

Frequency Analysis of Stream Flow Data Using L-moments of Probability Distributions for Estimation of Peak Flood Discharge

N. Vivekanandan

Central Water and Power Research Station, Pune, Maharashtra, India

ABSTRACT

Estimation of Peak Flood Discharge (PFD) at a desired location on a river is important for planning, design and management of hydraulic structures. This can be achieved through Flood Frequency Analysis (FFA) by fitting of probability distributions to the series of annual maximum discharge data. In the present study, Exponential, Extreme Value Type-1, Extreme Value Type-2, Generalized Pareto and Generalized Extreme Value (GEV) and Normal distributions are adopted in FFA for river Ganga at Allahabad and Varanasi sites. Method of L-Moments (LMO) is used for determination of parameters of the distributions. The adequacy of fitting of probability distributions to the recorded data is quantitatively assessed by applying Goodness-of-Fit (GoF) tests viz., Chi-square and Kolmogorov-Smirnov and diagnostic test using D-index. However, the diverging results based on GoF tests lead to adopt qualitative test to aid the selection of most suitable distribution for estimation of PFD. The study presents the GEV is found to be better suited probability distribution for estimation of PFD for river Ganga at Allahabad and Varanasi.

Keywords: Chi-square, D-index, Generalized Extreme Value, Kolmogorov-Smirnov, L-Moments, Peak Flood

I. INTRODUCTION

Estimation of Peak Flood Discharge (PFD) with a specified return period is crucial for the design of hydraulic structures such as bridges, barrages, culverts, dams and drainage systems. Since the hydrologic phenomena governing the PFD are highly stochastic in nature, the PFD can be effectively determined by fitting of probability distributions to the series of recorded Annual Maximum Discharge (AMD) data. An AMD is the highest instantaneous discharge value at a definite cross-section of a natural stream throughout an entire hydrologic year (water year). The longer the period of observation, the greater would be the length of the recorded series that may offer better results of the Flood Frequency Analysis (FFA).

Number of probability distributions such as Exponential (EXP), Extreme Value Type-1 (EV1), Extreme Value Type-2 (EV2), Generalized Extreme Value (GEV), Generalized Pareto (GPA) and Normal (NOR) are used

in FFA [1]. Generally, Method of Moments (MoM) is used for determination of parameters of the probability distributions. 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 by the 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. To address the aforesaid shortcomings, the application of alternative approach, namely L-Moments (LMO) discussed in this paper is used for FFA [2].

In the recent past, number of studies has been carried out by researchers adopting probability distributions for FFA. Kjeldsen et al. [3] applied LMO in Regional FFA (RFFA) for KwaZulu-Natal Province of South Africa.

Kumar et al. [4] carried out RFFA adopting twelve probability distributions (using LMO) and found that the GEV is better suited distribution for eight gauging sites. Yue and Wang [5] applied LMO to identify the suitable probability distribution for modelling of annual stream flow in different climatic regions of Canada. Kumar and Chatterjee [6] employed the LMO to define homogenous regions within 13 gauging sites of the north Brahmaputra region of India. Atiem and Harmancioglu [7] carried out RFFA using the index flood LMO approach for 14 gauged sites on the tributaries of river Nile. Study by Saf [8] revealed that the Pearson Type-III distribution is better suited for modelling of extreme values in Antalya and Lower-West Mediterranean sub-regions whereas the Generalized Logistic distribution for the Upper-West Mediterranean sub-region.

Bhuyan et al. [9] applied generalized version of LMO (LH-moments) for RFFA of river Brahmaputra. They have found the RFFA based on the GEV distribution by using level one LH-moment give better results over LMO. It was reported by Malekinezhad et al. [10] that GEV (using LMO) is better suited for modelling AMD of three different regions in Iran. Badreldin and Feng [11] carried out the RFFA for the Luanhe basin using LMO and cluster techniques. Haberlandt and Radtke [12] carried out FFA using rainfall and peak flows for three mesoscale catchments in northern Germany. Thus, the studies reported didn't suggest applying a particular distribution for FFA for different region or country. This apart, when different distributions are used for estimation of PFD, a common problem is encountered as regards the issue of best model fits 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 was made from the plot of the recorded and estimated PFD. For quantitative assessment on PFD within in the recorded range, Chi-square (χ^2) and Kolmogorov-Smirnov (KS) tests are applied. A diagnostic test of D-index is used for the selection of most suitable probability distribution for FFA. In the present study, comparison of six probability distributions is made which also illustrates the applicability of GoF and diagnostic tests procedures in identifying the best suitable distribution for estimation of PFD for river Ganga at Allahabad and Varanasi sites.

II. METHODOLOGY

The study is to assess the Probability Distribution Function (PDF) for FFA. Thus, it is required to process and validate the data for application such as (i) select the PDFs for FFA (say, EXP, EV1, EV2, GEV, GPA and NOR); (ii) determine the parameters of the distributions using LMO; (iii) select quantitative GoF and diagnostic tests and (iv) conduct FFA and analysis of results obtained thereof.

Theoretical Description of LMO

LMOs are summary statistics for probability distributions and data samples. They are analogous to ordinary moments, which provide measures of location, dispersion, skewness, kurtosis, and other aspects of the shape of probability distributions or data samples. But, LMOs are computed from linear combinations of the ordered data values [13]. LMO can be used as the basis of a unified approach to the statistical analysis adopting probability distributions. According to CWC [14], LMOs possess the following advantages:

- i) They characterize a wider range of probability distributions than conventional moments.
- ii) They are less sensitive to outliers in the data.
- iii) They approximate their asymptotic normal distribution more closely.
- iv) They are nearly unbiased for all combinations of sample sizes and populations.

LMO will thus particularly useful in providing accurate quantile estimates of hydrological data in developing countries where small sample size typically exist. LMO is a linear combination of probability weighted moments. Let Q_1, Q_2, \dots, Q_N be a conceptual random sample of size N and $Q_{1N} \leq Q_{2N} \leq \dots \leq Q_{NN}$ denote the corresponding order statistics.

The $r+1^{\text{th}}$ LMO defined by Hosking and Wallis [15] is:

$$l_{r+1} = \sum_{k=0}^r \frac{(-1)^{r-k} (r+k)!}{(k!)^2 (r-k)!} b_k \quad \dots (1)$$

where, l_{r+1} is the $r+1^{\text{th}}$ sample moment and b_k is an unbiased estimator with

$$b_k = N^{-1} \sum_{i=k+1}^N \frac{(i-1)(i-2)\dots(i-k)}{(N-1)(N-2)\dots(N-k)} Q_{iN} \quad \dots (2)$$

The first two sample LMOs are expressed by:

$$l_1 = b_0 \text{ and } l_2 = 2b_1 - b_0 \quad \dots (3)$$

Table 1 gives the details of quantile functions and parameters of six probability distributions considered in the study.

Goodness-of-Fit Tests

GoF tests are essential for checking the adequacy of probability distributions to the recorded series of AMD 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 as follows:

χ^2 statistic:

$$\chi^2 = \sum_{j=1}^{NC} \frac{(O_j(Q) - E_j(Q))^2}{E_j(Q)} \quad \dots (4)$$

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. The rejection region of χ^2 statistic at the desired significance level (η) is given by $\chi_C^2 \geq \chi_{1-\eta, NC-m}^2$. Here, m denotes the number of parameters of the distribution.

KS statistic:

$$KS = \text{Max}_{i=1}^N (F_e(Q_i) - F_D(Q_i)) \quad \dots (5)$$

where, $F_e(Q_i) = (i - 0.44) / (N + 0.12)$ is the empirical CDF of Q_i and $F_D(Q_i)$ is the computed CDF of Q_i [16]. If the computed values of GoF tests statistic given by the distribution are less than that of the theoretical values at the desired significance level, then the distribution is considered to be acceptable for estimation of PFD.

TABLE 1
QUANTILE FUNCTIONS AND PARAMETERS OF PROBABILITY DISTRIBUTIONS

S.No.	Distribution	Quantile function (Q_T)	Parameters
1	EXP	$Q_T = \psi - \mu \ln(1 - F)$	ψ (known); $\mu = l_1$
2	EV1	$Q_T = \xi - \alpha \ln(-\ln(1 - F))$	$\xi = l_1 - 0.5772157 \alpha$ $\alpha = l_2 / \ln(2)$
3	EV2	$Q_T = \alpha e^{(-\ln(-\ln(F)))/k}$	By using the logarithmic transformation of the recorded data, parameters of EV1 are initially obtained by MoM and LMO; and further used to determine the parameters of EV2 from $\alpha = e^\xi$ and $k = 1/(\text{scale parameter of EV1})$.
4	GEV	$Q_T = \xi + \alpha(1 - (-\ln F)^k) / k$	$z = (2/(3 + t_3) - (\ln 2 / \ln 3))$; $k = 7.8590z + 2.9554z^2$; $\alpha = l_2 k / (1 - 2^{-k}) \Gamma(1 + k)$; $\xi = l_1 + (\alpha(\Gamma(1 + k) - 1) / k)$
5	GPA	$Q_T = \xi + \alpha(1 - (1 - F)^k) / k$	$\xi = l_1 + l_2(k + 2)$ $k = (4/(t_3 + 1)) - 3$ $\alpha = (1 + k)(2 + k)l_2$
6	NOR	$Q_T = \mu + \sigma \phi^{-1}(F)$	$\mu = l_1$; $\sigma = l_2 \sqrt{\pi}$

where, P is the probability of exceedance ($F(Q)=1/T$); $F(Q)$ (or F) is the Cumulative Distribution Function (CDF) of Q; ϕ^{-1} is the inverse of the standard normal distribution function and $\phi^{-1} = (P^{0.135} - (1-P)^{0.135}) / 0.1975$; ξ, α, k are the location, scale and shape parameters

respectively; μ (or \bar{Q}), σ (or S_Q) and C_s (or Ψ) are the average, standard deviation and coefficient of skewness of the recorded AMD data; $\text{sign}(k)$ is plus or minus 1 depending on the sign of k; Q_T is the estimated PFD by probability distribution for a return period T [17].

Diagnostic Test

The selection of a suitable probability distribution for estimation of PFD is usually carried out through D-index, which is defined as:

$$D\text{-index} = \left(\frac{1}{\bar{Q}} \right) \sum_{i=1}^6 |Q_i - Q_i^*| \quad \dots (6)$$

where, \bar{Q} is the average of the recorded AMD, Q_i 's ($i=1$ to 6) are the first six highest sample values in the series and Q_i^* is the estimated value by PDF. The distribution having the least D-index is identified as better suited

distribution in comparison with the other distributions for estimation of PFD [18].

III. APPLICATION

In this paper, a study was carried out to estimate the PFD adopting six probability distributions (using LMO) for river Ganga at Allahabad and Varanasi sites. Based on the water year (June to May), stream flow data for the period 1986 to 2005 is used. The series of AMD is derived from the daily stream flow data and further used in FFA. Table 2 gives the descriptive statistics of AMD.

TABLE 2
DESCRIPTIVE STATISTICS OF AMD FOR RIVER GANGA AT ALLAHABAD AND VARANASI

Gauging site	Statistical parameters (SD: Standard Deviation; CV: Coefficient of Variation)				
	Mean (m ³ /s)	SD (m ³ /s)	CV (%)	Skewness	Kurtosis
Allahabad	26644.2	8532.4	32.0	-0.004	-1.236
Varanasi	27953.7	7776.5	27.8	-0.115	-0.143

IV. RESULTS AND DISCUSSIONS

The procedures described above for estimating PFD have been implemented adopting computer codes and used in FFA. The program computes the parameters of the six probability distributions (using LMO) and PFD estimates for different return periods. It also performs the GoF tests statistic and D-index values for the data under study.

Estimation of Peak Flood Discharge

The parameters of six probability distributions are determined by LMO and used for estimation of PFD at Allahabad and Varanasi, and given in Tables 3 and 4. These PFD estimates are further used to develop the flood frequency curves and presented in Figures 1 and 2 respectively.

TABLE 3
PFD ESTIMATES FOR DIFFERENT RETURN PERIODS FOR ALLAHABAD

Return period (year)	Estimated PFD (m ³ /s)					
	NOR	EXP	EV1	EV2	GPA	GEV
2	26644	23575	25124	23861	26635	26585
5	34104	32739	33301	31611	35643	34310
10	38003	39671	38715	38081	38654	38234
20	41223	46604	43908	45528	40162	41298
50	44847	55767	50630	57371	41070	44451
100	47263	62700	55667	68224	41373	46331
200	49474	69632	60685	81079	41525	47873
500	52154	78795	67306	101816	41616	49499
1000	54033	85728	72310	120939	41647	50479

TABLE 4
PFD ESTIMATES FOR DIFFERENT RETURN PERIODS FOR VARANASI

Return period (year)	Estimated PFD (m ³ /s)					
	NOR	EXP	EV1	EV2	GPA	GEV
2	27954	25206	26593	25610	28159	28065
5	34632	33410	33913	32652	36051	34884
10	38123	39617	38760	38349	38513	38230
20	41006	45823	43409	44746	39677	40774
50	44250	54027	49427	54635	40334	43317
100	46413	60233	53937	63454	40539	44789
200	48393	66439	58430	73656	40636	45966
500	50792	74644	64358	89667	40691	47172
1000	52475	80850	68838	104038	40708	47878

Analysis Based on GoF Tests

In the present study, the degree of freedom (NC-m-1) was considered as one for 3-parameter distributions (GEV and GPA) and two for 2-parameter distributions (EXP, EV1, EV2 and NOR) while computing the χ^2 statistic values for Allahabad and Varanasi. GoF tests statistic is computed using Eqs. (4) and (5). The GoF tests results are presented in Table 5.

TABLE 5
COMPUTED VALUES OF GoF TESTS STATISTIC

Distribution	Allahabad		Varanasi	
	χ^2	KS	χ^2	KS
NOR	2.000	0.111	5.000	0.077
EXP	1.500	0.166	4.500	0.208
EV1	3.000	0.147	5.500	0.133
EV2	6.500	0.190	5.500	0.189
GPA	4.000	0.077	3.500	0.100
GEV	2.000	0.105	2.500	0.084

From Table 5, it may be noted that the computed values of χ^2 statistic for EV2 and GPA distributions are greater than the theoretical values ($\chi_{0.05,1}^2=3.84$ for GEV and GPA; and $\chi_{0.05,2}^2=5.99$ for EXP, EV1, EV2 and NOR) at 5% significance level and thus these two distributions are not acceptable at 5% level for estimation of PFD at Allahabad. On the other hand, the computed values of χ^2 statistic by the distributions are lesser than the theoretical values at 5% significance level and thus all six distributions are acceptable at 5% level for estimation of PFD at Varanasi. Also, from Table 5, it

may be noted that the computed values of KS statistic by six probability distributions are lesser than the theoretical values (0.294) at 5% significance level, and at this level, all six distributions are acceptable for estimation of PFD at Allahabad and Varanasi.

Analysis Based on Diagnostic Test

For the selection of the best suitable distribution for estimation of PFD, the D-index values of six probability distributions are computed from Eq. (6) and given in Table 6.

TABLE 6
D-INDEX VALUES OF PROBABILITY DISTRIBUTIONS

Gauging station	Distribution					
	NOR	EXP	EV1	EV2	GPA	GEV
Allahabad	0.369	0.865	0.629	0.924	0.196	0.351
Varanasi	0.195	0.347	0.236	0.455	0.349	0.217

By using the diagnostic test results presented in Table 6, the following observations are drawn:

- 1) The indices of D-index of 0.196 (using GPA) for Allahabad and 0.195 (using NOR) for Varanasi are comparatively minimum when LMO is applied for determination of parameters of the distributions.
- 2) For Allahabad and Varanasi, the D-index values of GEV are found to be the second minimum next to the corresponding D-index values of GPA (for Allahabad) and NOR (for Varanasi).
- 3) Based on D-index values, GPA is found to be suitable for estimation of PFD at Allahabad whereas NOR for Varanasi.

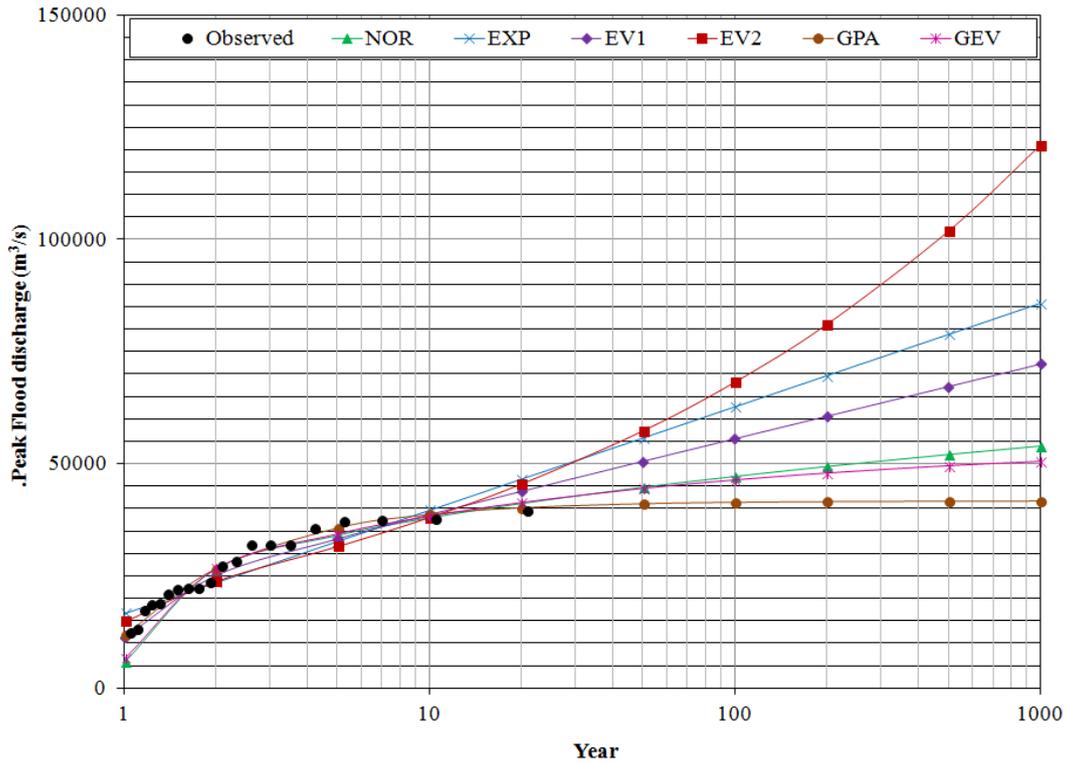


Figure 1. Plots of recorded and estimated PFD by six probability distributions for Allahabad

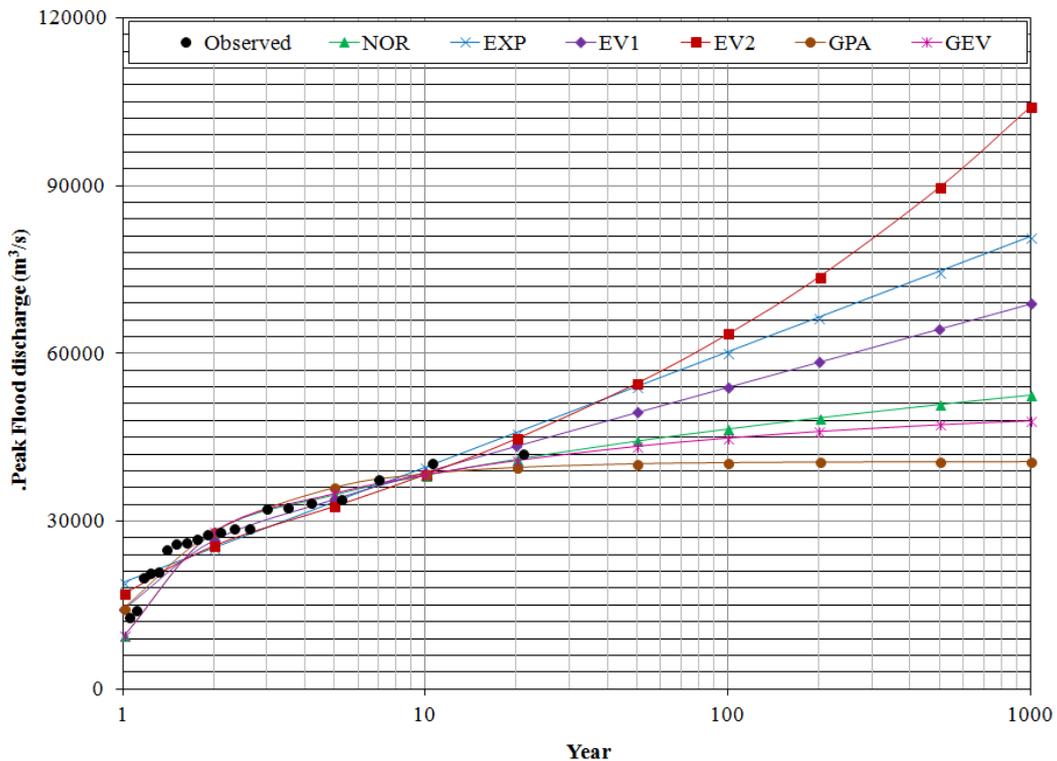


Figure 2. Plots of recorded and estimated PFD by six probability distributions for Varanasi

But, the PFD estimates given by GPA (for Allahabad) and NOR (for Varanasi) are less accurate when compared to GEV because of the characteristics of moment estimators, as described earlier. By considering

the trend lines of the fitted curves using estimated PFD values, the study identifies the GEV distribution is found to be a good choice for estimation of PFD at Allahabad and Varanasi sites.

V. CONCLUSIONS

The paper describes briefly the study carried out for estimation of PFD by adopting FFA using a computer aided procedure for determination of parameters of six probability distributions (using LMO) for river Ganga at Allahabad and Varanasi. The following conclusions are drawn from the study:

- i) For the return period of 20-year and above, it is found that the estimated PFD using EV2 distribution is higher than the corresponding values of other distributions for Allahabad and Varanasi.
- ii) The χ^2 test results showed that the EV2 and GPA distributions are not acceptable for estimation of PFD at Allahabad.
- iii) The χ^2 test results showed that all six distributions adopted in the study are acceptable for estimation of PFD at Varanasi.
- iv) The KS test results indicated that the six distributions are acceptable for estimation of PFD at Allahabad and Varanasi.
- v) By considering the trend lines of the fitted curves using estimated MFD values, the study presented that the GEV distribution is better suited amongst six distributions studied for estimation of PFD at Allahabad and Varanasi.
- vi) The study suggested that the PFD values for different return periods computed by GEV distribution (using LMO) could be considered as the design parameter for planning and design of water resources projects on river Ganga at Allahabad and Varanasi.

VI. ACKNOWLEDGEMENT

The author is grateful to the Director, Central Water and Power Research Station, Pune, for providing the research facilities to carry out the study.

VII. REFERENCES

- [1] Naghavi, B., Yu, F.X. and Singh, V.P. (1993); Comparative evaluation of frequency distributions for Louisiana extreme rainfall, *Water Resources Bulletin*, Vol. 29, No. 2, pp. 211–219.
- [2] Hosking, J.R.M. (1990); L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *Royal Statistical Society (Series B)*, Vol. 52, No. 1, pp. 105-124.
- [3] Kjeldsen, T.R., Smithers, J.C. and Schulze, R.E. (2002); Regional flood frequency analysis in the KwaZulu-Natal province, South Africa, using the index-flood method, *Journal of Hydrology*, Vol. 255, Nos. 1-4, pp. 194–211.
- [4] Kumar, R., Chatterjee, C., Kumar, S., Lohani, A.K. and Singh, R.D. (2003); Development of regional flood frequency relationships using L-moments for Middle Ganga Plains Subzone 1(f) of India, *Water Resources Management*, Vol. 17, No. 4, pp. 243–257.
- [5] Yue, S. and Wang, C.Y. (2004); Possible regional probability distribution type of Canadian annual stream flow by L-moments, *Water Resources Management*, Vol. 18, No. 5, pp. 425–438.
- [6] Kumar, R. and Chatterjee, C. (2005); Regional flood frequency analysis using L-Moments for north Brahmaputra region of India, *Hydrologic Engineering*, Vol. 10, No. 1, pp. 1–7.
- [7] Atiem, I.A. and Harmancioglu, N.B. (2006); Assessment of regional floods using L-moments approach: the case of the River Nile, *Water Resources Management*, Vol. 20, No. 5, pp. 723–747.
- [8] Saf, B. (2009); Regional flood frequency analysis using L-Moments for the West Mediterranean Region of Turkey, *Water Resources Management*, Vol. 23, No. 3, pp. 531-551.
- [9] Bhuyan, A., Borah, M. and Kumar, R. (2010); Regional flood frequency analysis of North-Bank of the River Brahmaputra by using LH-Moments, *Water Resources Management*, Vol. 24, No. 9, pp. 1779-1790.
- [10] Malekinezhad, H., Nachtnebel, H.P. and Klik, A. (2011); Regionalization approach for extreme flood analysis using L-moments, *Agricultural Science and Technology (Iran)*, Vol. 13, Supplementary Issue, pp. 1183–1196.
- [11] Badreldin, G.H.H. and Feng, P. (2012); Regional rainfall frequency analysis for the Luanhe Basin using L-moments and cluster techniques, *International Conference on Environmental Science and Development*, 5-7 January 2012, Hong Kong, Vol. 1, pp. 126–135.
- [12] Haberlandt, U. and Radtke, I. (2014); Hydrological model calibration for derived flood frequency analysis using stochastic rainfall and probability distributions of peak flows, *Hydrology and Earth System Sciences*, Vol. 18, No. 1, pp. 353-365
- [13] Vogel, R.M. and Wilson, I. (1996); The probability distribution of annual maximum, minimum and average stream flow in the United States, *Journal of Hydrologic Engineering*, Vol. 1, No. 1, pp. 69-76.
- [14] Central Water Commission (CWC) (2010); Development of Hydrological Design Aids (Surface water) under Hydrology Project II: State of the Art Report, Consulting Engineering Services (India) in association with HR Wallingford.
- [15] Hosking, J.R.M. and Wallis, J.R. (1993); Some statistics useful in regional frequency analysis, *Water Resources Research*, Vol. 29, No. 2, pp. 271-281.
- [16] Zhang, J. (2002); Powerful goodness-of-fit tests based on the likelihood ratio, *Journal of Royal Statistical Society*, Vol. 64, No. 2, pp. 281-294.
- [17] Rao, A.R. and Hameed, K.H. (2000); Flood frequency analysis, CRC Press, Boca Raton, Florida, USA.
- [18] United States Water Resources Council (USWRC) (1981); Guidelines for determining flood flow frequency, *Bulletin No. 17B*.