551.579.4 : 551.577

INTERCOMPARISON OF ANNUAL RAINFALL ESTIMATES OF TARAI REGION USING STATISTICAL DISTRIBUTIONS

1. Analysis of rainfall is one of the important governing factors in planning the agricultural program for any region. Knowledge of rainfall pattern and its distribution would be very much helpful to the decision makers to identify the optimal cropping pattern and effective water management plan and to design irrigation drainage, erosion and flood control structures. Approaches like deterministic and probabilistic are commonly available to study about the rainfall pattern and its distribution in the region. Deterministic approach

TABLE 1

PDF and rainfall estimator (X_T) of six statistical distributions

No.	Distribution	PDF	X_T
1	EV I	$f(x;\alpha,m)=\dfrac{e^{-(x-\alpha)/m}e^{-e^{-(x-\alpha)/m}}}{m}$ $-\infty< x <\infty, -\infty<\alpha<\infty, m >0$	$X_T = \alpha + Y_T m$
2	G 2	$f(x;\alpha,\lambda)=\dfrac{\lvert\alpha\rvert e^{-\alpha x}(\alpha x)^{\lambda-1}}{\Gamma(\lambda)}$, $x>0, \alpha, \lambda>0$	$X_T = \left(\dfrac{1}{\alpha}\right)\left(K_T\sqrt{\lambda}+\lambda\right)$
3	N 2	$f(x;\alpha,\lambda)=\left(1/\lambda\sqrt{2\pi}\right)\exp\left[-\dfrac{(x-\alpha)^2}{2\lambda^2}\right]$, $\alpha, \lambda>0$	$X_T = \alpha + \lambda K_T$
4	LN 2	$f(x;\alpha,\lambda)=\left(1/\lambda x\sqrt{2\pi}\right)\exp\left[-\dfrac{(\ln x-\alpha)^2}{2\lambda^2}\right]$	$X_T = \mathrm{Exp}(\alpha + \lambda K_T)$
5	P III	$f(x;\alpha,\lambda,m)=\dfrac{\lvert\alpha\rvert}{\Gamma(\lambda)}e^{-\alpha(x-m)}\left[\alpha\left(x-m\right)\right]^{\lambda-1}$ $x>0, \alpha, \lambda>0, -\infty<m<\infty$	$X_T = m + \left(\dfrac{\lambda + K_P\sqrt{\lambda}}{\alpha}\right)$
6	LP III	$f(x;\alpha,\lambda,m)=\dfrac{\lvert\alpha\rvert}{\Gamma(\lambda)}\left(\dfrac{e^{\alpha m}}{x^{1+\alpha}}\right)\left[\alpha\left(\ln x-m\right)\right]^{\lambda-1}$ if $\alpha>0$ then $e^m \le x < \infty$; if $\alpha<0$ then $0 \le x < e^m$, $\lambda>0$, $-\infty < m < \infty$	$X_T = 10^{m+\left[\left(\lambda+K_P\sqrt{\lambda}\right)/\alpha\right]}$

exemplifies that the catchments with contrasting hydrogeology and physical characteristics and modelled using several deterministic models. Probabilistic approach involves fitting of standard probability distributions to the recorded rainfall data. In this paper, analysis based on probabilistic approach is used.

Probability analysis is the expedient method of resolving uncertainty through working out the magnitude and frequency of damaging events, which is vital in hazard management. With the problem of engineering design being an economic one, the trick is to avoid the excess cost associated with either under or over-design (WMO, 1986). This is often achieved by a quantitative probabilistic perception of the frequency and magnitude of rainfall events. Analytical procedures for such studies involve interpreting the past record of events in terms of future probabilities of occurrence using frequency analysis.

Frequency analysis enables estimation of the probability of occurrence of a certain hydrological event of practical importance by fitting a theoretical probability distribution to one that is empirically obtained from recorded data. The three main steps involved in frequency analysis are: selection of a sample in the form of a data series that satisfies certain statistical criteria; fitting the best theoretical probability distribution to represent the sample, using the best fitting technique available for the distribution; and using the fitted distribution to make statistical inferences about the underlying population (Bobee and Ashkar, 1991).

In the present study, comparison of annual rainfall for different return periods in Tarai region using recorded daily rainfall data relating to Gadarpur and Rudrapur sites was carried out. Goodness-of-Fit (GoF) tests like Anderson Darling and Kolmogorov Smirnov statistic were employed for checking the adequacy of fitting of the statistical distributions to the recorded data. Diagnostic

TABLE 2

Parameters of six distributions for Gadarpur and Rudrapur

Site	Parameters	Distribution					
		EV I	G 2	N 2	LN 2	P III	LP III
Gadarpur	α	1058.151	102.432	1211.529	7.060	350.249	-0.029
	λ	-	11.828	340.840	0.280	1.595	98.829
	m	265.659	-	-	-	652.808	9.923
Rudrapur	α	1118.514	118.881	1283.137	7.114	296.053	-0.069
	λ	-	10.793	365.830	0.300	2.590	19.012
	m	285.136	-	-	-	516.350	8.434

(α: Scale parameter; λ: Shape parameter; m: Location parameter)

test, involving D-index statistic, was used for evaluating the applicability of appropriate distribution for estimation of annual rainfall for the region under study.

2.1. *Estimation of annual rainfall* - Rational-theoretical analyses of extreme hydrologic phenomena has led to identification of Extreme Value Type I (EV I), commonly known as Gumbel distribution, as a standard distribution for frequency analysis of recorded rainfall events. Singh *et al*. (1990) applied the EV I distribution for modelling rainfall events at Bombay. Singh (1989) expressed that the distributions of 2-parameter Gamma (G 2), 2-parameter Normal (N 2), 2-parameter Lognormal (LN 2), Pearson Type III (P III) and Log Pearson Type III (LP III) could be used as an alternative choices for statistical analysis of hydrometeorological data such as rainfall, flood, etc. This paper explores the use of six statistical distributions for estimation of annual rainfall for different return periods in Tarai region. Method of maximum likelihood (MLM) was used to determine the parameters of the distributions. Table 1 gives the probability density function (PDF) and rainfall estimator of six distributions used in the study.

In Table 1, α, λ and m are scale, shape and location parameters respectively; Y_T is a reduced variate for EV I and $Y_T = -\mathrm{Ln}\{-\mathrm{Ln}[1-(1/T)]\}$; K_T is the frequency factor corresponding to the coefficient of skewness (C_s) and $C_s = 2/\sqrt{\lambda}$ for G 2, $C_s = 0.0$ for N 2 and LN 2; K_P is the frequency factor corresponding to C_s of the original and log-transformed series of the recorded data for P III and LP III respectively (Bobee, 1975; Matalas and Wallis, 1973; Stedinger, 1980).

2.2. *Anderson Darling test* - The empirical distribution function $[F_N(x)]$ of the sample is defined by: $F_N(x)$ = (Number of observations $\leq x)/N$, $-\infty < x < \infty$. $F_N(x)$ can be seen to be a step function calculated from the data. As 'x' increases, it takes a step of height $(1/N)$. $F_N(x)$ records the proportion of the observations less than or equal to 'x'. The Anderson Darling test statistic (A^2) is defined by:

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

For a given sample of 'N' values, $Z_{(i)} = F(x_i)$, for $i = 1,2,3,\ldots.N$; and $x_1<x_2< \ldots.x_N$. The distribution of A^2 statistic doesn't depend on $F(x)$, but on the set of 'N' sample values. D'Agostino and Stephens (1986) expressed that the rejection region of A^2 statistic at the desired significance level 'η' is $A^2_C > A^2_{1-\eta}$.

2.3. *Kolmogorov Smirnov test* - The test statistic (K) is defined by

$$K = \mathop{\mathrm{Max}}_{i=1}^{N}[F_e(x_i) - F_D(x_i)] \quad (2)$$

where, $F_e(x_i) = (i-0.35)/N$ is empirical cumulative distribution function of x_i, $F_D(x_i)$ is the computed

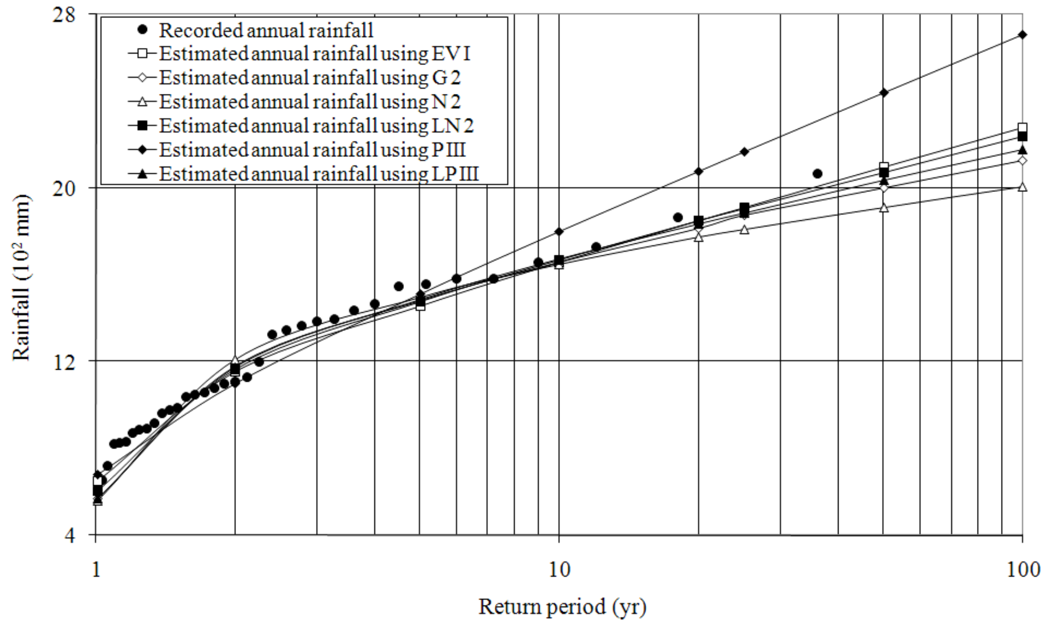

Fig. 1. Probability plot of recorded and estimated annual rainfall for different return periods using six distributions for Gadarpur

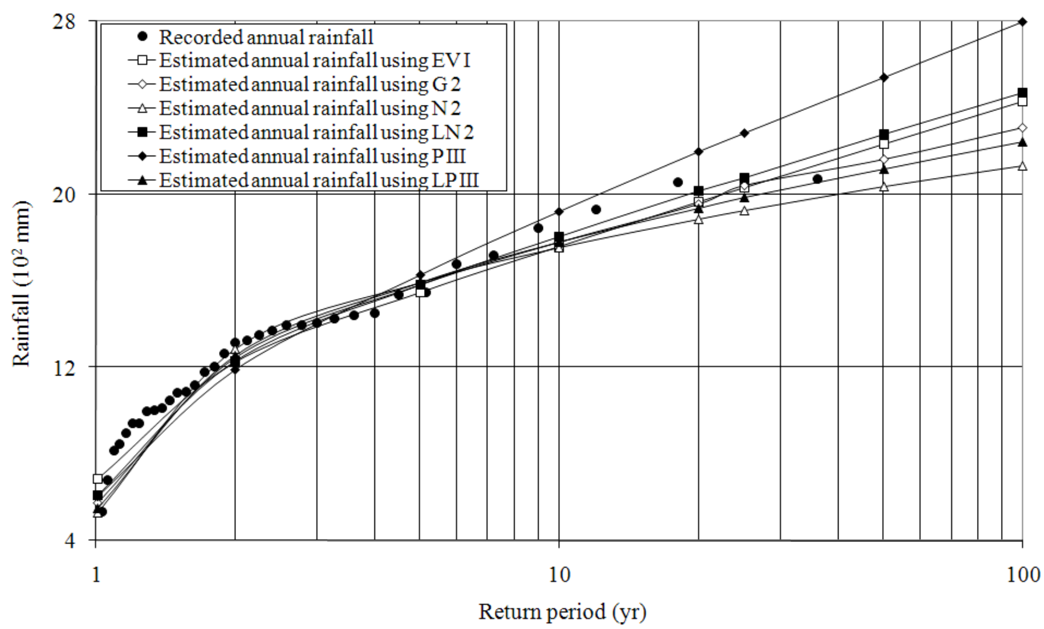

Fig. 2. Probability plot of recorded and estimated annual rainfall for different return periods using six distributions for Rudrapur

TABLE 3

GoF test statistics of six distributions for Gadarpur and Rudrapur

GoF test statistics	Site	Computed values of GoF test statistics for					
		EV I	G 2	N 2	LN 2	P III	LP III
A^2	Gadarpur	0.477	0.331	0.562	0.354	0.693	0.339
	Rudrapur	0.440	0.209	0.288	0.256	0.748	0.194
K	Gadarpur	0.123	0.099	0.138	0.106	0.122	0.097
	Rudrapur	0.120	0.087	0.096	0.103	0.123	0.081

TABLE 4

D-index values of six distributions for Gadarpur and Rudrapur

Site	D-index for					
	EV I	G 2	N 2	LN 2	P III	LP III
Gadarpur	0.208	0.164	0.356	0.170	0.617	0.186
Rudrapur	0.432	0.269	0.480	0.296	0.500	0.326

cumulative distribution function of x_i. Horn (1977) expressed that the rejection region of K statistic at the desired significance level 'η' is $K_C > K_{N,1-\eta}$.

If the computed values (A_C^2 and K_C) of GoF test statistics of the distribution are less than that of theoretical value at the desired significance level 'η' then the selected distribution is accepted to be adequate than any other distribution.

2.4. *Diagnostic test* - A qualitative assessment of the goodness of fit is ascertainable from the probability plot of the recorded and estimated rainfall data. For quantitative assessment in the upper tail region, D-index is used (USWRC, 1981) as a diagnostic statistic. D-index in upper tail level is given as:

$$\text{D-index} = (1/\bar{x})\sum_{i=1}^{6}\left|x_i - \hat{x}_i\right| \quad (3)$$

where, x_i and $\hat{x}_i$ are the i^{th} highest observed and estimated rainfall by different distributions, and $\bar{x}$ is the series mean of the recorded data. Essentially, the D-index gives weightage to the upper six data points only, rather than the data points at lower levels, with the designated objective of checking the suitability for modelling the rainfall events for extrapolation. The distribution having the minimum value for D-index is considered as the better method for estimation of rainfall.

2.5. *Data used* - Daily rainfall data in respect of Gadarpur and Rudrapur Tehsil of Udam Singh Nagar located in Tarai region of the Uttaranchal for the period 1968-2002 (35 years) was used to estimate annual rainfall for different return periods adopting six distributions.

3. A computer program was developed and used to fit the recorded rainfall data to the six distributions described earlier. The program computes the parameters of the distributions, annual rainfall estimates for different return periods, GoF (A^2 and K) and diagnostic (D-index) test statistics. Table 2 gives the parameters of six distributions for the data under study. The parameters were further used to estimate annual rainfall for different return periods of 2, 5, 10, 20, 50, 100, 200, 500 and 1,000

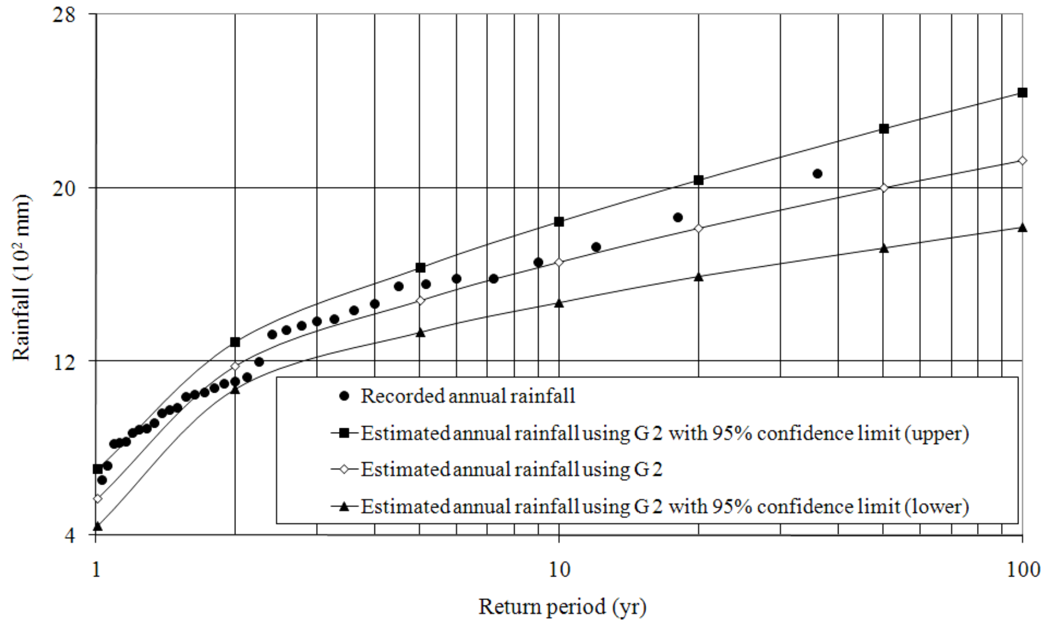

Fig. 3. Probability plot of recorded and estimated annual rainfall using G 2 with 95 percent confidence limits for different return periods of Gadarpur

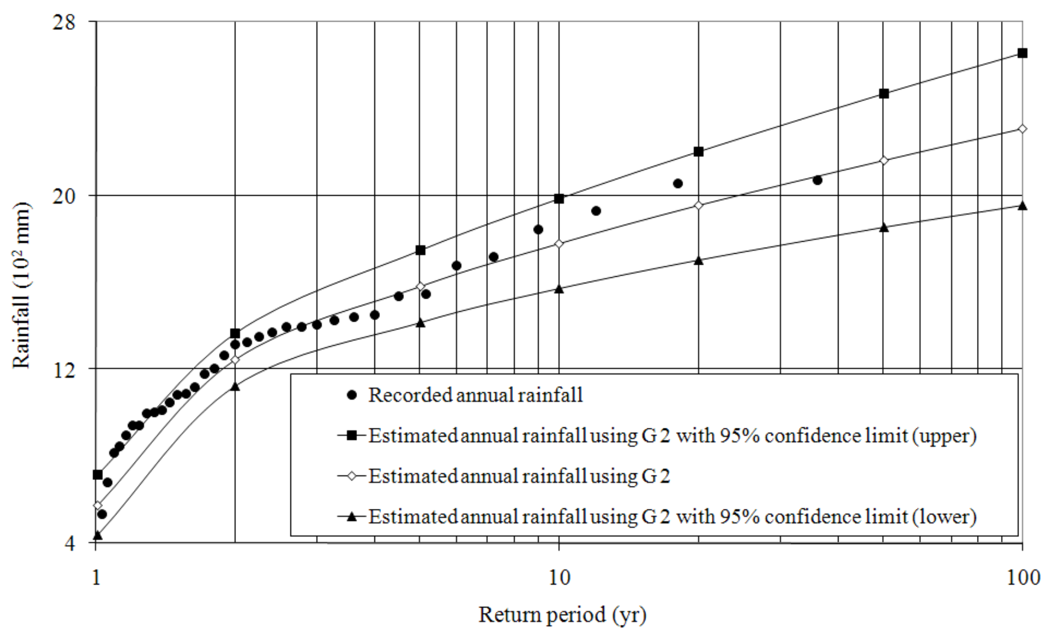

Fig. 4. Probability plot of recorded and estimated annual rainfall using G 2 with 95 percent confidence limits for different return periods of Rudrapur

years at Gadarpur and Rudrapur sites and are given in Figs. 1 and 2 respectively. It is observed that the annual rainfall estimates for different return periods from 5 to 100-year (yr) obtained using P III are relatively higher when compared with the corresponding estimates of other distributions for both the data sets.

Qualitative analysis shows that the fitted curve using P III shows convergence in the upper tail region, indicating a better fit as regards historical rainfall events for Gadarpur and Rudrapur. For the assessment on fitting of statistical distributions to the rainfall data, GoF and diagnostic tests were carried out. GoF test statistics of six distributions were computed by using Eqns. 1 and 2, and are given in Table 3.

From Table 3, it may be noted that the computed values of A^2 statistic of all six distributions are lesser than the theoretical value of 0.757 at five percent level of significance, and hence at this level, these six distributions are accepted to fit the rainfall data recorded at Gadarpur and Rudrapur sites. Also, from Table 3, it may be noted that the computed values of K statistic of six distributions are not greater than the theoretical value of 0.230 at five percent level of significance, and at this level, all six distributions are fitted well to the rainfall data recorded at the respective sites. For quantitative analysis, D-index was computed by using Eqn. (3) for six distributions and is given in Table 4.

From Table 4, it may be noted that the values of D-index of G 2 are minimum when compared with the corresponding indices of other five distributions for both the data sets. So, from the results of the quantitative assessment, G 2 is considered as the best among six distributions for estimation of annual rainfall for the data under study. Figs. 3 and 4 give the probability plots of recorded and estimated annual rainfall for different return periods using G 2 together with lower and upper confidence limits at 95 percent level for Gadarpur and Rudrapur sites respectively.

From the results of the analysis, it is observed that the highest recorded annual rainfall of 2,067 mm at Gadarpur falls within the range of 1,725 mm to 2,273 mm, which are the lower and upper confidence limits for 50-yr return period by G 2. Similarly, the highest annual rainfall of 2,071 mm recorded at Rudrapur falls within the range of 1,853 mm to 2,469 mm, which are the lower and upper confidence limits for 50-yr return period by G 2. Also, from Figs. 3 and 4, it can be seen that about 98 percent of the recorded rainfall of both data sets falls within the confidence limits of estimated annual rainfall given by G 2. The results showed that the estimated 50-yr return period rainfall of 1,999 mm for Gadarpur is about 4 percent less than the highest recorded annual rainfall. On the other hand, the estimated 50-yr return period rainfall of 2,164 mm is about 5 percent more than the highest annual rainfall recorded at Rudrapur. By considering the variation of magnitude in the recorded and estimated annual rainfall, it is suggested that 100-yr return period rainfall obtained using G 2 may be considered as the expected annual rainfall in Tarai region. So, the estimated annual rainfall for once in 100-yr for Gadarpur and Rudrapur, obtained using G 2 are 2,129 mm and 2,307 mm respectively. The study showed that the theoretical curve using G 2 distribution gives a comparatively better fit when extrapolation is involved.

4. The paper presents the results of a study for estimation of annual rainfall for Gadarpur and Rudrapur sites in Tarai region by six statistical distributions. The paper also details the results of GoF (A^2 and K) and diagnostic (D-index) tests. The study shows the 2-parameter Gamma (G 2) distribution appears to be the best suited among six distributions for estimation of annual rainfall for different return periods for the region under study. The study gives the 100-yr return period rainfall obtained using G 2 to be considered as the expected annual rainfall in Gadarpur and Rudrapur sites of Tarai region.

References

Bobee, B., 1975, "The Log Pearson Type 3 Distribution and its Application in Hydrology", *Water Resources Research*, **11**, 681-689.

Bobee, B. B. and Ashkar, F., 1991, "The Gamma Family and Derived Distributions Applied in Hydrology", *Water Recourses Publication*, 40-45.

D'Agostino, B. Ralph and Stephens, A. Michael, 1986, "Goodness-of-Fit Techniques", Marcel Dekkar, Inc., 270 Madison Avenue, New York 10016, USA, 145-148.

Horn, S. D., 1977, "Goodness of Fit Tests for Discrete Data; A Review and an Application to a Health Impairment Scale", *Biometrics*, **33**, 237-248.

Matalas, N. C. and Wallis, J. R., 1973, "Eureka: It fits a Pearson Type 3 Distribution", *Water Resources Research*, **9**, 281-289.

Singh, N., 1989, "Statistical Distribution of Extremes of Rainfall at Colaba, Bombay", **40**, 85-90.

Singh, N., Krishnakumar, K. and Abraham, O., 1990, "The EV_1 Distribution for Modelling Extreme Rain Events at Bombay", *Mausam*, **41**, 427-432.

Stedinger, J.R., 1980, "Fitting Log Normal Distribution to Hydrological Data", *Water Resources Research*, **16**, 481-490.

United States Water Resources Council (USWRC), 1981, "Guidelines for Determining Flood Flow Frequency", Bulletin No. 17B, 15-19.

World Meteorological Organization (WMO), 1986, "Manual for Estimation of Probable Maximum Flood", Report No. 332, 22-29.

N. VIVEKANANDAN

Central Water and Power Research Station, Pune, India
(*7 January 2008, Modified 13 April 2009*)
e mail : anandaan@rediffmail.com