551.579.4 : 556.5.072

INTERCOMPARISON OF DETERMINATION OF PARAMETERS OF LP III FOR ESTIMATION OF DESIGN FLOOD

1. Estimation of design flood of desired return period is one of the pre-requisite for rational and economic design of hydraulic structures. Moreover, hydraulic design of river structures such as dams, bridges, barrages, river training networks, etc. is primarily based on the design flood corresponding to different return periods depending upon the operational requirements and importance of the structures. Frequency analysis procedures involving determination of parameters of the distribution is one of the techniques for estimation of flood from the recorded data.

Number of methods based on standard probability distributions like Normal, Log-normal, Pearson Type III, Log Pearson Type III (LP III), Extreme Value Type I, etc., are commonly used for estimation of design flood of different return periods. The estimation procedures are carried out on the basis of the assumption that the annual daily maximum discharge at a certain location follows one of the standard probability distribution functions. United States Water Resources Council (USWRC, 1976) recommended LP III distribution for estimation of flood which was supported by the Institution of Engineers of Australia (1977). Bobee (1975) suggested that LP III could be used for estimating the flood using the recorded annual maximum discharge data while Bobee and Ashkar (1988) adopted Pearson Type III distribution. National Environmental Research Council (NERC, 1975) recommended the use of Extreme Value Type I for flood frequency analysis. Though no exhaustive and conclusive study has been carried out to adopt the particular distribution for flood frequency analysis, LP III is widely accepted as an appropriate distribution for analysing the annual maximum discharge data. Therefore, in the present study, efforts have been made to estimate the design flood for different return periods using six different parameter estimation methods of LP III for rivers Narmada at Mortakka, Vamsadhara at Srikakulam and Yamuna at Poiya Ghat sites.

Parameter estimation methods such as direct method of moments (MOM), indirect method of moments based on coefficient of skewness [IMM (CS1) and IMM (CS2)], sundry averages method (SAM), method of mixed moments (MMM) and maximum likelihood method (MLM) were used for determination of parameters of LP III (Hazen, 1924 and Rao, 1983). Number of research studies has been carried out by different researchers on determination of parameters of LP III using different methods (Hoshi and Burges, 1981; Griffis *et al.*, 2004 and Phien and Hira, 1983). But, there is no general agreement in adopting particular method for determination of parameters of LP III. The objective of the study is to compare the design flood estimates obtained from different parameter estimation methods of LP III and to identify the appropriate method for determination of parameters of LP III for estimation of flood for the river basins considered in the study. Anderson-Darling (A^2) test was used for checking the goodness-of-fit of the LP III distribution to the recorded data. Diagnostic analysis using relative absolute error (D_a) and relative mean square error (D_r) was carried out to evaluate the efficacy of an appropriate method employed for determination of parameters of LP III for estimation of design flood.

2.1. *Design flood estimation using LP III* - As described by Benson (1968) and Bobee (1973), the probability distribution function of LP III is given by:

$$f(x;\alpha,\lambda,m)=\frac{|\alpha|}{\Gamma(\lambda)}\left(\frac{e^{\alpha m}}{x^{1+\alpha}}\right)\left[\alpha(\ln x-m)\right]^{\lambda-1},\ x>0,\ \lambda>0,\ -\infty<m<+\infty \quad (1)$$

Where α, λ and m are scale, shape and position parameters of LP III respectively. The parameters of LP III by six different methods are determined and further used to compute the design flood for different return periods using

$$X_T=10^{m+(\lambda+K_T\sqrt{\lambda})/\alpha} \quad (2)$$

Where, X_T is the estimated design flood corresponding to T-year return period and K_T is the frequency factor (or standardised LP III variable) corresponding to positive value of coefficient of skewness (C_S) of logarithmically transformed data and probability of exceedance ($1/T$).

2.2. *Goodness of fit test* - Number of goodness-of-fit (GoF) tests like Chi-square, Kolmogorov-Smirnov and Anderson Darling (A^2) are commonly used to judge the suitability of the distribution to the recorded data. Horn (1977) argued that many hydrologists discourage the use of the Chi-square and Kolmogorov-Smirnov tests when testing hydrologic frequency distributions in flood studies, due to the importance of the tails of the distribution and the insensitivity of these statistical tests in the tail of the distribution. So, A^2 test was used for checking the GoF of LP III to the data under study.

The Empirical Distribution Function (EDF), $F_N(x)$ of the sample is defined by : $F_N(x)$ = (No. of observations $\leq x)/N$, $-\infty<x<+\infty$. $F_N(x)$ can be seen to be a step function calculated from the data. As 'x' increases, it takes a step of height ($1/N$). $F_N(x)$ records the proportion of the observations less than or equal to 'x'. D'Agostino *et al*. (1986) has given the relation for estimation of A^2 as:

$$A^2=(-N)-(1/N)\sum_{i=1}^{N}\Big\{(2i-1)\log\big(Z_{(i)}\big)+(2N+1-2i)\log\big(1-Z_{(i)}\big)\Big\} \quad (3)$$

For a given sample of 'N' values, $Z_{(i)}=F(x_i)$, for i = 1, 2, 3,….N; and $x_1<x_2<\ldots x_N$.

The distribution of A^2 statistic doesn't depend on $F(x)$, but on the set of 'N' sample values. The distribution theory of ordered statistic gives the percentage points for testing of A^2 statistic. If the percentage points are used with small samples, the corresponding EDF statistic may be multiplied by $\left[1+\left(0.2/\sqrt{N}\right)\right]$. The upper tail percentage values for A^2 statistic are given below:

Statistic	*Level of significance*			
	0.10	0.05	0.025	0.01
A^2	0.637	0.757	0.877	1.038

TABLE 1

Details of data availability and summary statistics of annual maximum discharge

River	Site	Data availability	Statistical parameters			
			$\bar{x}$ (m^3/s)	σ_x (m^3/s)	C_S	C_K
Narmada	Mortakka	1949-1994 (46 years)	20,668.50 (*4.28710*)	7,972.72 (*0.15700*)	1.510 (*0.12640*)	3.743 (*0.38535*)
Vamsadhara	Srikakulam	1971-2001 (31 years)	1,267.11 (*2.97827*)	1,199.51 (*0.32664*)	3.199 (*0.16863*)	12.966 (*0.30940*)
Yamuna	Poiya Ghat	1971-2002 (32 years)	2,288.92 (*3.28746*)	1,595.44 (*0.24440*)	2.858 (*0.44265*)	11.577 (*0.43069*)

(Here, $\bar{x}$ is mean, σ_x is standard deviation, C_S is the coefficient of skewness and C_K is the coefficient of kurtosis. The summary statistics of the log-transformed series of the recorded data are given in *italic*)

TABLE 2

A^2 statistic values of different parameter estimation methods for rivers Narmada, Vamsadhara and Yamuna

River	Site	A^2 statistic values of					
		MOM	IMM CS1	IMM CS2	SAM	MMM	MLM
Narmada	Mortakka	0.652	0.712	0.745	0.659	0.723	0.600
Vamsadhara	Srikakulam	0.721	0.705	0.746	0.729	0.738	0.703
Yamuna	Poiya Ghat	0.405	0.443	0.395	0.378	0.435	0.343

TABLE 3

Index values of D_a for different parameter estimation methods for rivers Narmada, Vamsadhara and Yamuna

River	Site	Index values of D_a for					
		MOM	IMM (CS1)	IMM (CS2)	SAM	MMM	MLM
Narmada	Mortakka	0.072	0.064	0.063	0.068	0.066	0.060
Vamsadhara	Srikakulam	0.148	0.134	0.128	0.133	0.136	0.125
Yamuna	Poiya Ghat	0.204	0.198	0.177	0.187	0.209	0.166

TABLE 4

Index values of D_r for different parameter estimation methods for rivers Narmada, Vamsadhara and Yamuna

River	Site	Index values of D_r for					
		MOM	IMM (CS1)	IMM (CS2)	SAM	MMM	MLM
Narmada	Mortakka	0.099	0.093	0.090	0.094	0.096	0.087
Vamsadhara	Srikakulam	0.171	0.158	0.154	0.157	0.159	0.150
Yamuna	Poiya Ghat	0.256	0.251	0.216	0.230	0.267	0.204

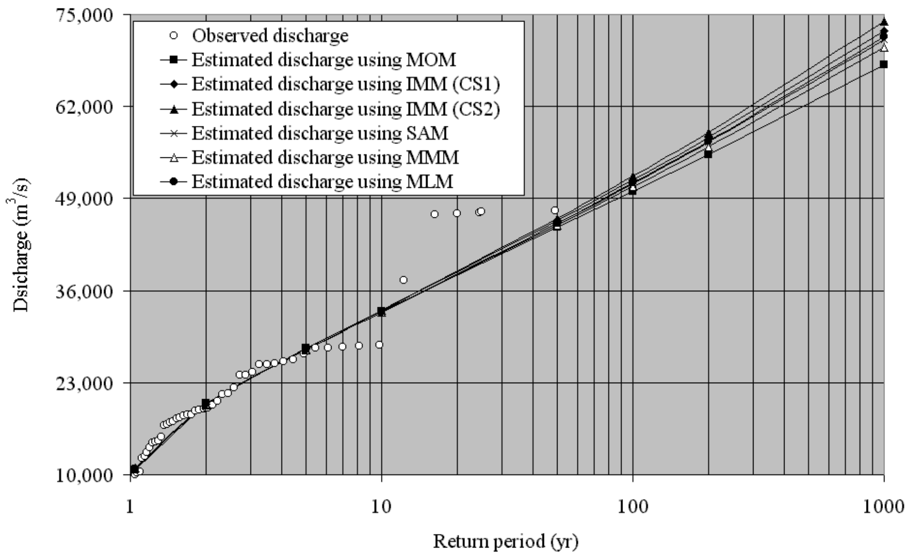

Fig. 1. Probability plot of observed and estimated discharge using LP III from six different parameter estimation methods for river Narmada at Mortakka site

If the EDF statistics computed is less than the theoretical value at the specified significance level, then the null hypothesis that the selected distribution of '*x*' is accepted.

In the present study, A^2 statistic is used to quantitatively assess, within specified levels of significance, whether the annual maximum discharge values relating to the region of interest can be deemed to come from the sample defined by LP III. A qualitative assessment of GoF was conducted by probability plotting of the estimated discharge.

2.3. *Diagnostic analysis* - Singh (1987) expressed that the diagnostic analysis using relative absolute error (D_a) and mean square error (D_r) could be employed for evaluating the applicability of an appropriate method for determination of parameters of LP III for estimation of flood for different return periods and are given by:

$$D_a = \frac{1}{N}\sum_{i=1}^{N}\left|\left(\frac{X_o(T_i) - X_e(T_i)}{X_o(T_i)}\right)\right| \tag{4}$$

$$D_r = \left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_o(T_i) - X_e(T_i)}{X_o(T_i)}\right)^2\right)^{1/2} \tag{5}$$

Where X_o (*T*) and X_e (*T*) are the observed and estimated discharge respectively for a given return period (*T*), *N* is the number of observations. The least index values of D_a and D_r are considered to be the deciding factor for selecting of an appropriate method for determination of parameters of LP III for modelling of flood data.

2.4. *Data used* - Annual maximum discharge data in respect of river Narmada at Mortakka, Vamsadhara at Srikakulam and Yamuna at Poiya Ghat sites are used. Table 1 gives the details of data availability and summary statistics of the annual maximum discharge for the river basins considered in the study.

3. A computer program was developed and used to determine the parameters of LP III by six different methods for the data under study. The parameters were further used to compute the design flood estimates for different return periods of 2, 5, 10, 25, 50, 100, 200, 500, and 1,000-year (yr) using Eqn. (2) and are given in Figs. 1-3. It is observed that the IMM (CS2) gave higher estimates for higher order return periods such as 100, 200, 500 and 1000-yr and MOM gave lower estimates for different return periods from 2 to 1,000-yr as compared to the other parameter estimation methods of LP III.

Figs. (1-3) give the probability plots of observed and estimated discharge for different return periods computed by six different parameter estimation methods of LP III for rivers Narmada at Mortakka, Vamsadhara at Srikakulam and Yamuna at Poiya Ghat sites respectively.

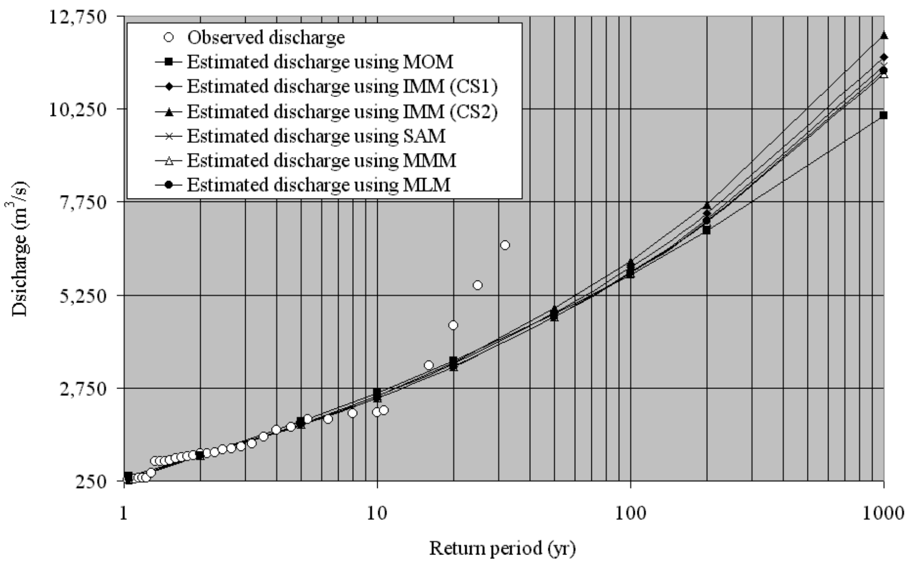

Fig. 2. Probability plot of observed and estimated discharge using LP III from six different parameter estimation methods for river Vamsadhara at Srikakulam site

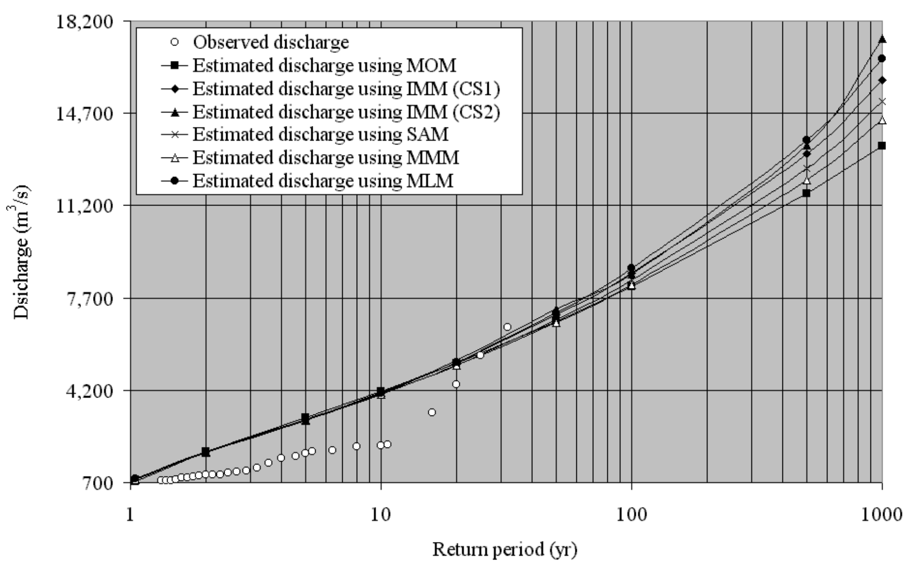

Fig. 3. Probability plot of observed and estimated discharge using LP III from six different parameter estimation methods for river Yamnua at Poiya Ghat site

By using the Eqn. (3), A^2 statistic values for different parameter estimation methods of LP III were computed and are given in Table 2.

From Table 2, it may be noted that all of these A^2 statistic values are less than the theoretical value of 0.757 at the 5 per cent level of significance. As such, at this level, the sample would not be rejected as coming from LP III distribution. The results of A^2 statistic also confirm that all six methods are acceptable to determine the parameters of LP III for estimation of flood for the data under study.

Diagnostic statistics using D_a and D_r was carried out for selection of an appropriate method for determination of parameters of LP III using six different methods for estimation of design flood. By using the Eqns. 4 and 5, the index values of D_a and D_r for different parameter estimation methods of LP III were computed and are given in Tables 3 and 4 respectively.

From the Tables 3 and 4, it may be noted that the indices of D_a and D_r for MLM are to be minimum when compared to the corresponding indices relating to parameter estimation methods of MOM, IMM based on CS1 and CS2, SAM and MMM for all three river basins. From the results of the analysis, it may be further noted that the indices of D_a and D_r for IMM (CS2) are the second minimum only when compared to the corresponding indices of MLM even though the method gave higher estimates for the river basins considered in the study. On the basis of the results of diagnostic analysis, and also from the quantitative and qualitative assessment, MLM is considered as the best method and suggested for determination of parameters of LP III for estimation of flood for different return periods for the data under study.

4. The study gives the results of design flood estimates for different return periods obtained by using MOM, IMM based on CS1 and CS2, SAM, MMM and MLM of LP III. The study shows the A^2 test is used for checking the goodness of fit of LP III distribution to the recorded data and Diagnostic analysis using D_a and D_r are carried out for selection of an appropriate method for determination of parameters of LP III for estimation of flood for different return periods. The study presents the MLM appears to be the most suitable method among six methods for determination of parameters of LP III for estimation of flood for the rivers Narmada, Vamsadhara and Yamuna.

References

Benson, M. A., 1968, "Uniform flood frequency estimating methods for federal agencies", *Water Resources Research*, **4**, 891-908.

Bobee, B., 1973, "Sample of T-year events computed of bias in estimation of coefficient of skewness", *Water Resources Research*, **9**, 1264-1270.

Bobee, B., 1975, "The Log Pearson Type 3 distribution and its application in Hydrology", *Water Resources Research*, **11**, 681-689.

Bobee, B. and Ashkar, Fahim, 1988, "The generalised method of moments applied to the Log Pearson type 3 distribution", *J. Hydraulic Engineering*, **114**, 899-909.

D'Agostino, B. Ralph and Stephens, A. Michael, 1986, "Goodness of fit test statistic", M/s Marcel Dekkar Inc., 270 Madison Avenue, New York 10016, USA.

Griffis, V. W., Stedinger, J. R. and Cohn, T. A., 2004, "LP3 Quantile estimators with regional skew information and low outlier adjustments", *Water Resources Resource*, **40**, 54-66

Hazen, A., 1924, "Discussion on theoretical frequency curves", by *J. Transactions*, **87**, 174-179.

Horn, S. D., 1977, "Goodness of Fit Tests for Discrete Data: A Review and an application to a health impairment scale", *Biometrics,* **33**, 237-248.

Hoshi, K. and Burges, S. J., 1981, "Sampling properties of parameter estimates for the Log Pearson type 3 distribution using moments in real space", *J. Hydrology*, **53**, 305-316.

Institution of Engineers, Australia, 1977, "Australian rainfall and runoff, flood analysis and design", ACI, Canbena, 150-153.

National Environmental Research Council (NERC), 1975, "Flood studies report", 1-5

Phien, H. N. and Hira, M. A., 1983, "Log-Pearson Type 3 Distribution: Parameter Estimation", *J. Hydrology*, **64**, 25-37.

Rao, D. V., 1983, "Estimating Log Pearson Parameters by Mixed Moments", *J. Hydraulics Division*, **109**, 1119-1132.

Singh, V. P., 1987, "On statistical inter-comparison of EV I estimators by Monte Carlo simulation", *Advanced Water Resources Research*, **10**, 87-105.

United States Water Resources Council (USWRC), 1976, "Guidelines for determining flood flow frequency", *Bulletin*, **17**, Washington. D.C.

N. VIVEKANANDAN

Central Water and Power Research Station, Pune, India
(*Received 9 April 2007, Modified 4 December 2007*)
e mail: anandaan@rediffmail.com