Scope
Volume 14 Number 03 September 2024

Modeling Flood Peak Heights using q-Generalized Extreme Value Distribution in Mahanadi River Basin, India

Nagesh. S[a]* & Laxmi. B. Dharmannavar[b]
[a]Department of Statistics, Karnatak University Dharwad, Karnataka
[b]Department of Statistics and Data Science, CHRIST(Deemed to be University), Bangalore, Karnataka, India

Corresponding Author: **Nagesh. S**

Abstract: The Generalized Extreme Value (GEV) distribution was identified as a good model for flood frequency analysis in hydrology. In this study, annual daily maximum flood heights data from 1970 to 2017 were modeled for five hydrometric sites in the Mahanadi River Basin, first time using the q-Generalized Extreme Value distribution (q-GEV) over GEV distribution model. The target of the study was met by estimating the parameters of the distributions using method of maximum likelihood estimation and performing Kolmogorov-Smirnov, Anderson-Darling and Cramer-von Mises goodness of fit tests; information based criteria AIC and BIC for q-GEV distribution and made comparison with GEV distribution. The simulation study was also conducted for checking the suitability of the model. The results revealed that q-GEV distribution performed better than GEV distribution in modeling the extreme events. This model is more helpful to water practitioners for predicting the extreme events and in taking necessary preparations to mitigate the bad effects of flooding on livings, crops and assets in the associated region.

Keywords: q-GEV, GEV, Mahanadi River Basin, Flood frequency analysis.

1. Introduction

Flooding is the most common natural hazard that people face frequently especially in India. Extreme flood is one of the deadly natural disasters which cause greater damages to human society though it occurs once in many years. Odisha and Gujarat states were repeatedly experienced severe floods in Mahanadi River Basin (MRB) [Beura, (2015)]. As the International Federation of Red Cross (IFRC) quoted pre-disastrous preparations have huge humanitarian impact than post disaster relief operations. Some of the sites experienced severe floods and caused huge economic loss, number of deaths in the MRB of Odisha state. Hydrological extremes such as floods can be described using extreme value theory by estimating high quantiles of extreme flood levels and their return periods.

The GEV distribution was identified as a good model for flood frequency analysis in hydrology. Because the GEV model is a limiting distribution, it may be insufficient in practice; nevertheless, its generalizations should offer greater flexibility in modeling. Provost et al. (2018) proposed the extended model q-GEV distribution which is a q analogue of GEV distribution, q is the additional parameter and provides more flexibility in modeling extreme events than GEV distribution.

Predictions of extreme flood heights and their return periods are important to take precaution in minimizing the adverse effects of disasters on the lives and livelihood. Such rare events can be modeled for prediction using extreme value theory. Fisher and Tippet (1928) found three limiting distributions for maxima namely; Frechet, Gumbel and Weibull distribution that were also called as Type I, Type II and Type III distributions respectively. The GEV distribution was obtained by Jenkinson (1955), which is a single expression consisting of three limiting distributions Frechet, Gumbel and Weibull distributions.

Under the Block Maxima approach the GEV distribution was shown that most suitable flood frequency model for obtaining high quantiles and their return period at some hydrometric sites of the river basin. In relation to this, some of the recent studies can be found in the literature for MRB as well as other river basin are Sukla et al. (2014), Guru and Jha (2015), Chakraborty and Sarma(2017), Singh (2016), Kadhum and Abdulah (2021) and Panigrahi et al. (2020). The GEV distribution was identified as a good flood frequency model for twelve hydrometric sites in the MRB [Nagesh and Laxmi, 2021]. In this study our aim was to apply q-GEV distribution for the first time to the annual daily maximum flood heights data at five sites of MRB comparing with GEV distribution for justifying whether q-GEV is better than GEV distribution in modeling extreme flood heights.

The rest of the paper is structured as follows: Section 2 presents the methods and materials, Section 3 contains the findings and discussions, and Section 4 contains the conclusions of the study.

2. Materials and Methods

This part describes the data, q-GEV model and methods used to obtain the results.

2.1 Data used

The real time data on flood heights recorded thrice a day at five hydrometric sites in MRB were collected from the Central Water Commission (CWC), Bhubaneswar, Odisha, India for the period from 1972 to 2017. The annual daily maximum flood heights were obtained by considering sequential steps.

2.2 q - Generalized Extreme Value Model

The GEV distribution has been identified as a good flood frequency model in analysis of hydrological extremes. The q-GEV distribution is an extension of GEV distribution. The q-GEV distribution is a q analogue of generalized extreme value distribution; q is the additional parameter and provides more flexibility in modeling extreme events than GEV distribution. The distribution function of q-GEV distribution is given by

$$F(x; s, m, \xi, q) = \left[1 + q\left(1 + \xi(xs - m)\right)^{-\frac{1}{\xi}}\right]^{-\frac{1}{q}} \;;\quad \xi \neq 0, q \neq 0 \qquad (1)$$

Probability density function of q-GEV distribution is given by

$$f(x; s, m, \xi, q) = s\left(1 + \xi(xs - m)\right)^{\left(-\frac{1}{\xi}\right)-1} \left[1 + q\left(1 + \xi(xs - m)\right)^{(-1/\xi)}\right]^{(-1/q)-1} ;\quad \xi \neq 0, q \neq 0 \qquad (2)$$

$m = \frac{\mu}{\sigma}$ and $s = \frac{1}{\sigma}$, where μ is location parameter, σ is scale parameter and ξ is shape parameter.

$$x \in \begin{cases} \left(\frac{m}{s} - \frac{1}{\xi s}, \infty\right) & \xi > 0, \quad q > 0 \\ \left(-\infty, \frac{m}{s} - \frac{1}{\xi s}\right) & \xi < 0, q > 0 \\ \left(\frac{(-q)^{\xi} - 1}{\xi s} + \frac{m}{s}, \infty\right) & \xi > 0, q < 0 \\ \left(\frac{(-q)^{\xi} - 1}{\xi s} + \frac{m}{s}, \frac{m}{s} - \frac{1}{\xi s}\right) & \xi < 0, q < 0 \\ (-\infty, \infty) & \xi \to 0, q > 0 \\ \left(\frac{m + \ln(-q)}{s}, \infty\right) & \xi > 0, q < 0 \end{cases}$$

2.3. Maximum Likelihood Estimation

The parameters of q-GEV distribution were estimated by making use of maximum likelihood estimation technique. Firstly log-likelihood function was obtained for the distribution and maximize it in respect of the model parameters. If x_i, i=1,2,..., n are the observations, then the log-likelihood function of the q-GEV distribution is given by

$$l(s, m, \xi, q) = n * \log(s) + \left(-\frac{1}{q} - 1\right) \sum_{i=1}^{n} \log\left[q\left(\xi(x_i s - m) + 1\right)^{-1/\xi} + 1\right] + \left(-\frac{1}{\xi} - 1\right) \sum_{i=1}^{n} \log[\xi(x_i s - m) + 1] \qquad (3)$$

Maximum likelihood estimates of q-GEV were obtained by solving the non-linear system of set of equations.

The goal was achieved by applying the maximum likelihood estimation approach to estimate the parameters of the q-GEV and GEV distributions, followed by goodness of fit analysis using Anderson-Darling (AD), Kolmogorov-Smirnov (KS) and Cramer-von Mises (CvM) goodness of fit based analysis.

The model quality was assessed using Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). For estimated parameters, 95% confidence intervals were determined. For various return times, different return levels were obtained. A simulation study was also carried out to test the model's appropriateness, the expected return level plot is also drawn.

3. Results and discussion

This section discusses the results of analysis for extreme flood heights in the MRB.

3.1. Estimation of parameters

One of the most often used approaches for calculating flood frequency distribution parameters is the maximum likelihood estimation technique (Dombry, 2015; Ferriera and De Haan, 2015). Table 1 provides the GEV and q-GEV distribution parameter estimations for sites. The standard errors of parameter estimations were also computed and recorded in brackets.

The Newton-Raphson Method in MLE technique was used to achieve the results in Table 1. The GEV estimates have larger standard errors than q-GEV estimates. This indicates q-GEV distribution is better than the GEV distribution in modeling extreme flood heights.

Table 1: Parameter estimates of GEV and q-GEV distribution for sites

Site name	Distribution	Parameters			
		s	m	ξ	q
Bamnidhi	GEV	1.0729 (0.1394)	4.4408 (0.1607)	-0.4917 (0.1735)	--
	q-GEV	1.2414 (0.0270)	5.8275 (0.1853)	-0.5750 (0.1489)	1.7019 (0.1703)
Kotni	GEV	0.4716 (0.3074)	3.5896 (0.3867)	-0.3843 (0.1631)	--
	q-GEV	0.5658 (0.0150)	5.3983 (0.1856)	-0.5996 (0.0951)	1.2471 (0.1392)
Pathardhi	GEV	0.5874	3.3248	-0.5181	--

		(0.2641)	(0.335)	(0.131)	
	q-GEV	0.7061 (0.0159)	4.5105 (0.1303)	-0.4998 (0.1862)	1.8411 (0.2192)
Seorinarayan	GEV	0.4375 (0.3237)	4.3037 (0.4369)	-0.391 (0.1147)	--
	q-GEV	0.4902 (0.0156)	6.9208 (0.3814)	-1.2899 (0.2883)	1.1590 (0.1672)
Alipingal	GEV	0.4135 (0.1989)	4.3792 (0.1056)	-0.9582 (0.5438)	--
	q-GEV	0.4373 (0.0029)	5.0169 (0.0274)	-1.2679 (0.2354)	3.7157 (0.3739)

3.2. Goodness of fit analysis

Anderson-Darling, Kolmogorov-Smirnov and Cramer-von Mises tests were used to determine goodness of fit. The p-values of the q-GEV distribution using AD, KS, and CvM are higher than those of the GEV distribution using AD, KS, and CvM. We focus on AD based results because we are dealing with extreme values. Because AD is more sensitive to the tail of the distribution, KS and CvM tests yield the same results. Table 2 shows the p-values of the AD, KS, and CvM tests of GEV and q-GEV distribution at the Bamnidhi site. Test statistic values and p values of AD, KS, and CvM test statistic are given in Table 2. Values in brackets show p values.

Table 2: Goodness of fit analysis for Bamnidhi site

Distribution	AD	KS	CvM
GEV	0.9417 (0.1171)	0.1174 (0.5353)	0.1290 (0.1453)
q-GEV	0.3824 (0.5964)	0.6245 (0.8189)	3.7595 (0.3016)

Note: AD-Anderson-Darling test, KS-Kolmogorov Smirnov test, CvM-Cramer-von Mises test

AIC and BIC of q-GEV are 113.9608 and 121.3614 respectively. AIC and BIC of GEV are 122.9148 and 128.4652 respectively. AIC and BIC of q-GEV distribution is smaller than that of GEV distribution. From Goodness of fit test and Information based criteria AIC and BIC we can say that q-GEV distribution is better fit than GEV for Bamnidhi site.
Goodness of fit analysis tests results for other four sites are given in Table 3.

Table 3: Goodness of fit analysis (p values) for four sites

Site name	Distributions	p values	

		AD	**KS**	**CvM**	**BIC**
Kotni	GEV q-GEV	0.3937 0.8282	0.1057 0.7825	0.0568 0.8775	185.2982 179.7159
Pathardhi	GEV q-GEV	0.2305 0.3782	0.0847 0.3000	0.1302 0.3049	123.1048 116.0306
Seorinarayan	GEV q-GEV	0.3374 0.3542	0.0928 0.1718	0.0518 0.2867	154.6220 147.5300
Alipingal	GEV q-GEV	0.2791 0.6492	0.1010 0.4719	0.1056 0.4565	159.7636 147.9823

Table 3 shows the Anderson-Darling, Kolmogorov-Smirnov, and Cramer-von Mises test statistic's p-values for four different sites for GEV and q-GEV distributions. The p-values of the q-GEV distribution are greater than the GEV distribution for all sites. For both the distributions we do not reject H_{01}: Data follows GEV distribution, H_{02}: Data follows q-GEV distribution. Because q-GEV has higher p-values than GEV, it is evident that q-GEV fits the data better than GEV distribution.

3.3. Confidence intervals and return levels

The 95% Confidence intervals for each parameter results are given in Table 4.

Table 4: Confidence intervals for parameter estimates

Site name	**Distribution**	**s**	**m**	**ξ**	**q**
Bamnidhi	GEV	(1.049, 1.096)	(4.415, 4.466)	(-0.518, -0.465)	
	q-GEV	(1.231, 1.251)	(5.801, 5.854)	(-0.599, -0.551)	(1.676, 1.727)
Kotni	GEV	(0.437, 0.506)	(3.550, 3.628)	(-0.409, -0.359)	
	q-GEV	(0.558, 0.573)	(5.371, 5.425)	(-0.619, -0.580)	(1.223, 1.270)
Pathardhi	GEV	(0.555, 0.619)	(3.288, 3.361)	(-0.541, -0.495)	
	q-GEV	(0.698, 0.714)	(4.489, 4.531)	(-0.527, -0.472)	(1.812, 1.870)
Seorinarayan	GEV	(0.414, 0.461)	(4.262, 4.345)	(-0.412, -0.369)	
	q-	(0.482, 0.498)	(6.882,	(-1.323, -1.256)	(1.133, 1.184)

	GEV		6.959)		
Alipingal	GEV	(0.385, 0.441)	(4.358, 4.399)	(-1.004, -0.912)	
	q-GEV	(0.434, 0.440)	(5.006, 5.027)	(-1.298, -1.237)	(3.677, 3.754)

Table 4 shows the 95 % confidence intervals for calculated parameters of the q-GEV distribution, which are shorter than those of the GEV distribution. Table 5 shows the expected return values for various return periods.

Table 5: Expected return periods and return levels of maximum flood heights

Site Name	**Expected Return Period**						
	20	50	100	200	250	500	1000
	Expected return levels (in meters)						
Bamnidhi	5.8348	5.9452	5.9953	6.0284	6.0365	6.0559	6.0688
Kotni	11.9824	12.2024	12.3011	12.3653	12.3808	12.4176	12.4418
Pathardhi	8.5639	8.8147	8.9358	9.0202	9.0416	9.0944	9.1317
Seorinarayan	15.6642	15.6894	15.6956	15.6981	15.6985	15.6993	15.6996
Alipingal	13.2288	13.2626	13.2706	13.2738	13.2744	13.2753	13.2757

Return levels computed using q-GEV are higher than return levels calculated using GEV distribution, as seen in Table 5. Based on the findings, we conclude that for MRB's 5 hydrometric stations, q-GEV distribution is superior to GEV distribution. For example, for higher accuracy when modeling severe flood heights, q-GEV distribution can be utilized instead of GEV distribution for the sites.

Return level plot for Bamnidhi site is shown in Figure 1.

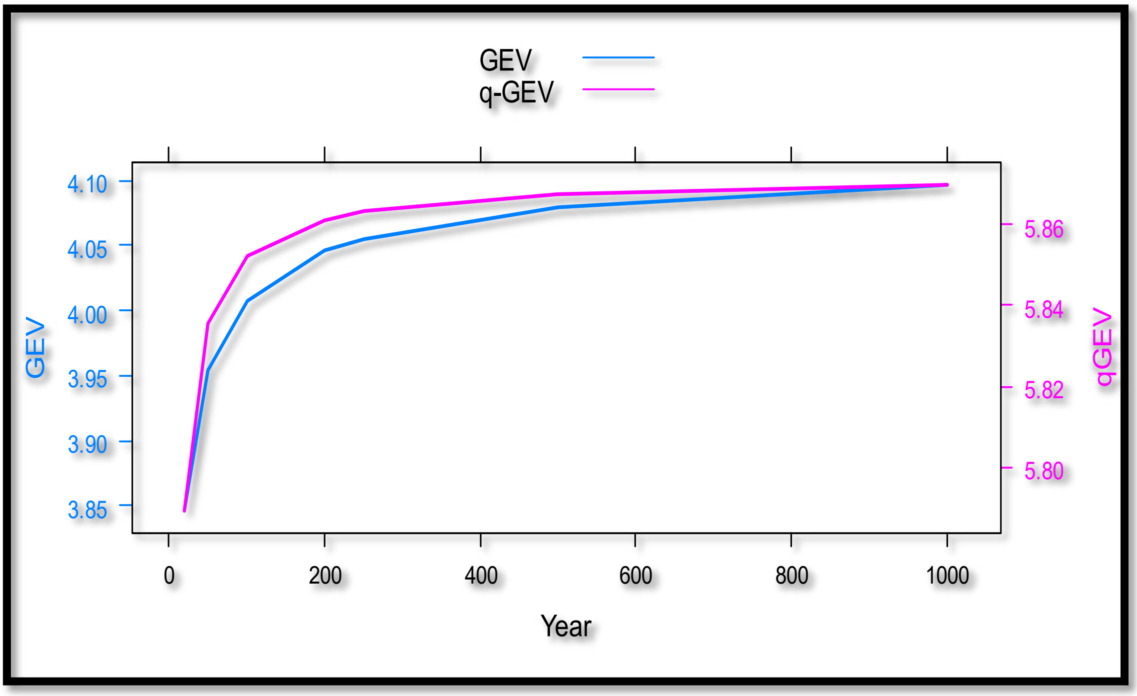

Figure 1: Return level plot for Bamnidhi site

3.4. Simulation study

The results of a simulation study for q-GEV distribution at the Bamnidhi site with different sample sizes of 50, 100, 200, and 500 are shown in Table 6. The suitability of the MLE approach was tested through a simulation study. Using the inverse transformation method, random numbers were generated. Result indicates that as the sample size increases bias and MSE are decreasing.

Table 6: Simulation results of q-GEV distribution for Bamnidhi site

n	Actual values				Bias				MSE			
	S	M	ξ	q	s	m	ξ	q	s	m	ξ	Q
50	1.5	6.5	-0.5	3	0.0044	0.1747	-0.0283	0.0422	0.0000	0.0305	0.0008	0.0018
	1.5	6	-0.6	2.5	-0.007	0.1607	-0.0241	0.0405	0.000	0.0258	0.0006	0.0016
	2	5.5	-0.5	2	0.0059	0.2327	-0.0282	0.0564	0.0000	0.0541	0.0008	0.0032
	2	6	-0.3	3.5	0.0421	0.2718	-0.0405	0.0661	0.0018	0.0739	0.0016	0.0044
	1	7	-0.5	3	0.003	0.1163	-0.0282	0.0282	0.000	0.0135	0.0008	0.0008
	1	6.5	-0.	2.5	0.0117	0.1261	-0.0342	0.0303	0.0001	0.0159	0.0012	0.0009

			4									
100	1.5	6.5	-0.5	3	0.0221	0.0841	0.0395	0.0074	0.0005	0.0071	0.0016	0.0001
	1.5	6	-0.6	2.5	0.0220	0.0755	0.0299	-0.004	0.0005	0.0057	0.0009	0.000
	2	5.5	-0.5	2	0.0292	0.1119	0.0397	1e-02	0.0008	0.0125	0.0016	1e-04
	2	6	-0.3	3.5	0.0429	0.1461	0.0459	0.0421	0.0018	0.0214	0.0021	0.0018
	1	7	-0.5	3	0.0147	0.0559	0.0396	0.0048	0.0002	0.0031	0.0016	0.0000
	1	6.5	-0.4	2.5	0.0170	0.0637	0.0447	0.0129	0.0003	0.0041	0.0020	0.0002
200	1.5	6.5	-0.5	3	0.0852	0.0129	0.0318	0.0141	0.0073	0.0002	0.0010	0.0002
	1.5	6	-0.6	2.5	0.0891	0.0019	0.0207	0.0076	0.0079	0.0000	0.0004	0.0001
	2	5.5	-0.5	2	0.1136	0.0170	0.0318	0.0188	0.0129	0.0003	0.0010	0.0004
	2	6	-0.3	3.5	0.1101	0.0511	0.0476	0.0392	0.0121	0.0026	0.0023	0.0015
	1	7	-0.5	3	0.0568	0.0088	0.0316	0.0093	0.0032	0.0001	0.0010	0.0001
	1	6.5	-0.4	2.5	0.0552	0.0166	0.0409	0.0144	0.0030	0.0003	0.0017	0.0002
500	1.5	6.5	-0.5	3	-0.002	0.0027	0.0014	-0.0054	0.0000	0.0000	0.0000	0.0000
	1.5	6	-0.6	2.5	-0.006	-8e-04	0.0063	-0.0058	0.0000	0.0000	0.0000	0.0000
	2	5.5	-0.5	2	-0.002	0.0036	0.0014	-0.0054	0.0000	0.0000	0.0000	0.0000
	2	6	-0.3	3.5	0.0140	0.0159	-0.013	-0.0093	0.0002	0.0003	0.0002	0.0001
	1	7	-0.5	3	-0.0012	0.0017	0.0015	-0.0053	0.0000	0.0000	0.0000	0.0000

	1	6.5	-0.4	2.5	0.0026	0.0048	-0.0055	-0.0068	0.0000	0.0000	0.0000	0.0000

Similar results are obtained for q-GEV distribution by conducting simulation study for other four sites.

4. Conclusions

The generalization of probability distribution offers more flexibility and accuracy in applications. In the present study, an extended version of GEV distribution known as the q-GEV distribution was considered to apply for maximum flood heights. The parameter estimates of GEV and q-GEV were also calculated using Maximum likelihood estimation approach including standard error of the estimates. The standard errors of q-GEV are smaller than standard errors of GEV distributions for all the sites. Goodness of fit analysis was also performed using Anderson-Darling test, Kolmogorov-Smirnov test and Cramer-von Mises test

For all of the sites, the p-values for the AD test of q-GEV are lower than those of the GEV distribution (the test statistic of q-GEV distribution is significant). As a result, we may conclude that the q-GEV distribution is the most appropriate for all sites. The Anderson-Darling test was adopted since our investigation involved extreme values and this test more sensitive to the tails of the distribution. The results of Akaike Information Criteria and Bayesian Information Criteria add the superiority of q-GEV over GEV distribution.

Based on the results of analysis this work concludes that q-GEV distribution fit well to the 5 hydrometric stations of MRB, which is more flexible and accurate than GEV distribution for modeling maximum flood heights. That is q-GEV can be used for modeling in place of GEV distribution.

5. Acknowledgements

The authors thanked Central Water Commission (CWC), Bhuvaneshwar, the authority for water resource management in India under the Ministry of Jal Shakti for providing hydrometric data used in this study.

6. Conflict of interest

The authors declare that no conflict of interest exists regarding publication of the paper.

References

1. Beura, D. (2015). Floods in Mahanadi river, Odisha, India: its causes and management. International Journal of Engineering and Applied Sciences, 2(2), 257992.
2. Chakraborty, Subrata and Girish Ch. Sarma (2017). Doubly Truncated Extreme Value Distribution of Type-I. International Journal of Agricultural and Statistical Science, 13(2), 379-386.
3. Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In Mathematical proceedings of the Cambridge philosophical society Cambridge University Press, 24(2), 180-190.
4. Guru, N., & Jha, R. (2015). Flood frequency analysis of Tel Basin of Mahanadi river system, India using annual maximum and POT flood data. Aquatic Procedia, 4, 427-434.
5. Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society, 81(348), 158-171.
6. Kadhum, M. K., & Abdulah, E. K. (2021). Statistical Computation and Application with Generalized Poisson Distribution. International Journal of Agricultural and Statistical Sciences, 17(1), 1401-1405.
7. Nagesh. S & Laxmi, Dharmannavar. (2021). Identifying Suitable Probability Models for Extreme Flood Heights in Mahanadi River Basin. International Journal of Agricultural and Statistical Science, 17(1), 197-207.
8. Panigrahi, B., Paramjita, D., Giri, M., & Paul, J. C. (2020). Frequency Analysis for Prediction of Maximum Flood Discharge in Mahanadi River Basin. Int. J. Curr. Microbiol. App. Sci, 9(8), 3626-3639.
9. Provost, S. B., Saboor, A., Cordeiro, G. M., & Mansoor, M. (2018). On the q-generalized extreme value distribution. REVSTAT-Statistical Journal, 16(1), 45-70.
10. Singh, Yashbir (2016). On a Goodness of Fit Test. International Journal of Agricultural and Statistical Science, 12(1), 157-159.
11. Sukla, M. K., Mungaraj, A. K., & Sahoo, L. N. (2014). An investigation on the stochastic modelling of daily rainfall amount in the Mahanadi Delta region, India. Research Journal of Mathematical and Statistical Sciences, 2 (9), 1–8.