



# Comparison of Probability Distributions for Estimation of Peak Flood Discharge

**N. Vivekanandan**

Central Water and Power Research Station, Pune, India

Email: [anandaan@rediffmail.com](mailto:anandaan@rediffmail.com)

Received 10 April 2014; revised 20 May 2014; accepted 1 July 2014

Copyright © 2014 by author and OALib.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Estimation of peak flood discharge (PFD) for a given return period is of utmost importance for planning and design of hydraulic structures in the project site. The peak value of a flood discharge having a pre-defined average return period is determined by frequency analysis, known as flood frequency analysis (FFA). The paper compares the eight probability distributions used for estimation of PFD for Malakkara and Neeleswaram. Maximum likelihood method is used for determination of parameters of the probability distributions. Goodness-of-Fit tests such as Anderson-Darling and Kolmogorov-Smirnov are applied for checking the adequacy of fitting of the distributions to the recorded annual maximum discharge. A diagnostic test of D-index is used for the selection of a most suitable distribution for FFA. Based on GoF and diagnostic test results, the study shows the EV1 distribution is better suited for estimation of PFD for Malakkara whereas LP3 for Neeleswaram.

## Keywords

Anderson-Darling, D-Index, Kolmogorov-Smirnov, Peak Flood, Probability Distribution

**Subject Areas:** Applied Statistical Mathematics, Civil Engineering, Probability Theory

---

## 1. Introduction

Water management and design of irrigation and drainage projects are based on extreme values rather than on average values. Annual peak flood discharge (PFD) corresponding to return periods varying from 2 to 1000 years is generally used by the design engineers and hydrologists for economic planning and design of minor and major hydraulic structures such as barrages, bridges, culverts and dams, etc. The peak value of a flood for a certain return period can be determined by frequency analysis applied to the recorded series of annual maximum discharge (AMD). An annual flood peak is the highest instantaneous discharge value at a definite cross-section of a natural stream throughout an entire hydrologic year (water year), which can be interpreted as the maximum

flood peak occurring in a water year [1] [2]. The longer the period of observation, the greater the length of the recorded sample series, and the more realistic the results of the flood frequency analysis (FFA). Values of the parameters of the probability distribution functions estimated using the available sample series should be unbiased and close to their population values. In the present study, maximum likelihood method (MLM) is used for determination of parameters of the probability distributions.

Probability distributions such as normal (NOR), log-normal (LN2), exponential (EXP), Gamma (GAM), Pearson Type-3 (PR3), Log Pearson Type-3 (LP3), Extreme Value Type-1 (EV1) and Frechet (EV2) are commonly adopted for FFA [3]-[5]. Salami [6] studied the NOR, LN2, square-root-normal and cube-root-normal frequency distributions of meteorological data for Texas. The results showed that the square-root-normal distribution is suitable for precipitation data while all of the frequency distribution for evaporation and temperature data. Lee [7] expressed that the PR3 distribution is better suited amongst five distributions studied for analyzing the rainfall distribution characteristics of Chia-Nan plain area. Bhakar *et al.* [8] studied the frequency analysis of consecutive day's maximum rainfall at Banswara, Rajasthan, India. Fang *et al.* [9] proposed an approach based on the peak-over-threshold sampling method and a non-identical Poisson distribution to model the flood occurrence within each season. Chen *et al.* [10] proposed the use of a copula function to jointly model the distributions of flood magnitude and date of occurrence. Allamano *et al.* [11] analyse the magnitude of under (or over-) estimation of design events in the presence of seasonality by using the peak-over-threshold approach. Bowers *et al.* [12] presents a statistical procedure to partition river flow data and focuses on power transformation and log-normal distribution to describe the constructed seasonal river flows. Vijayagopal *et al.* [13] applied Gumbel and Frechet distributions for development of intensity-duration-frequency relationships for Mandla and Jabalpur regions. But there is no general agreement in applying particular distribution for FFA for different region or country. Moreover, when different distributional models are used for modelling AMD data series, 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 was made from the plot of the recorded and estimated AMD. For quantitative assessment on AMD within in the recorded range, GoF tests such as Anderson-Darling ( $A^2$ ) and Kolmogorov-Smirnov ( $KS$ ) are applied. A diagnostic test of D-index is used for the selection of a most suitable probability distribution for estimation of PFD. The study compares the eight probability distributions used in FFA, and illustrates the applicability of GoF and diagnostic tests procedures in identifying which distributional model is best for estimation of PFD for Malakkara and Neeleswaram.

## 2. Methodology

**Table 1** gives the probability density function (PDF) with the corresponding flood quantile estimator ( $x_T$ ) of eight probability distributions used in FFA.

Here,  $\mu_x$  and  $\sigma_x$  are the mean and standard deviation of the series of AMD,  $\mu_y$  and  $\sigma_y$  are the mean and standard deviation of the log-transformed series of AMD.  $\alpha$ ,  $\beta$  and  $\gamma$  are the scale, location and shape parameters respectively. For EV1 and EV2 distributions, the reduced variable ( $R_T$ ) corresponding to the return period ( $T$ ) is defined by  $R_T = -\ln(-\ln(1-(1/T)))$ .  $K_T$  is the frequency factor corresponding to the return period and coefficient of skewness ( $C_S$ )  $C_S = 2/\sqrt{\beta}$  for GAM,  $C_S = 0.0$  for NOR and LN2;  $K_p$  is the frequency factor corresponding to the  $C_S$  of the original and log-transformed series of AMD for PR3 and LP3 distributions respectively [14].

### 2.1. Goodness-of-Fit Tests

GoF tests such as Anderson-Darling ( $A^2$ ) and Kolmogorov-Smirnov ( $KS$ ) are applied for checking the adequacy of fitting of probability distributions to the recorded AMD [15]. The  $A^2$  statistic is defined by:

$$A^2 = (-N) - (1/N) \sum_{i=1}^N \{ (2i-1) \ln(Z_i) + (2N+1-2i) \ln(1-Z_i) \} \quad (1)$$

Here,  $Z_i = F(X_i)$ , for  $i = 1, 2, 3, \dots, N$  with  $X_1 < X_2 < \dots < X_N$ ,  $F(x_i)$  is the CDF of  $i^{\text{th}}$  sample ( $X_i$ ) and  $N$  is the sample size. The critical value ( $A_C^2$ ) of  $A^2$  statistic for different sample size ( $N$ ) at 5% significance

**Table 1.** PDF with flood quantile estimator of probability distributions.

Distribution	PDF	$x_T$
NOR	$f(x; \mu_x, \sigma_x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \mu_x}{\sigma_x} \right)^2}, x, s_x > 0$	$x_T = \mu_x + K_T \sigma_x$
LN2	$f(x; \mu_y, \sigma_y) = \frac{1}{\sigma_y x \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\ln(x) - \mu_y}{\sigma_y} \right)^2}, x, s_y > 0$	$x_T = \text{Exp}(\mu_y + K_T \sigma_y)$
EXP	$f(x; \alpha, \varepsilon) = \frac{1}{\alpha} e^{-\left( \frac{x - \varepsilon}{\alpha} \right)}, x, \alpha > 0$	$x_T = \varepsilon + \alpha \ln(T)$
GAM	$f(x; \alpha, \varepsilon) = \frac{1}{\alpha^\beta \Gamma \beta} x^{\beta-1} e^{-\left( \frac{x}{\alpha} \right)}, x, \alpha, \beta > 0$	$x_T = \alpha \beta + K_T \sqrt{\beta \alpha^2}$
PR3	$f(x; \alpha, \beta, \gamma) = \frac{1}{\alpha \Gamma \beta} \left( \frac{x - \gamma}{\alpha} \right)^{\beta-1} e^{-\left( \frac{x - \gamma}{\alpha} \right)}, x, \alpha, \beta > 0$	$x_T = \alpha \beta + \gamma + K_T \sqrt{\beta \alpha^2}$
LP3	$f(x; \alpha, \beta, \gamma) = \frac{1}{\alpha x \Gamma \beta} \left( \frac{\ln(x) - \gamma}{\alpha} \right)^{\beta-1} e^{-\left( \frac{\ln(x) - \gamma}{\alpha} \right)}$	$x_T = \text{Exp}((\gamma + \alpha \beta) + K_T \alpha \sqrt{\beta})$
EV1	$f(x; \alpha, \beta) = \frac{e^{-(x-\beta)/\alpha} e^{-e^{-(x-\beta)/\alpha}}}{\alpha}, -\infty < x < \infty, \alpha > 0$	$x_T = \beta + \alpha R_T$
EV2	$f(x; \alpha, \beta) = \frac{\beta}{\alpha} \left( \frac{\alpha}{x} \right)^{\beta+1} e^{-(\alpha/x)^\beta}, -\infty < x < \infty, \alpha > 0$	$x_T = \alpha \text{Exp}(R_T / \beta)$

level is computed from

$$A_C^2 = 0.757 \left( 1 + \left( 0.2 / \sqrt{N} \right) \right) \tag{2}$$

Similarly, the KS statistic is defined by:

$$KS = \text{Max}_{i=1}^N \left( F_e(X_i) - F_D(X_i) \right) \tag{3}$$

Here,  $F_e(X_i) = (i - 0.44) / (N + 0.12)$  is the empirical CDF of  $x_i$  and  $F_D(X_i)$  is the computed CDF of  $X_i$ . If the computed value of GoF tests statistics given by the distribution is less than that of critical value at the desired significance level, then the distribution is considered to be acceptable for FFA.

### 2.2. Diagnostic Test

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

$$\text{D-index} = (1/\bar{x}) \sum_{i=1}^6 \left| x_i - x_i^* \right| \tag{4}$$

Here,  $\bar{x}$  is the average value of the recorded AMD,  $x_i$  is the  $i^{\text{th}}$  sample of the first six highest values in the series of AMD and  $x_i^*$  is the corresponding estimated value by probability distribution. The distribution having the least D-index is considered as the better suited distribution for estimation of PFD [16].

### 3. Application

An attempt has been made to estimate the PFD by eight probability distributions for Malakkara and Neeleswaram. Based on the water year (June-May), stream flow data for the period 1985-86 to 2012-13 for Malakkara and 1971-72 to 2012-13 for Neeleswaram is used. The series of AMD is derived from the daily stream flow data

and further used for FFA. **Table 2** gives the summary statistics of AMD.

## 4. Results and Discussions

By applying the procedures, as described above, a computer program was developed and used for FFA. The program computes the parameters of the probability distributions and GoF tests statistics, PFD estimates for different return periods and D-index values for the stations under study.

### 4.1. Analysis Based on GoF Tests

By applying the procedures of GoF tests, quantitative assessment on fitting of probability distributions to the data series of AMD was carried out; and the results are given in **Table 3**.

From **Table 3**, it may be noted that the computed values of  $A^2$  statistic by EXP and PR3 distributions are greater than the theoretical value of 0.786 at 5% significance level, and at this level, EXP and PR3 distributions are not acceptable for modelling AMD for Malakkara. For Neeleswaram, it may be noted that the computed values of  $A^2$  statistic by EXP, PR3, EV1 and EV2 distributions are greater than the theoretical value of 0.780 and therefore these four distributions are not acceptable for modelling AMD. Also, from **Table 3**, it may be noted that the computed values of  $KS$  statistic by seven probability distributions other than EXP are lesser than the theoretical values (0.250 for Malakkara and 0.205 for Neeleswaram) at 5% significance level, and at this level, the seven distributions other than EXP are found to be acceptable for modelling AMD for Malakkara and Neeleswaram.

### 4.2. Estimation of Peak Flood Discharge

Parameters of eight probability distributions were determined by MLM and used for estimation of PFD for different return periods for Malakkara and Neeleswaram, which are given in **Table 4** and **Table 5** respectively. From the PFD estimates, it may be noted that the EXP and EV2 distributions gave higher values for return period of 5-yr and above when compared to the corresponding values of other seven distributions for Malakkara and Neeleswaram.

**Figure 1** shows the plots of recorded and estimated PFD by five probability distributions for Malakkara and Neeleswaram. Since, the  $A^2$  test results rejected the use of EXP, PR3 and EV2 distributions for both the stations, the corresponding frequency curves are not shown in **Figure 1**.

### 4.3. Analysis Based on Diagnostic Test

For the selection of a most suitable distribution for estimation of PFD for Malakkara and Neeleswaram, the D-index values of eight probability distributions were computed and given in **Table 6**.

From **Table 6**, it may be noted that the D-index values given by EV2 (for Malakkara) and LP3 (for Neeleswaram) distributions are found to be minimum when compared to the corresponding indices of other distributions. But, the  $A^2$  test result showed that the EV2 distribution is not acceptable for estimation of PFD for Malakkara. After eliminating the EV2 distribution from the group of eight probability distributions used in FFA, it may be observed that the D-index value of EV1 is the second minimum next to the EV2; and therefore EV1 distribution is considered to be the most appropriate method for estimation of PFD for Malakkara. On the basis of GoF and diagnostic test results, EV1 distribution is identified as better suited for estimation of PFD for Malakkara whereas LP3 for Neeleswaram.

**Table 2.** Summary statistics of AMD.

Gauging station	Statistical parameters				
	Average (m <sup>3</sup> /s)	SD (m <sup>3</sup> /s)	CV (%)	Skewness	Kurtosis
Malakkara	1007.2	326.3	32.4	1.064	1.632
Neeleswaram	2026.2	663.2	32.7	0.247	-0.864

SD: Standard Deviation; CV: Coefficient of Variation.

**Table 3.** Computed values of GoF tests statistics.

Probability distribution	Computed values of GoF tests statistics			
	Malakkara		Neeleswaram	
	$A^2$	KS	$A^2$	KS
NOR	0.444	0.100	0.564	0.124
LN2	0.182	0.068	0.589	0.128
EXP	6.378	0.437	8.754	0.378
GAM	0.205	0.074	0.454	0.114
PR3	0.958	0.166	1.263	0.123
LP3	0.214	0.078	0.398	0.103
EV1	0.188	0.073	1.015	0.145
EV2	0.746	0.127	1.926	0.168

**Table 4.** Peak flood discharge estimates given by probability distributions for Malakkara.

Return period (yr)	Estimated peak flood discharge (m <sup>3</sup> /s)							
	NOR	LN2	EXP	GAM	PR3	LP3	EV1	EV2
2	1007.2	960.9	698.2	972.6	864.7	952.3	954.6	914.1
5	1276.9	1241.1	1621.1	1265.4	1273.3	1232.2	1237.7	1195.7
10	1417.9	1418.7	2319.3	1439.2	1587.8	1418.7	1425.1	1428.4
20	1534.3	1584.4	3017.4	1593.7	1904.6	1599.1	1604.8	1694.1
50	1665.3	1794.1	3940.4	1780.0	2325.6	1836.4	1837.5	2112.6
100	1752.6	1949.1	4638.5	1911.5	2645.1	2017.8	2011.9	2492.6
200	1832.6	2102.7	5336.7	2037.1	2965.3	2202.8	2185.6	2939.3
500	1929.5	2305.1	6259.7	2196.1	3389.4	2454.9	2414.8	3653.4
1000	1997.4	2458.6	6957.8	2312.2	3710.8	2651.9	2588.1	4306.0

**Table 5.** Peak flood discharge estimates given by probability distributions for Neeleswaram.

Return period (yr)	Estimated peak flood discharge (m <sup>3</sup> /s)							
	NOR	LN2	EXP	GAM	PR3	LP3	EV1	EV2
2	2026.2	1917.2	1404.5	1947.8	1773.6	1988.0	1918.5	1813.3
5	2577.7	2550.0	3261.1	2574.9	2631.5	2595.0	2497.4	2446.3
10	2866.0	2960.1	4665.6	2949.8	3249.7	2920.0	2880.7	2982.7
20	3104.0	3348.0	6070.0	3284.6	3855.0	3185.9	3248.3	3607.4
50	3371.9	3845.6	7926.6	3689.4	4642.9	3477.1	3724.1	4614.1
100	3550.6	4217.9	9331.1	3976.1	5232.7	3664.2	4080.7	5548.7
200	3714.0	4590.0	10735.6	4250.4	5818.6	3829.0	4436.0	6668.1
500	3912.1	5085.3	12592.2	4598.3	6588.5	4019.2	4904.7	8497.7
1000	4051.1	5464.3	13996.7	4852.6	7168.3	4145.7	5259.0	10206.7

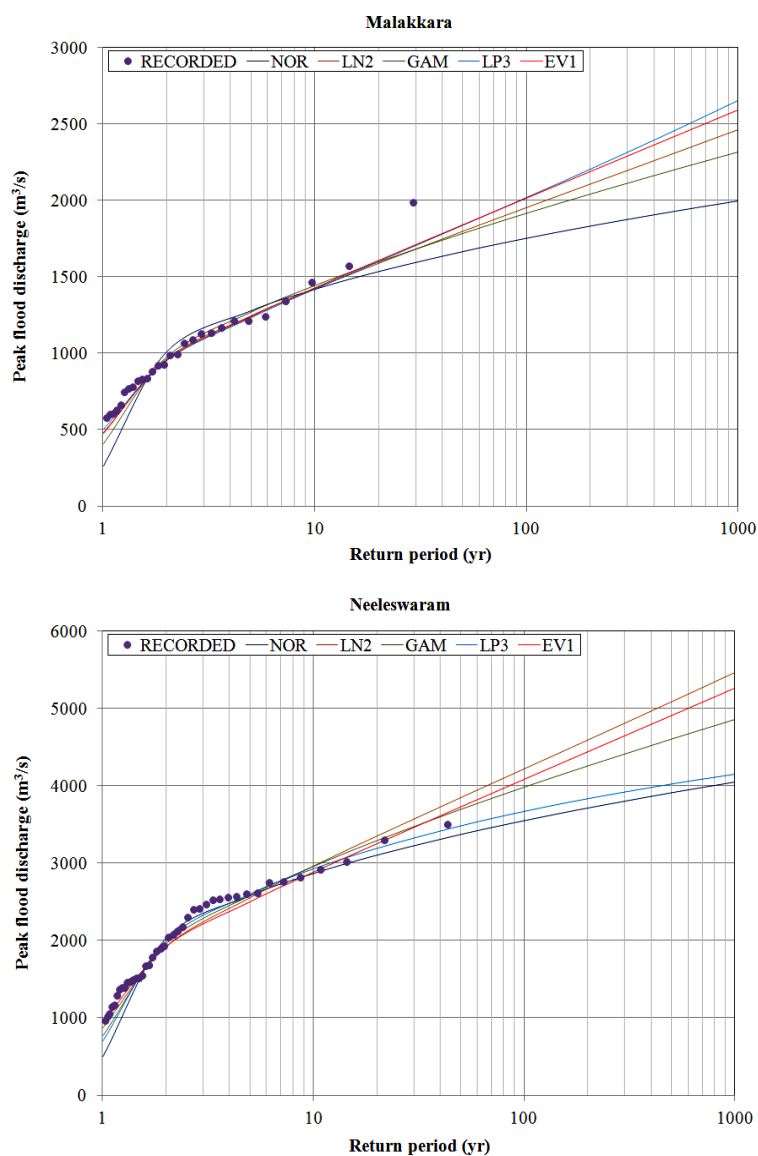
**Table 6.** Indices of D-index for nine probability distributions.

Gauging station	D-index							
	NOR	LN2	EXP	GAM	PR3	LP3	EV1	EV2
MAL	0.668	0.482	4.887	0.513	0.629	0.439	0.427	0.231
NEL	0.218	0.309	7.076	0.185	1.505	0.134	0.148	0.882

MAL: Malakkara; NEL: Neeleswaram.

## 5. Conclusions

The paper compared the eight probability distributions used in estimation of PFD for Malakkara and Neeleswaram. The paper described that the selection of a suitable distribution was performed through GoF and diagnostic



**Figure 1.** Plots of recorded and estimated peak flood discharge by five probability distributions for Malakkara and Neeleswaram.

tests. The  $A^2$  test results showed that the distributions such as NOR, LN2, GAM and LP3 are uniformly acceptable for fitting AMD of both the stations. The  $KS$  test results supported the use of seven distributions other than EXP for fitting AMD of Malakkara and Neeleswaram. The study showed that the EV1 distribution is better suited amongst eight distributions used for estimation of PFD for Malakkara whereas LP3 for Neeleswaram. The 1000-yr return period PFD for Malakkara and Neeleswaram was computed as about 2588 m<sup>3</sup>/s (using EV1) and 4146 m<sup>3</sup>/s (using LP3) respectively. The study suggested that the results presented in the paper would be helpful to the stakeholders for planning and design of irrigation and drainage projects, and major and minor hydraulic structures in Malakkara and Neeleswaram.

### Acknowledgements

The author is grateful to the Director, Central Water and Power Research Station, Pune, for providing the research facilities to carry out the study. The author is also thankful to the Executive Engineer, Central Water Commission, Kerala and Dr. C. Ramesh, Senior Research Officer, CWPRS, Pune, for the supply of stream flow data.

## References

- [1] Neslihan, S., Recep, Y., Tefaruk, H. and Ahmet, D. (2010) Comparison of Probability Weighted Moments and Maximum Likelihood Methods Used in Flood Frequency Analysis for Ceyhan River Basin. *Arabian Journal of Science and Engineering*, **35**, 49-69.
- [2] Baratti, E., Montanari, A., Castellarin, A., Salinas, J.L., Viglione, A. and Bezzi, A. (2012) Estimating the Flood Frequency Distribution at Seasonal and Annual Time Scales. *Hydrological Earth System Science*, **16**, 4651-4660. <http://dx.doi.org/10.5194/hess-16-4651-2012>
- [3] Aksoy, H. (2000) Use of Gamma Distribution in Hydrological Analysis. *Turkey Journal of Engineering and Environmental Sciences*, **24**, 419-428.
- [4] May, W. (2004) Variability and Extremes of Daily Rainfall during the Indian Summer Monsoon in the Period 1901-1989. *Journal of Global & Planetary Change*, **44**, 83-105. <http://dx.doi.org/10.1016/j.gloplacha.2004.06.007>
- [5] Sharda, V.N. and Das, P.K. (2005) Modelling Weekly Rainfall Data for Crop Planning in a Sub-Humid Climate of India. *Journal of Agricultural and Water Management*, **76**, 120-138. <http://dx.doi.org/10.1016/j.agwat.2005.01.010>
- [6] Salami, A.W. (2004) Prediction of the Annual Flow Regime along ASA River Using Probability Distribution Models. AMSE Periodicals. *Modelling, Measurement and Control*, **65**, 41-56.
- [7] Lee, C. (2005) Application of Rainfall Frequency Analysis on Studying Rainfall Distribution Characteristics of Chianan Plain area in Southern Taiwan. *Journal of Crop, Environment & Bioinformatics*, **2**, 31-38.
- [8] Bhakar, S.R., Bansal, A.K., Chhajed, N. and Purohit R.C. (2006) Frequency Analysis of Consecutive Days Maximum Rainfall at Banswara, Rajasthan, India. *ARPN Journal of Engineering and Applied Sciences*, **1**, 64-67.
- [9] Fang, B., Guo, S., Wang, S., Liu, P. and Xiao, Y. (2007) Non-Identical Models for Seasonal Flood Frequency Analysis. *Journal of Hydrological Sciences*, **52**, 974-991. <http://dx.doi.org/10.1623/hysj.52.5.974>
- [10] Chen, L., Guo, S., Yan, B., Liu, P. and Fang, B. (2010) A New Seasonal Design Flood Method Based on Bivariate Joint Distribution of Flood Magnitude and Date of Occurrence. *Journal of Hydrological Sciences*, **55**, 1264-1280. <http://dx.doi.org/10.1080/02626667.2010.520564>
- [11] Allamano, P., Laio, F. and Claps, P. (2011) Effects of Disregarding Seasonality on the Distribution of Hydrological Extremes. *Journal of Hydrology and Earth System Sciences*, **15**, 3207-3215. <http://dx.doi.org/10.5194/hess-15-3207-2011>
- [12] Bowers, M.C., Tung, W.W. and Gao, J.B. (2012) On the Distributions of Seasonal River Flows: Lognormal or Power Law? *Journal of Water Resources Research*, **48**, Paper No. W05536.
- [13] Vijayagopal, P., Vivekanandan, N. and Kannan, S. (2013) Assessing Adequacy of Probability Distribution for Development of IDF Relationships for Mandla and Jabalpur. *Journal of Scientific Research and Reviews*, **2**, 99-114.
- [14] Rao, A.R. and Hamed, K.H. (2000) Flood Frequency Analysis. CRC Publications, New York.
- [15] Zhang, J. (2002) Powerful Goodness-of-Fit Tests Based on the Likelihood Ratio. *Journal of Royal Statistical Society Series B*, **64**, 281-294. <http://dx.doi.org/10.1111/1467-9868.00337>
- [16] United States Water Resources Council (USWRC) (1981) Guidelines for Determining Flood Flow Frequency. Bulletin No. 17B, 15-19.

## Bibliography

N. Vivekanandan is currently working as Assistant Research Officer in Central Water and Power Research Station. He did post graduated in mathematics from Madurai Kamaraj University in 1991. He obtained post graduate degree in hydrology from University of Roorkee and Master of Philosophy degree in mathematics from Bharathiar University. His research interest includes hydrometeorological studies using probabilistic approach, prediction of hydrometeorological variables using soft computing techniques and optimization of hydrometric network using spatial regression approach for various water resources projects.