Rainfall Frequency Analysis using Order Statistics Approach of Extreme Value Distributions

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

ABSTRACT: Estimation of rainfall for a desired return period is of utmost importance for planning, design and management of the hydraulic structures in the project site. This can be achieved by fitting of probability distributions to the series of annual 1-day maximum rainfall. This paper illustrates the use of extreme value distributions for estimation of rainfall for Fatehabad, Hansi, Hissar and Tohana stations. Order Statistics Approach is used for determination of parameters of the extreme value 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 data. A diagnostic test of D-index is used for the selection of a suitable probability distribution for rainfall estimation. Based on GoF and diagnostic tests results, the study shows the Gumbel distribution is better suited for rainfall estimation for the stations under study.

Keywords **-** Frechet, Gumbel, Kolmogorov-Smirnov, D-index, Order Statistics, Return period

1. INTRODUCTION

Technical and engineering appraisal of large infrastructure projects such as dams, bridges, barrages, culverts, etc needs to be carried out during the planning and formulation stages of such projects. In a hydrological context, it is well recognised that whatsoever extreme the design-loading, more severe conditions are likely to be encountered in nature [1]. For the reason, Extreme Value Analysis (EVA) of recorded rainfall relating to the geographical region where the project is located is a basic requirement for assessing such phenomena, and arriving at structural and other design parameters for the project. Depending on the size, life time and design criteria of the structure, different return periods are generally stipulated for adopting EVA results. Atomic Energy Regulatory Board (AERB) [2] guidelines indicated that the 1000-year (yr) return period Mean+SE (where Mean denotes the estimated rainfall and SE the Standard Error) is considered to arrive at the design rainfall depth that a structure must withstand during its lifetime. For arriving at such design values, a standard procedure is to analyse historical rainfall data over a period of time (yr) and arrive at statistical estimates.

In probabilistic theory, the extreme value distributions (EVDs) include Generalised Extreme Value (GEV), Gumbel, Frechet and Weibull are generally adopted for modelling the rainfall and stream flow data [3]. EVDs arise as limiting distributions for the sample of independent, identically distributed random variables, as the sample size increases. In addition to the above, Atomic Energy Regulatory Board (AERB) guidelines described that the Order Statistics Approach (OSA) can also be adopted for determination of parameters of Gumbel and Frechet distributions because of the OSA estimators are unbiased and having minimum variance. In this paper, GEV and Weibull distributions are not considered for EVA of rainfall due to non-existence of OSA for determination of distributional parameters. Number of studies has been carried out by different researchers on EVA of rainfall and the reports indicated that there is no unique distribution is available for modelling the rainfall data for a region or country [4-6]. This paper describes the procedures involved in rainfall estimation adopting Gumbel and Frechet probability distributions (using OSA) for Fatehabad, Hansi, Hissar and Tohana stations. Goodness-of-Fit (GoF) tests such as Anderson-Darling (A^2) and Kolmogorov-Smirnov (KS) are applied for checking the adequacy of fitting of the distributions to the recorded rainfall data. A diagnostic test of D-index is used for the selection of a suitable probability distribution for estimation of rainfall. The methodology adopted in determining the parameters of Gumbel and Frechet distributions (using OSA), computation of GoF tests statistic and D-index are briefly described in the ensuing sections.

2. METHODOLOGY

2.1 Probability Distributions

The Cumulative Distribution Functions (CDFs) of Gumbel and Frechet distributions are given by:

$$F(R) = e^{-e^{-\left(\frac{R_G-\alpha_G}{\beta_G}\right)}},\quad \alpha_G, \beta_G>0 \text{ (Gumbel)} \qquad \ldots (1)$$

$$F(R) = e^{-\left(\frac{R_F}{\beta_F}\right)^{(-\lambda_F)}},\quad \alpha_F, \beta_F>0 \text{ (Frechet)} \qquad \ldots (2)$$

Here, α_G and β_G are the location and scale parameters of Gumbel distribution. The rainfall estimates (R_G) adopting Gumbel distribution are computed from

$$R_G = \alpha_G + Y_T\beta_G \text{ and } Y_T = -\ln(-\ln(1-(1/T)))$$

Similarly, β_F and λ_F are the scale and shape parameters of Frechet distribution. Based on extreme value theory, Frechet distribution can be transformed to Gumbel distribution through logarithmic transformation [7]. Under this transformation, the rainfall estimates (R_F) adopting Frechet distribution are computed from $R_F = \mathrm{Exp}(R_G)$, $\beta_F = \mathrm{Exp}(\alpha_G)$ and $\lambda_F = 1/\beta_G$.

2.2 Theoretical Description of OSA

OSA is based on the assumption that the set of extreme values constitutes a statistically independent series of observations. The parameters of Gumbel distribution are given by:

$$\alpha_G = r^*\alpha_M^* + r'\alpha_M^{'} ; \beta_G = r^*\beta_M^* + r'\beta_M^{'} \qquad \ldots (3)$$

Here, r^* and $r^{'}$ are proportionality factors, which can be obtained from the selected values of k, n and n′ using the relations $r^* = kn/N$ and $r^{'} = n'/N$. Also, N is the sample size contains basic data that are divided into k sub groups of n elements each leaving n′ remainders; and N can be written in the form of N=kn+ n′.

In OSA, α_M^* and β_M^* are the distribution parameters of the groups and $\alpha_M^{'}$ and $\beta_M^{'}$ are the parameters of the remainders, if any. These can be computed from the following equations:

$$\alpha_M^* = (1/k)\sum_{i=1}^{n}\alpha_{ni}S_i \text{ and } \alpha_M^{'} = \sum_{i=1}^{n'}\alpha_{n'i}R_i \qquad \ldots (4)$$

$$\beta_M^* = (1/k)\sum_{i=1}^{n}\beta_{ni}S_i \text{ and } \beta_M^{'} = \sum_{i=1}^{n'}\beta_{n'i}R_i \qquad \ldots (5)$$

Here, $S_i = \sum_{i=1}^{k} R_{ij}$, j=1,2,3,..,n. R_i is the i^{th} observation in the remainder group having n′ elements, R_{ij} is the i^{th} observation in the j^{th} group having n elements. Table 1 gives the weights of α_{ni} and β_{ni} used in determination of parameters of Gumbel and Frechet distributions [8].

Table 1. Weights of α_{ni} and β_{ni} for determination of distributional parameters

α_{ni} (or) β_{ni}	i					
	1	**2**	**3**	**4**	**5**	**6**
α_{2i}	0.91637	0.08363				
α_{3i}	0.65632	0.25571	0.08797			
α_{4i}	0.51099	0.26394	0.15368	0.07138		
α_{5i}	0.41893	0.24628	0.16761	0.10882	0.05835	
α_{6i}	0.35545	0.22549	0.16562	0.12105	0.08352	0.04887
β_{2i}	-0.72135	0.72135				
β_{3i}	-0.63054	0.25582	0.37473			
β_{4i}	-0.55862	0.08590	0.22392	0.24879		
β_{5i}	-0.50313	0.00653	0.13046	0.18166	0.18448	
β_{6i}	-0.45927	-0.03599	0.07319	0.12672	0.14953	0.14581

The parameters are used to estimate the expected rainfall for different return periods. The standard error (SE) on the estimated rainfall is computed by:

$$SE = [\mathrm{Var}(R_T)]^{1/2} \text{ and } \mathrm{Var}(R_T) = r^*R_n + r'R_{n'} \qquad \ldots (6)$$

$$r^* = \frac{1}{k}\left(\frac{kn}{N}\right)^2 \text{ and } r' = \left(\frac{n'}{N}\right)^2 \qquad \ldots (7)$$

Here, R_T denotes the estimated rainfall by either R_G or R_F. R_n and $R_{n'}$ are defined by the general form as $R_n = \left(A_n Y_T^2 + B_n Y_T + C_n\right)\beta_G^2$. The values of A_n, B_n, and C_n are given in Table 2.

Table 2. Variance determinators for R_n

n	A_n	B_n	C_n
2	0.71186	-0.12864	0.65955
3	0.34472	0.04954	0.40286
4	0.22528	0.06938	0.29346
5	0.16665	0.06798	0.23140
6	0.13196	0.06275	0.19117

2.3 Goodness-of-Fit Tests

The adequacy of fitting of probability distributions to the recorded rainfall is tested by GoF tests using A^2 and KS statistic, which are defined by:

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

$$KS = \underset{i=1}{\overset{N}{\mathrm{Max}}}(F_e(R_i) - F_D(R_i)) \quad \ldots (9)$$

Here, $Z_i = F_e(R_i) = (i-0.44)/(N+0.12)$ is the empirical CDF of R_i with $R_1<R_2<R_3,\ldots,<R_N$ and $F_D(R_i)$ is the computed CDF of R_i [9]. If the computed values of GoF tests statistic given by the distribution are less than that of theoretical value at the desired significance level, then the distribution is found to be acceptable for modelling the rainfall data.

2.4 Diagnostic Test

The selection of a suitable distribution for rainfall estimation is performed through D-index, which is defined by:

$$\text{D-index} = (1/\overline{R})\sum_{i=1}^{6}\left|R_i - R_i^*\right| \quad \ldots (10)$$

Here, $\overline{R}$ is the average value of the series of the recorded rainfall, $R_i{}'s$ (for i=1 to 6) are the six highest values in the series of recorded rainfall and R_i^* is the estimated rainfall by probability distribution. The distribution having the least D-index is considered as the better suited distribution for rainfall estimation [10].

3. APPLICATION

An attempt has been made to fit the series of annual 1-day maximum rainfall (AMR) recorded at Fatehabad, Hani, Hissar and Tohana rain-gauge stations adopting Gumbel and Frechet distributions (using OSA). Daily rainfall data recorded at Fatehabad and Hansi for the period 1954-2011, Hissar for the period 1969-2011 and Tohana for the period 1951-2011 are used. The series of AMR is derived from the daily rainfall data and further used for EVA. Table 3 gives the summary statistics of the AMR recorded at the stations under study.

Table 3. Summary statistics of AMR

Station	Average (mm)	SD (mm)	CV (%)	Skewness	Kurtosis
Fatehabad	62.0	31.2	50.2	0.850	0.473
Hansi	62.4	51.0	81.7	2.225	4.868
Hissar	93.8	56.4	60.1	1.631	2.320
Tohana	73.2	39.6	54.1	0.932	0.119
SD: Standard Deviation; CV: Coefficient of Variation					

4. RESULTS AND DISCUSSIONS

By applying the procedures described above, a computer program was developed and used to fit the AMR recorded at Fatehabad, Hansi, Hissar and Tohana stations. The program computes the OSA estimators of the distributions, GoF tests statistic and D-index values. Tables 4 and 5 give the rainfall estimates together with standard error for different return periods obtained from Gumbel and Frechet distributions for the stations under study.

Table 4. Estimated rainfall (mm) together with standard error (mm) using Gumbel and Frechet distributions for Fatehabad and Hansi

Return period (yr)	Fatehabad				Hansi			
	Gumbel		Frechet		Gumbel		Frechet	
	R_G	SE	R_F	SE	R_G	SE	R_F	SE
2	57.5	3.4	50.4	3.5	56.9	4.3	45.3	3.7
5	82.1	5.4	82.2	9.3	88.5	6.9	80.7	10.9
10	98.5	7.1	113.6	17.2	109.4	9.1	118.3	21.4

20	114.1	8.9	154.9	29.7	129.5	11.4	170.7	39.3
50	134.4	11.2	231.4	57.4	155.6	14.3	274.5	82.1
100	149.6	13.0	312.7	91.5	175.0	16.6	391.9	138.7
200	164.8	14.7	422.1	143.0	194.5	18.9	558.6	229.9
500	184.7	17.1	626.9	252.7	220.1	21.9	891.8	438.7
1000	199.8	18.9	845.3	383.8	239.5	24.2	1270.0	706.0

Table 5. Estimated rainfall (mm) together with standard error (mm) using Gumbel and Frechet distributions for Hissar and Tohana

Return period (yr)	Hissar				Tohana			
	Gumbel		Frechet		Gumbel		Frechet	
	R_G	SE	R_F	SE	R_G	SE	R_F	SE
2	85.3	7.0	74.7	6.9	66.8	4.8	58.1	4.8
5	129.5	11.2	129.9	19.6	102.6	7.6	105.2	14.1
10	158.7	14.8	187.4	38.1	126.4	10.0	155.8	28.2
20	186.7	18.4	266.3	69.1	149.1	12.5	227.3	52.2
50	223.0	23.3	419.6	142.0	178.6	15.7	370.3	110.4
100	250.2	27.0	590.0	237.0	200.7	18.2	534.0	188.4
200	277.3	30.7	828.5	388.1	222.7	20.7	768.9	315.3
500	313.0	35.6	1296.6	728.6	251.7	24.1	1243.9	609.3
1000	340.0	39.3	1818.9	1158.3	273.6	26.6	1789.2	990.1

From Tables 4 and 5, it may be noted that the estimated rainfall using Frechet distribution is relatively higher than the corresponding values of Gumbel for the return periods of 5-yr and above for the stations under study.

4.1 Rainfall Frequency Curves (RFCs)

The parameters of Gumbel and Frechet distributions were used to develop the plots of CDFs and presented in Figure 1. Similarly, the rainfall estimates obtained from Gumbel and Frechet distributions were used to develop the RFCs and presented in Figures 2 and 3. From these figures, it can be seen that the RFCs using Gumbel distribution are in the form of linear whereas the RFCs using Frechet distribution are in the form of exponential for the stations under study.

4.2 Analysis Based on GoF Tests

The values of A^2 and KS statistic for the series of AMR adopting Gumbel and Frechet distributions were computed from Eqs. (8-9) and given in Table 6.

Table 6. Computed and theoretical values of A^2 and KS statistic

Station	A^2			KS		
	Computed values		Theoretical values at 5% level	Computed values		Theoretical values at 5% level
	Gumbel	Frechet		Gumbel	Frechet	
Fatehabad	0.615	2.497	0.777	0.076	0.148	0.175
Hansi	2.401	0.894	0.777	0.124	0.094	0.175
Hissar	0.976	0.941	0.780	0.100	0.145	0.203
Tohana	0.596	1.082	0.776	0.088	0.137	0.171

From A^2 test results, it may be observed that the Gumbel distribution is acceptable for modelling the AMR of Fatehabad and Tohana whereas the KS test supports the use of both Gumbel and Frechet distributions for all four stations under study. The A^2 test results also didn't support the use of Frechet distribution for modelling the AMR recorded at the four stations.

4.3 Analysis Based on Diagnostic Test

For the selection of a suitable probability distribution, D-index values of the distributions were computed by Eq. (10) and given in Table 7. From the diagnostic test results, it may be noted that the D-index values given by Gumbel distribution are comparatively minimum when compared to the corresponding values of Frechet for Fatehabad and Tohana. Also, from Table 7, it may be noted that the D-index values of Frechet are minimum when compared to the corresponding values of Gumbel for Hansi and Hissar. But, the A^2 test results didn't support the use of Frechet distribution for modelling the AMR of Hansi and Hissar.

Table 7. D-index values of Gumbel and Frechet distributions

Station	Indices of D-index	
	Gumbel	Frechet
Fatehabad	1.395	3.028
Hansi	6.069	3.857
Hissar	2.577	2.221
Tohana	1.562	7.006

Based on GoF and diagnostic tests results, the study showed that the Gumbel distribution is better suited for estimation of rainfall though KS test supports the use of Gumbel and Frechet distributions for modelling the AMR recorded at the stations under study.

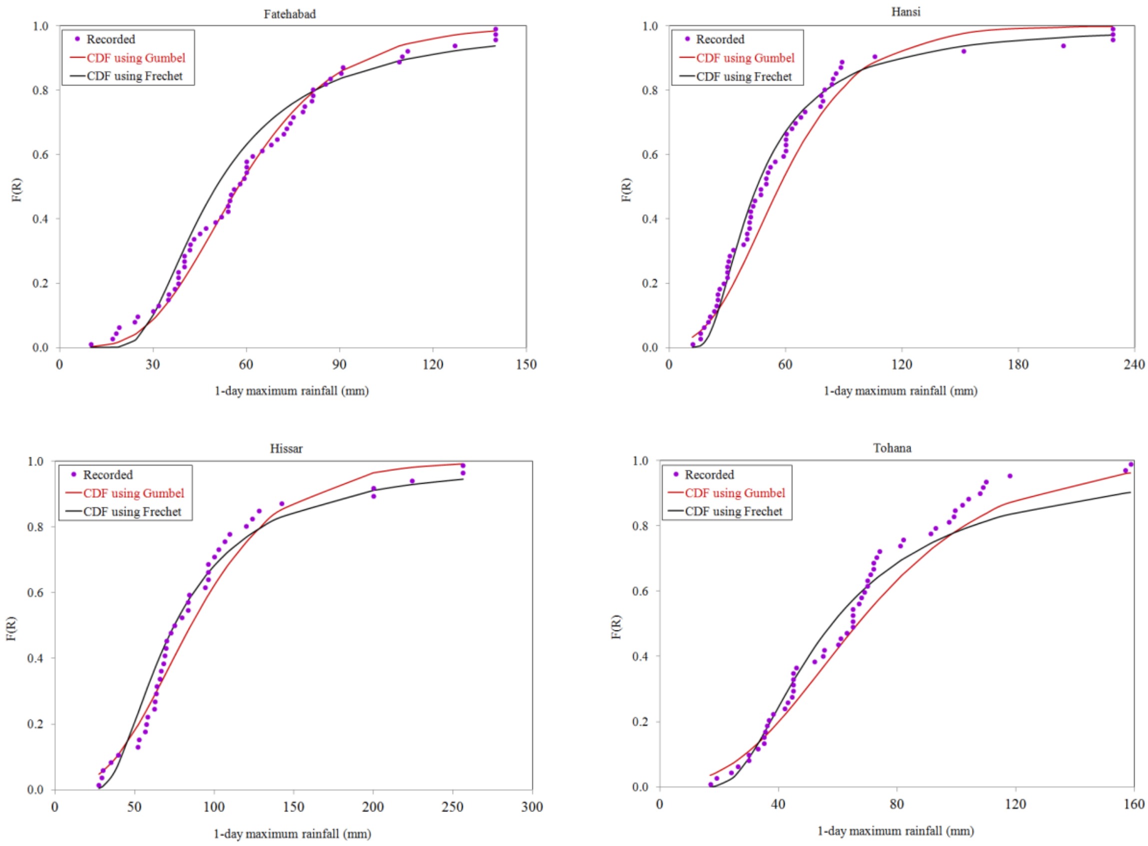

Figure 1: Plots of CDFs using Gumbel and Frechet distributions

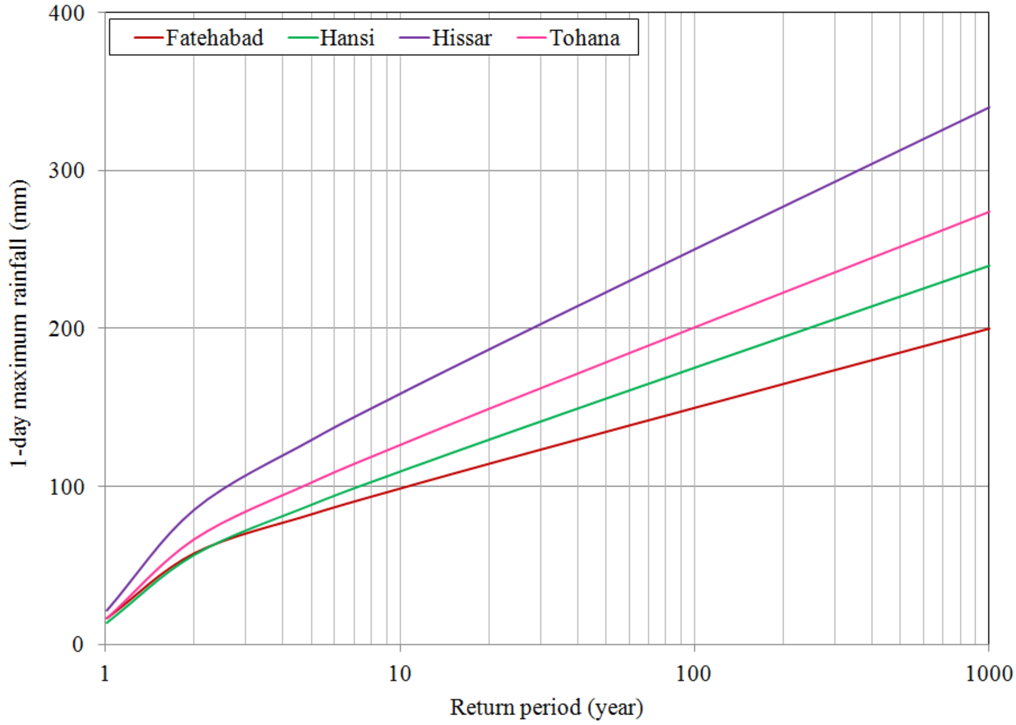

Figure 2. Rainfall frequency curves using Gumbel distribution

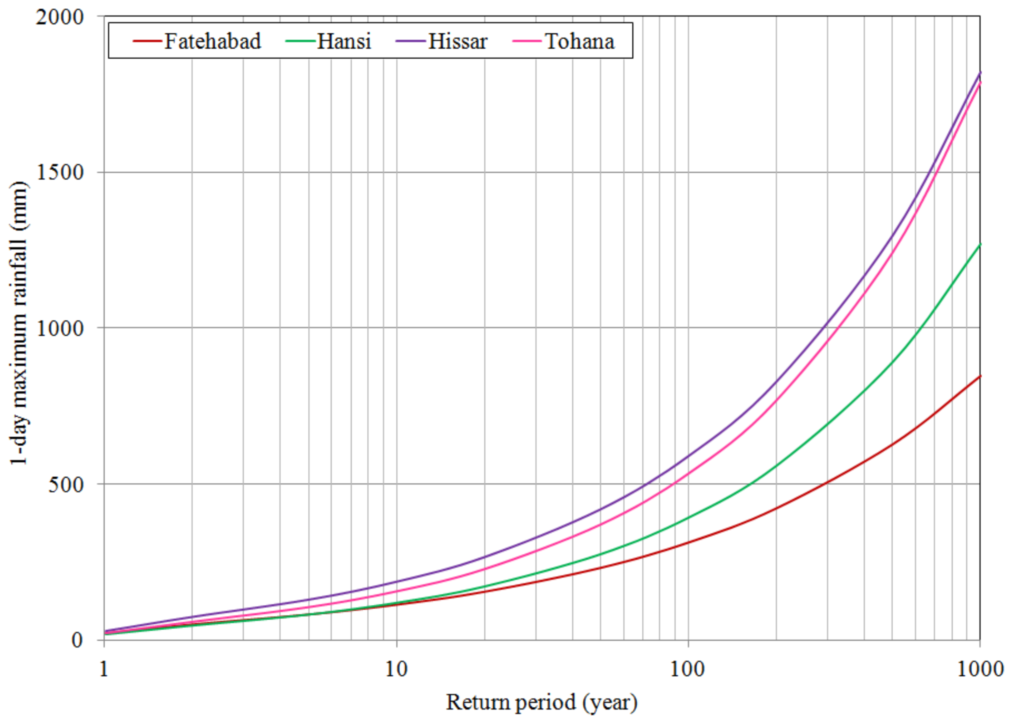

Figure 3. Rainfall frequency curves using Frechet distribution

5. CONCLUSIONS

The paper presented a computer aided procedure for estimation of rainfall for Fatehabad, Hansi, Hissar and Tohana. The A^2 test results supported the use of Gumbel distribution for modelling AMR of Fatehabad and Tohana whereas KS test results confirmed the use of both Gumbel and Frechet distributions for all four stations. Based on GoF and diagnostic tests results, the study showed that the Gumbel distribution is better suited for estimation of rainfall for the stations under study. The study showed that the 1000-yr return period Mean+SE (where Mean denotes the estimated rainfall and SE the Standard Error) values of about 219 mm for Fatehabad, 264 mm for Hani, 379 mm for Hissar and 300 mm for Tohana given by Gumbel distribution (using OSA) could be adopted while planning and design of hydraulic structures in the respective stations.

ACKNOWLEDGMENTS

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 M/s Nuclear Power Corporation of India Limited, Mumbai and India Meteorological Department, Pune, for the supply of rainfall data.

REFERENCES

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

[2] S.R. Bhakar, A.K. Bansal, N. Chhajed and R.C. Purohit, Frequency analysis of consecutive days maximum rainfall at Banswara, Rajasthan, India, *ARPN journal of Engineering and Applied Sciences*, *1(1)*, 2006, 64-67.

[3] Atomic Energy Regulatory Board (AERB) (2008), Extreme values of meteorological parameters (Guide No. NF/SG/ S-3).

[4] L.S. Esteves, Consequences to flood management of using different probability distributions to estimate extreme rainfall, *Environmental Management*, *115(1)*, 2013, 98-105.

[5] F. Uboldi, A.N. Sulis, C. Lussana, M. Cislaghi and M. Russo, A spatial bootstrap technique for parameter estimation of rainfall annual maxima distribution, *Hydrology and Earth System Sciences Discussions*, *10(10)*, 2013, 11755–11794.

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

[7] N. Mujere, Flood frequency analysis using the Gumbel distribution, *Computer Science and Engineering*, *3(7)*, 2011, 2774-2778.

[8] A.H. Ang and W.H. Tang, Probability concepts in engineering planning and design, Vol. 2, John Wiley & Sons, 1984.

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

[10] United States Water Resources Council (USWRC) (1981), Guidelines for determining flood flow frequency, Bulletin No.17B, 15-19.