

Intercomparison of Estimators of Gumbel Distribution using Goodness-of-Fit Tests for Estimation of Extreme Rainfall

R. S. Bharadwaj^{1*}, (Mrs.) A. D. Thube², N. Vivekanandan³, C. Srishailam⁴

¹M.Tech. Scholar, Department of Civil Engineering, College of Engineering, Pune, Maharashtra, India

²Associate Professor, Department of Civil Engineering, College of Engineering, Pune, Maharashtra, India

³Scientist-B, Central Water and Power Research Station, Pune, Maharashtra, India

⁴Scientist-C, Central Water and Power Research Station, Pune, Maharashtra, India

*Corresponding author's E-mail: rohan_bharadwaj90@yahoo.co.in

ABSTRACT

Estimation of extreme rainfall for a given return period is of utmost importance for planning, design and management of hydraulic structures and riverfront development projects. This can be achieved by Extreme Value Analysis (EVA) that involves fitting of Gumbel probability distribution to the series of Annual 1-day Maximum Rainfall (AMR). Standard parameter estimation procedures viz., Method of Moments (MoM), Maximum Likelihood Method (MLM) and Probability Weighted Moments (PWM) are applied for determination of parameters of the Gumbel distribution. This paper presents a study on comparison of MoM, MLM and PWM estimators of Gumbel distribution adopted in EVA of rainfall for Kalyan, Thane and Ulhasnagar sites of Ulhas river basin. Goodness-of-Fit tests viz., Anderson–Darling, Kolmogorov–Smirnov and Mean Absolute Percentage Error are used for checking the adequacy of fitting of three methods of Gumbel probability distribution to the AMR data. Based on the GoF tests results, the MLM is identified as better-suited method amongst three methods applied for determination of parameters of Gumbel distribution for estimation of extreme rainfall at Kalyan, Thane and Ulhasnagar sites

Keywords: Anderson–Darling test, Extreme Value Analysis, Gumbel distribution, Kolmogorov Smirnov test, Mean Absolute Percentage Error, Probability Weighted Moments, Rainfall

I. INTRODUCTION

Rainfall analysis for any river basin is essential for planning and design of hydraulic structures and riverfront development projects. As the distribution of rainfall varies over space and time, it is required to analyse the data covering long periods and recorded at various locations to obtain reliable information. For example, the analysis of consecutive days of rainfall is more relevant for the

drainage design of agricultural lands whereas 1-day maximum rainfall for a desired return period (T) is needed for arriving at a design parameter for planning, design and management of hydraulic structures. Likewise, weekly rainfall data analysis is relevant for planning cropping patterns whereas analysis of monthly, seasonal, and annual rainfall data is more useful for water management practices [1].

Out of a number of probability distributions, the family of Extreme Value Distributions (EVDs) includes Generalized Extreme Value, Extreme Value Type-1 (Gumbel), Extreme Value Type-2 (Frechet), Extreme Value Type-3 (Weibull) and Generalized Pareto is widely adopted for Extreme Value Analysis (EVA) of rainfall [2]. EVDs arise as limiting distributions for the sample of independent, identically distributed random variables, as the sample size increases. In the group of EVDs, Gumbel distribution has no shape parameter as when compared to other distributions and this means that there is no change in the shape of Probability Distribution Function (PDF). Moreover, the Gumbel distribution has the advantage of having only positive values, since the data series of rainfall are always positive (greater than zero); and therefore Gumbel distribution is important in statistics. Deka and Borah [3] have derived the best-fitted distribution amongst five EVDs used to describe the annual series of maximum rainfall data of nine distantly located stations in north east India. Sharma and Singh [4] analyzed the series of annual, seasonal, monthly and weekly maximum rainfall data of Pantnagar region and identified the best fitted probability distribution among the sixteen distributions used in the study. Mujere [5] applied Gumbel distribution for modelling flood data for Nyanyadzi River, Zimbabwe. Esteves [6] applied Gumbel distribution to estimate the extreme rainfall depths at different rain gauge stations in southeast United Kingdom. Vivekanandan [7] applied Gumbel distribution for modelling the seasonal and annual rainfall for Krishna and Godavari river basins. Rasel and Hossain [8] applied Gumbel distribution for development of intensity-duration-frequency curves for seven divisions in Bangladesh. Mohammed and Azhar [9] stated that conventional flood frequency analysis involving Log Normal, Gumbel and Log Pearson Type-3 distributions could be applied to estimate extreme flows. In view of the above, the Gumbel distribution is used for EVA of

rainfall. Standard parameter estimation procedures viz., Method of Moments (MoM), Maximum Likelihood Method (MLM) and Probability Weighted Moments (PWM) are generally applied for the determination of the parameters of the Gumbel distribution [10]. A number of studies have been carried out by different researchers on analysing the characteristics of parameter estimation methods for the Gumbel distribution. Research reports indicate that the MoM is a natural and relatively easy parameter estimation method [11]. MLM is considered as the most efficient method as it provides the smallest sampling variance of the estimated parameters and hence of the estimated quantiles compared to other methods [12]. Phien [13] and Swami et al. [14] studied that PWM and method of least squares are much less complicated, and the computations are simpler. Parameter estimates from small samples using PWM is sometimes more accurate than MLM estimates for the Gumbel distribution. But, there is no general agreement concerning the application of a particular method for determining the parameters of the Gumbel distribution for estimation of rainfall. In this paper, a study on comparison of MoM, MLM and PWM estimators of Gumbel distribution was carried out to identify a best suitable method for EVA of rainfall for Kalyan, Thane, Ulhasnagar sites of Ulhas river basin through Goodness-of-Fit (GoF) tests such as Anderson-Darling, Kolmogorov-Smirnov and Mean Absolute Percentage Error [15]. The procedures adopted in determining the parameters of Gumbel distribution by three methods and GoF tests are briefly described in the ensuing sections.

II. METHODOLOGY

The PDF and Cumulative Distribution Function (CDF) of Gumbel distribution is given as follows:

$$\text{PDF: } f(r) = \frac{e^{-(r-\alpha)/\beta} e^{-e^{-(r-\alpha)/\beta}}}{\beta}, \alpha, \beta > 0 \quad \dots (1)$$

$$\text{CDF: } F(r) = e^{-e^{-(r-\alpha)/\beta}} \quad \dots (2)$$

where, r is the recorded data, α is the location parameter and β is the scale parameter [16]. The parameters are computed by MoM, MLM and PWM; and used to estimate the extreme rainfall (R_T) for different return periods from $R_T = \hat{\alpha} + Y_T \hat{\beta}$ wherein $Y_T = -\text{Ln}(-\text{Ln}(1-(1/T)))$, $\hat{\alpha}$ is the estimator of location parameter and $\hat{\beta}$ is the estimator of scale parameter.

Method of Moments

$$\hat{\alpha} = \bar{R} - 0.5772157\hat{\beta} \text{ and } \hat{\beta} = (\sqrt{6}/\pi)S_R \quad \dots (3)$$

where, \bar{R} and S_R are the mean and standard deviation of the recorded data [17].

Maximum likelihood method

$$\hat{\alpha} = -\hat{\beta} \log \left[\sum_{i=1}^N \exp(-r_i/\hat{\beta}) / N \right] \text{ and}$$

$$\hat{\beta} = \bar{R} - \left[\sum_{i=1}^N r_i \exp(-r_i/\hat{\beta}) / \sum_{i=1}^N \exp(-r_i/\hat{\beta}) \right] \quad \dots (4)$$

where, $\bar{R} = \frac{1}{N} \sum_{i=1}^N r_i$ and N is the sample size .

Probability Weighted Moments

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

$$\hat{\beta} = (M_{100} - 2M_{101}) / \ln 2 \quad \dots (5)$$

where $M_{100} = \bar{R}$ 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.

Computation of Standard Error and Confidence Limits

Standard Error (SE) on the estimated extreme rainfall using MoM, MLM and PWM is computed from the following equation:

$$SE = \frac{\hat{\beta}}{\sqrt{N}} (A + BY_T + CY_T^2)^{0.5} \quad \dots (6)$$

Table 1 gives the coefficients used in computation of SE by MoM, MLM and PWM methods. The lower and upper confidence limits (LCL and UCL) of the estimated rainfall are obtained from the linear expressions viz., $LCL=ER-1.96(SE)$ and

$UCL=ER+1.96(SE)$. Here, SE is the Standard Error on the estimated ER.

Table 1: Coefficients used in computation of SE by MoM, MLM and PWM

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

Goodness-of-Fit tests

Generally, GoF tests viz., Anderson-Darling (A^2) and Kolmogorov-Smirnov (KS) tests are applied for checking the adequacy of fitting of Gumbel distribution. Theoretical descriptions of GoF tests are as follows:

A^2 test statistic is defined as below:

$$A^2 = (-N) - (1/N) \sum_{i=1}^N \left\{ \frac{(2i-1) \ln(Z_i) + (2N+1-2i) \ln(1-Z_i)}{2} \right\} \quad \dots (7)$$

Here, $Z_i = F(r_i)$ for $i=1,2,3,\dots,N$ with $r_1 < r_2 < \dots < r_N$ and $F(r_i)$ is the CDF of r_i .

KS test statistic is defined as below:

$$KS = \text{Max}_{i=1}^N |F_e(r_i) - F_D(r_i)| \quad \dots (8)$$

Here, $F_e(r_i)$ is the empirical CDF of r_i and $F_D(r_i)$ is the derived CDF of r_i by probability distribution. In this study, Weibull plotting position formula is used for computation of empirical CDF. The theoretical values of A^2 and KS tests statistic for different sample size (N) at either 5% or 1% significance level are available in the technical note on ‘Goodness-of-Fit Tests for Statistical Distributions’ by Charles Annis [18]. If the computed value of GoF test statistic given by the probability distribution is less than that of its theoretical values at the desired significance level then the distribution is assumed to be suitable for EVA of rainfall at that level of significance.

Mean Absolute Percentage Error

$$MAPE(\%) = \left(\frac{1}{N} \sum_{i=1}^N \left| \frac{r_i - r_i^*}{r_i} \right| \right) * 100 \quad \dots (9)$$

where, r_i is the recorded value of AMR of i^{th} sample and r_i^* is the estimated value of AMR of i^{th} sample. The test criteria indicates the method having minimum MAPE value is considered as better-suited method for determination of estimators of Gumbel probability distribution for EVA of rainfall.

III. APPLICATION

In this paper, a study on comparison of estimators of Gumbel distribution for estimation of extreme rainfall at Kalyan, Thane and Ulhasnagar sites of Ulhas river basin was carried out. The Ulhas River is one of the west-flowing rivers in Maharashtra that drains into the Arabian Sea. The Ulhas rises from Sahyadri hill ranges in the Raigad district of Maharashtra at an elevation of 600 meter above mean sea level. The boundary of the basin consists

of the main Sahyadri hills on the east, westerly off shoots on the north and south and on the west, a narrow opening at the end leading to the sea. The Ulhas river basin lies between the latitudes of $18^{\circ} 44' N$ to $19^{\circ} 42' N$ and longitudes of $72^{\circ} 45' E$ to $73^{\circ} 48' E$. The drainage area of Ulhas basin is $4,637 \text{ km}^2$ which lies completely in Maharashtra. The total length of the west flowing river from its origin to its outfall in to the Arabian Sea is 122 km. The average rainfall in the Ulhas basin is 2,943 mm. The basin receives 99% of rainfall from the south-west monsoon during June to October. The average maximum and minimum temperatures are recorded as $38.9^{\circ} C$ and $12.4^{\circ} C$ respectively. Figure 1 gives the location map of the study area. In this paper, the daily rainfall data observed at Kalyan, Thane and Ulhasnagar sites during the period 1986 to 2017 was used. The AMR series was extracted from the daily rainfall data and used for EVA of rainfall.

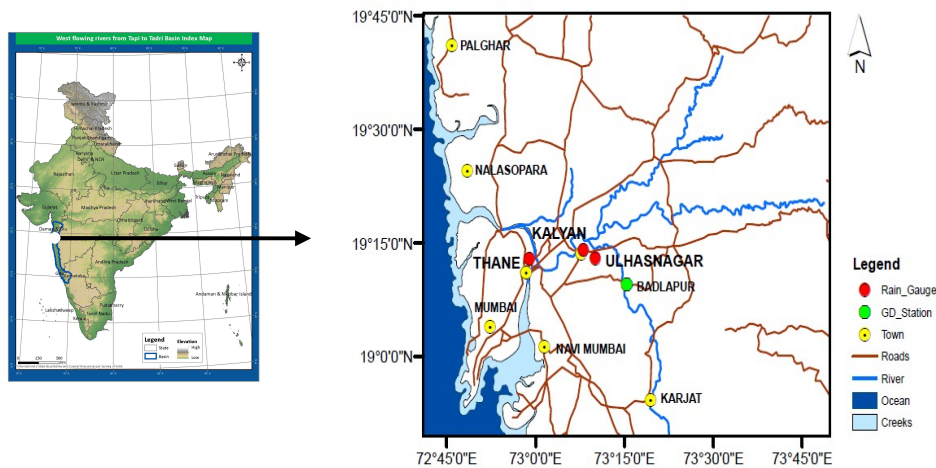


Figure 1: Location map of the study area

IV. RESULTS AND DISCUSSIONS

By applying the procedures, as described above, parameters of the Gumbel distribution were determined by MoM, MLM and PWM methods and used for EVA. The EVA results of Kalyan, Thane and Ulhasnagar sites are presented in Tables 2 to 4

respectively. From Tables 2 and 4, it may be noted that the rainfall estimates given by PWM are comparatively higher when compared with the corresponding estimates obtained from MoM and MLM for Kalyan and Ulhasnagar. Also, from Tables 2 to 4, it may be noted that the SE on the estimated rainfalls are minimum when MLM estimator is used

for estimation of extreme rainfall. The plots of estimated rainfalls by three methods of Gumbel distribution with the recorded rainfall for Kalyan, Thane and Ulhasnagar sites are developed and presented in Figures 2 to 4 respectively.

Analysis Based on GoF Tests

GoF tests such as A², KS and MAPE were used for checking the adequacy of fitting of three methods of Gumbel distribution to the AMR series of Kalyan, Thane and Ulhasnagar sites. The GoF tests values were computed from Eqs. (7 to 9) and presented in Table 5. From GoF tests results, it may be noted that the computed values of A² test statistic of MoM and

PWM methods are higher than the theoretical values i.e. 0.757 at 5% significance level and hence these two methods are not found to be acceptable to fit the AMR series recorded at Kalyan, Thane and Ulhasnagar sites. Also, A² test results indicated that the MLM method is acceptable for EVA of rainfall for the sites considered in the study. From Table 5, it may be noted that the computed values of KS test statistic by MoM, MLM and PWM methods are less than its theoretical value i.e. 0.241 at 5% significance level and hence all three methods are found to be acceptable to fit the AMR series recorded at Kalyan, Thane and Ulhasnagar sites.

Table 2. Extreme rainfall estimates with SE by three methods of Gumbel distribution for Kalyan

Return period (yr)	Estimated rainfall (mm) with SE (mm) using					
	MoM		MLM		PWM	
	R _T	SE	R _T	SE	R _T	SE
2	190.0	15.7	191.7	13.0	200.3	17.2
5	274.5	26.5	258.2	20.0	290.8	27.7
10	330.4	35.8	302.2	25.6	350.6	36.3
20	384.0	45.2	344.5	31.4	408.1	45.1
50	453.4	57.8	399.2	39.0	482.4	56.7
100	505.4	67.3	440.2	44.8	538.1	65.5
200	557.2	76.9	481.0	50.7	593.7	74.4
500	625.6	89.6	534.9	58.4	666.9	86.2
1000	677.3	99.2	575.6	64.3	722.2	95.2

Table 3. Extreme rainfall estimates with SE by three methods of Gumbel distribution for Thane

Return period (yr)	Estimated rainfall (mm) with SE (mm) using					
	MoM		MLM		PWM	
	R _T	SE	R _T	SE	R _T	SE
2	198.5	19.3	199.6	13.6	207.1	18.4
5	302.2	32.5	271.5	20.8	303.9	29.6
10	370.8	44.0	319.0	26.7	368.0	38.9
20	436.7	55.6	364.6	32.7	429.5	48.2
50	521.9	71.0	423.7	40.7	509.1	60.7
100	585.8	82.7	467.9	46.7	568.7	70.2
200	649.4	94.4	512.0	52.8	628.2	79.7
500	733.4	110.0	570.1	60.9	706.6	92.3
1000	796.8	121.8	614.1	67.1	765.8	101.9

Table 4. Extreme rainfall estimates with SE by three methods of Gumbel distribution for Ulhasnagar

Return period (yr)	Estimated rainfall (mm) with SE (mm) using					
	MoM		MLM		PWM	
	R _T	SE	R _T	SE	R _T	SE
2	151.0	13.0	152.9	9.2	162.0	13.9
5	221.1	22.0	201.7	14.2	235.3	22.4
10	267.5	29.7	234.0	18.2	283.9	29.5
20	312.0	37.5	264.9	22.2	330.4	36.5
50	369.6	48.0	305.0	27.6	390.7	46.0
100	412.8	55.9	335.1	31.8	435.9	53.1
200	455.8	63.8	365.0	35.9	480.9	60.3
500	512.5	74.3	404.5	41.4	540.3	69.9
1000	555.4	82.3	434.4	45.6	585.2	77.1

Table 5: Computed values of GoF tests by three methods of Gumbel distribution

Site	A ² test			KS test			MAPE (%)		
	MoM	MLM	PWM	MoM	MLM	PWM	MoM	MLM	PWM
Kalyan	1.937	0.502	1.383	0.190	0.097	0.139	8.9	6.4	12.1
Thane	1.573	0.715	1.397	0.183	0.095	0.133	13.7	7.8	13.2
Ulhasnagar	1.668	0.536	1.932	0.142	0.051	0.102	13.2	6.6	14.0

From Table 5, it is also noted that the MAPE given by MLM is considerably minimum when compared with the corresponding values of MoM and PWM of Gumbel distribution. Based on GoF tests results, it is identified that the MLM is better suited method for determining the estimators of Gumbel distribution that are used for estimation of extreme rainfall at Kalyan, Thane and Ulhasnagar sites. The plots of recorded and estimated extreme rainfalls by

MLM of Gumbel distribution together with 95% confidence limits for Kalyan, Thane and Ulhasnagar sites are presented in Figures 5 to 7 respectively. From Figure 5 to 7, it can be seen that the percentages of recorded rainfall values within the confidence limits of the estimated extreme rainfalls are about 75% for Kalyan whereas about 95% for Thane and about 90% for Ulhasnagar.

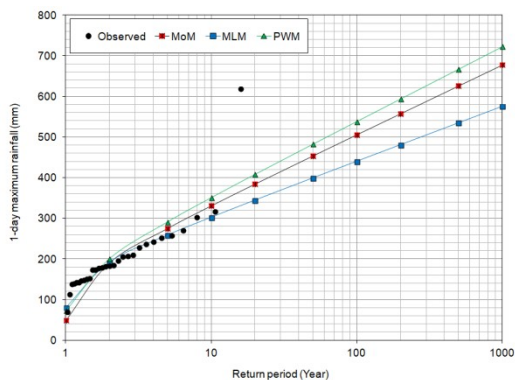


Figure 2. Plots of observed and estimated rainfall by Gumbel distribution (using MoM, MLM and PWM) for Kalyan

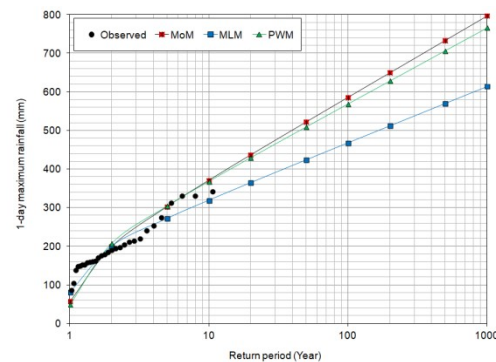


Figure 3. Plots of observed and estimated rainfall by Gumbel distribution (using MoM, MLM and PWM) for Thane

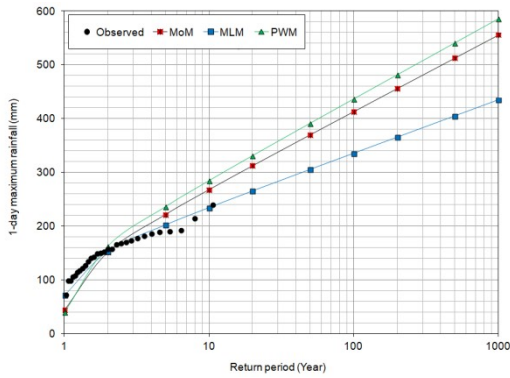


Figure 4. Plots of observed and estimated rainfall by Gumbel distribution (using MoM, MLM and PWM) for Ulhasnagar

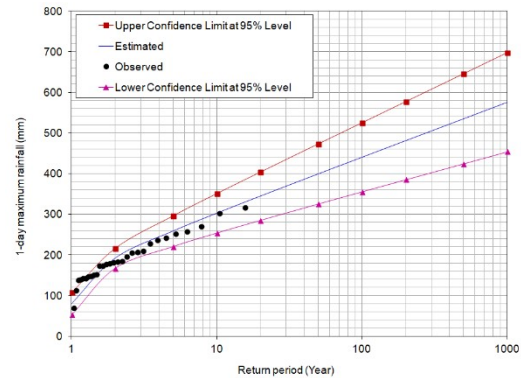


Figure 5. Plots of observed and estimated rainfall by Gumbel distribution (using MLM) with 95% confidence limits for Kalyan

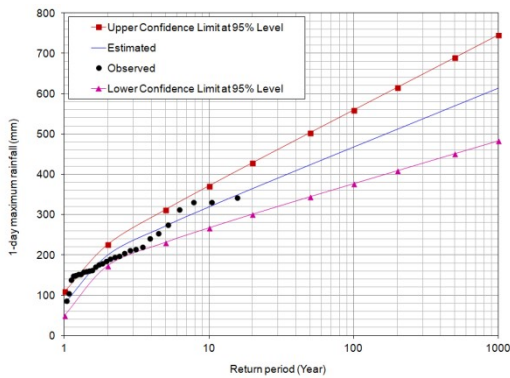


Figure 6. Plots of observed and estimated rainfall by Gumbel distribution (using MLM) with 95% confidence limits for Thane

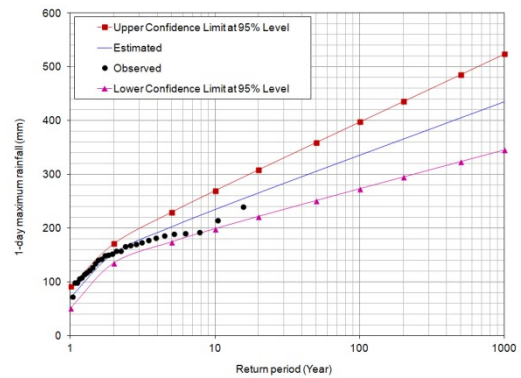


Figure 7. Plots of observed and estimated rainfall by Gumbel distribution (using MLM) with 95% confidence limits for Ulhasnagar

V. CONCLUSIONS

The paper presents the study carried out for intercomparison of MoM, MLM and PWM estimators of Gumbel distribution using GoF tests to identify the best suitable method for estimation of extreme rainfall at Kalyan, Thane and Ulhasnagar sites by adopting EVA. Based on GoF tests results, the following conclusions were drawn from the study:

- i. The estimated extreme rainfall by Gumbel (using PWM) is higher than the corresponding values obtained from MoM and MLM methods used in EVA of rainfall for Kalyan and Ulhasnagar.

- ii. A^2 test results confirmed the applicability of MLM of Gumbel distribution for EVA of rainfall.
- iii. KS test results supported the use of MoM, MLM and PWM methods of Gumbel distribution for EVA of rainfall.
- iv. MAPE values indicated that the MLM is identified as better suited method for determination of parameters of Gumbel distribution for EVA of rainfall.
- v. The study suggested that the 1000-yr return period $\text{Mean} + (1.96)\text{SE}$ (where Mean denotes the estimated ER and SE the Standard Error) values of about 702 mm at Kalyan, 746 mm at Thane and 524 mm at Ulhasnagar given by MLM of Gumbel distribution could be

considered for designing the hydraulic structures having a design life of 1000-years.

VI. ACKNOWLEDGEMENTS

The authors are grateful to Dr. B. B. Ahuja, Director, College of Engineering, Pune for his continuous encouragement during the course of study. The authors are thankful to Dr. (Mrs.) V. V. Bhosekar, Director, Central Water and Power Research Station, Pune for providing research facilities to carry out the study. The authors are also thankful to Indian Meteorological Department, Pune for the supply of rainfall data used in this paper.

VII. REFERENCES

- [1] V.N. Sharda and P.K. Das. 2005. Modelling weekly rainfall data for crop planning in a sub-humid climate of India. *Agricultural Water Management*, 76(2), 120–138.
- [2] M. C. Casas., R. Rodriguez, M. Prohom, A. Gazquez and A. Redano. 2011. Estimation of the probable maximum precipitation in Barcelona (Spain). *Journal of Climatology*, 31(9), 1322–1327.
- [3] S. Deka and M. Borah. 2009. Distribution of annual maximum rainfall series of North East India. *European Water Publications*, 27, 3–14.
- [4] M.A. Sharma and J.B. Singh. 2010. Use of probability distribution in rainfall analysis. *New York Science Journal*, 3(9), 40–49.
- [5] N. Mujere. 2011. Flood frequency analysis using the Gumbel distribution. *Journal of Computer Science and Engineering*, 3(7), 2774–2778.
- [6] L.S. Esteves. 2013. Consequences to flood management of using different probability distributions to estimate extreme rainfall. *Journal of Environmental Management*, 115(1), 98–105.
- [7] N. Vivekanandan. 2014. Modelling annual rainfall of Krishna and Godavari river basins using Extreme Value Type-1 distribution. *i-manger Journal of Structural Engineering*, 3(1), 7–12.
- [8] M.M. Rasel and S.M. Hossain. 2015. Development of rainfall intensity duration frequency equations and curves for seven divisions in Bangladesh. *International Journal of Scientific and Engineering Research*, 6(5), 96–101.
- [9] S. Mohammed and H. Azhar. 2017. Estimation of design flood at Kol dam using hydrometeorological approach. *International Journal of Environmental Sciences & Natural Resources*, 4(1), 1–6.
- [10] K. Arora and V.P. Singh. 1987. On statistical intercomparison of EVI estimators by Monte Carlo simulation. *Advances in Water Resources*, 10(2), 87–107.
- [11] J.M. Landwehr, N.C. Matalas and J.R. Wallis. 1979. Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research*, 15(5), 1055–1064.
- [12] J.A. Ranyal and J.D. Salas. 1986. Estimation procedures for the type-1 extreme value distribution. *Journal of Hydrology*, 87(3&4), 315–336.
- [13] H.N. Phien. 1987. A review of methods of parameter estimation for the extreme value type-1 distribution. *Journal of Hydraulics*, 90(3&4), 251–268.
- [14] B.L.P. Swami, K. Seetaramulu and K.K. Chaudhry. 1986. A critical review of the methods for correcting the sampling errors in the extreme wind speeds. *Journal of Structural Engineering*, 12(4), 143–148.
- [15] J. Zhang. 2002. Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society*, 64(2), 281–294.

- [16] E.J. Gumbel. 1960. Statistics of extremes (2nd edition). New York: Columbia University Press.
- [17] D. Manik and S.K. Datta. 1998. A comparative study of estimation of extreme value. Journal of River Behaviour & Control, 25(1), 41–47.
- [18] Annis Charles P.E. 2009. Goodness-of-Fit tests for statistical distributions.

Cite this article as:

R. S. Bharadwaj, (Mrs.) A. D. Thube, N. Vivekanandan, C. Srishailam, "Intercomparison of Estimators of Gumbel Distribution using Goodness-of-Fit Tests for Estimation of Extreme Rainfall", International Journal of Scientific Research in Civil Engineering (IJSRCE), ISSN : 2456-6667, Volume 3 Issue 3, pp. 38-46, May-June 2019. URL:<http://ijsrce.com/IJSRCE 19337>