Assessment of Probable Maximum Precipitation Using Gumbel Distribution and Hershfield Method

N. Vivekanandan and Dr. S.K. Roy

Abstract---*Assessment of Probable Maximum Precipitation (PMP) has utmost importance for planning, design, management and risk analysis of hydraulic and other structures in a region. This paper details the procedures involved in estimation of Extreme Rainfall (ER) for Bhavnagar region using five parameter estimation methods of Gumbel distribution. Goodness-of-Fit test involving Kolmogorov-Smirnov (KS) statistics is used for checking the adequacy of fitting of the method for determination of parameters of the distribution. Root Mean Square Error (RMSE) is used for selection of a suitable method for estimation of ER. The paper presents that the Probability Weighted Moments (PWM) is better suited for modelling daily maximum rainfall and Order Statistics Approach (OSA) for 24-hour maximum rainfall for the region under study. The results obtained from Gumbel distribution are compared with PMP value given by Hershfield method. The study shows that the Mean+SE (where Mean denotes the estimated ER and SE the standard error) value of 1000-year return period one-day ER given by PWM may be considered for design purposes for Bhavnagar region.*

Keywords---*Hershfield Method, Gumbel, Mean Square Error, Probable Maximum Precipitation*

I. INTRODUCTION

ESTIMATION of rainfall for a desired return period and different durations is often required for design and risk analysis of hydraulic and other structures in a region. Annual maximum rainfall estimates likely to occur for different return periods are very often important inputs for design purposes. These extreme events are also essential in the post commissioning stage, wherein the assessment of failure of hydraulic structures needs to be carried out. The estimation of extreme values is also an important aspect of the siting and design of nuclear power projects [1]. Research studies showed that hourly rainfall data is needed for designing the storm water drainage around the site, while the daily rainfall data are needed for generating design basis flood water level at inland sites, which are often situated near a river course or dam. Hourly rainfall data may be made use of to arrive at running average data for any desired duration. For example, if the point of interest is one-day rainfall, then the 24-hour rainfall value may be considered in the absence of daily rainfall data.

A theoretical analysis of extreme hydrologic phenomena has led researchers to identify Gumbel distribution as a standard distribution for frequency analysis of recorded extreme hydrologic events, such as rainfall, flood, temperature, evaporation, etc; and hence used in the present study [2]. A number of parameter estimation methods such as Method of Moments (MOM), Maximum Likelihood Method (MLM), Method of Least Squares (MLS), OSA and PWM are available for determination of parameters of Gumbel distribution [3].

MOM is a natural and relatively easy parameter estimation method. MLM is considered the most efficient method, since it provides the smallest sampling variance of the estimated parameters and hence of the estimated quantiles compared to other methods. But, the method has the disadvantage of frequently giving biased estimates and often failed to give the desired accuracy in estimating extremes from hydrological data. It may produce quality estimators in small samples, especially when the random variable is restricted to an interval that depends on the parameters [4-5]. PWM is much less complicated, and the computations are simpler. Parameter estimates from small samples using PWM are sometimes more accurate than the ML estimates for Gumbel distribution [6-7].

In this paper, MOM, MLM, MLS, OSA and PWM are used for determination of parameters of Gumbel distribution for estimation of ER for different return periods for Bhavnagar region. Goodness-of-Fit test involving KS test is employed for checking the adequacy of fitting of the method to the recorded rainfall data. RMSE is used to evaluate the performance of the estimated annual daily and 24-hr maximum rainfall using five methods of Gumbel distribution for Bhavnagar. According to AERB guidelines, the Mean+SE (which indicates the upper confidence limit will represent the value that will not be exceeded by 84.13 % of the events having a desired return period) value is generally considered in arriving at a design value. In the present study, the Mean+SE values of 1000-year (yr) return period ERs (based on hourly and daily rainfall data) given by five methods of Gumbel distribution are compared with the values obtained from Hershfield method to arrive at design parameter for Bhavnagar region. The paper presents the procedures involved in determination of parameters of Gumbel distribution (using MOM, MLM, MLS, OSA and PWM) for estimation of ER, estimation of PMP by Hershfield method and the results obtained thereof.

N. Vivekanandan, Assistant Research Officer, Central Water and Power Research Station, Pune 411024. E-mail: anandaan@rediffmail.com

Dr. S.K. Roy, Chief Research Officer, Central Water and Power Research Station, Pune 411024.

DOI: 10.9756/BIJDM.10149

II. METHODOLOGY

A) Probability Distribution

The probability density function [f(R)] and Cumulative Distribution Function [CDF; F(R)] of Gumbel distribution is given by:

$$f(R) = \frac{e^{-(R_i-\alpha)/\beta} e^{-e^{-(R_i-\alpha)/\beta}}}{\beta}, R_i, \beta > 0 \quad \ldots (1)$$

$$F(R) = e^{-e^{-(R_i-\alpha)/\beta}} \quad \ldots (2)$$

where α and β are location and scale parameters [8]. The parameters are computed by MOM, MLM, MLS, OSA and PWM; and used to estimate ER (R_T) for different return periods using the relation $R_T = \alpha + Y_T\beta$ where $Y_T = -Ln(-Ln(1-(1/T)))$.

A1) Determination of Parameters of Gumbel Distribution

Method of Moments

$$\alpha = \overline{R} - 0.5772157\beta \text{ and } \beta = \left(\sqrt{6}/\pi\right)S_R \quad \ldots (3)$$

where $\overline{R}$ and S_R are the mean and standard deviation of the recorded rainfall data.

Maximum Likelihood Method

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

$$\alpha = -\beta \log\left[\sum_{i=1}^{N} \exp(-R_i/\beta) \Big/ N\right] \quad \ldots (4)$$

Method of Least Squares

$$\beta = \left(\left(\sum_{i=1}^{N} R_i\right)^2 - \left(N\sum_{i=1}^{N} R_i^2\right)\right) \Big/ \left(\left(N\sum_{i=1}^{N} R_i(\ln(-\ln(P_i)))\right) - \left(\sum_{i=1}^{N} R_i\right)\left(\sum_{i=1}^{N} \ln(-\ln(P_i))\right)\right) \quad \ldots (5)$$

$$\alpha = \overline{R} + \left(\left(\sum_{i=1}^{N} \ln(-\ln(P_i))\right)\beta \Big/ N\right) \quad \ldots (6)$$

where P_i=(i-0.44)/(N+0.12) and $\ln(-\ln(P_i))$ defines the cumulative probability of non-exceedance for each R_i.

Order Statistics Approach

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

$$\alpha = r^*\alpha_M^* + r'\alpha_M'; \quad \beta = r^*\beta_M^* + r'\beta_M' \quad \ldots (7)$$

where r^* and r' are proportionality factors, which can be obtained from the selected values of k, n and n' using the relations as follows:

$$r^* = kn/N \text{ and } r' = n'/N \quad \ldots (8)$$

Here N is the sample size containing the basic data that are divided into k sub groups of n elements each leaving n' remainders. α_M^* and β_M^* are the distribution parameters of the groups, and α_M' and β_M' are the parameters of the remainders, if any. These can be computed from the following equations:

$$\alpha_M^* = (1/k)\sum_{i=1}^{n} \alpha_{ni}S_i; \; \alpha_M' = \sum_{i=1}^{n'} \alpha_{n'i}R_i;$$

$$\beta_M^* = (1/k)\sum_{i=1}^{n} \beta_{ni}S_i \text{ and } \beta_M' = \sum_{i=1}^{n'} \beta_{n'i}R_i \quad \ldots (9)$$

where $S_i = \sum_{i=1}^{k} R_{ij}$, j=1,2,3,..,n. The values of the weights α_{ni} and β_{ni} are given in Table 1 [9].

Probability Weighted Moments

$$\alpha = M_{100} - 0.5772157\beta \text{ and}$$

$$\beta = (M_{100} - 2M_{101})/\ln 2 \quad \ldots (10)$$

where $M_{100} = \overline{R}$ and $M_{101} = \sum_{i=1}^{N} R_i(N-i)/(N(N-1))$. Here 'i' is the rank assigned to each sample arranged in ascending order and N is the sample size.

Table 1: Weights α_{ni} and β_{ni} for Computation of OSA Estimators

α_{ni}/ β_{ni}	i					
	1	2	3	4	5	6
α_{2i}	0.9164	0.0836				
α_{3i}	0.6563	0.2557	0.0880			
α_{4i}	0.5109	0.2639	0.1537	0.0714		
α_{5i}	0.4189	0.2463	0.1676	0.1088	0.0584	
α_{6i}	0.3555	0.2255	0.1656	0.1211	0.0835	0.0489
β_{2i}	-0.7214	0.7214				
β_{3i}	-0.6305	0.2558	0.3747			
β_{4i}	-0.5586	0.0859	0.2239	0.2488		
β_{5i}	-0.5031	0.0065	0.1305	0.1817	0.1845	
β_{6i}	-0.4593	-0.0359	0.0732	0.1267	0.1495	0.1458

A2) Computation of Standard Error

The values of SE of the estimated ER by MOM, MLM, MLS and PWM may be computed from Eq. (11) and given by:

$$SE = \left(\beta/\sqrt{N}\right)\left(A + BY_T + CY_T^2\right)^{0.5} \quad \ldots (11)$$

The coefficients used in computation of SE by MOM, MLM, MLS and PWM are given in Table 2.

Table 2: Coefficients Used in Computation of SE by MOM, MLM, MLS and PWM

Parameter Estimation Method	Coefficients used in computation of SE		
	A	B	C
MOM and MLS	1.1589	0.1919	1.1000
PWM	1.1128	0.4574	0.8046
MLM	1.1087	0.5140	0.6079

The SE of the estimated ER by OSA can be obtained from

$$SE = \left(r^{*}W_{n} + r^{'}W_{n'}\right)^{1/2} \quad \dots (12)$$

where $r^{*} = (1/k)(kn/N)^{2}$ and $r^{'} = (n'/N)^{2}$. W_n and $W_{n'}$ are defined by general form as $W_n = \left(A_n Y_T^2 + B_n Y_T + C_n\right)\beta^2$. The weights α_{ni} and β_{ni} used in determining OSA estimators, and values of A_n, B_n, and C_n used in computing the SE for OSA, are given in AERB safety guide [10-11].

B) Goodness-of-Fit Test

The KS statistics is defined by:

$$KS = \underset{i=1}{\overset{N}{Max}}\left(F_e(R_i) - F_D(R_i)\right) \quad \dots (13)$$

Here $F_e(R_i)$=(i-0.44)/(N+0.12) is the empirical CDF of R_i and $F_D(R_i)$ is the computed CDF of R_i. If the computed value of KS statistics by the method is less than that of critical value at the desired significance level, then the method is acceptable for determination of parameters of Gumbel distribution [12].

C) Performance Analysis

The performance of estimated ER using five methods of Gumbel distribution is analyzed by using RMSE and given by

$$RMSE = SQRT\left(\frac{1}{N}\sum_{i=1}^{N}\left(R_i - R_i^{*}\right)^2\right) \quad \dots (14)$$

where R_i and R_i^{*} are the recorded and estimated annual daily/ 24-hour (hr) maximum rainfall of i[th] event [13].

III. APPLICATION

An attempt has been made to estimate PMP for Bhavnagar region using five methods of Gumbel distribution and Hershfield method. Bhavnagar is a coastal city on the eastern coast of Saurashtra, Gujarat located between 21° 77′N and 72° 15′ E and situated to the west of the Gulf of Khambhat. Daily rainfall for the period 1969 to 2005 and hourly rainfall for the period 1975 to 2005 recorded at Bhavnagar region are used [14].

IV. RESULTS AND DISCUSSIONS

A) Estimation of ER Using Gumbel Distribution

By applying the methodology as detailed above, frequency analysis of annual daily and 24-hr maximum rainfall data was carried out. Table 3 gives the statistical parameters of the series of recorded annual daily and 24-hr maximum rainfall at Bhavnagar. Tables 4 and 5 give the ER estimates for different return periods based on daily and 24-hr maximum rainfall series given by five methods of Gumbel distribution together with the SEs.

Table 3: Statistical Parameters of Annual Daily and 24-hr Maximum Rainfall for Bhavnagar

Data series	Mean (mm)	Standard deviation (mm)	Coefficient of skewness	Coefficient of kurtosis
Daily	114.1	79.8	2.216	5.836
24-hr	124.3	94.8	2.141	4.338

Table4: ER Estimates Together With SEs for Different Return Periods Based on Annual Daily Maximum Rainfall Series for Bhavnagar

Return period (year)	Estimated ER (mm) with SE using									
	MOM		MLM		MLS		OSA		PWM	
	R_T	SE	R_T	SE	R_T	SE	R_T	SE	R_T	SE
2	101.0	12.0	100.2	8.7	100.7	13.0	100.6	10.0	102.4	10.7
5	171.5	20.3	151.3	13.4	177.2	22.0	159.0	16.0	165.3	17.3
10	218.2	27.4	185.2	17.2	227.9	29.7	197.7	21.2	206.9	22.7
20	263.0	34.6	217.7	21.0	276.5	37.5	234.8	26.4	246.8	28.2
50	321.0	44.2	259.8	26.1	339.4	48.0	282.9	33.4	298.5	35.5
100	364.4	51.5	291.3	30.0	386.6	55.9	318.9	38.8	337.2	41.0
200	407.7	58.8	322.8	33.9	433.6	63.8	354.8	44.1	375.8	46.6
500	464.8	68.5	364.2	39.1	495.5	74.3	402.1	51.2	426.7	54.0
1000	508.0	75.8	395.5	43.1	542.4	82.3	437.9	56.6	465.2	59.8

Table 5: ER Estimates Together With SEs for Different Return Periods Based on Annual 24-hr Maximum Rainfall Series for Bhavnagar

Return period (year)	Estimated ER (mm) with SE using									
	MOM		MLM		MLS		OSA		PWM	
	R_T	SE	R_T	SE	R_T	SE	R_T	SE	R_T	SE
2	108.7	15.6	107.4	10.7	108.1	17.4	108.4	12.2	110.7	13.7
5	192.5	26.3	164.7	16.4	201.5	29.3	173.7	19.4	184.1	22.1
10	247.9	35.5	202.7	21.0	263.3	39.6	216.9	25.6	232.6	29.0
20	301.1	44.9	239.2	25.7	322.6	50.0	258.4	31.8	279.3	36.0
50	369.9	57.3	286.4	32.0	399.3	63.9	312.1	40.2	339.6	45.3
100	421.5	66.8	321.7	36.8	456.8	74.4	352.3	46.5	384.8	52.3
200	472.9	76.3	357.0	41.6	514.1	85.0	392.4	52.9	429.9	59.4
500	540.7	88.8	403.4	47.9	589.8	99.1	445.2	61.4	489.3	68.8
1000	591.9	98.4	438.5	52.8	646.9	109.7	485.2	67.8	534.2	76.0

From Tables 4 and 5, it may be noted that the ER estimates using MLM are comparatively lower and higher while using MLS when compared to the corresponding estimates for other methods. Also, from Tables 4 and 5, it may be noted that the SEs on the estimated values are minimum, when MLM is used and maximum when MLS used.

B) Analysis Based on KS Test and RMSE

For the assessment on fitting of five methods of Gumbel distribution to the observed annual daily and 24-hr maximum rainfall data series, KS statistics were computed. The performance of the methods used for estimation of annual daily and 24-hr maximum rainfall was also judged by RMSE. The values of RMSE and KS statistics are computed from Eqs. (13-14) and given in Table 6.

Table 6: Computed Values of KS Statistics and RMSE

Data series	Computed values of KS statistics using					RMSE (mm) given by				
	MOM	MLM	MLS	OSA	PWM	MOM	MLM	MLS	OSA	PWM
Daily	0.178	0.137	0.199	0.134	0.147	24.1	31.5	25.3	27.0	25.1
24-hr	0.203	0.167	0.233	0.153	0.158	34.3	43.0	36.7	38.3	35.1

From Table 6, it may be noted that the computed values of KS statistic of all five methods are less than the theoretical values (0.224 for daily rainfall series and 0.240 for 24-hr rainfall series) at 5% level of significance, and hence at this level, all five methods are accepted to fit the data relating to annual daily and 24-hr maximum rainfall recorded at Bhavnagar. Also, from Table 6, it may be seen that the RMSE of annual daily and 24-hr maximum rainfall using MLM is relatively high when compared with the corresponding values of other four methods. The results of RMSE indicated that there is a marginal difference on RMSE computed by MOM, MLS, OSA and PWM for both data sets.

C) Estimation of PMP using Hershfield Method

PMP is the theoretically greatest depth of precipitation for a given duration that is physically possible over a particular drainage area at a certain time of year. In other words, it is the magnitude of rainfall, which will yield flood flows, of which there is virtually no risk of being exceeded. PMP can be estimated using Hershfield method as

$$PMP = \overline{R} + KS_R \quad \ldots (15)$$

The constant K for a station can be obtained by using the equation:

$$K = \left(R_{max} - \overline{R_1}\right) / S_{R1} \quad \ldots (16)$$

Where R_{max} is maximum value of the annual daily maximum rainfall data series and $\overline{R_1}$ and S_{R1} are mean and standard deviation of the series respectively excluding R_{max} [15].

For Bhavnagar region, the one-day PMP is computed as:

K= (424.8-105.5)/61.0 =5.234 and

PMP=114.1+ (79.8*5.234) =531.8mm ≅ 532 mm

The one-day PMP for Bhavnagar region given by Hershfield method is 532 mm, which is five percent less than the regional PMP value of 560 mm [16]. Table 7 gives the percentage of variation on the estimated 1000-year (yr) return period Mean+SE values of extreme daily/ 24-hr rainfall with reference to the PMP of 532 obtained from Hershfield method.

Table 7: Percentage of Variation on the Estimated ER With Reference to PMP Given by Hershfield Method

Data series	Percentage of variation on the estimated ER using				
	MOM	MLM	MLS	OSA	PWM
Daily	9.7	17.6	17.4	7.1	1.4
24-hr	29.8	7.7	42.2	4.0	14.7

From Table 7, it may be noted that the percentage of variation on the estimated 1000-yr return period Mean+SE value of ER given by PWM is 1.4% when daily rainfall series used for estimation of rainfall adopting five methods of Gumbel distribution. Also, from Table 7, it may be noted that the percentage of variation on the estimated ER given by OSA is 4% when 24-hr rainfall series used for rainfall estimation. By considering the percentage of variation in magnitude, it is suggested that PWM could be used for modelling daily maximum rainfall and OSA for 24-hr maximum rainfall for Bhavnagar region. The results showed that the 1000-yr return period Mean+SE value of one-day ER of 525 mm given by PWM is about six percent less than the regional value of 560 mm. The study suggested that the 1000-yr return period Mean+SE value of 525 mm may be considered for design purposes in Bhavnagar region.

V. CONCLUSIONS

The paper presents results of a study for assessing the frequency of ER at Bhavnagar, Gujarat adopting five parameter estimation methods of Gumbel distribution based on daily and 24-hr rainfall data. The study showed that the KS test results and RMSE values support the use of PWM for modelling daily maximum rainfall and OSA for 24-hr maximum rainfall for estimation of PMP. The results of daily rainfall data showed that the 1000-yrreturn period Mean+SE value of estimated rainfall by PWM for Bhavnagar region is 525 mm. The results of hourly rainfall data showed that the 1000-yr return period Mean+SE value of estimated rainfall by OSA is 553 mm for the region under study. By considering the percentage of variation on the estimated rainfall with reference to PMP given by Hershfield method, it is suggested that the 1000-yr return period Mean+SE value of 525 mm may be considered for design purposes in Bhavnagar region.

ACKNOWLEDGMENTS

The authors are grateful to the Director, CWPRS, Pune, for providing the research facilities to carry out the study. The authors are thankful to India Meteorological Department, Pune, for making available the rainfall data for Bhavnagar region.

REFERENCES

[1] AERB Safety Guide (2008), Extreme Values of Meteorological Parameters. (AERB/NF/SG/S-3), Atomic Energy Regulatory Board, Mumbai, India.

[2] A. Zaharim, A.M. Razali, R.Z. Abidin and K. Sopian, "Fitting of Statistical Distributions to Wind Speed Data in Malaysia", *European Journal of Scientific Research*, Vol. 26, No. 1, pp 6-12, 2009.

[3] J.A. Ranyal, and J.D. Salas, "Estimation Procedures for the Type-1 ExtremeValue Distribution",*Journal of Hydrology*,Vol. 87, Nos. 3-4, Pp 315-336, 1986.

[4] H.N. Phien, "A Review of Methods of Parameter Estimation for the ExtremeValue Type–1 Distribution",*Journal of Hydraulics*, Vol. 90, Nos. 3-4, Pp 251-268, 1987.

[5] V.P. Singh, "On Statistical Intercomparison of EVI Estimators by Monte Carlo Simulation", *Advanced Water Resources Publications*, Vol. 10, Pp 87-105, 1987

[6] J.M. Landwehr, N.C. Matalas and J.R. Wallis, "Probability Weighted Moments Compared with Some Traditional Techniques in Estimating GumbelParameters and Quantiles", *Water Resources Research*, Vol. 15, No. 5, Pp 1055-1064, 1979.

[7] P.F. Rasmussen and N. Gautam, "Alternative PWM-estimators of the Gumbel distribution", *Journal of Hydrology*, Vol. 280, Nos. 1-4, Pp 265-271, 2003.

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

[9] J. Lieblein, Note on Simplified Estimates for Type I Extreme Value Distribution, NBSIR 75-647, National Bureau of Standards, U.S. Department of Commerce, Washington DC, 1974.

[10] B.L.P. Swami, K. Seetaramulu and K.K. Chaudhry, "A Critical Review of the Methods for Correcting the Sampling Errors in the Extreme Wind Speeds", *Journal of Structural Engineering*, Vol.12, No.4, Pp 143-148, 1986

[11] J. Zhang, "Powerful Goodness-of-Fit Tests Based on the Likelihood Ratio", *Journal of Royal Statistical Society* Series-B, Vol. 64, Pp 281-294, 2002.

[12] P.P. Sreekala, S.V.B. Rao and M. Rajeevan, "Northeast Monsoon Rainfall Variability Over South Peninsular India and its Teleconnections",*Journal of Theoretical and Applied Climatology*, Vol. 108, Nos. 1-2, Pp 73-83, 2012.

[13] Y.P. Xu, C.Yu, X. Zhang, Q. Zhang and X. Xu, "Design Rainfall Depth Estimation through Two Regional Frequency Analysis Methods in Hanjiang River Basin, China",*Journal of Theoretical and Applied Climatology*, Vol. 107, Nos. 3-4, Pp 563-578, 2012.

[14] India Meteorological Department (IMD), Report on Availability of Meteorological Data, 2010.

[15] 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.

[16] M. Mohapatra, G.S. Mandal, B.K. Bandopadhyay, A. Tyagi and U.C. Mohanty, "Classification of Cyclone Hazard Prone Districts of India",*Natural Hazards*, Vol. 63, No.3, Pp 1601-1620, 2012.