© 2015 IJSRST | Volume 1 | Issue 2 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science

Quantitative Assessment on Fitting of Gumbel and Frechet Distributions for Extreme Value Analysis of Rainfall

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

ABSTRACT

Rainfall frequency analysis plays an important role in hydrologic and economic evaluation of water resources projects. It helps to estimate the return periods and their corresponding event magnitudes thereby creating reasonable design criteria. Depending on the size, life time and design criteria of the structure, different return periods are generally stipulated for adopting Extreme Value Analysis (EVA) results. This paper illustrates the use of quantitative assessment on fitting of Gumbel (EV1) and Frechet (EV2) probability distributions to the series of annual 1-day maximum rainfall (AMR) data using Goodness-of-Fit (GoF) and diagnostic tests. Order Statistics Approach (OSA) is used for determination of parameters of the distributions. Based on GoF (using Anderson-Darling and Kolmogorov-Smirnov) and diagnostic (using D-index) test results, the study identifies the EV1 distribution is better suited for EVA of rainfall for Fatehabad and Hissar.

Keywords: Anderson-Darling, D-Index, Frechet, Gumbel, Kolmogorov-Smirnov, Rainfall

I. INTRODUCTION

Estimation of rainfall for a desired return period is a pre-requisite for planning, design and operation of various hydraulic structures such as dams, bridges, barrages and storm water drainage systems. Depending on the size, life time and design criteria of the structure, different return periods are generally stipulated for adopting Extreme Value Analysis (EVA) results. For arriving at such design values, a standard procedure is to analyse historical annual 1-day maximum rainfall (AMR) 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 (EV1), Frechet (EV2) and Weibull (EV3) is generally adopted for EVA of rainfall [1-3]. EVDs arise as limiting distributions for the sample of independent, identically distributed random variables, as the sample size increases. Out of number of parameter estimation methods, Order Statistics Approach (OSA) is applied for determination of distributional parameters because of the OSA estimators are having minimum variance. In this paper, GEV and EV3 distributions are not considered for EVA of rainfall due to non-existence of OSA for determination of distributional parameters. Number of studies carried out different researchers illustrated that there is no unique distribution is available for EVA of rainfall for a region or country [4-10]. This apart, when different distributions are used for estimation of rainfall, a common problem is encountered as regards the issue of best model fits for a given set of data. This can be answered by quantitative assessment using Goodness-of-Fit (GoF) and diagnostic tests; and the results are quantifiable and reliable [11].

For quantitative assessment on rainfall within in the recorded range, Anderson-Darling (A^2) and Kolmogorov-Smirnov (KS) tests are applied for checking the adequacy of fitting of EV1 and EV2 distributions to the series of AMR data. A diagnostic test of D-index is used for the selection of suitable probability distribution for estimation of rainfall. In this paper, quantitative assessment on fitting of EV1 and EV2 probability distributions is made to identify the best suitable distribution for estimation of rainfall for Fatehabad and Hissar.

II. METHODS AND MATERIALS

The Cumulative Distribution Functions (CDFs) of EV1 and EV2 distributions are expressed by:

$$F(R)=e^{-e^{-\left(\frac{R_G-\alpha_G}{\beta_G}\right)}},\ \alpha_G,\beta_G>0 \text{ (for EV1)} \qquad \dots (1)$$

$$F(R)=e^{-\left(\frac{R_F}{\beta_F}\right)^{(-\lambda_F)}},\ \alpha_F,\beta_F>0 \text{ (for EV2)} \qquad \dots (2)$$

Here, α_G and β_G are the location and scale parameters of EV1 distribution. The rainfall estimates (R_G) adopting EV1 distribution are computed from $R_G=\alpha_G+Y_T\beta_G$ with $Y_T=-\ln(-\ln(1-(1/T)))$. Similarly, β_F and λ_F are the scale and shape parameters of EV2 distribution. Based on extreme value theory, EV2 distribution can be transformed to EV1 distribution through logarithmic transformation. Under this transformation, the rainfall estimates (R_F) adopting EV2 distribution are computed from $R_F=\mathrm{Exp}(R_G)$, $\beta_F=\mathrm{Exp}(\alpha_G)$ and $\lambda_F=1/\beta_G$ [12].

Theoretical Descriptions of OSA

OSA is based on the assumption that the set of extreme values constitutes a statistically independent series of observations. The parameters of EV1 distribution are given by:

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

where 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 $r'=n'/N$. Here, N is the sample size of the 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:

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

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

where $S_i=\sum_{i=1}^{k}R_{ij}$, j=1,2,3,..,n. Here, 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 the distributions [13]. The parameters are further used to estimate the rainfall for different return periods. The Standard Error (SE) on the estimated rainfall is computed by:

$$\left.\begin{array}{l} SE=[Var(R_T)]^{1/2} \text{ and } Var(R_T)=r^*R_n+r'R_{n'} \\ r^*=\frac{1}{k}\left(\frac{kn}{N}\right)^2 \text{ and } r'=\left(\frac{n'}{N}\right)^2 \end{array}\right\} \qquad \dots (6)$$

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

TABLE 1
WEIGHTS OF α_{ni} AND β_{ni} FOR COMPUTATION 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

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

Goodness-of-Fit Tests

The adequacy of fitting of probability distributions to the series of recorded AMR is evaluated by quantitative assessment using GoF tests statistic. Theoretical description of A^2 test statistic is as follows:

$$A^2=(-N)-(1/N)\sum_{i=1}^{N}\left\{\begin{array}{l}(2i-1)\ln(Z_i)\\+(2N+1-2i)\ln(1-Z_i)\end{array}\right\} \qquad \dots (7)$$

Here $Z_i=F(R_i)$ for i=1,2,3,…,N with $R_1<R_2<\dots<R_N$, $F(R_i)$ is the CDF of i^{th} sample (R_i) and N is the sample size. The theoretical value (A_C^2) of A^2 statistic for different sample size (N) at 5% percent significance level is computed from $A_C^2=0.757(1+(0.2/\sqrt{N})$.

The KS statistic is defined by:

$$KS = \underset{i=1}{\overset{N}{Max}} \; (F_e(R_i) - F_D(R_i)) \qquad \ldots (8)$$

Here $F_e(X_i)$ is the empirical CDF of X_i and $F_D(X_i)$ is the computed CDF of X_i (Zhang, 2002). The theoretical value KS statistic for different sample size (N) at 5% significance level is available in the technical note on 'Goodness-of-Fit Tests for Statistical Distributions book' [14].

Test criteria: If the computed values of GoF tests statistic given by probability distribution are less than that of theoretical values at the desired significance level then the distribution is considered to be acceptable for EVA of rainfall at that level.

Diagnostic Test

The selection of a suitable probability distribution for EVA of rainfall is performed through D-index test [12], which is defined as below:

$$\text{D-index} = \left(1/\overline{R}\right)\sum_{i=1}^{6}\left|R_i - R_i^*\right| \qquad \ldots (9)$$

Here, $\overline{R}$ is the average value of the recorded data whereas R_i and R_i^* are the highest recorded and corresponding estimated values by EV1 and EV2. The distribution having the least D-index is considered as better suited distribution for rainfall estimation [15].

III. APPLICATION

In this paper, a study was carried out to estimate the rainfall for different return periods for Fatehabad and Hissar adopting EV1 and EV2 distributions (using OSA). Daily rainfall data recorded at Fatehabad for the period 1954 to 2011 and Hissar for the period 1969 to 2011 was used. From the scrutiny of the daily rainfall data, it was observed that the data for the intermittent period for Fatehabad and Hansi (1966 and 1967) and Hissar (2002) are missing. So, the AMR for the missing years were imputed by the series maximum value of 140 mm (for Fatehabad) and 256.5 mm (for Hissar) in accordance with Atomic Energy Regulatory Board guidelines and used for EVA. Table 3 gives the descriptive statistics of AMR recorded at Fatehabad and Hissar.

TABLE 3
DESCRIPTIVE STATISTICS OF AMR

Region	Descriptive statistics			
	$\overline{R}$ (mm)	SD (mm)	Skewness	Kurtosis
Fatehabad	61.2	28.0	0.571	0.266
Hissar	90.0	51.0	1.674	2.909
SD: Standard Deviation				

IV. RESULTS AND DISCUSSIONS

By applying the procedures as described above, a computer program was developed and used to fit the AMR recorded at Fatehabad and Hissar. The program computes the rainfall estimates for different return periods adopting EV1 and EV2 distributions (using OSA), GoF tests statistic and D-index values. Table 4 gives the rainfall estimates (ER) together with Standard Error (SE) adopting EV1 and EV2 distributions for the stations under study. From Table 4, it may be noted that the estimated rainfall by EV2 distribution is relatively higher than the corresponding values of EV1 for Fatehabad and Hissar.

TABLE 4
ESTIMATED RAINFALL WITH STANDARD ERROR ADOPTING EV1 AND EV2 DISTRIBUTIONS (USING OSA) FOR FATEHABAD AND HISSAR

Return period (yr)	Estimated rainfall (mm) with standard error (mm) for							
	Fatehabad				Hissar			
	EV1		EV2		EV1		EV2	
	ER	SE	ER	SE	ER	SE	ER	SE
2	57.5	3.4	50.4	3.5	85.3	7.0	74.7	6.9
5	82.1	5.4	82.2	9.3	129.4	11.2	129.9	19.6
10	98.5	7.1	113.6	17.2	158.7	14.8	187.4	38.1
20	114.1	8.9	154.9	29.7	186.7	18.4	266.3	69.1
50	134.4	11.2	231.4	57.4	223.0	23.3	419.6	142.0
100	149.6	13	312.7	91.5	250.2	27.0	590.0	237.0
200	164.8	14.7	422.1	143.0	277.3	30.7	828.5	388.1
500	184.7	17.1	626.9	252.7	313.0	35.6	1296.6	728.6
1000	199.8	18.9	845.3	383.8	340.0	39.3	1818.9	1158.3
2000	214.9	20	1139.8	435.2	367.0	43.7	2551.5	2226.2
5000	234.9	23.1	1691.9	980.6	402.7	48.0	3990.5	3294.1
10000	249.9	24.9	2281.1	1453.4	429.7	51.8	5597.0	5112.9

Analysis Based on GoF Tests

For quantitative assessment on fitting of EV1 and EV2 distributions to the recorded AMR data, GoF tests statistic values were computed from Eqs. (7) and (8), and given in Table 5.

TABLE 5
COMPUTED AND THEORETICAL VALUES OF GOF TESTS STATISTIC

Region	A^2			KS		
	Computed values		Theoretical value at 5% level	Computed values		Theoretical value at 5% level
	EV1	EV2		EV1	EV2	
Fatehabad	0.599	2.433	0.777	0.050	0.131	0.175
Hissar	0.947	0.913	0.780	0.070	0.122	0.203

From the GoF tests results given in Table 5, it may be noted that the KS test confirmed the use of EV1 and EV2 distributions (using OSA) for EVA of rainfall (Fatehabad and Hissar). Similarly, A^2 test confirmed the use of EV1 distribution for EVA of rainfall for Fatehabad. As regards EVA of rainfall for Hissar, A^2 test suggested the EV1 and EV2 distributions were not acceptable.

Analysis Based on Diagnostic Test

For the selection of a suitable probability distribution, D-index values of EV1 and EV2 distributions are computed from Eq. (9) and given in Table 6. From the results, it may be noted that the D-index values of EV1 distribution are minimum when compared with the corresponding values of EV2 for the stations under study.

TABLE 6
D-INDEX VALUES OF EV1 AND EV2

Region	D-index	
	EV1	EV2
Fatehabad	1.373	4.844
Hissar	2.138	4.104

Based on quantitative assessment using GoF and diagnostic tests, the study showed that the EV1 distribution is better suited for estimation of rainfall for Fatehabad and Hissar. Figures 1 and 2 give the plots of recorded and estimated rainfall using EV1 (OSA) with confidence limits at 84.13 percentage level for Fatehabad and Hissar.

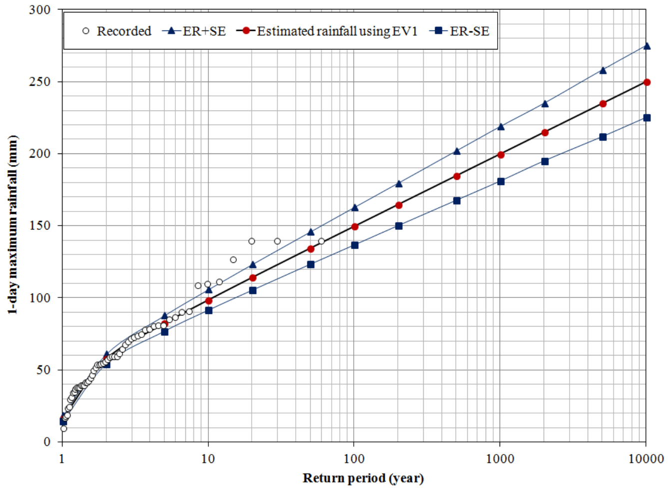

FIGURE 1: RECORDED AND ESTIMATED 1-DAY MAXIMUM RAINFALL USING EV1 (OSA) DISTRIBUTION WITH 84.13 PERCENT LOWER AND UPPER CONFIDENCE LIMITS FOR FATEHABAD

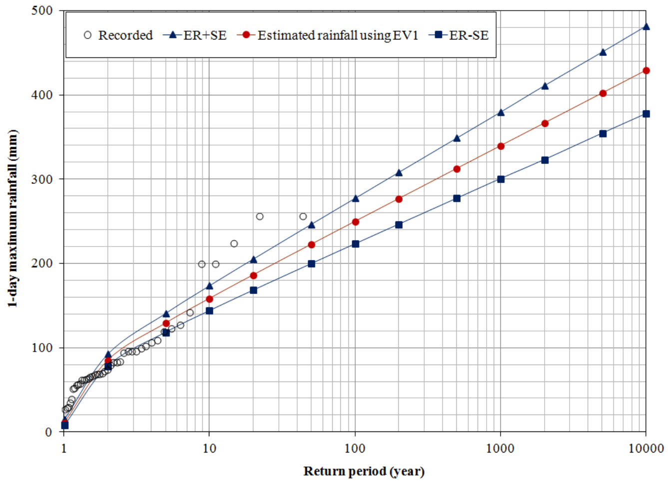

FIGURE 2: RECORDED AND ESTIMATED 1-DAY MAXIMUM RAINFALL USING EV1 (OSA) DISTRIBUTION WITH 84.13 PERCENT LOWER AND UPPER CONFIDENCE LIMITS FOR HISSAR

V. CONCLUSIONS

The paper presented the procedures involved in quantitative assessment on fitting of EV1 and EV2 distributions (using OSA) for EVA of rainfall for Fatehabad and Hissar. The KS test results confirmed the fitting of EV1 and EV2 distributions to the series of AMR recorded at the stations under study. The A^2 test results suggested the use of EV1 distribution for EVA of rainfall for Fatehabad. The diagnostic analysis showed that the EV1 distribution is better suited for estimation of rainfall for Fatehabad and Hissar. By considering the design-life of the structure over the entire intended economic lifetime, the 10000-yr return period Mean+SE (where Mean denotes the estimated rainfall and SE the Standard Error) values of about 275 mm (for Fatehabad) and 482 mm (for Hissar) computed from EV1 (OSA) distribution were suggested for design purposes.

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

REFERENCES

[1] E.J. Gumbel, Statistic of Extremes, 2nd Edition, Columbia University Press, New York, 1960.

[2] J.A. Carta and P. Ramirez, "Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions", Journal of Renewable Energy, Vol. 32, No.3, pp. 518-531, 2007.

[3] N. Mujere, "Flood frequency analysis using the Gumbel distribution", Journal of Computer Science and Engineering, Vol. 3, No. 7, pp. 2774-2778, 2011.

[4] M.C. Casas, R. Rodriguez, M. Prohom, A. Gazquez, and A. Redano, "Estimation of the probable maximum precipitation in Barcelona (Spain)", Journal of Climatology, Vol. 31, No. 9, pp. 1322–1327, 2011.

[5] E. Baratti, A. Montanari, A. Castellarin, J.L. Salinas, A. Viglione, and A. Bezzi, "Estimating the flood frequency distribution at seasonal and annual time scales", Hydrological Earth System Science, Vol. 16, No. 12, pp. 4651–4660, 2012.

[6] A. Peck, P. Prodanovic and S.P. Simonovic, "Rainfall intensity duration frequency curves under climate change: city of London, Ontario, Canada", Canadian Water Resources Journal, Vol. 37, No. 3, pp. 177–189, 2012.

[7] L.S. Esteves, "Consequences to flood management of using different probability distributions to estimate extreme rainfall", Journal of Environmental Management, Vol. 115, No. 1, pp. 98-105, 2013.

[8] 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)", Journal of Engineering and Technology, Vo1. 3, No.2, pp. 94-98, 2013.

[9] V. Rahmani, S.L. Hutchinson, J.M.S. Hutchinson and A. Anandhi, "Extreme daily rainfall event distribution patterns in Kansas," Journal of Hydrologic Engineering, Vol. 19, No. 4, pp. 707–716, 2014.

[10] Zhanling Li, Zhanjie Li, Wei Zhao, and Yuehua Wang, "Probability Modeling of Precipitation Extremes over Two River Basins in Northwest of China," Advances in Meteorology, Vol. 13, Article ID 374127, pp. 1-13, 2015.

[11] J. Zhang, "Powerful goodness-of-fit tests based on the likelihood ratio", Journal of Royal Statistical Society, Vol. 64, No. 2, pp. 281-294, 2002.

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

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

[14] P.E. Charles Annis, Goodness-of-Fit Tests for Statistical Distributions, [http://www.statistical engineering.com/goodness.html], 2009.

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