T\beta$
EV2 (Scale, Shape)	$F(q;\beta,\gamma)=\exp\left(-\left(\frac{q}{\beta}\right)^{-\gamma}\right)$, q>0, β>0	$q_T=\beta\exp\left(\frac{Y_T}{\gamma}\right)$
GEV (Location, Scale, Shape)	$F(q;\alpha,\beta,\gamma)=\exp\left\{-\left[1-\gamma\left(\frac{q-\alpha}{\beta}\right)\right]^{(1/\gamma)}\right\}$, q>0, β>0, γ≠0	$q_T=\alpha+\left(\frac{\beta}{\gamma}\right)\left[1-\left(-\ln\left(1-\left(\frac{1}{T}\right)\right)\right)^{\gamma}\right]$
NOR (Location, Scale)	$F(q;\alpha,\beta)=\frac{1}{2}\left[1+\text{erf}\left(\frac{q-\alpha}{\beta\sqrt{2}}\right)\right]$, q>0, β>0	$q_T=\alpha+\beta\sqrt{2}\,\text{erf}^{-1}(2F-1)$

In Table 1, F(q) is the CDF of variable D (i.e., PFD), α is the location parameter, β is the scale parameter, γ is the shape parameter, erf is the error function and T is the return period. For EV1 and EV2 distributions, the reduced variate (Y_T) is defined by Y_T = -ln(-ln(1-(1/T))). The parameters of the distributions are determined by MoM and LMO, and used to estimate the PFD by the quantile functions of the PDs, as given in Table 1. Theoretical descriptions of the determination of parameters of PDs by MoM and LMO are available in the text book titled 'Flood Frequency Analysis' by Rao and Hamed (2000).

Goodness-of-Fit Tests

GoF tests are essential for checking the adequacy of probability distributions to the PFD data series in the estimation of PFD. Out of a number GoF tests available, the widely accepted GoF tests are χ^2 and KS, which are used in the study. The theoretical descriptions of GoF tests statistic are given as below:

$$\chi^2=\sum_{j=1}^{NC}\frac{(O_j(q)-E_j(q))^2}{E_j(q)} \quad (1)$$

where, $O_j(q)$ is the observed frequency value of j^{th}class, $E_j(q)$ is the expected frequency value of j^{th}class and NC is the number of frequency classes [15].The rejection region of χ^2 statistic at the desired significance level (η) is given by $\chi_C^2\geq\chi_{1-\eta,NC-m-1}^2$.

Here, m denotes the number of parameters of the distribution and χ_C^2 is the computed value of statistic for the PDs.

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

where, $F_e(q_i)$=i/(N+1) is the empirical CDF of q_i, $F_D(q_i)$ is the computed CDF of q_i by PDs, q_i is the observed PFD for i^{th} observation and N is the number of observations [16]. If the computed values of GoF tests statistic given by the distribution are less than that of the theoretical values at the desired level of significance then the distribution is considered to be acceptable for estimation of PFD at that level.

Diagnostic Tests

Sometimes the GoF test results would not offer a conclusive inference thus posing a problem for the user in selecting a suitable PD or method for their application. In such cases, a diagnostic test in adoption to GoF is applied for making inference. The selection of a suitable PD for estimation of PFD is performed through MAPE and D-index test which is defined as below:

$$\text{MAPE}(\%)=(1/N)\sum_{i=1}^{N}\left|\frac{q_i-q_i^*}{q_i}\right|*100 \quad (3)$$

Where, q_i is the observed PFD for i^{th} observation, q_i^* is the estimated PFD for i^{th} observation by PDs, $\bar{q}$ is the average value of observed PFD and $\overline{q^*}$ is the average value of estimated PFD [17].

$$\text{D-index} = \left(1/\bar{q}\right)\sum_{i=1}^{6}\left|q_i - q_i^*\right| \quad (4)$$

Here, $\bar{q}$ is the average value of the observed PFD whereas q_i (i= 1 to 6) and q_i^* are the six highest observed and corresponding estimated values by different PDs [18]. The distribution with minimum D-index value and least MAPE is identified as better suited distribution in comparison with the other distributions for estimation of PFD.

Application

In this paper, a comparative study on estimation of PFD at Badlapur GD site using MoM and LMO estimators of five PDs viz., EXP, EV1, EV2, GEV and NOR was carried out. The latitude and longitude of Badlapur GD site is 19° 09′ 44″ N and 73° 15′ 16″ E respectively. The catchment area of the site is 785 km^2. Badlapaur GD site is located on Ulhas River, which is one of the west-flowing rivers in Maharashtra that drains into the Arabian Sea. The Ulhasriver rises from Sahyadri hill ranges in the Raigad district of Maharashtra at an elevation of 600 meter above mean sea level. The Ulhas river basin lies between the latitudes of 18° 44′ N to 19° 42′ N and longitudes of 72° 45' E to 73° 48′ E. Figure 1 gives the location map of the study area. The drainage area of Ulhas basin is 4,637 km^2 which lies completely in Maharashtra. The total length of the west flowing river from its origin to its outfall into the Arabian Sea is 122 km. The average rainfall in the Ulhas basin is 2,943 mm. The basin receives 99% of rainfall from the south-west monsoon during June to October. The average maximum and minimum temperatures are recorded as 38.9 °C and 12.4 °C respectively. The observed PFD data series for the period 1986 to 2016 was extracted from the daily discharge data and used for EVA. Table 2 gives the descriptive statistics of the observed PFD data.

Table2. *Descriptive statistics of the PFD data*

Site	Average (m^3/s)	Standard Deviation (m^3/s)	Coefficient of Skewness	Coefficient of Kurtosis	Maximum (m^3/s)	Minimum (m^3/s)
Badlapur	2409.0	888.7	0.609	- 0.388	4483.0	1103.0

RESULTS AND DISCUSSIONS

Based on the parameter estimation procedures of EXP, EV1, EV2, GEV and NOR PDs, parameters were determined by using MoM and LMO with the aid of statistical software viz., Hydrognomon and used for EVA of discharge data. The estimated PFD at Badlapur given by five PDs (using MoM and LMO) are presented in Table 3. The CDF plots of observed and estimated PFD for Badlapur is presented in Figure 2. The plots of estimated PFD obtained from five PDs (using MoM and LMO) together with observed data for different return periods for Badlapur are presented in Figure 3.

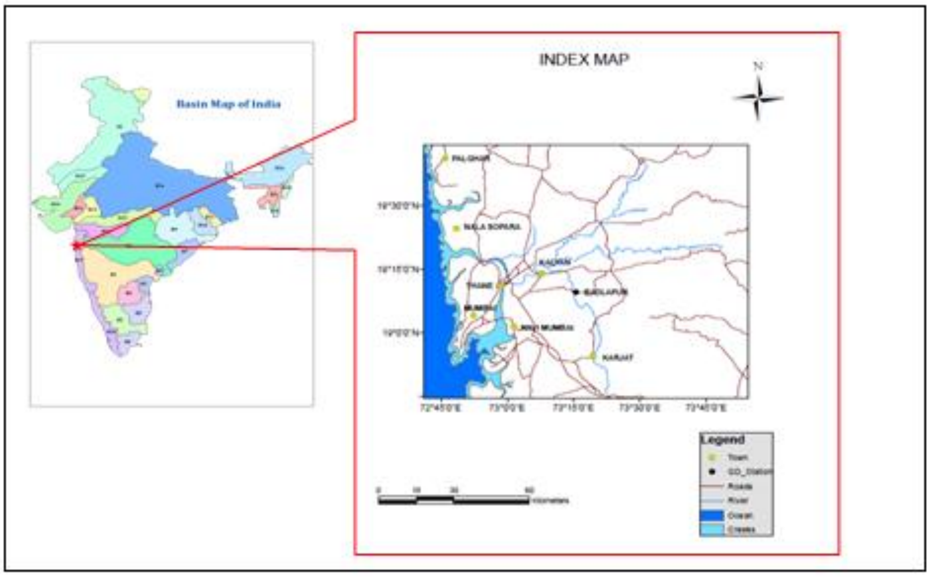

Figure1. *Location map of the study area*

Table3. *Estimated PFD by five PDs (using MoM and LMO) for different return periods for Badlapur*

Return Period (year)	Estimated PFD (m^3/s)									
	EXP		EV1		EV2		GEV		NOR	
	MoM	LMO	MoM	LMO	MoM	LMO	MoM	LMO	MoM	LMO
2	2136.3	2096.6	2262.9	2254.3	2190.5	2111.8	2313.9	2266.9	2409.0	2409.0
5	2950.6	3029.3	3048.6	3086.5	2829.9	2889.2	3116.3	3098.5	3156.9	3168.1
10	3566.5	3734.8	3568.7	3637.5	3352.8	3555.4	3596.6	3637.0	3547.8	3565.0
20	4182.5	4440.3	4067.7	4166.0	3945.0	4338.4	4022.5	4144.7	3870.7	3892.7
50	4996.8	5373.0	4713.6	4850.1	4869.5	5613.4	4527.6	4789.0	4234.1	4261.6
100	5612.8	6078.5	5197.5	5362.7	5701.7	6808.9	4874.5	5262.7	4476.3	4507.5
200	6228.8	6784.1	5679.8	5873.5	6672.3	8253.1	5195.3	5726.9	4698.0	4732.5
500	7043.1	7716.7	6316.0	6547.4	8210.1	10637.4	5583.9	6327.8	4966.7	5005.2
1000	7659.1	8422.2	6796.8	7056.7	9603.3	12886.5	5853.3	6773.3	5155.2	5196.5

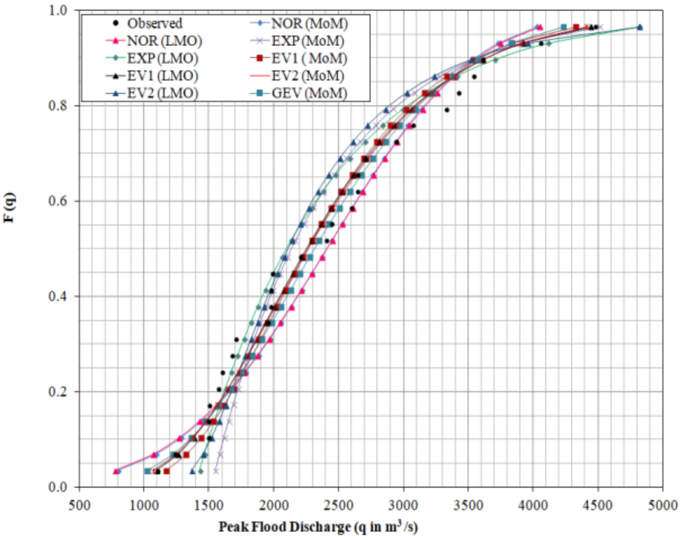

Figure2. *CDF plots of observed and estimated PFD by five PDs (using MoM and LMO) for Badlapur*

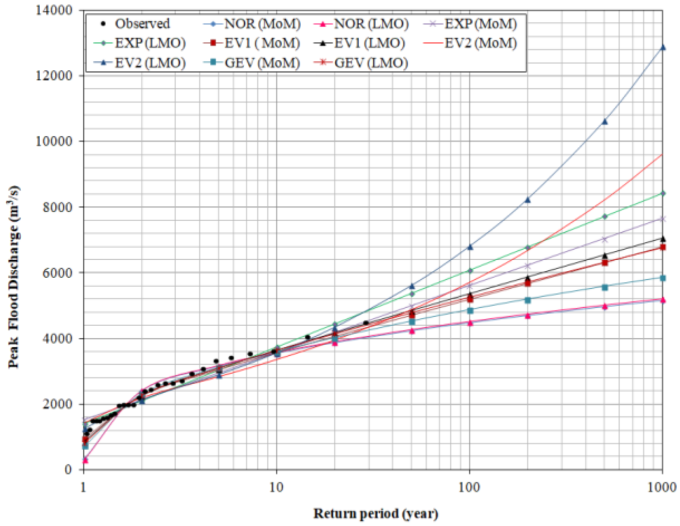

Figure3. *Plots of observed and estimated PFD by five PDs (using MoM and LMO) for Badlapur*

From Table 3, it is noted that the estimated PFD obtained from EV2 (using LMO) is comparatively higher than the corresponding values of other PDs (or methods) for return periods from 50-year to 1000-year for Badlapur. From Figure 3, it can be seen that the fitted lines of the estimated PFD by EV2 (using LMO) is in the form of exponential growth curve.

Analysis Based on GoF Test

By using the estimators of PDs (using MoM and LMO), GoF tests statistic values were computed from Eqs. (1) and (2), and the GoF tests results are presented in Table 4. From χ^2 test results, it is noted that the computed values of χ^2 test statistic by EXP, EV1, EV2, GEV and NOR (using MoM and LMO) are less than its theoretical values (3.84 for GEV, 5.99 for NOR, EXP, EV1 and EV2) at 5% significance level, and at this level, all these five distributions are found as acceptable for EVA of discharge data for Badlapur. From Table 4, it is also noted that the computed values of KS test statistic by EXP, EV1, EV2, GEV and NOR distributions (using MoM and LMO) are less than its theoretical value of 0.238 at 5% significance level, and at this level, all five PDs are found as acceptable for EVA of discharge data for Badlapur.

Table4. *Computed values of GoF tests statistics by five PDs (using MoM and LMO)*

GoF Test	EXP		EV1		EV2		GEV		NOR	
	MoM	LMO	MoM	LMO	MoM	LMO	MoM	LMO	MoM	LMO
χ^2	2.000	0.857	1.429	1.429	2.571	2.571	2.413	2.829	0.857	0.857
KS	0.152	0.113	0.097	0.078	0.186	0.134	0.101	0.082	0.130	0.128

Analysis Based on Diagnostic Test

The selection of an appropriate PD with parameter estimation method for EVA of discharge data was carried out by MAPE (%) and D-Index though GoF tests results confirmed the applicability of EXP, EV1, EV2 and GEV and NOR distributions for EVA of discharge data for Badlapur. The MAPE (%) and D-Index values were computed by using Eqs. (3) and (4), and presented in Table 5.

Table5. *MAPE (%) and D-Index values given by five PDs for Badlapur*

Diagnostic test	EXP		EV1		EV2		GEV		NOR	
	MoM	LMO	MoM	LMO	MoM	LMO	MoM	LMO	MoM	LMO
MAPE(%)	9.1	6.9	5.3	4.0	11.8	8.2	4.8	4.1	7.5	7.5
D-INDEX	0.538	0.505	0.504	0.338	0.959	0.718	0.460	0.345	0.583	0.542

From Table 5, it is noted that the MAPE and D-index values obtained from EV1 (using LMO) distribution are minimum when compared with the corresponding values given by EXP, EV1, EV2 and GEV (using MoM and LMO). By considering the diagnostic tests results and quantitative assessment through GoF tests, it is identified that EV1 (using LMO) is better suited distribution for EVA of discharge data for Badlapur. The study suggested that the estimated PFDs for different return periods obtained from EV1 (using LMO) distribution could be used as a design value for designing civil and hydraulic structures.

Conclusions

The paper presents the study carried out for EVA of discharge data for Badlapur by adopting five PDs such as EXP, EV1, EV2, GEV and NOR. The parameters of the distributions were determined by MoM and LMO, and used for estimation of PFD. The intercomparison of the results was carried out and the following conclusions were drawn from the study:

- The χ^2 and KS test results confirmed the applicability of EXP, EV1, EV2, GEV and NOR distributions (using MoM and LMO) for EVA of discharge data.
- Qualitative assessment through plot indicated that the fitted line of the estimated PFD by EV2 (using LMO) for Badlapur is in the form of exponential curve.
- For the return period of 50-year and above, it was found that the estimated PFD by EV2 (using LMO) is comparatively higher than the corresponding values of other PDs (or methods).
- On the basis of quantitative assessment using GoF and diagnostic tests, and qualitative assessment using fitted curves, the study suggested that the 1000-year return period PFD of about 7057 m^3/s could be used as a design value for designing civil and hydraulic structures having a design life of 1000-year.
- The study also suggested that by considering the design life of the proposed structure, the stakeholders may adopt the appropriate estimated PFD values for design purposes.

Acknowledgements

The authors are grateful to Dr. B. B. Ahuja, Director, College of Engineering, Pune for his continuous encouragement during the course of study. The authors are thankful to Dr. (Mrs.) V. V. Bhosekar, Director, Central Water and Power Research Station, Pune for providing research facilities to carry out the study. The authors are also thankful to Central Water Commission for the supply of stream flow data used in the study.

REFERENCES

[1] Rao AR, Hamed KH. Flood frequency analysis. CRC Publications, New York; 2000.

[2] CEH. FLOODS version 1.1, Regional flood frequency analysis software manual. Water Resources Section, Centre for Ecology and Hydrology (CEH), Wallingford, U.K.; 2001.

[3] Hosking JRM. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Royal Statistical Society (Series B), 1990; 52(1): 105-124.

[4] Topaloglu F. Determining suitable probability distribution models for flow and precipitation series of the Seyhan River basin. Turkish Journal of Agriculture and Forestry, 2002; 26(4): 187-194.

[5] Guevara E. Engineering design parameters of storms in Venezuela. Hydrology Days, 2003; pp. 80-91.

[6] Saf B, Dikbas F, Yasar M. Determination of regional frequency distributions of floods in West Mediterranean River Basins in Turkey. Fresenius Environment Bulletin, 2007; 16(10): 1300–1308.

[7] Mujere N. Flood frequency analysis using the Gumbel distribution. Journal of Computer Science and Engineering, 2011; 3(7):2774-2778.

[8] Baratti E, Montanari A, Castellarin A, Salinas JL, Viglione A, Bezzi A. Estimating the flood frequency distribution at seasonal and annual time scales. Hydrological Earth System Science, 2012; 16(12): 4651–4660.

[9] Olumide BA, Saidu M, Oluwasesan A. Evaluation of best fit probability distribution models for the prediction of rainfall and runoff volume (Case Study Tagwai Dam, Minna-Nigeria). Journal of Engineering and Technology, 2013; 3(2): 94-98.

[10] Haberlandt U, Radtke I. Hydrological model calibration for derived flood frequency analysis using stochastic rainfall and probability distributions of peak flows. Hydrology and Earth System Sciences, 2014; 18(1): 353-365.

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

[12] Suhartanto E, Lily ML, Dina N, Febri IH, Dwi AK. Estimation of Design Flood with Four Frequency Analysis Distributions. Asian Journal of Applied Science and Technology, 2018; 2(1): 13-27.

[13] Kolbjørn E, Donna W, Péter B, Lars R, Erik H. Use of historical data in flood frequency analysis: a case study for four catchments in Norway. Hydrology Research, 2018; 49(2): 466-486.

[14] Winter B, Schneeberger K, Dung NV, Huttenlau MS, Achleitner J, Stötter BM, Vorogushyn S. A continuous modelling approach for design flood estimation on sub-daily time scale. Hydrological Sciences Journal, 2019; 64(5): 539-554.

[15] Charles Annis, P.E. Goodness-of-Fit tests for statistical distributions; 2009.

[16] Zhang J. Powerful goodness-of-fit tests based on the likelihood ratio. Journal of Royal Statistical Society, 2011; 64(2): 281-294.

[17] Singh RD, Mishra SK, Chowdhary H. Regional flow duration models for 1200 ungauged Himalayan watersheds for planning micro-hydro projects. ASCE Journal of Hydrologic Engineering, 2001; 6(4): 310-316.

[18] USWRC. Guidelines for deter-mining flood flow frequency. Bulletin No. 17B, United States Water Resources Council (USWRC); 1981.

Citation: *R. S. Bharadwaj, (Mrs.) A. D. Thube, N. Vivekanandan, C. Srishailam, "A Comparative Study on Estimation of Peak Flood Discharge using MoM and LMO Estimators of Probability Distributions", International Journal of Emerging Engineering Research and Technology, 7(4), 2019, pp.4-10*