*International Journal of Applied Engineering and Technology ISSN: 2277-212X (Online)*
*An Online International Journal Available at http://www.cibtech.org/jet.htm*
*2012 Vol. 2 (1) January-March, pp. 30-37/Vivekanandan*
***Research Article***

# INTERCOMPARISON OF EXTREME VALUE DISTRIBUTIONS FOR ESTIMATION OF ADMR

***N. Vivekanandan**
*Central Water and Power Research Station, Pune 411024*
**Author for correspondance*

## ABSTRACT
Estimation of rainfall for a desired return period is one of the pre-requisite for planning, design and management of various hydraulic structures. This paper illustrates the use of Gumbel (EV1) and Frechet (EV2) distributions for estimation of annual one-day maximum rainfall (ADMR) for different return periods for Bhira, Khalapur, Mangaon and Murud regions. Goodness-of-Fit (GoF) tests like Anderson-Darling, Chi-square and Kolmogorov-Smirnov are applied for checking the adequacy of fitting of the distributions to the recorded rainfall data. Model Performance Indicators (MPIs) such as correlation coefficient and relative mean square error are used for selection of suitable distribution for estimation of ADMR. The results of GoF tests and MPIs showed that the EV1 distribution is better suited for estimation of ADMR for the regions under study.

***Key Words**: Correlation; Frechet; Goodness-of-Fit; Gumbel; Mean Square Error; Rainfall; Return Period*

## INTRODUCTION
Estimation of rainfall for a desired return period is a pre-requisite for various hydraulic structures such as dams, bridges, barrages and design of storm water drainage systems. Rainfall depth also becomes an important input in derivation of the flood discharges that includes standard project flood and probable maximum flood. In practice, hydrological and stream flow series of a significant duration are generally available for large river basins. However, for ungauged basins or comparatively smaller industrial, thermal power or similar plant areas, not much data are available except the rainfall data (AERB, 2008). Depending on the size and the proposed life of the structure/plant, the design rainfall depth corresponding to a certain return period is used. Such an approach is generally adopted while working with ungauged catchments.
Probability distributions of Extreme Value Type-1/2/3, Gamma, Normal, Lognormal, Pearson Type-3 and Log Pearson Type-3 are commonly available for fitting hydrometeorological variables such as rainfall, streamflow, temperature, wind speed, etc. In probabilistic theory, generalized extreme value distribution is identified as a family of continuous probability distributions that include Gumbel (EV1), Frechet (EV2) and Weibull (EV3). As defined by the extreme value theorem in statistics, the asymptotic distributions of the extremes tend to converge on certain limiting forms for large sample; specifically to the double exponential form, or to two different single exponential forms. Since the extreme values of a random variable are invariably associated with the tails of its probability density function, the convergence of the distribution function of its extreme value to a particular limiting form will depend on the behaviour at tail end of the initial distribution in the direction of the extreme (Aksoy, 2000; May, 2004; Sharda and Das, 2005; Carta and Ramirez, 2007). In view of the above, an attempt has been made to compare the rainfall estimation procedures using EV1 and EV2 distributions with a specific objective to identify the most suitable distribution for modelling rainfall data recorded at Bhira, Khalapur, Mangaon and Murud regions. Maximum likelihood method (MLM) is used for estimation of parameters of EV1 and EV2 distributions. Goodness-of-Fit (GoF) tests like Anderson-Darling ($A^2$), Chi-square ($\chi^2$) and Kolmogorov-Smirnov (KS) are applied for checking the adequacy of fitting of the distributions to the recorded rainfall data. Model Performance Indicators (MPIs) such as correlation coefficient (CC) and relative mean square error (RMSE) are used for selection of suitable distribution for estimation of ADMR for different return

periods for the regions under study. The paper presents the methodology adopted in estimating the rainfall by EV1 and EV2 distributions, GoF tests and MPIs, and the results obtained from the study.

## MATERIALS AND METHODS

Fitting of probability distributions to the recorded rainfall data provides the estimates of ADMR for different return periods such as 2-year (yr), 5-yr, 10-yr, 20-yr, 50-yr and 100-yr. The theoretical descriptions of probability distributions, GoF tests and MPIs are briefly described in the ensuing sections.

***Probability distributional models***

The cumulative distribution function (CDF) of EV1 and EV2 distributions is given by:

$$F(R) = e^{-e^{-\left(\frac{R_i-\alpha_G}{\beta_G}\right)}}, \quad \beta_G>0 \text{ and } \infty<R_i<\infty \quad \text{(for EV1)}$$

$$F(R) = e^{-\left(\frac{R_i}{\beta_F}\right)^{(-\lambda_F)}}, \quad \beta_F>0 \text{ and } \infty<R_i<\infty \quad \text{(for EV2)}$$

where, $\alpha_G$ and $\beta_G$ are the location and scale parameters of Gumbel distribution. The MLM estimators of EV1 distribution are given by:

$$\beta = \sum_{i=1}^{N}\frac{R_i}{N} - \left[\sum_{i=1}^{N} R_i \exp(-R_i/\beta) \Big/ \sum_{i=1}^{N}\exp(-R_i/\beta)\right]$$

$$\alpha = -\beta \log\left[\sum_{i=1}^{N}\exp(-R_i/\beta) \Big/ N\right]$$

The estimates of ADMR adopting Gumbel distribution ($R_G$) for different return periods (T) are computed from $R_G = \alpha_G + Y_T\beta_G$ with $Y_T = -\ln(-\ln(1-(1/T))$. Similarly, $\beta_F$ and $\lambda_F$ are the scale and shape parameters of Frechet distribution (Suhaila and Jemain, 2007). Based on extreme value theory, Frechet distribution can be transformed to Gumbel distribution through logarithmic transformation. Under this transformation, the estimates of ADMR adopting Frechet distribution ($R_F$) for different return periods are computed from $R_F = \text{Exp}(R_G)$, $\beta_F = \text{Exp}(\alpha_G)$ and $\lambda_F = 1/\beta_G$ (IAEA, 2003). The standard error (SE) on the estimated rainfall for a return period (T) is computed by:

$$SE = \frac{\beta}{\sqrt{N}}\left(1.1589 + 0.1919 Y_T + 1.1 Y_T^2\right)^{0.5}$$

The lower and upper confidence limits (LCL and UCL) of the estimated ADMR are obtained by using $LCL = R_T - 1.96SE$ and $UCL = R_T + 1.96SE$

***Goodness-of-Fit tests***

$$\chi^2 = \sum_{j=1}^{NC}\frac{\left(O_j(R) - E_j(R)\right)^2}{E_j(R)} \qquad (1)$$

where $O_j(R)$ is the recorded frequency value of ADMR of $j^{th}$ class, $E_j(R)$ is the expected frequency value of ADMR of $j^{th}$ class, NC is the number of frequency classes, p is the number of parameters of the distribution.

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

For a given sample of 'N' values, $Z_{(i)}=F(R_i)$, for i=1,2,3,….N; and $R_1<R_2<$ ….$R_N$. The distribution of $A^2$ statistics doesn't depend on F(R), but on the set of 'N' sample values. The distribution theory of ordered statistic gives the percentage points for testing $A^2$ statistics.

$$KS = \underset{i=1}{\overset{N}{\mathrm{Max}}}\left(F_e(R_i) - F_D(R_i)\right) \quad (3)$$

Here, $F_e(R_i) = (i-0.35)/N$ is the empirical CDF of $R_i$, $F_D(R_i)$ is the computed CDF of $R_i$ and N is the number of observations.

The rejection region of $\chi^2$, $A^2$ and KS statistics at the desired significance level 'η' are $\chi^2_C > \chi^2_{NC-p-1,1-\eta}$, $A^2_C > A^2_{1-\eta}$ and $KS_C > KS_{N,1-\eta}$. If the computed values ($\chi^2_C$, $A^2_C$ and $KS_C$) of GoF test statistics of the distribution are less than that of critical value at the desired significance level 'η' then the selected distribution is accepted to be adequate than any other distribution (Zhang, 2002).

***Model performance indicators***

Following, Chen and Adams (2006), the performance of estimated ADMR using EV1 and EV2 distributions are analyzed by CC and RMSE, and are:

$$CC = \frac{\sum_{i=1}^{N}\left(R_i - \overline{R}\right)\left(R_i^* - \overline{R^*}\right)}{\sqrt{\left(\sum_{i=1}^{N}\left(R_i - \overline{R}\right)^2\right)\left(\sum_{i=1}^{N}\left(R_i^* - \overline{R^*}\right)^2\right)}} \quad (4)$$

$$RMSE = \left(\frac{1}{N}\right)\sum_{i=1}^{N}\left(\frac{R_i - R_i^*}{R_i}\right)^2 \quad (5)$$

where $R_i$ is the recorded ADMR, $R_i^*$ is the estimated ADMR using EV1 and EV2, $\overline{R}$ is the series mean of recorded ADMR and $\overline{R^*}$ is the series mean of estimated ADMR.

**APPLICATION**

An attempt has been made to estimate the ADMR for different return periods using EV1 and EV2 distributions. The daily rainfall data recorded at Bhira, Khalapur, Mangaon and Murud regions related to the period 1968-1996 are used. Table 1 gives the statistical parameters of ADMR for the regions under study.

**Table 1**: **Statistical parameters of ADMR**

| Region | Mean (mm) | Std. Dev. (mm) | Skewness | Kurtosis |
|---|---|---|---|---|
| Bhira | 284.0 | 69.5 | 0.183 | -0.981 |
| Khalapur | 197.3 | 47.6 | 0.412 | 0.074 |
| Mangaon | 220.5 | 52.7 | 0.992 | 0.290 |
| Murud | 187.2 | 42.5 | 0.211 | -0.848 |

**RESULTS AND DISCUSSION**

By describing the procedures detailed above, a computer program was developed and used to fit the recorded ADMR data for rainfall estimation using EV1 and EV2 distributions. The program computes the parameters of the distributional model, one-day maximum rainfall estimates for different return periods, GoF test statistics and MPIs. Table 2 gives the parameters of EV1 and EV2 distributions. The distributional parameters were used to develop CDF curves and delineated in Figure 1.

**Table 2**: **Parameters of EV1 and EV2 distributions**

| Region | EV1 (Gumbel) | | EV2 (Frechet) | |
|---|---|---|---|---|
| | Location ($\alpha_G$) | Scale ($\beta_G$) | Scale ($\beta_F$) | Shape ($\lambda_F$) |
| Bhira | 250.461 | 59.932 | 246.705 | 5.203 |
| Khalapur | 176.215 | 41.835 | 172.046 | 5.331 |
| Mangaon | 197.249 | 40.320 | 194.617 | 5.793 |
| Murud | 168.392 | 32.576 | 164.708 | 5.633 |

Tables 3 and 4 give the annual one-day maximum rainfall estimates using EV1 and EV2 distributions together with standard error on the estimated rainfall for the regions under study.

**Table 3**: **One-day maximum rainfall estimates with standard error using EV1 distribution**

| Return period (yr) | Estimated rainfall (mm) with standard error using EV1 distribution for | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Bhira | | Khalapur | | Mangaon | | Murud | |
| | Rainfall | SE | Rainfall | SE | Rainfall | SE | Rainfall | SE |
| 2 | 272 | 10.2 | 192 | 7.2 | 212 | 8.0 | 180 | 6.9 |
| 5 | 340 | 17.1 | 239 | 12.1 | 258 | 13.5 | 217 | 11.5 |
| 10 | 385 | 23.1 | 270 | 16.3 | 288 | 18.2 | 242 | 15.6 |
| 20 | 428 | 29.2 | 300 | 20.6 | 317 | 23.0 | 265 | 19.6 |
| 50 | 484 | 37.3 | 339 | 26.5 | 355 | 29.4 | 296 | 25.1 |
| 100 | 526 | 43.5 | 369 | 30.6 | 383 | 34.2 | 318 | 29.2 |

**Table 4**: **One-day maximum rainfall estimates with standard error using EV2 distribution**

| Return period (yr) | Estimated rainfall (mm) with standard error using EV2 distribution for | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Bhira | | Khalapur | | Mangaon | | Murud | |
| | Rainfall | SE | Rainfall | SE | Rainfall | SE | Rainfall | SE |
| 2 | 265 | 10.0 | 184 | 7.0 | 207 | 7.9 | 176 | 6.7 |
| 5 | 329 | 16.9 | 228 | 11.8 | 252 | 13.3 | 215 | 11.3 |
| 10 | 380 | 22.8 | 262 | 15.9 | 287 | 18.0 | 246 | 15.2 |
| 20 | 437 | 28.8 | 300 | 20.1 | 325 | 22.7 | 279 | 19.2 |
| 50 | 522 | 36.8 | 358 | 25.6 | 382 | 29.0 | 329 | 24.5 |
| 100 | 597 | 42.8 | 408 | 29.9 | 431 | 33.8 | 373 | 28.6 |

From Tables 3 and 4, it is noticed that the ADMR estimates for different return periods above 10-yr using EV2 are relatively higher, when compared with the corresponding estimates given by EV1 for all four data sets.

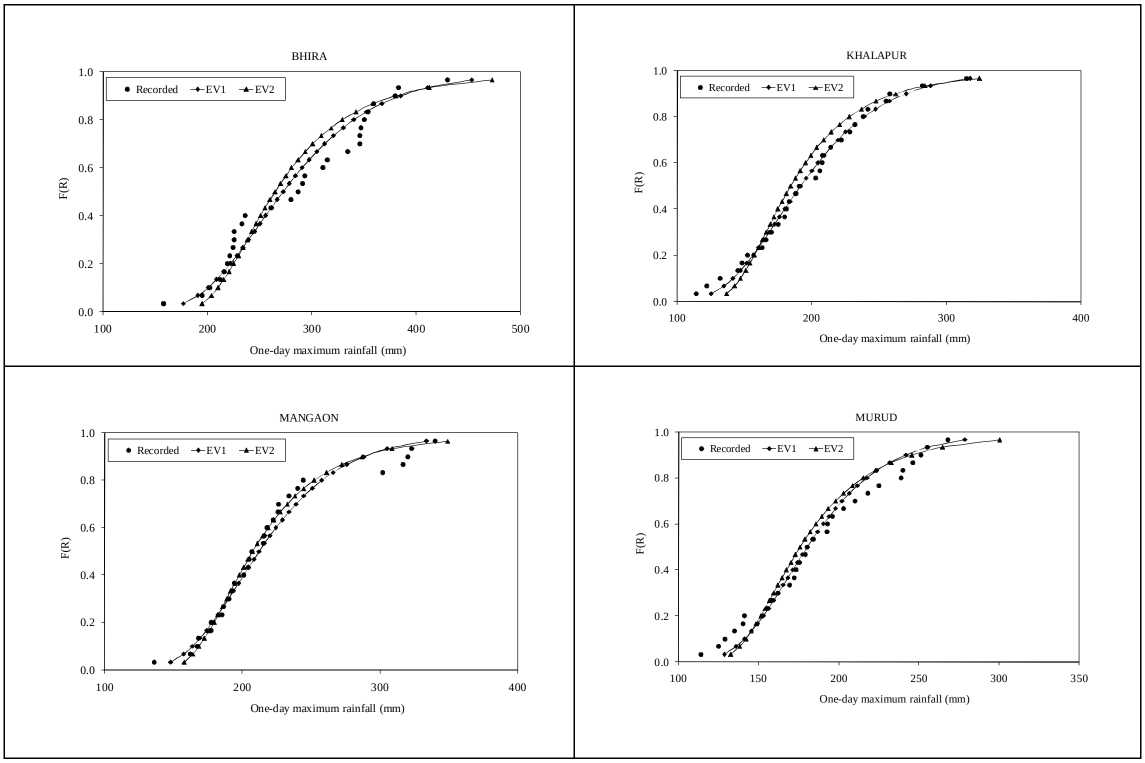


**Figure 1**: **CDF plots of recorded and estimated rainfall using EV1 and EV2 distributions**

Figure 2 shows the plots of recorded and estimated rainfall for different return periods using EV1 and EV2 distributions for Bhira, Khalapur, Mangaon and Murud regions.

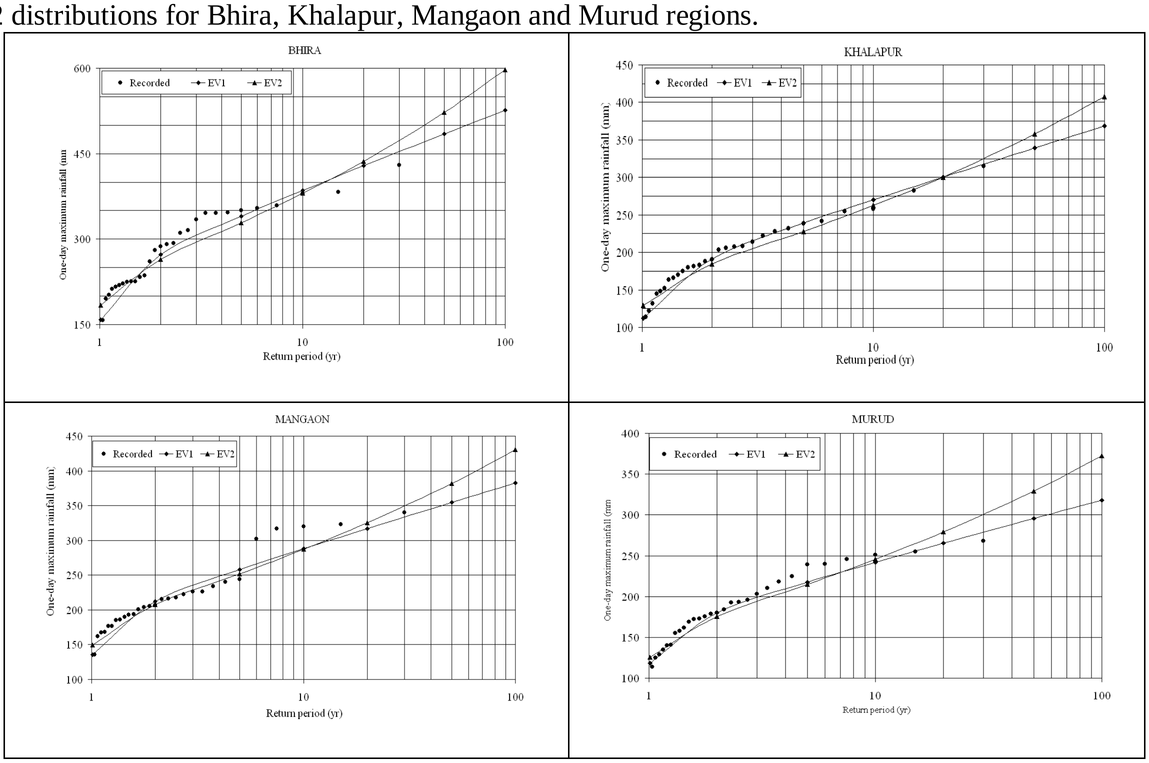


**Figure 2**: **Plots of recorded and estimated rainfall for different return periods using EV1 and EV2 distributions**

***Analysis based on GoF tests***

For the assessment on fitting of EV1 and EV2 distributions to the recorded rainfall data, GoF test statistics were computed by using Eqs. (1-3) and are given in Table 5. In the present study, degrees of freedom for both the distributions were considered as four while computing $\chi^2$ values.

**Table 5**: **Computed values of GoF test statistics using EV1 and EV2 distributions**

| Region | EV1 | | | EV2 | | |
|---|---|---|---|---|---|---|
| | $\chi^2$ | $A^2$ | KS | $\chi^2$ | $A^2$ | KS |
| Bhira | 3.590 | 0.678 | 0.134 | 7.030 | 1.335 | 0.160 |
| Khalapur | 1.520 | 0.174 | 0.073 | 2.210 | 1.048 | 0.144 |
| Mangaon | 1.170 | 0.457 | 0.109 | 0.480 | 0.373 | 0.097 |
| Murud | 1.520 | 0.532 | 0.108 | 2.550 | 1.065 | 0.116 |

From Table 5, it is noted that the computed values of $\chi^2$ statistics by EV1 and EV2 distributions are less than the critical ($\chi^2_{4,0.05}$) value of 9.488 at 5% level of significance, and hence at this level, both distributions are accepted to fit the ADMR data recorded at the respective sites. Also from Table 5, it is noted that the computed values of KS statistics given by both distributions are also less than the critical value of 0.240 at 5% level of significance, and at this level, EV1 and EV2 distributions are fitted well to the ADMR data recorded at the respective sites. On the other hand, $A^2$ statistics doesn't support the use of EV2 distribution for modelling ADMR data recorded at Bhira, Khalapur and Murud regions because of the computed values of $A^2$ statistics are greater than critical value of 0.757 at 5% level of significance.

***Analysis based on MPIs***

For the selection of an appropriate distribution for modelling ADMR data, MPIs were computed from Eq. (4-5) and given in Table 6.

**Table 6: Computed values of MPIs using EV1 and EV2 distributions**

| Region | CC | | RMSE | |
|---|---|---|---|---|
| | EV1 | EV2 | EV1 | EV2 |
| Bhira | 0.972 | 0.944 | 0.056 | 0.080 |
| Khalapur | 0.994 | 0.978 | 0.034 | 0.066 |
| Mangaon | 0.972 | 0.970 | 0.049 | 0.053 |
| Murud | 0.983 | 0.957 | 0.053 | 0.066 |

From Table 6, it is noticed that the values of RMSE given by EV1 are minimum when compared with the corresponding values of EV2 distribution though there is a good correlation between the recorded and estimated values given by both distributions. The values of CC given by EV1 and EV2 distributions are varied from 0.944 to 0.994. Based on GoF tests results and MPIs, EV1 is considered to be good choice for estimation of ADMR for different return periods for the regions under study. Figure 3 shows the plots of recorded and estimated rainfall using EV1 distribution together with confidence limits at 95% level for Bhira, Khalapur, Mangaon and Murud regions.

From Figure 3, it can be seen that the recorded data falls within the line of agreement of confidence limits of the estimated rainfall given by EV1. The study showed that the percentages of uncertainty in rainfall estimation by EV1 for the regions varied from 3% to 6%, which is less than the acceptable tolerance limit of 10 % (ISO 5168, 1978). From the results of data analysis, it is suggested that EV1 (Gumbel) distribution could be used uniformly for estimation of ADMR for the regions considered in the study.

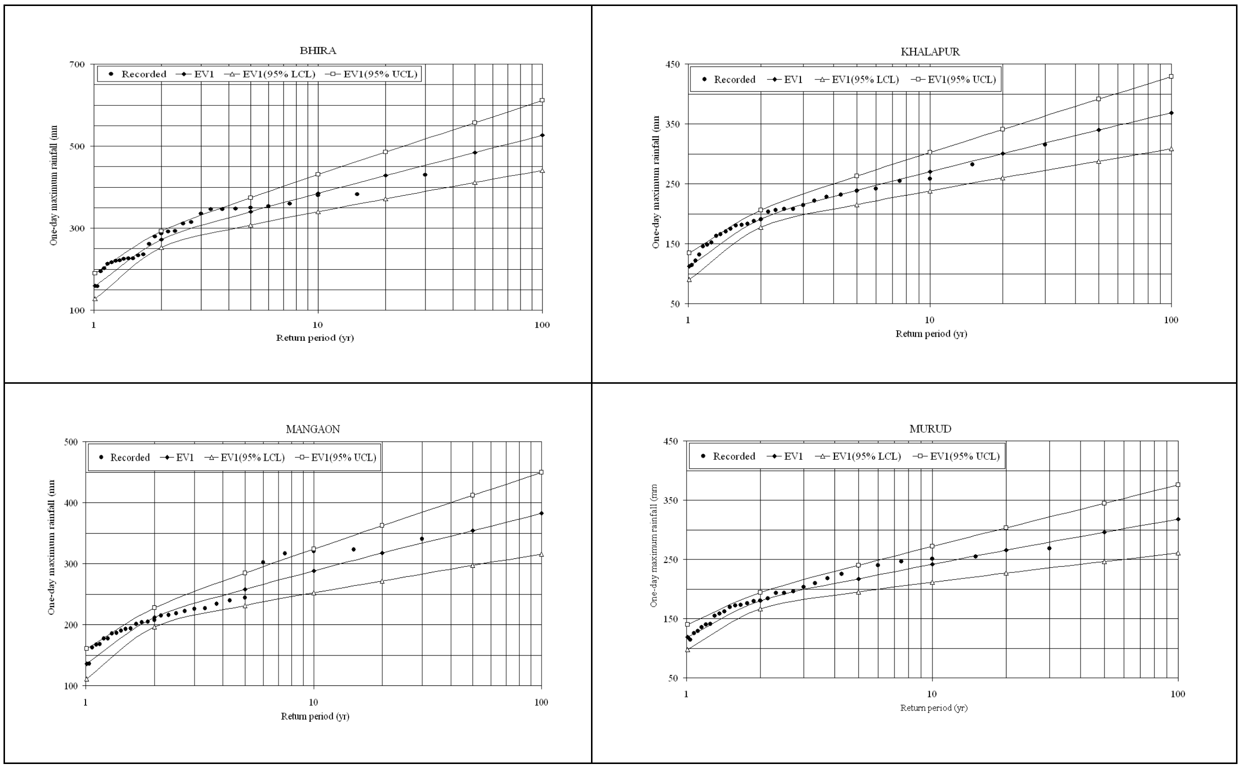


**Figure 3**: **Plots of recorded and estimated rainfall using EV1 distribution together with *95% confidence limits***

## CONCLUSIONS

The paper presented a study on rainfall estimation procedures using extreme value distributions of EV1 and EV2 for modelling rainfall data. The results of GoF tests and MPIs indicated that EV1 (Gumbel) distribution is better suited for estimation of ADMR for different return periods for Bhira, Khalapur, Mangaon and Murud regions. The study showed that the values of CC given by both EV1 and EV2 distributions for the regions varied from 0.944 to 0.994. The study also showed that the RMSE on the estimated rainfall using EV1, with reference to the recorded rainfall is less than the corresponding values given by EV2 for all four sites. From the results of GoF tests and MPIs, it is identified that EV1 is better suited for modelling ADMR data for the regions under study. The methodology reported in the paper is expected to be of assistance to stakeholders to arrive at a design parameter for planning, design and management of hydraulic structures at the regions.

## ACKNOWLEDGEMENTS

The author is thankful to India Meteorological Department, Pune, for making available the rainfall data. The author is grateful to the Director, Central Water and Power Research Station, Pune 411024, for providing the research facilities to carry out the study.

**REFERENCES**

**Aksoy H (2000).** Use of Gamma Distribution in Hydrological Analysis, Turkey Journal of Engineering Environmental Sciences, **24** (6), 419-428.

**Atomic Energy Regulatory Board (AERB, 2008).** Extreme values of meteorological parameters, AERB Safety Guide AERB/ NF/ SG/S-3.

**Carta J A and Ramirez P (2007).** Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions, Renewable Energy, **32** (3), 518-531.

**Chen J and Adams B J (2006).** Integration of Artificial Neural Networks with Conceptual Models in Rainfall-Runoff Modelling, Journal of Hydrology, **318** (3), 232-249.

**International Atomic Energy Agency (IAEA, 2003).** Meteorological events in site evaluation for Nuclear Power Plants – Safety Guide, No. Ns-G-3.4, International Atomic Energy Agency, Vienna.

**ISO 5168 (1978).** Measurement of fluid flow – Estimation of uncertainty of a flow rate measurement, Bureau of Indian Standards.

**May W (2004).** Variability and extremes of daily rainfall during the Indian summer monsoon in the period 1901-1989, Global and Planetary Change, **44** (1-2), 83-105.

**Sharda V N and Das P K (2005).** Modelling weekly rainfall data for crop planning in a sub-humid climate of India, Agricultural Water Management, **76** (2), 120-138.

**Suhaila J and Jemain A A (2007).** Fitting Daily Rainfall Amount in Peninsular Malaysia using Several Types of Exponential Distributions, Journal of Applied Sciences Research, **3 (10)**, 1027-1036.

**Zhang J (2002).** Powerful goodness-of-fit tests based on the likelihood ratio, Journal of Royal Statistical Society Series-B, **64** (2), 281-294.