Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2320-7051.7189

ISSN: 2320 – 7051

Int. J. Pure App. Biosci. **6 (6):** 1309-1316 (2018)

Research Article

International Journal of Pure & Applied Bioscience

OPEN ACCESS

Analysis of Extreme Rainfall Events and Calculation of Return Levels using Generalised Extreme Value Distribution

M. R. Namitha[1*] and V. Ravikumar[2]

[1]Assistant Professor, Department of Agriculture Engineering, Sethu Institute of Technology, Anna University

[1] Professor and Head, Department of Soil and Water Conservation and Agricultural Structures, Tamil Nadu Agricultural University

*Corresponding Author E-mail: namithamadhu93@gmail.com

Received: 6.11.2018 | Revised: 10.12.2018 | Accepted: 16.12.2018

ABSTRACT

The analysis of 27 years rainfall data of Kumulur region was conducted using two types of probability distributions, viz Gumbel distribution and generalised extreme value distribution. The method of L- moments was used for the analysis. Annual one day maximum and 2, 3,4, 5 and 7 consecutive days maximum rainfall data for 27 years was analysed and the return levels for 2, 5, 10 and 25-years were calculated using the proposed probability distribution functions. Chi-square test was conducted for comparison of the observed and expected return levels obtained using both the distributions. The statistical analysis revealed that, the annual maxima rainfall data for one day maxima and consecutive days maxima of Kumulur region fits best with the generalised extreme value distribution.

Key words: *Generalised Extreme Value distribution, Gumbel distribution, Chi-square test, L-Moments.*

INTRODUCTION

Extreme rainfall events are a primary cause of flooding hazards worldwide. Analysis of consecutive days maximum rainfall of different return periods is a basic tool for safe and economical planning and design of small dams, bridges, culverts, irrigation and drainage work etc. Though the nature of rainfall is erratic and varies with time and space, yet it is possible to predict design rainfall fairly accurately for certain return periods using various probability distributions[1].

Design Engineers and Hydrologists require one day maximum rainfall at different return periods for appropriate planning and design of small and medium hydraulic structures like small dams, bridges, culverts, etc.[2]. Analysis of consecutive day's maximum rainfall is more relevant for drainage design of agricultural lands[1,3]. Analysis of weekly rainfall data is more useful for planning cropping pattern and its management.

At present a few studies have been done in India and these studies were mainly carried out to validate the statistical procedure of different types of probability distribution function, viz., Normal, Log Normal and Gamma, keeping in view the importance of watershed development programme.

Cite this article: Namitha, M. R. and Ravikumar, V., Analysis of Extreme Rainfall Events and Calculation of Return Levels using Generalised Extreme Value Distribution, *Int. J. Pure App. Biosci.* **6(6):** 1309-1316 (2018). doi: http://dx.doi.org/10.18782/2320-7051.7189

MATERIAL AND METHODS

Location of study

AEC & RI, Kumulur campus which is located in Lalgudy taluk in Trichy district of Tamilnadu is chosen as the study area. The latitude and longitude of Kumulur is found to be 10.55’29.34”N and 78.49’35.61”E. The average annual rainfall of the area was found to be 85.8 cm.

Data collection

The daily rainfall data for past 27 years (1991-2017) was collected from the meteorological observatory in AEC & RI, Kumulur campus.

Methodology

The daily data in a particular year is converted to 2 to 5 consecutive days rainfall by summing up the rainfall of corresponding previous days. The maximum amount of one week and 2 to 5 consecutive days rainfall for each year was taken for analysis.

Table 1: Annual maximum rainfall for 1 day and 2-7 consecutive days

Sl. No.	Year	Annual Maximum Rainfall for consecutive days					
		1 day	2 days	3 days	4 days	5 days	7 days
1	1991	151	181.5	190	197.4	197.4	197.4
2	1992	85.1	137.4	172.7	196.7	225.7	235.6
3	1993	106.2	133.8	157.4	175.4	176.4	180.4
4	1994	71.8	73.4	88.2	88.2	102.8	113.8
5	1995	100	130	130	130	130	130
6	1996	122.6	122.6	137.1	137.1	151	201.8
7	1997	69.7	72.9	106.3	107.3	118.1	134.1
8	1998	120	136.3	197.2	217	233.3	263.9
9	1999	205.8	217.4	219.8	224.4	224.4	227.1
10	2000	90.5	177.0	184.4	206	213.4	213.4
11	2001	84	123	123	123	150.5	189.5
12	2002	156.6	208.9	260.1	260.1	260.1	260.1
13	2003	57.6	114.2	130.6	131.6	131.6	131.6
14	2004	115.8	127.5	163.2	182.2	182.2	206
15	2005	115.2	204.9	216.2	254	263	277.8
16	2006	69.9	84.2	91.9	91.9	91.9	104.2
17	2007	158	192	209.2	223	223	246
18	2008	149	287.5	390.7	419.7	427.2	430.4
19	2009	176	191.2	191.2	242.2	257.4	257.4
20	2010	136.6	139.6	180	186.2	193.6	212.6
21	2011	78.5	114	132.8	153.8	157.2	161.4
22	2012	94.4	110.6	122.6	130.6	141	168.2
23	2013	80	84.5	95.2	99.7	99.7	112.8
24	2014	67	67	70.2	86.6	97.1	98.4
25	2015	61.4	75.7	79.7	92.5	97.9	97.9

Table 2: Statistical parameters of annual 1 day and consecutive days maximum rainfall

Sl. No.	Parameters	1 day	2 day	3 day	4 day	5 day	7 days
1.	Minimum (mm)	57.6	67	70.2	86.6	91.9	97.9
2.	Maximum (mm)	205.8	287.5	390.7	419.7	427.2	430.4
3.	Mean (mm)	108.90	140.28	161.59	174.26	181.84	194.07
4.	Standard deviation, σ (mm)	7.93	11.02	13.78	15.07	15.10	15.03
5.	Coefficient of skewness	0.74	0.80	1.5	1.38	1.36	1.13
6.	Kurtosis	-0.18	0.49	3.90	3.29	3.23	2.63

The statistical parameters of annual 1 day as well as consecutive days maximum rainfall are shown in Table 2.

Gumbel distribution and Generalized Extreme Value distribution were used for the analysis of extreme rainfall events and the calculation of return periods. One day to 1week maximum rainfall data were fitted to the corresponding distributions.

Fitting the Distributions for the Extreme Rainfall Analysis

Based on theoretical probability distributions, it could be possible to forecast the incoming rainfall of various magnitudes with different return periods. The extreme value distributions, used for the study of extreme hydrologic events (e.g. extreme rainfall, peak flow etc.) are analysed. The analysis of extreme events involves the selection of a sequence of the largest or smallest observations from sets of data.

A large problem in working with the Extreme Value distributions is determining whether to use Type 1, 2 or 3. EV3, which has a negative shape parameter is often appealing as it has a finite upper limit, which the general belief of observed flood magnitudes[4]. In general, a distribution with a larger number of flexible parameters, for instance GEV, will be able to model the input data more accurately than a distribution with a lesser number of parameters. EV1 is effective for small sample sizes, however if the size is greater than 50, GEV shows a better overall performance[4].

The GEV and Extreme Value Type 1 distributions have a wide variety of applications for estimating extreme values of given data sets. They are commonly used in hydrological applications.

a. Generalized Extreme Value Distribution (GEV)

The GEV distribution is a family of continuous probability distributions that combines the Gumbel (EV1), Fréchet and Weibull distributions. GEV makes use of 3 parameters: location, scale and shape. The location parameter describes the shift of a distribution in a given direction on the horizontal axis. The scale parameter describes how spread out the distribution is, and defines where the bulk of the distribution lies. As the scale parameter increases, the distribution will become more spread out. The third parameter in the GEV family is the shape parameter, which strictly affects the shape of the distribution, and governs the tail of each distribution. The shape parameter is derived from skewness, as it represents where the majority of the data lies, which creates the tail(s) of the distribution. When shape parameter (k) = 0, this is the EV1 distribution.

The shape parameter for GEV can greatly affect the results. A positive shape parameter will result in the distributions being upper bounded. This phenomenon is undesirable in practical applications as this produces very minimal differences in magnitudes between large return periods. A negative shape parameter assures that the distribution is unbounded and that results in an increase in magnitudes, as the return period gets larger. When designing for extreme events, we are looking for these large values.

The CDF of GEV is defined in[5] as:

$$F(x) = \exp\{-\left(1 - \frac{k(x-\xi)}{\alpha}\right)^{\frac{1}{k}}\} \quad \text{..... (1)}$$

where, ξ is the location parameter, α is the scale parameter, and κ is the shape parameter.

a. Gumbel Distribution (EV1)

Gumbel distribution also referred as Extreme Value Type-1 distribution is used for the study of extreme hydrologic events (eg: extreme rainfall, peak flow etc.) The EV1 distribution uses only 2 parameters, location (ξ) and scale (α).

The CDF for Gumbel distribution as defined in[5] is:

$$F(x) = exp\left[-exp\left(-\frac{x-\xi}{\alpha}\right)\right] \quad \text{..... (2)}$$

where, ξ is the location parameter, α is the scale parameter.

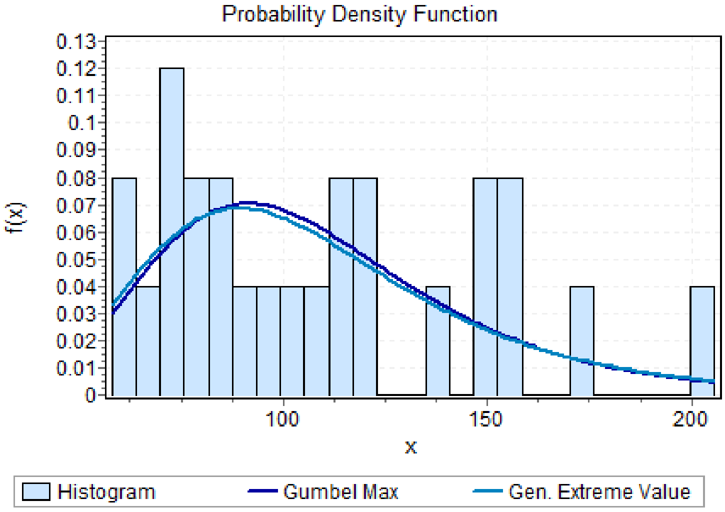

Fig. 1: Gumbel and GEV distributions fitted to the daily maximum rainfall

Parameter estimation for the distributions

Several methods of parameter estimation techniques are available namely Maximum Likelihood Estimation (MLE), Method of Moments (MOM), method of L-Moments etc. L-Moments are based on probability-weighted moments (PWMs) however provide a greater degree of accuracy and ease. PWMs use weights of the cumulative distribution function (F(x))[5]. L-Moments are a modification of the PWMs, as they use the PWMs to calculate parameters that are easier to interpret and that can be used in the calculation of parameters for statistical distributions. L-Moments are based on linear combinations of data that have been arranged in ascending order. They provide an advantage, as they are easy to work with, and more reliable as they are less sensitive to outliers. The MOM techniques only apply to a limited range of parameters, whereas L-Moments can be more widely used, and are also nearly unbiased[8].

Probability Weighted Moments Equations

PWMs are needed for the calculation of L-Moments. The data first must be arranged in ascending order, and then apply the following equations from[4]

$$M_{100} = 1/N \sum_{i=1}^{N} Q_i \quad ......(3)$$

$$M_{110} = 1/N \sum_{i=1}^{N} \frac{(i-1)}{(N-1)} Q_i \quad ......(4)$$

$$M_{120} = 1/N \sum_{i=1}^{N} \frac{(i-1)(i-2)}{(N-1)(N-2)} Q_i \quad ......(5)$$

$$M_{130} = 1/N \sum_{i=1}^{N} \frac{(i-1)(i-2)(i-3)}{(N-1)(N-2)(N-1)} Q_i \quad ......(6)$$

where, N is the sample size, Q is the data value, and i is the rank of the value in ascending order.

L-Moment Equations

The following L-Moments are defined in[4]:

$$\lambda_1 = L1 = M_{100} \quad ......(7)$$

$$\lambda_2 = L2 = 2M_{110} - M_{100} \quad ......(8)$$

$$\lambda_3 = L3 = 6M_{120} - 6M_{100} + M_{100} \quad ......(9)$$

$$\lambda_4 = L4 = 20M_{130} - 30M_{120} + 12M_{110} - M_{100} \quad ......(10)$$

The 4 L-Moments (λ_1, λ_2, λ_3, λ_4) are all derived using the 4 PWMs. Other useful ratios are L-CV (τ_2), L-Skewness (τ_3) and L-Kurtosis (τ_4).

L-CV is similar to the normal coefficient of variation (CV). The standard equation for $CV = \frac{Standard\ Deviation}{Mean}$ and shows how the data set varies. The larger the CV value, the larger the variation of the data set from the mean. For example, in arid regions that receive few storm events, the variation will be large, as one storm will deviate greatly from the low mean.

$$\tau_2 = L - CV = \frac{L2}{L1} \qquad \text{......(11)}$$

L-Skewness is a measure of the lack of symmetry in a distribution. If the value is negative, the left tail is long compared with the right tail, and if the value is positive, the right tail is longer. For GEV frequency analysis, a positive L-Skewness value is desired, as we are interested in the extreme events that occur in the right-side tail of the distribution.

$$\tau_3 = L - Skewness = \frac{L3}{L2} \qquad \text{......(12)}$$

L-Kurtosis is difficult to interpret, however is often described as the measure of "peakedness" of the distribution[5]. L-kurtosis is much less biased than ordinary kurtosis.

$$\tau_4 = L - Kurtosis = \frac{L4}{L2} \qquad \text{......(13)}$$

a. Generalized Extreme Value Distribution

GEV distribution uses three parameters: ξ is the location parameter, α is the scale parameter and κ is the shape parameter. The parameters are defined from[5] as:

$$k = 7.8590c + 2.9554c^2 \qquad \text{......(14)}$$

where,

$$c = \frac{2}{3+\tau_3} - \frac{\ln 2}{\ln 3} \qquad \text{......(15)}$$

$$\text{Scale factor } (\alpha),\ \alpha = \frac{\lambda_2 k}{(1-2^{-k})\Gamma(1+k)} \qquad \text{......(16)}$$

$$\text{Location factor } (\xi),\ \xi = \lambda_1 - \alpha\{1 - \Gamma(1+k)\}/k \qquad \text{......(17)}$$

where, Γ is the gamma function

a. Gumbel (EV1) Distribution

The EV1 Parameters are defined in[5]:

$$\text{Scale factor } (\alpha),\ \alpha = \frac{\lambda_2}{\log 2} \qquad \text{......(18)}$$

$$\text{Location factor } (\xi),\ \xi = \lambda 1 - (\alpha\gamma) \qquad \text{......(19)}$$

where, γ = 0.5772 (Euler's Constant)

The estimated parameters of GEV and Gumbel using the method of L-Moments are:

Table 3: Parameters for GEV distribution

Sl. No:	Parameters for GEV	Annual Maximum Rainfall for consecutive days					
		1 day	2 days	3 days	4 days	5 days	7 days
1	Scale, α	31.69	45.23	51.57	57.09	57.82	62.89
2	Location, ξ	89.66	114.70	130.54	105.68	147.30	161.68
3	Shape, k	-0.0297	0.0144	-0.0299	-0.028	-0.019	0.0667

Table 4: Parameters for Gumbel distribution

Sl. No:	Parameters for Gumbel	Annual Maximum Rainfall for consecutive days					
		1 day	2 days	3 days	4 days	5 days	7 days
1	Scale, α	32.59	44.64	53.05	58.65	58.91	59.34
2	Location, ξ	90.09	114.52	130.97	106.43	148.46	157.77

Return Periods and Return Levels:

Return Period (T) also known as a *recurrence interval* (sometimes *repeat interval*) is an average length of time in years for an event (e.g. flood or river level) of given magnitude to be equalled or exceeded at least once. The return period for an event can be calculated by the following formula:

$$T = \frac{N+1}{m} \quad \text{.....(11)}$$

where, N is the total number of years of record and R is the rank of observed rainfall values arranged in descending order. Return levels represents the amount of rainfall equalled or exceeded at the given return period. In this study, the return levels of rainfall are calculated for the assumed return periods of 2, 5, 10 and 25 years.

Calculation of Return Levels

With the help of theoretical probability distributions, it could be possible to forecast the incoming rainfall of various magnitudes with different return periods. The probability distributions used in this study are Gumbel (EV1) and Generalised Extreme Value distribution. Chow[7] suggested that rainfall analysis by theoretical probability distributions can be done by using frequency factor 'K' which is based on some statistical parameters. Methods used for assessing probability distribution are as follows:

1. *Gumbel distribution:*

The equation for fitting the Gumbel distribution to observed series of flood flows at different return periods T is[6]:

$$X_T = \bar{X} + K\sigma \quad \text{.....(12)}$$

where, X_T denotes the magnitude of the T-year flood event, K is the frequency factor $\bar{X}$ and σ are the mean and the standard deviation of the maximum instantaneous flows respectively.

The frequency factor expresses as:

$$K = -\frac{\sqrt{6}}{\pi}(0.5772 + \ln\left(ln\frac{T}{T-1}\right)) \quad \text{.....(13)}$$

1. *Generalised Extreme Value Distribution:*

The return value is defined as a value that is expected to be equalled or exceeded on average once every interval of time (T) (with a probability of 1/T). Therefore, CDF of the GEV distribution (i.e., equation (1)) = 1-1/T, which implies:

$$X_T = \xi + \frac{\alpha}{k}[1 - (-\ln\left(1 - \frac{1}{T}\right))^k] \quad \text{.....(14)}$$

where, T is the return period, X_T is the return level at T years.

Table 5: Observed and Expected return levels for one day maximum rainfall

S. No.	Return Period	Observed rainfall for one day maximum rainfall	Expected Return Level for one day maximum rainfall	
			Gumbel	GEV
1	2	100	107.59	101.3382
2	5	149	114.61	138.2678
3	10	158	119.25	163.4113
4	25	205.8	125.12	195.9922

Table 6: Observed and Expected return levels for consecutive days maximum rainfall

S. No.	Return Period, T (Years)		2	5	10	25
1	2 Days Maximum Return Level	Observed	130.00	192.00	208.90	287.50
		Gumbel	138.47	148.21	154.66	162.81
		GEV	131.32	183.28	218.15	262.75
2	3 Days Maximum Return Level	Observed	157.40	209.20	219.80	390.70
		Gumbel	159.32	171.51	179.58	189.77
		GEV	149.55	209.65	250.58	303.63
3	4 Days Maximum Return Level	Observed	175.40	224.40	254.00	419.70
		Gumbel	171.78	185.11	193.93	205.08
		GEV	126.71	193.13	238.29	296.71
4	5 days Maximum Return Level	Observed	176.40	233.30	260.10	427.20
		Gumbel	179.37	192.66	201.46	212.57
		GEV	168.57	235.27	280.24	337.97
5	7 Days Maximum Return Level	Observed	197.40	257.40	277.80	430.40
		Gumbel	191.60	204.89	213.69	224.80
		GEV	185.01	260.89	314.38	385.90

Goodness of fit

The goodness of fit between the observed and the expected return levels were analysed using Chi-square test.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \quad \text{.....(15)}$$

where, O_i is the observed rainfall and E_i is the expected return level using probability distribution functions.

RESULTS AND DISCUSSIONS

The return levels of extreme rainfall for one day maxima and 2, 3, 4, 5 and 7-days maxima for 2, 5, 10 and 20 years return period was calculated using the cumulative distribution functions of both Gumbel and generalised extreme value distributions. Chi-square test was conducted for comparison of the results with observed data. The expected return levels using generalised extreme value distribution was found to have a good agreement with the observed data. The chi-squared test results for one day and consecutive days maximum rainfall revealed that, a minimum value of chi-square is obtained for generalised extreme value distribution than that of Gumbel distribution. Therefore, it can be concluded that, the rainfall data of the study area fits best with generalised extreme value distribution.

CONCLUSION

The result of the analysis shows that the rainfall of the study area fits best with generalised extreme value distribution.

REFERENCES

1. Upadhaya, A., Singh, S. R., Estimation of consecutive days maximum rainfall by

various methods and their comparison. *Indian J Soil Conser* **26(2):** 193–201 (1998).
2. Aggarwal, M. C., Katiyar, V. S., Babu, R., Probability analysis of annual maximum daily rainfall of UP *Himalayan. Indian J Soil Cons* **16(1):** 35-42 (1988).
3. Bhattacharya, A. K. and Sarkar, T. K., Analysis of Rainfall Data for Agricultural Land Drainage Design. *Journal of Agricultural Engineering*, **19(1):** 15-25 (1982).
4. Cunnane, C., Statistical distributions for flood frequency analysis. WMO Operational Hydrology Report no. 33. World Meteorological Organization, Geneva, Switzerland. 49-95 (1989).
5. Hosking, J. K. M. and Wallis, J. R., Regional frequency analysis, an approach based on &moments, Cambridge university press (1997).
6. Sarma, P., Flood risk zone mapping of Dikrong sub basin in Assam (1999).
7. Chow, V. T., Hand Book of Applied Hydrology: Chapter 8, Section 1, McGraw Hill Book Co. Inc., New York (1964).
8. Rowinski, P. M., Strupczewski, W. G., Singh, V. P., A note on the applicability of log-Gumbel and log-logistic probability distributions in hydrological analyses, *J Sci Hydrol*, **47:** (2001).