

Regional flood frequency analysis of Tripura based on L-moment

Abhijit Bhuyan and Munindra Borah*

Department of Mathematical Sciences, Tezpur University Naapam, Tezpur-784028, INDIA

*Corresponding author: E-mail: mborah@tezu.ernet.in

Abstract: The annual maximum discharge data of six gauging sites have been considered for L-moment based regional flood frequency analysis of Tripura, India. Homogeneity of the region has been tested based on heterogeneity measure (H) using method of L-moment. Based on heterogeneity measure it has been observed that the region consist of six gauging sites is homogeneous. Different probability distributions viz. Generalized extreme value (GEV), Generalized Logistic (GLO), Generalized Pareto (GPA), Generalized Normal (GNO), Pearson Type III (PE3) and Wakebay (WAK) have been considered for this investigation. PE3, GNO and GEV have been identified as the candidate distributions based on the L-moment ratio diagram and $|Z^{DIST}|$ -statistics criteria. Regional growth curves for three candidate distributions have been developed for gauged and ungauged catchments. Monte Carlo simulations technique has also been used to estimate accuracy of the estimated regional growth curves and quantiles. From simulation study it has been observed that PE3 distribution is the robust one.

Keywords: Frequency analysis, L-moments, Monte carlo simulation, PE3 distribution, Quantile estimates

INTRODUCTION

Flood frequency estimation is an important task for proper planning of water resources project. In design of hydraulic structures such as dams, spillways etc. flood frequency estimation for large return periods is essential at the site of interest for proper planning. The reliable estimation of flood frequency is depends mostly on the data availability and the method used. At a single site, frequency analysis method may suffer from sampling variability due to data availability, especially for estimating return period that exceed the length of observed data. The regional frequency analysis based on L-moment parameter estimation method proposed by Hosking (1990) overcomes difficulties faced for at site estimation due to data availability. In regional frequency analysis, index flood procedure first introduced in hydrology literature by Dalrymple (1960) is used with L-moments by Hosking and Wallis (1997).

Recently, regional flood frequency method has been used world wide for estimation of design flood in various countries. In India, L-moments based regional flood frequency analysis was carried out by Paradia *et al.* (1998) and Kumar *et al.* (1999, 2003 and 2005) to developed flood frequency relationship for both gauged and ungauged catchments for different regions. In addition to this, some other recent application of regional flood frequency analysis includes Atiem and Harmancioglu (2006), Saf (2008) and Hussain and Pasha (2008). The present study has been undertaken up to estimates quantiles, based

on L-moment approach for gauged and ungauged catchments of Tipura, India and to measure the accuracy of the estimated quantiles, Monte Carlo simulation procedure is used. For this various distributions viz. Generalized Extreme value (GEV), Generalized Normal (GNO), Generalized Pareto (GLP), Generalized Logistic (GLO) and Pearson Type III (PE3) have been used.

L-moments: Hosking (1990) introduce L-moments of a random variable as a linear combination of Probability Weighted Moments (PWMs). The PWMs, defined by Greenwood *et al.* (1979) as:

$$\beta_r = E[x\{F(x)\}^r]$$

which can be rewritten as:

$$\beta_r = \int_0^1 x(F) F^r dF$$

Where, $F = F(x)$ is the cumulative distribution function (CDF) for x , $x(F)$ is the inverse CDF of x evaluated at the probability F , and $r = 0, 1, 2, \dots$, is a nonnegative integer. When $r = 0$, β_0 is equal to the mean of the distribution $\mu = E[x]$.

Hosking (1990) defined r th L-moments related to the r th PWM as:

$$\lambda_{r+1} = \sum_{k=0}^r \beta_k (-1)^{r-k} \binom{r}{k} \binom{r+k}{k}$$

For any distribution, the first four L-moments can be

calculated from:

$$\lambda_1 = \beta_0$$

$$\lambda_2 = 2\beta_1 - \beta_0$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$$

The L-moment ratio defined by Hosking (1990) as:

$$\text{L-coefficient of Variation, L-CV } (\tau_2) = \lambda_2 / \lambda_1$$

$$\text{L-coefficient of skewness, L-skew } (\tau_3) = \lambda_3 / \lambda_2$$

$$\text{L-coefficient of kurtosis, L-kurtosis } (\tau_4) = \lambda_4 / \lambda_2$$

STUDY AREA AND DATA AVAILABILITY

Tripura is one of the North Eastern states of India. The state extends between 22°56'N and 24°32'N and 90°09'E and 92°10'E. Annual maximum peak flood data of six gauging sites of Tripura and varying over 11 years to 28 years in record length have been considered for the study. The catchments area, sample statistics, sample size and discordancy measure of each gauging stations under considerations have been given in Table 1. The mean annual peak floods of these sites vary from 97.28m³/s to 762.2442 m³/s and the catchments areas of these sites vary from 414 km² to 2492 km².

REGIONAL FLOOD FREQUENCY ANALYSIS AND ASSESSMENT OF ACCURACY OF ESTIMATED QUANTILES

Discordancy measure for screening the data: Hosking and Wallis (1997) proposed a discordancy measure (D_i) based on L-moments, to recognize those sites that are grossly discordant with the group as a whole. Discordancy measure, D_i for site i defined by Hosking and Wallis (1997) as:

$$D_i = \frac{1}{3}(u_i - \bar{u})^T S^{-1}(u_i - \bar{u}),$$

where, $u_i = [t^{(i)}, t_3^{(i)}, t_4^{(i)}]^T$ vector containing the L-CV, L-skew and L-kurtosis for a site i , S = covariance matrix of u_i , \bar{u} =mean of vector u_i .

A site is regarded as discordant if D_i exceeds the critical value given in tabular form Hosking and Wallis (1997). From Table 1 it has been observed that D_i values of all the six gauging sites are less than the critical value of D_i for 6 sites i.e. 1.648. Therefore we consider all the six gauging sites for our further study.

Homogeneity test: Hosking and Wallis (1997) proposed a heterogeneity measure (H) for identification of homogeneous region and H is defined by

$$H = \frac{V - \mu_v}{\sigma_v},$$

where, V = weighted standard deviation of the at site sample L-CVs

μ_v, σ_v = the mean and standard deviation N_{sim} values of V , and N_{sim} = no. of simulation.

A region is acceptably homogeneous if heterogeneity measure $H < 1$, possibly heterogeneous if $1 \leq H < 2$ and definitely heterogeneous if $H \geq 2$.

The heterogeneity measure H was computed for the study area by carrying out 500 simulations using the four parameter Kappa distribution (KAP) and obtained the value of H as 0.79. Thus the region comprising six gauging sites is identified as homogeneous region.

Goodness-of-fit measure

Z-Statistics criteria: This measure, Z, decides how well the simulated L-skewness and L-kurtosis of a fitted distribution matches the regional average L-skewness and L-kurtosis values obtained from the observed data. The goodness-of-fit measure for a distribution proposed by Hosking and Wallis (1997) is calculated by using the formula defined as

$$Z^{DIST} = (\tau_4^{DIST} - t_4^R + B_4) / \sigma_4,$$

where,

τ_4^{DIST} = L-kurtosis of the fitted distribution,

t_4^R = average value t_4 obtained from the data of a given region,

B_4 = bias of t_4^R , and

σ_4 = standard deviation of the obtained from simulation. Declare the fit to be adequate if Z^{DIST} is sufficiently close to zero, a reasonable criterion being $|Z^{DIST}| \leq 1.64$. While a number of distributions may qualify the goodness-of-fit criteria, the most potential will be the one that has the minimum $|Z^{DIST}|$ value.

The $|Z^{DIST}|$ -statistic values for various distributions are given in Table 2. It has been observed that the $|Z^{DIST}|$ -statistic values are lower than 1.64 for the three distributions namely GEV, GNO and PE3. So, these three distributions are suitable candidates for the regional distribution and can be used to obtain future quantile estimates. Again, the $|Z^{DIST}|$ -statistic value is lowest i.e. 0.70 for PE3 distribution. Therefore, it can be concluded that PE3 distribution is the most suitable distribution for this region. The regional parameters for the candidate distributions, which have $|Z^{DIST}|$ -statistic value less than 1.64, are given in Table 3.

L-moment ratio diagram: The L-moment ratio diagram is a plot of L-kurtosis and L-skewness, which is a useful guideline for the selection of an appropriate probability distribution for a region. In L-moment ratio diagram, the

Table 1. Sample statistics and discordancy measure.

Site	Sample size	Catchments area(km)	Mean annual peak flood	L-CV	L-Skew	L-Kurtosis	Discordancy measure
Manu	12	2278	762.2442	0.2466	0.2221	0.3088	1.59
Dhalai	11	630	186.4400	0.1718	0.2587	0.1244	1.49
Khowai	19	1328	283.8453	0.2933	0.1874	0.0614	0.95
Burima	14	414	97.2886	0.2787	0.0916	0.0031	0.70
Gumti	24	2492	560.7533	0.2100	0.0438	0.0949	1.08
Muhuri	28	576	360.1432	0.2083	0.1302	0.1045	0.20

Table 2. $|Z^{DIST}|$ -statistic for various distributions.

S No.	Distribution	-statistic
1	General Logistic (GLO)	2.26
2	Generalized Extreme Value (GEV)	0.99
3	Generalized Normal (GNO)	0.97
4	Pearson Type III (PE3)	0.70
5	Generalized Pareto (GPA)	-1.67

distinct relationship between L-kurtosis and L-skewness proposed by Hosking and Wallis (1997) are used for theoretical curves of each of the probability distributions and are plotted together with average L-skewness and L-kurtosis of the region.

The regional average values of L-skewness i.e. $\tau_3 = 0.1394$ and L-kurtosis i.e. $\tau_4 = 0.1064$ are plotted along with theoretical distributions in Fig. 1. It has been observed from L-moment ratio diagram that regional average L-skewness and L-kurtosis point lies closest to the PE-3 distribution but also not far away from both GNO and GEV distributions. Therefore, the three distributions viz. PE3, GNO and GEV can be considered as suitable regional distributions for our study area.

Thus, based on both L-moment ratio diagram and $|Z^{DIST}|$ -statistic criteria, the PE3 distribution is identified as the most suitable distribution for this region.

ESTIMATION OF QUANTILES

For gauged catchments: The quantile estimates $\hat{Q}(F)$ with nonexceedance probability F at each gauged catchments of the study area can be expressed as

$$\hat{Q}(F) = \bar{Q} \hat{q}(F) \quad (1)$$

Where, $\hat{q}(F)$ is the regional quantile estimates with nonexceedance probability F and \bar{Q} is the mean of the gauged catchments.

The regional quantile estimates with nonexceedance probability F for each of the candidate distributions are given in Table 4. The Growth curves for each of the

candidate distributions PE3, GNO and GEV along with WAK distribution presented in Fig 2. It has been observed that the growth curves of PE3, GNO and GEV distributions are almost identical up to till 200 years return period, that is $F=0.995$. From then onwards, both the GNO and GEV has higher growth curves, in comparison to PE3 distribution. The WAK distribution has lower growth curve than PE3 distribution from $F=0.990$ onwards.

For ungauged catchments: For ungauged catchments a relationship between the mean annual peak flood of gauged catchments in the region and their physiographic catchments characteristics is developed, which is used to estimate mean annual peak flood for an ungauged site. The relationship developed for the region in log domain using least squares approach based on the data of the study area

$$\bar{Q} = 0.881(A)^{0.8} \quad (2)$$

where, A is the catchments area, in sq. km. and \bar{Q} is the mean annual peak flood in m^3/s .

Now by coupling the the quantile estimates $\hat{Q}(F)$ with nonexceedance probability F at ungauged catchments of the region can be expressed as

$$\hat{Q}(F) = (A)^{0.8} \hat{q}^*(F) \quad (3)$$

Where, A is the catchments area in km^2 and $\hat{q}^*(F)$ is the regional quantile estimates for ungauged catchments with nonexceedance probability F.

Assessment of accuracy of quantile estimates:

Assessment of regional frequency analysis based on Monte Carlo simulation technique has been provided by Hosking and Wallis (1997). The procedure starts with development of simulated region. The simulated region should have same number of sites, record length at each site, and regional average L-moment ratios as the actual data. The L-moments ratios at the individual sites should be chosen for simulated region such that whose heterogeneity is consistent with the heterogeneity measure calculated from the actual data. In the simulation procedure, quantile estimates are calculated for various nonexceedance probabilities. At the mth repetition, site-*i* quantile estimate and estimated regional growth curve

Table 3. Regional parameters of three candidate distributions

Distribution	Parameters of the distribution		
PE3	$\mu = 1.000$	$\sigma = 0.423$	$\lambda = 0.849$
GNO	$\mu = 0.942$	$\sigma = 0.400$	$k = -0.287$
GEV	$\mu = 0.813$	$\sigma = 0.351$	$k = 0.048$

for nonexceedance probability F are $\hat{Q}_i^{[m]}(F)$ and $\hat{q}^{[m]}(F)$, respectively. The relative error of the quantile estimate for nonexceedance probability F is $\{ \hat{Q}_i^{[m]}(F) - Q_i(F) \} / Q_i(F)$. The relative root mean square error (relative RMSE) can be approximated by squaring and averaging over all M repetitions of $\hat{Q}_i^{[m]}(F)$ and given by

$$R_i(F) = \left[M^{-1} \sum_{m=1}^M \left\{ \frac{\hat{Q}_i^{[m]}(F) - Q_i(F)}{Q_i(F)} \right\}^2 \right]^{1/2}$$

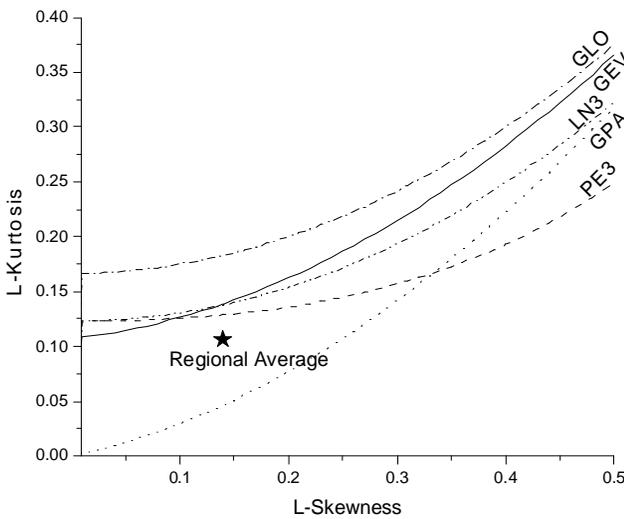


Fig. 1. L-moment ratio diagram for Tripura region.

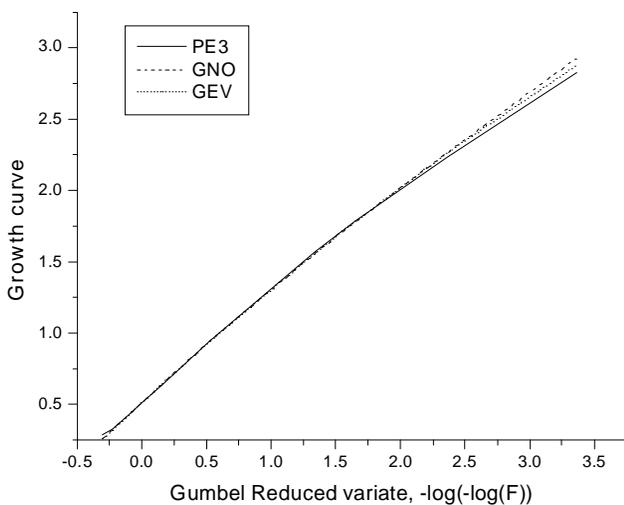


Fig. 2. Growth curves of three candidate distributions for Tripura region.

For estimating the accuracy of estimation over all of the sites in the region, we compute the regional average relative RMSE of the estimated quantile as

$$R^R(F) = N^{-1} \sum_{i=1}^N R_i(F)$$

The above equations can be used for the growth curve estimates but with and replaced by and respectively.

The 90% error bound for are given by Hosking and Wallis (1997) as

$$\frac{\hat{Q}(F)}{U_{.05}(F)} \leq Q(F) \leq \frac{\hat{Q}(F)}{L_{.05}(F)}$$

Where and are the values between which approximately 90% of the distribution of simulated values of $\hat{Q}(F)/Q(F)$ lies.

In this study the Monte Carlo simulation procedure suggested by Hosking and Wallis (1997) has been used for accuracy measures for the estimated regional growth curves to each of the candidate distributions on the six

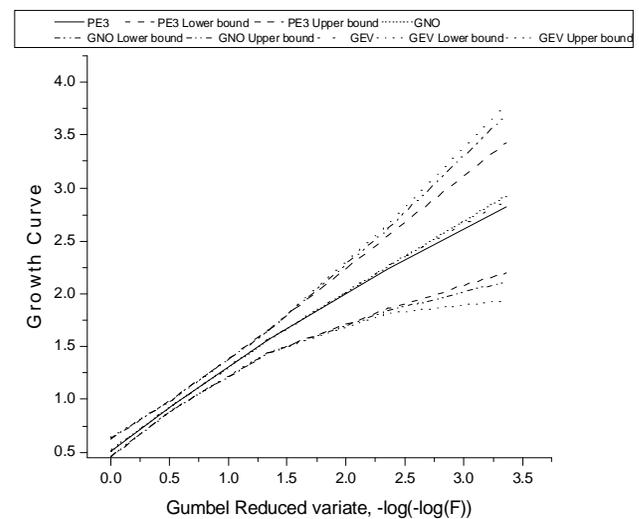


Fig. 3. Growth curves with their 90% error bounds for three candidate distribution of Tripura.

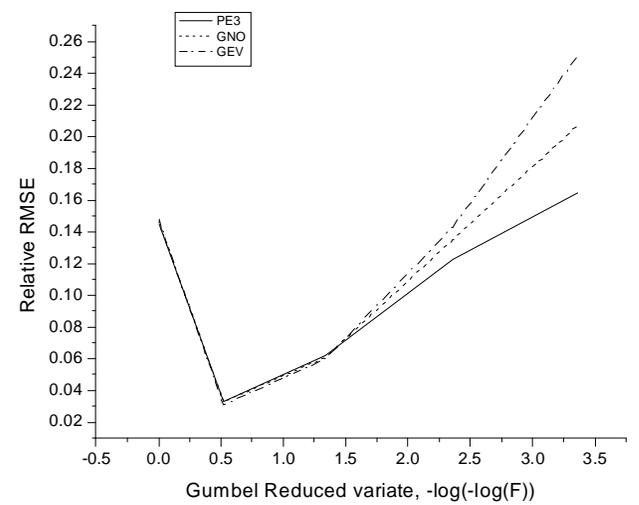


Fig. 4. RMSE of regional growth curves of three candidate distributions of Tripura.

Table 4. Quantile estimates with nonexceedance probability F for gauged catchments.

Distributions	F									
	0.010	0.020	0.050	0.100	0.200	0.500	0.900	0.950	0.990	0.999
PE3	0.282	0.333	0.420	0.511	0.638	0.941	1.566	1.782	2.236	2.825
GNO	0.263	0.321	0.417	0.513	0.643	0.942	1.560	1.781	2.263	2.928
GEV	0.257	0.318	0.418	0.514	0.644	0.941	1.562	1.785	2.263	2.877
WAK	0.129	0.303	0.491	0.558	0.638	0.924	1.604	1.795	2.091	2.303

Table 5. Quantiles estimates with nonexceedance probability F for ungauged catchments.

Distributions	F									
	0.010	0.020	0.050	0.100	0.200	0.500	0.900	0.950	0.990	0.999
PE3	0.2484	0.2934	0.3700	0.4501	0.5621	0.8290	1.3796	1.5699	1.9699	2.4888
GNO	0.2317	0.2828	0.3674	0.4520	0.5666	0.8299	1.3744	1.5691	1.9937	2.5796
GEV	0.2264	0.2802	0.3683	0.4528	0.5674	0.8290	1.3761	1.5726	1.9937	2.5346

Table 6. RMSE and lower and upper bound for three candidate distributions.

F	PE3 distribution				GNO distribution				GEV distribution			
	RMSE				RMSE				RMSE			
0.100	0.511	0.145	0.453	0.623	0.513	0.147	0.454	0.632	0.514	0.148	0.455	0.636
0.500	0.941	0.033	0.898	0.999	0.942	0.033	0.898	0.998	0.941	0.031	0.899	0.993
0.900	1.566	0.062	1.437	1.654	1.560	0.061	1.437	1.649	1.562	0.060	1.444	1.647
0.990	2.236	0.123	1.859	2.553	2.263	0.135	1.838	2.618	2.263	0.143	1.811	2.659
0.999	2.825	0.165	2.200	3.431	2.928	0.207	2.113	3.692	2.877	0.251	1.935	3.791

sites of Tripura, India. For this, 100 simulations of correlated PE3 regions with record lengths the same as the actual regions and L-CV values vary over a range of 0.0838, from 0.1868 at site 1 to 0.2706 at site 6 has been performed and H value obtained as 0.77. This amount of heterogeneity of the simulated region resembles with the heterogeneity value 0.79 of the actual region. Ten thousand realization of this region have been made and the procedures repeated for each of the candidate distributions of the region. The regional average relative root mean square errors and 90% error bounds of the regional growth curves are calculated for each of the candidate distributions and are given in Table 6. The estimated regional growth curves and 90% error bound of each of the candidate distributions are shown in Figure 3. The regional average relative root mean square errors of regional growth curves of each of the candidate distributions are shown in Fig. 4.

It has been observed from Table 6 and Fig. 4 that the RMSE value of all the three candidate distributions are almost same upto F=0.900 and from F=0.900 onwards the RMSE values of PE3 distribution become the lowest among all the candidate distributions. Therefore, PE3 distribution performs relatively better in terms of RMSE in upper tail of the distribution. Again, Fig. 3 shows that in the extreme upper tails, the PE3 distribution has the shortest 90% error bounds compared to GNO and GEV distribution. Therefore, PE3 distribution is the robust distribution for our study area.

Conclusion

The following conclusion was drawn from the regional flood frequency analysis of our study area.

In the step of initial screening of the data the discordancy measure was used, the discordancy measure (Table 1) showed that data of all gauging sites of our study area were suitable for using regional flood frequency analysis. For testing homogeneity of the region the L-moment based heterogeneity measure was used, the L-moment based heterogeneity measure showed that the region is homogeneous.

The regional flood frequency analysis was performed using the various frequency distributions: viz. GLO, GEV, GNO, PE3 and GPA and KAP. The L-moment ratio diagram and $|z^{DIST}|$ -statistic criteria (Table 2) were used to identify suitable candidate distributions PE3, GNO and GEV for the region.

The regional flood frequency relationship for gauged and ungauged catchment has been developed for the region. The regional quantile estimates with non exceedance probability F for PE3, GNO and GEV distributions have been calculated. The quantile estimates for each gauged sites and ungauged sites for known catchments areas was obtained by using the Eq.1 and Eq.2. However, more accurate results for ungauged sites can be obtained if more physiographic characteristics other than catchments area are available.

A Monte Carlo simulation study provided by Hosking

and Wallis (1997) has been carried out to calculate the accuracy measures of the estimated regional growth curves for the three candidate distributions of the study area. Simulation study showed that PE3 distribution is the robust distribution for our study area and can be used for regional quantile estimates for gauged and ungauged catchments.

REFERENCES

- Atiem, I.A. and Harmancioglu, N.B. (2006). Assessment of regional floods using L-moments approach: the case of the River Nile. *Water Resour. Manage.*, 20, 723-747.
- Dalrymple, T. (1960). Flood frequency analysis, *USGS water supply paper*. US Govt. printing office, Washington, DC, 1543, 11-51-A.
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Wat. Resour. Res.*, 15:1049-1054.
- Hosking, J.R.M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J.R. Stat. Soc. Ser. B. Methodol.*, 52(2), 105-124.
- Hosking, J.R.M., and Wallis, J.R. (1997). Regional frequency analysis- An approach based on L-moments, Cambridge University Press, New York.
- Hussain, Z. and Pasha, G.R. (2008). Regional flood frequency analysis of the seven sites of Punjab, Pakistan, using L-moments. *Water Resour. Manage.*, DOI:10.1007/s11269-008-9360-7.
- Kumar, R., Singh, R.D. and Seth, S.M. (1999). Regional flood formulas for seven subzones of zone 3 of India. *J. Hydrologic Eng.*, 4(3), 240-244.
- Kumar, R., Chatterjee, C., Kumar, S. and Lohani, A.K. (2003). Development of regional flood frequency relationship using L-moments for Middle Ganga Plains Subzone 1(f) of India. *Water Resour Manage.*, 17, 243-257.
- Kumar, R. and Chatterjee, C. (2005). Regional flood frequency analysis using L-moments for North Brahmaputra Region of India. *J. Hydrologic Eng.*, 4(3), 240-244.
- Paradia, B.P., Kachroo, R.K. and Shrestha, D.B. (1998). Regional Flood frequency analysis of Mahi-Sabarmati Basin (Subzone 3-a) using Index Flood procedure with L-moments. *Water Resour Manage.*, 12: 1-12, 1998.
- Saf, B. (2008). Regional Flood frequency analysis using L-moments for the West Mediterranean region of Turkey. *Water Resour Manage.*, DOI: 10.1007/s11269-008-9287-z.