International Journal of Current Science Research
www.drbgrpublications.in

Volume: 1; Issue: 6; November-2015; pp 116-126. ISSN: 2454-5422

# Comparison of EV1 and LP3 Distributions using Goodness-of-Fit and Diagnostic Tests for Estimation of Design Flood

**N. Vivekanandan**

Central Water and Power Research Station, Pune 411024, Maharashtra, India.

Corresponding Author E-mail: anandaan@rediffmail.com

## Abstract

Estimation of design flood at a desired location on a river is important for planning, design and management of hydraulic structures. This can be carried out through frequency analysis by fitting of probability distribution to the series of recorded annual Peak Flood Discharge (PFD) data. In this paper, Extreme Value Type-1 (EV1) and Log Pearson Type-3 distributions are adopted for frequency analysis of annual PFD data of Giri and Jalal rivers. Maximum likelihood method is used for determination of parameters of the distributions. Goodness-of-Fit tests viz., Anderson-Darling and Chi-square are applied for checking the adequacy of fitting of probability distributions to the recorded data. Diagnostic test, viz., D-index is used for the selection of most suitable distribution for estimation of PFD. The GoF test results and D-index values show the estimated PFD from LP3 distribution could be used for design of hydraulic structures in Giri and Jalal rivers and design of flood in river protection works of Yamuna river basin.

**Keywords: Anderson-Darling, Chi-square, Design flood, D-index, Frequency, Log Pearson**

## Introduction

Estimation of design flood at a desired location on a river is required for effective flood-plain management, and for efficient design of attenuation storages, bridges, culverts, embankments, and flood-protection structures. A design flood is used in comprehensive flood management to assess the Peak Flood Discharge (PFD) and to protect human lives and properties within a watershed (Stedinger and Griffis, 2012). In hydrological context, estimation of PFD can be carried out by two different approaches: (i) statistical approach by applying probability distribution to historical flow data to determine the design flood; and (ii) the design storm method uses the rainfall-runoff model and consider the rainfall quantile determined by frequency analysis as the input data (Rogger et al., 2012). Statistical approach using recorded annual PFD data is considered the standard method for estimation of flood quantile for a given return period and hence used in the present study.

A number of probability distributions such as Extreme Value Type-1 (EV1), Generalized Extreme Value (GEV), 2-parameter Log Normal (LN2), Log Pearson Type-3 (LP3) and Normal are widely applied for frequency analysis of stream flow data (Saf, 2007; Casas et al., 2011; Lee et al, 2012; Daneshfaraz et al., 2013). Olumide et al. (2013) applied normal and EV1 distributions for prediction of rainfall and runoff at Tagwai dam site in Minna, Nigeria. They have also expressed that the normal distribution is better suited for rainfall prediction while Log-Gumbel for runoff. Izinyon and Ajumuka (2013) carried out Flood Frequency Analysis (FFA) for three tributaries of upper Benue river basin, Nigeria adopting EV1, LN2 and LP3 distributions. Das and Qureshi (2014) evaluated the probability distributions of GEV, LN2 and LP3 adopted in FFA through D-index and found that the LP3 is better suited distribution for estimation of PFD for Jiya Dhol river basin. Rasel and Hossain (2015) applied EV1 distribution for development of intensity duration frequency curves for seven divisions in Bangladesh. In view of the above, EV1 and LP3 distributions are used in the present study. Parameters of the distributions are determined by Maximum Likelihood Method (MLM) and used to estimate the PFD. For quantitative assessment on PFD data within the recorded range, Goodness-of-Fit (GoF) tests such as Anderson-Darling ($A^2$) and Chi-square ($\chi^2$) are applied. Diagnostic test, viz., D-index is used for the selection of most suitable distribution for estimation of PFD. The methodology adopted in frequency analysis of annual PFD data through GoF tests and diagnostic index are briefly described in the following sections.

**Methodology**

The study is to estimate design flood of Giri and Jalal rivers of Yamuna River Basin (YRB). Thus, it is required to process and validate the data series of annual PFD such as (i) determine the parameters of EV1 and LP3 distributions (using MLM); (ii) assess the adequacy of fitting of EV1 and LP3 distributions to the series of annual PFD using GoF tests; (iii) selection of most suitable probability distribution for estimation of PFD using D-index and (iv) analyse the results obtained thereof.

**PDF and CDF of Probability Distributions**

The Probability Density Function (PDF; f(X)) and Cumulative Distribution Function (CDF; F(X)) of EV1 distribution is given as below:

$$\left. \begin{aligned} f(X) &= \frac{e^{-(X-\alpha)/\beta} e^{-e^{-(X-\alpha)/\beta}}}{\beta} \\ F(X) &= e^{-e^{-(X-\alpha)/\beta}}, \beta > 0 \end{aligned} \right\} \quad \text{... (1)}$$

where, α and β are the location and scale parameters of the distribution (Gumbel, 1960). The parameters are computed by MLM through Equations (2) and (3), and used to estimate the PFD ($X_T$) for different return periods from $X_T = \alpha + Y_T\beta$, where $Y_T = -\ln(-\ln(1-(1/T)))$.

$$\alpha = -\beta \ln\left[\sum_{i=1}^{N} \exp(-X_i/\beta) \Big/ N\right] \quad \text{... (2)}$$

$$\beta = \overline{X} - \left[\sum_{i=1}^{N} X_i \exp(-X_i/\beta) \Big/ \sum_{i=1}^{N} \exp(-X_i/\beta)\right] \quad \text{... (3)}$$

$$SE(X_T) = \left(\beta/\sqrt{N}\right)\left(1.15894 + 0.19187Y_T + 1.1Y_T^2\right)^{0.5} \quad \text{... (4)}$$

where, $X_i$ is the recorded PFD of $i^{th}$ sample, $\overline{X}$ is the average value of PFD and $SE(X_T)$ is the standard error of the estimated PFD ($X_T$) for a given return period (T).

The PDF and CDF of Log Pearson Type-3 (LP3) distribution is given by:

$$f(X;\alpha,\beta,\gamma) = \frac{1}{\beta X \Gamma\gamma}\left(\frac{\ln(X)-\alpha}{\beta}\right)^{\gamma-1} e^{-\left(\frac{\ln(X)-\alpha}{\beta}\right)}, \ \beta,\gamma>0 \text{ and } F(R) = \int f(X;\alpha,\beta,\gamma)dX \quad \text{... (5)}$$

where, α, β and γ are the location, scale and shape parameters of the LP3 distribution (Rao and Hameed 2000). The parameters are computed by MLM through Equation (6) and used to estimate the PFD ($X_T$) for different return periods from $X_T = \mathrm{Exp}((\alpha + \beta\gamma) + K_P\beta\sqrt{\alpha})$. Here, $K_P$ is the frequency factor corresponding to the probability of exceedance and coefficient of skewness based on the log transformed series of the recorded data for LP3.

$$\left. \begin{aligned} &\sum_{i=1}^{N} (\ln(X_i) - \alpha) = N\beta\gamma \\ &N\psi(\gamma) = \sum_{i=1}^{N} \ln\left[(\ln(X_i) - \alpha)/\beta\right] \\ &N = \beta(\gamma-1)\sum_{i=1}^{N}\left[1/(\ln(X_i) - \alpha)\right] \end{aligned} \right\} \quad \text{... (6)}$$

where, $\psi(\gamma) = \Gamma'(\gamma) / \Gamma(\gamma)$ is called a digamma function. The SE of estimated PFD adopting LP3 distribution (using MLM) is computed from Equation (7) and given by:

$$SE(X_T) = \frac{\beta}{N} \left\{ \begin{array}{l} 1 + KC_S + \left(\frac{K^2}{2}\right)\left(\frac{3C_S^2}{4} + 1\right) + 3K\frac{\partial K}{\partial C_S}\left(\frac{C_S^3}{4} + C_S\right) + \\ 3\left(\frac{\partial K}{\partial C_S}\right)^2\left(2 + 3C_S^2 + 5C_S + \frac{5C_S^4}{9}\right) \end{array} \right\} \qquad \text{... (7)}$$

where, $\frac{\partial K}{\partial C_S} \cong \left(\frac{Z^2 - 1}{6}\right) + \frac{4(Z^3 - 6Z)}{6^3}C_S - \frac{3(Z^2 - 1)}{6^3}C_S^2 + \frac{4Z}{6^4}C_S^3 - \frac{10}{6^6}C_S^4$ and Z is the standard normal variate.

The lower and upper confidence limits (LCL and UCL) of the estimated PFD value at 95% level are obtained from $LCL = X_T - 1.96(SE(X_T))$ and $UCL = X_T + 1.96(SE(X_T))$.

## Goodness-of-Fit Tests

GoF tests are essential for checking the adequacy of probability distribution to the recorded series of annual PFD for estimation of PFD. Out of a number GoF tests available, the widely accepted GoF tests are $A^2$ and $\chi^2$, which are used in the study. The theoretical descriptions of GoF tests statistic (Charles Annis, 2009) are as follows:

$A^2$ statistic:

$$A^2 = (-N) - (1/N)\sum_{i=1}^{N}\{(2i-1)\ln(Z_i) + (2N + 1 - 2i)\ln(1 - Z_i)\} \qquad \text{… (8)}$$

Here, $Z_i = F(X_i)$ for i=1,2,3,…,N with $X_1 < X_2 < \ldots < X_N$, $F(X_i)$ is the CDF of $i^{th}$ sample ($X_i$) and N is the sample size.

$\chi^2$ statistic:

$$\chi^2 = \sum_{j=1}^{NC}\frac{(O_j(X) - E_j(X))^2}{E_j(X)} \qquad \text{... (9)}$$

where, $O_j(X)$ is the observed frequency value of $j^{th}$ class, $E_j(X)$ is the expected frequency value of $j^{th}$ class and NC is the number of frequency class (Zhang, 2002). The rejection region of $\chi^2$ statistic at the desired significance level ($\eta$) 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 $\chi_C^2$ is the computed value of $\chi^2$ statistic by PDF.

*Test criteria*: 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 found to be acceptable for modelling the series of annual PFD data.

**Diagnostic Test**

Diagnostic test, viz., D-index is used for the selection of most suitable probability distribution for estimation of PFD (USWRC, 1981), which is defined by:

$$\text{RMSE}=\frac{1}{N}\sum_{i=1}^{6}\left|X_i - X_i^*\right| \qquad \text{... (10)}$$

Here, $\overline{X}$ is the average value of the recorded PFD data whereas $X_i$ and $X_i^*$ are the six highest recorded and corresponding estimated PFD values by PDF. The distribution having the least D-index is considered as better suited distribution for estimation of PFD.

**Application**

In this paper, an attempt has been made to estimate the PFD for different return periods for Giri and Jalal rivers, which are the tributaries of YRB. The series of annual PFD of Giri river for the period 1991 to 2009 and Jalal river for the period 1993 to 2009 is used. Table 1 gives the descriptive statistics of annual PFD data of Giri and Jalal rivers.

Table 1: Descriptive statistics of annual PFD of Giri and Jalal rivers

| PFD series | Statistical parameters | | | | |
|---|---|---|---|---|---|
| | $\overline{X}$ ($m^3$/s) | $S_X$ ($m^3$/s) | CV (%) | $C_S$ | $C_K$ |
| Giri river | 940.4 | 918.2 | 97.6 | 1.841 | 2.919 |
| Jalal river | 140.3 | 86.3 | 61.5 | 0.515 | -1.010 |
| $S_X$: Standard Deviation; $C_S$: Coefficient of Skewness; $C_K$: Coefficient of Kurtosis | | | | | |

**Results and Discussions**

By applying the procedures of EV1 and LP3 distributions, parameters were determined by MLM and used for estimation of PFD. Tables 2 and 3 give the PFD estimates with 95% lower and upper confidence limits for different return periods adopting EV1 and LP3 distributions for Giri and Jalal rivers.

Table 2: PFD estimates with confidence limits using EV1 and LP3 distributions for Giri river

| Return period (year) | EV1 | | | | LP3 | | | |
|---|---|---|---|---|---|---|---|---|
| | $X_T$ ($m^3/s$) | SE ($X_T$) ($m^3/s$) | Confidence limits | | $X_T$ ($m^3/s$) | SE ($X_T$) ($m^3/s$) | Confidence limits | |
| | | | LCL | UCL | | | LCL | UCL |
| 2 | 776.7 | 145.9 | 490.7 | 1062.6 | 661.1 | 124.2 | 417.7 | 904.6 |
| 5 | 1374.3 | 223.9 | 935.4 | 1813.1 | 1414.1 | 230.4 | 962.5 | 1865.6 |
| 10 | 1769.9 | 287.2 | 1207.0 | 2332.9 | 2037.1 | 330.5 | 1389.2 | 2684.9 |
| 15 | 1993.2 | 324.7 | 1356.9 | 2629.5 | 2418.8 | 394.0 | 1646.6 | 3191.0 |
| 20 | 2149.5 | 351.4 | 1460.8 | 2838.1 | 2709.5 | 442.9 | 1841.5 | 3577.6 |
| 25 | 2269.9 | 372.1 | 1540.5 | 2999.2 | 2936.2 | 481.4 | 1992.7 | 3879.7 |
| 50 | 2640.8 | 436.9 | 1784.4 | 3497.1 | 3671.0 | 607.4 | 2480.5 | 4861.5 |
| 75 | 2856.3 | 475.0 | 1925.3 | 3787.4 | 4150.5 | 690.3 | 2797.6 | 5503.4 |
| 100 | 3008.9 | 502.1 | 2024.7 | 3993.1 | 4470.7 | 746.1 | 3008.4 | 5933.0 |

Table 3: PFD estimates with confidence limits using EV1 and LP3 distributions for Jalal river

| Return period (year) | EV1 | | | | LP3 | | | |
|---|---|---|---|---|---|---|---|---|
| | $X_T$ ($m^3/s$) | SE ($X_T$) ($m^3/s$) | Confidence limits | | $X_T$ ($m^3/s$) | SE ($X_T$) ($m^3/s$) | Confidence limits | |
| | | | LCL | UCL | | | LCL | UCL |
| 2 | 125.1 | 19.5 | 86.9 | 163.3 | 123.1 | 19.2 | 85.6 | 160.7 |
| 5 | 200.3 | 29.9 | 141.7 | 258.8 | 206.4 | 30.8 | 146.0 | 266.7 |
| 10 | 250.1 | 38.3 | 174.9 | 325.2 | 259.4 | 39.8 | 181.5 | 337.3 |
| 15 | 278.2 | 43.3 | 193.2 | 363.1 | 287.5 | 44.8 | 199.8 | 375.3 |
| 20 | 297.8 | 46.9 | 205.9 | 389.7 | 307.1 | 48.4 | 212.4 | 401.9 |
| 25 | 313.0 | 49.7 | 215.6 | 410.3 | 321.5 | 51.0 | 221.5 | 421.5 |
| 50 | 359.6 | 58.3 | 245.3 | 473.9 | 363.8 | 59.0 | 248.2 | 479.4 |
| 75 | 386.7 | 63.4 | 262.5 | 511.0 | 388.3 | 63.6 | 263.6 | 513.1 |
| 100 | 405.9 | 67.0 | 274.6 | 537.3 | 406.6 | 67.1 | 275.1 | 538.1 |

From Tables 2 and 3, it may be noted that the PFD estimates given by LP3 are consistently higher than the corresponding values of EV1 for return period above 2-year. The plots of recorded and estimated PFD using EV1 and LP3 distributions with 95% confidence limits for Giri and Jalal rivers are presented in Figures 1 to 4. These figures showed that the recorded PFD data are falling between the confidence limits of the estimated PFD obtained from EV1 and LP3 distributions for Giri and Jalal rivers.

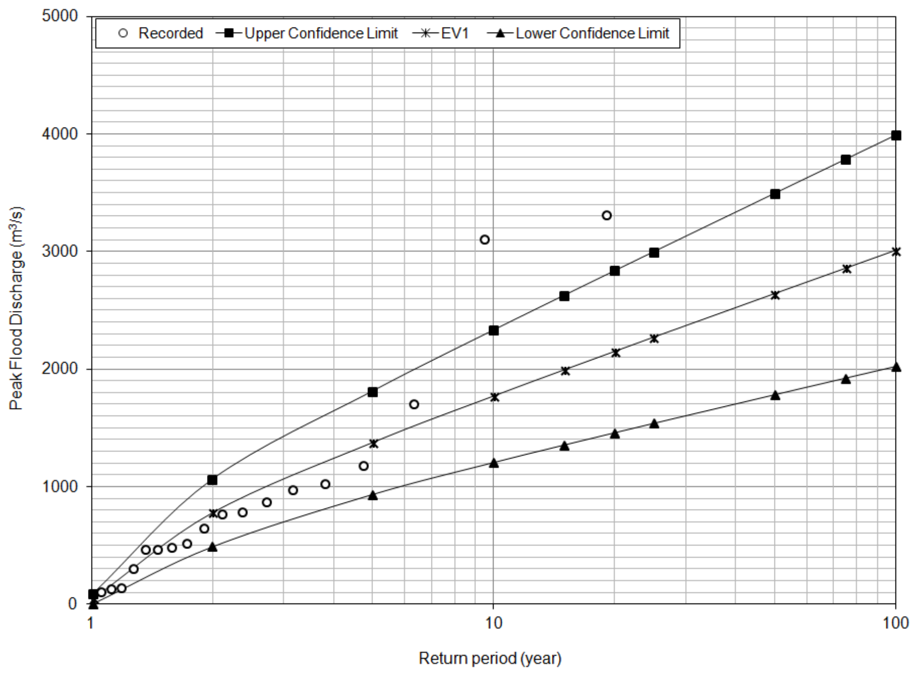


Figure 1: Plots of recorded and estimated PFD using EV1 distribution with 95% confidence limits for Giri river

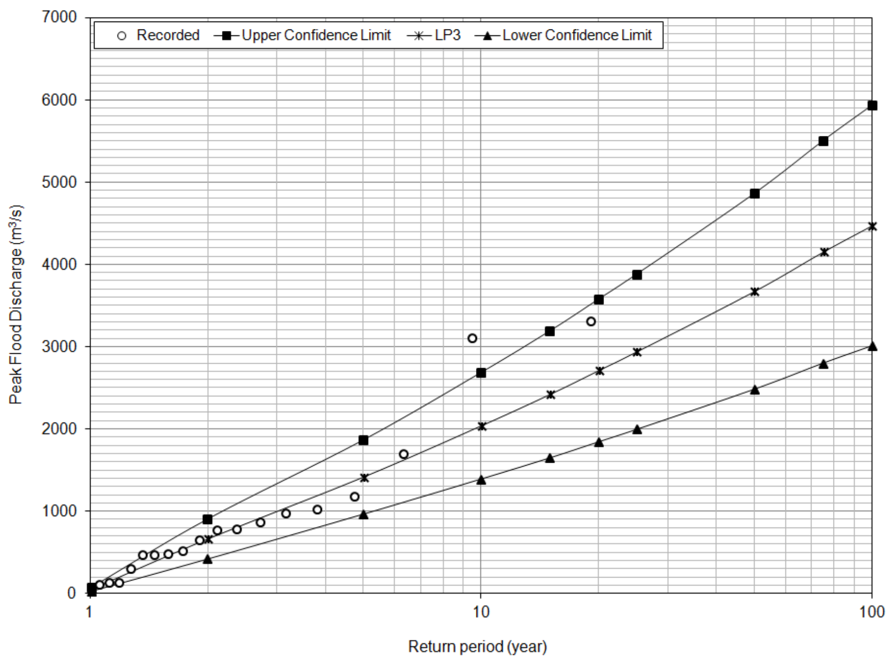


Figure 2: Plots of recorded and estimated PFD using LP3 distribution with 95% confidence limits for Giri river

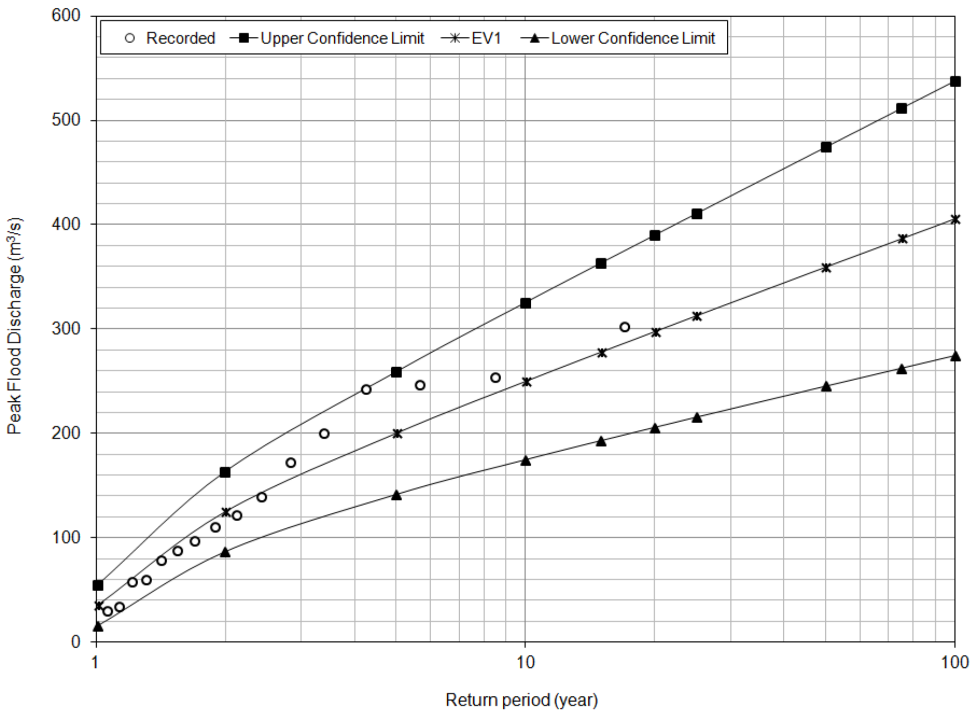


Figure 3: Plots of recorded and estimated PFD using EV1 distribution with 95% confidence limits for Jalal river

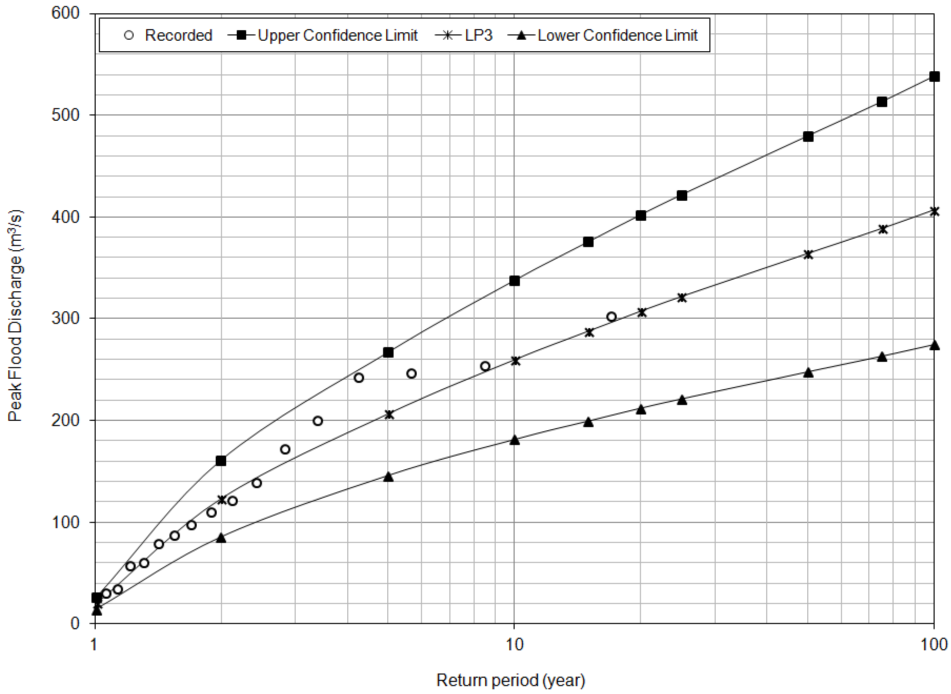


Figure 4: Plots of recorded and estimated PFD using LP3 distribution with 95% confidence limits for Jalal river

**Analysis Based on GoF Tests**

The adequacy of fitting of EV1 and LP3 distributions for frequency analysis of PFD data was performed by adopting GoF tests ($A^2$ and $\chi^2$) and the results are presented in Table 4.

Table 4: Computed values of GoF tests using EV1 and LP3 distributions

| PFD series | EV1 | | LP3 | |
|---|---|---|---|---|
| | $A^2$ | $\chi^2$ | $A^2$ | $\chi^2$ |
| Giri river | 0.673 | 1.333 | 0.329 | 2.333 |
| Jalal river | 0.329 | 0.500 | 0.236 | 0.500 |

From Table 4, it may be noted that the computed values are not greater than the theoretical values (0.757 for $A^2$ and 5.99 for $\chi^2$) at 5% significance level, and at this level, the GoF tests results supported the EV1 and LP3 distributions for modelling the series of annual PFD.

**Analysis Based on Diagnostic Test**

The selection of most suitable probability distribution amongst EV1 and LP3 distributions was made through D-index and the results are presented in Table 5.

Table 5: D-index values of EV1 and LP3 distributions

| PFD series | EV1 | LP3 |
|---|---|---|
| Giri river | 2.974 | 2.405 |
| Jalal river | 1.227 | 0.975 |

From Table 5, it may be noted that the D-index values of LP3 distribution are found to be minimum when compared to the corresponding values of EV1 and hence LP3 distribution was identified as better suited probability distribution for estimation of PFD.

**Conclusions**

The paper presents the study carried out for frequency analysis of annual PFD for Giri and Jalal rivers adopting EV1 and LP3 distributions (using MLM). GoF tests such as $A^2$ and $\chi^2$

were applied for checking the adequacy of fitting of probability distributions to the recorded annual PFD data. The selection of most suitable probability distribution was made through D-index. The following conclusions are drawn from the study:

i) From the frequency curves, it was found that the recorded annual PFD data are falling between the confidence limits of the estimated PFD from EV1 and LP3 distributions.
ii) The GoF test results supported the EV1 and LP3 distributions for modelling the series of annual PFD data.
iii) Based on D-index values, the LP3 distribution was identified as better suited distribution for estimation of PFD.
iv) The study suggested that the upper confidence limit of 100-year return period estimated PFD of about 5993 $m^3/s$ for Giri river and 538 $m^3/s$ for Jalal river from LP3 distribution (using MLM) could be used for design of hydraulic structures in the rivers.
v) The study also suggested that the results presented in the paper would be beneficial for stake holders for arriving at a design flood in river protection works of YRB.

**Acknowledgements**

The author is grateful to the Director, Central Water and Power Research Station (CWPRS), Pune, for providing the research facilities to carry out the study. The author is thankful to Dr. R.G. Patil, CWPRS, for the supply of stream flow data.

**References**

Stedinger J R and Griffis V W 2012 Flood frequency analysis in the United States: Time to update. J. Hydrologic Engineering. 13(4): 199-204

Rogger M; Kohl B; Pirkl H; Viglione A; Komma J; Kimbauer R; Merz R and Blöschl G 2012 Runoff models and flood frequency statistics for design flood estimation in Austria-Do they tell a consistent story?. J. Hydrology. 456-457: 30-43

Casas M C; Rodriguez R; Prohom M; Gazquez A and Redano A 2011 Estimation of the probable maximum precipitation in Barcelona (Spain). J. Climatology. 31(9): 1322-1327

Charles Annis P E 2009 Goodness-of-Fit tests for statistical distributions, [http://www.statistical engineering. com/goodness.html]

Das L M and Qureshi Z H 2014 Flood frequency analysis for Jiya Dhol river of Brahmaputra valley. J. Sci. Basic & Applied Res. 14(2): 14-24

Daneshfaraz R; Nemati S; Asadi H and Menazadeh M 2013 Comparison of four distributions for frequency analysis of wind speed: A case study. J. Civil Engineering & Urbanism. 3(1): 6-11

Gumbel E J 1960 Statistic of Extremes, 2nd edition, Columbia Univ. Press, New York

Izinyon O C and Ajumuka H N 2013 Probability distribution models for flood prediction in Upper Benue River Basin. J. Civil & Environ. Res. 3(2): 62-74

Lee B H; Ahn D J; Kim H G and Ha Y C 2012 An estimation of the extreme wind speed using the Korea wind map. Renewable Energy. 42(1): 4-10

Rasel M and Hossain S M 2015 Development of rainfall intensity duration frequency equations and curves for seven divisions in Bangladesh. Int. J. Scientific & Engineering Res. 6(5): 96-101

Rao A R and Hameed K H 2000 Flood Frequency Analysis, CRC Publications, Washington D.C., New York

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

Zhang J 2002 Powerful Goodness-of-Fit Tests Based on the Likelihood Ratio. J. Royal Statistical Society: Series B. 64(2): 281-294

United States Water Resources Council (USWRC) 1981 Guidelines for Determining Flood Flow Frequency, Bulletin No. 17B

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