Indian Journal of Ecology (2023) 50(1): 210-214
DOI: https://doi.org/10.55362/IJE/2023/3880

Manuscript Number: 3880
NAAS Rating: 5.79

Estimating the Extreme Flood Height Quantiles Using Bayesian Approach

Nagesh S. and Laxmi Dharmannavar

Department of Statistics, Karnatak University Dharwad, Karnataka-580 003, India
E-mail: nageshstatistics@gmail.com

Abstract: In the classical method of estimation, the GEV distribution was identified as the most appropriate model for estimating peak flood heights at twelve study sites in the Mahanadi River Basin. This paper presents that Bayesian parametric estimates of the GEV distribution are better compared to maximum likelihood estimates in estimating peak flood heights and their return periods at these sites. To arrive at this target Markov Chain Monte Carlo Bayesian technique is utilized to acquire parameters of GEV distribution. The estimates of Bayesian approach for peak flood heights and their return periods at the sites showed better predicted flood peak return periods with the Bayesian method. These were shorter than estimates obtained using the maximum likelihood method for all the sites.

Keywords: Bayesian, MLE, Mahanadi, Flood heights

Floods are frequent visitors to the Mahanadi River Basin (MRB) in India, causing grave harm to human society in the affected area. Hydrological extremes, such as floods, can be described using extreme value theory by estimating high quantiles of extreme flood levels and their return periods. In the classical method of estimation, Generalised Extreme Value (GEV) distribution was considered to be a good model for frequency analysis in hydrology (Nagesh and Laxmi 2021). The appropriate flood frequency distributions were identified for twenty six sites of MRB, in which the GEV distribution was found to be good probability model for twelve sites. The Maximum Likelihood Estimation (MLE) is one of the most widely used methods to estimate parameters of the flood frequency distributions (Dombry 2015, Ferreira and De Haan 2015). The likelihood based procedure is alluring, however the problem is the regularity conditions that are need for the normal asymptotic properties related with the MLE to be justified (Alam et al 2019). The Bayesian procedure is a popular method of parametric estimation technique alternate to MLE method and Bayesian estimates can be considered as an improved estimates over maximum likelihood estimates (Reis and Stedinger 2005, Chandra et al 2015, Alam et al 2019). Maposa et al (2014) showed that Bayesian-Based estimates were better than Maximum Likelihood estimates for GEV distribution at two sites in the lower Limpopo river basin, Muzambique. Bayesian approach allows comparison of other source of information by means of prior and posterior distribution. To estimate the parameters of a GEV distribution and make further predictions of the return levels and their related return periods, the researcher employs Bayesian Markov Chain Monte Carlo (MCMC) inference, which has the advantage of not requiring regularity constraints. These results and forecasts are contrasted to those obtained using a frequentist technique based on maximum likelihood GEV distribution estimations in a block maxima framework. Ferreira and De Haan (2015) revealed that the block maxima approach can outperform the Peaks over Threshold method in some circumstances. In this study our key objective is to check whether Bayesian estimates are improved estimates over MLE.

MATERIAL AND METHODS

This section explains how the data in this paper was analyzed using the methodologies employed in the study. In the Bayesian and frequentist paradigms, the methods include algorithms, prior distribution methods and the likelihood of the framework of block maxima.

Data: In the present work, daily water level (metres) data of the MRB recorded thrice a day at twelve hydrometric stations related to the period 1971-2017 were obtained from Central Water Commission (CWC), Bhubaneshwar.

Generalized extreme value model: The GEV distribution is one of the important extreme value distributions to determine the occurrence of the probability of rare event in the field of hydrology, climatology finance, insurance etc. Let the values x_1, x_2, x_3, …., x_n be the annual daily maximum flood height observations of n independent and identically distributed random variable X. As n sufficiently large, the annual daily maximum flood height observations approximate to GEV distribution. The Distribution Function of the GEV distribution is given by

$$F(x;\mu,\sigma,\xi)=\exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\};\xi\neq 0,-\infty<\mu<\infty,\sigma>0 \quad (1)$$

In which μ, σ and ξ are respectively the location, scale and shape parameters of the distribution and are estimated using MLE and MCMC Bayesian approach.

Bayesian model for flood frequency: The observation vector x= {x= x_1, x_2, x_3, …., x_n} consists of iid realizations of annual maximum flood heights and parameter vector θ={ μ, σ and ξ). The posterior distribution is computed using Bayer's Theorem

$$\pi(\theta|x)=\frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta} \quad (2)$$

which is usually written as

$$\pi(\theta|x)\propto f(x|\theta)\pi(\theta) \quad (3)$$

x is a vector of observations, θ is a parameter vector, π(θ) is the prior density function π(θvx) is the posterior distribution, F(θvx) is the density of x, interpreted as the conditional density of x given $_\theta$. The numerator is the joint density of θ and x and the denominator is the marginal density of x. The symbol θ now represents both a random variable and its value. When the parameter θ is discrete, the integral in the denominator of (2) is replaced by a sum. The conditional density π(θvx) of θ given x =x is called the posterior density, a quantification of our uncertainty about θ in the light of data (Ghosh et al 2006). A Bayesian can simply report posterior distribution, or report summary descriptive measures associated with posterior distribution. For example, for a real valued parameter$_\theta$, the posterior mean

$$E(\theta|x)=\int_{-\infty}^{\infty}\theta\pi(\theta|x)d\theta \quad (4)$$

and the posterior variance

$$Var(\theta|x)=\int_{-\infty}^{\infty}(\theta-E)(\theta|x)^2\pi(\theta|x)d\theta \quad (5)$$

Trivariate normal distribution

$$f(x)=\frac{1}{\sigma}\exp\left\{-\frac{1}{2}(\theta-\nu)^T\Sigma^{-1}(\theta-\nu)\right\}$$

where ϑ is mean vector, ∑ is symmetric positive definite covariance matrix.

Trivariate normal distribution is considered as prior distribution for MCMC Bayesian approach. All the outcomes of the present analysis were achieved through use of R software packages ismev and extRemes.

RESULTS AND DISCUSSION

The analysis for twelve hydrometric sites of MRB under the Block Maxima approach was done and estimated parameters of GEV distribution using both MLE and Bayesian approach (Table 1). The 95% credible interval for μ shows that true average of population of annual maximum flood heights fall in the range 3.80m-4.34m at the site with probability 0.95 (Table 1). According to the 95 percent confidence intervals for the ML estimates and 95 percent credible intervals for the Bayesian estimates, the shape parameter of the GEV distribution at Bamnidhi is not significantly different from zero, indicating that annual maximum flood heights can be modeled at the site by a light-tailed Gumbel family of distributions. The quantile function and the parameter estimates from Table 1 were used to construct Table 2.

Bayesian estimates of maximum flood heights are frequently greater than their corresponding ML estimates (Table 2), which is consistent with Table 1. The experimentally recorded maximum flood heights were compared to the anticipated flood heights using both Bayesian and frequentist techniques, and only 6m flood height that occurred during the disastrous floods of 1975 is greater than the 50-year flood level at Bamnidhi. In diagnostic plots displays empirical results for annual daily maximum flood heights for Bamnidhi site, reveal that annual peak flood heights are positively skewed (Fig. 1), with a maximum flood height of 6m occurring in 1975. The probability plot and probability density plot show that GEV Model is a good fit (Fig. 2). Because all of the points on the probability plot are extremely near to the fitted line, and the probability density plot shows that the GEV distribution imitates the empirical distribution form, as seen in Figure 2's histogram. Figure 3 depicts return level plot, which are calculated by Bayesian approach. Black line indicates return level for GEV distribution of observed values for Bamnidhi site. Red dotted lines indicate the interval for return levels considered different return periods. A simulation study of MCMC was also conducted to generate trace and subsequent marginal posterior densities estimated by the MLE method at Bamnidhi site (Figure 4). Fast convergence is observed in trace plots of Figure 4 and the results of posterior marginal densities indicate that posterior estimation of is very improbable to be below 3.80m and very unlikely to be above 4.34m for Bamnidhi site. Similar results are obtained for other eleven sites.

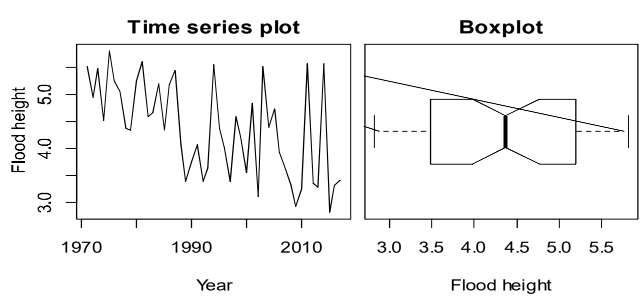

Fig. 1. Time series plot and Boxplot of Bamnidhi site

Bayesian estimates of the parameters are higher than ML estimates (Table 3). The confidence interval for the population average μ indicates that actual population average is the annual maximum heights of Dharamjaigarh, Kesinga, Kotni, Manendragarh, Mohana, Pathardhi, Rajim, Seorinarayan, Simga, Sundargarh and Alipilngal lies in the range of 4.76 - 5.53m, 6.98 – 8.76m, 6.74– 8.47m, 3.35 – 4.16m, 2.46 – 3.05m, 4.91m – 6.40m, 4.07m – 5.31m, 8.86 – 10.80m, 8.01 – 9.66m, 6.08 – 6.81m and 10.14 – 11.03m, respectively. The standard errors of parameter estimates

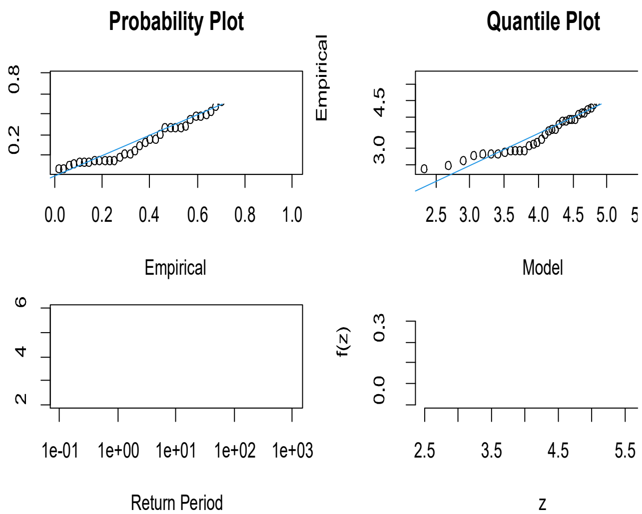

Fig. 2. Diagnostic plots using MLE for Bamnidhi site

Table 1. Parameter estimates of GEV distribution for Bamnidhi Site

ML estimates			
Parameter	Estimate	SE	95% CI
μ	4.1388	0.1607	(3.7807 4.4969)
σ	0.9319	0.1394	(0.6213, 1.2426)
ξ	-0.4917	0.1735	(-0.878, -0.1051)
Bayesian estimates			
Parameter	Estimate	SE	95% CI
μ	4.1399	0.0036	(3.8001, 4.3400)
σ	0.9900	0.0029	(0.7100, 1.1700)
ξ	-0.3541	0.0034	(-0.6300, 0.0000)

SE-Standard Error, CI-Confidence interval for ML estimates and Credible interval for Bayesian estimates

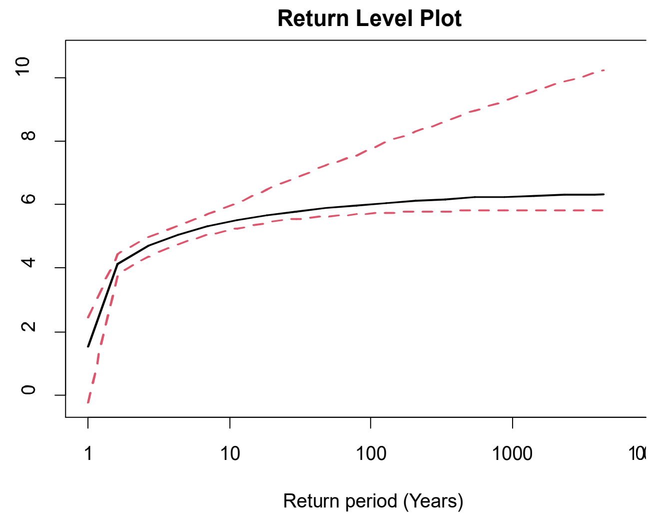

Fig. 3. Return level plot for Bamnidhi site

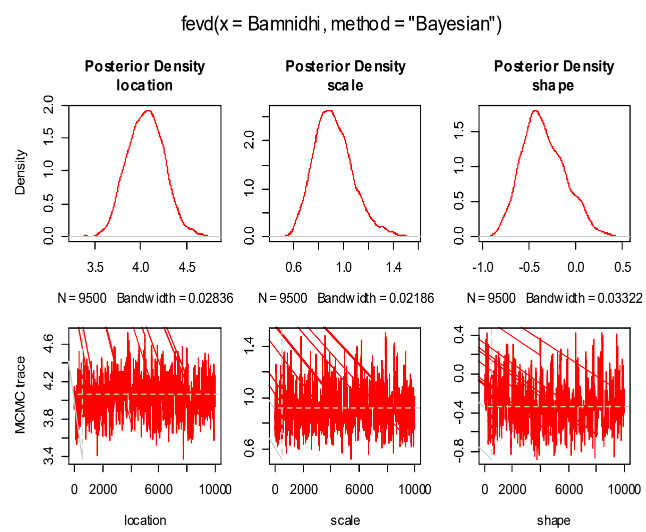

Fig. 4. Posterior density and trace plot for Bamnidhi site

Table 2. Estimation of tail quantile and expected return levels for Bamnidhi site

1-p	p	T	ML estimate (Exceedances)*	Bayesian estimate (Exceedances)*
0.9500	0.05	20	5.7422	5.9034
0.9800	0.02	50	5.8495	6.2753
0.9900	0.01	100	5.9031	6.3828
0.9950	0.005	200	5.9411	6.4665
0.9960	0.004	250	5.9508	6.4893
0.9980	0.002	500	5.9749	6.5497
0.9990	0.001	1000	5.9920	6.5968
0.9999	0.0001	10000	6.0205	6.6913

using the Bayesian technique are lower than those obtained using the MLE approach.

Expected return levels calculated using Bayesian approach is greater than expected return levels calculated using MLE (Table 4). Credible intervals are narrow than confidence interval which indicates Bayesian estimates of GEV distribution have narrow intervals than ML estimates which can be considered as one of the points to highlight the improvisation of Bayesian approach.

Due to heavy rain and cyclonic conditions, extreme flood heights occurred at twelve study sites: 6, 8, 9, 11 and 12 at Bamnidhi (1975), Rajim (1980), Manendagarh, Mohana, Sundargarh, and Pathardi (1990, 1990, 1998, and 2007, respectively); Dhrmarjgarh (1991) and Kotni, Kesinga, and Alipingal (1978, 2006, and 2011 respectively). Both methodologies in Table 2 and 4 were used to calculate the associated return periods of maximum flood heights. For the aforementioned sites, the findings of the Bayesian technique yielded return periods of 20, 10, 50, 50, 20, 200, 20, 20, 10, 2, 10 and 200 years, respectively, implying that these occurrences have a very low chance of being equalized or exceeded at least once in the above-mentioned years.

Table 3. Parameter estimates of GEV distribution for 11 sites

Site name		MLE			Bayesian		
		μ	σ	ξ	μ	σ	ξ
Dharamjaigarh	Estimates	5.150 (0.171)	0.839 (0.139)	0.224 (0.155)	5.191 (0.004)	0.939 (0.004)	0.2171 (0.0035)
	CI	(4.767, 5.532)	(0.529, 1.149)	(-0.121, 0.571)	(4.80, 5.61)	(0.66, 1.33)	(-0.07, 0.54)
Kesinga	Estimates	7.176 (0.399)	2.235 (0.311)	-0.399 (0.134)	7.873 (0.008)	2.190 (0.006)	-0.3701 (0.0031)
	CI	(6.987, 8.765)	(1.543, 2.92)	(-0.699, -0.099)	(6.93, 8.42)	(1.77, 2.82)	(-0.55, 0.00)
Kotni	Estimates	7.411 (0.386)	2.12 (0.307)	-0.384 (0.163)	7.518 (0.009)	2.183 (0.007)	-0.3252 (0.0037)
	CI	(6.749, 8.472)	(1.435, 2.805)	(-0.747, -0.021)	(6.70, 8.26)	(1.68, 2.89)	(-0.61, 0.05)
Manendragarh	Estimates	3.761 (0.182)	0.901 (0.135)	0.080 (0.129)	3.75 (0.004)	0.89 (0.003)	0.10 (0.0031)
	CI	(3.355, 4.166)	(0.599, 1.203)	(-0.209, 0.369)	(3.42, 4.13)	(0.73, 1.25)	(-0.09, 0.44)
Mohana	Estimates	2.758 (0.131)	0.637 (0.117)	0.394 (0.173)	2.777 (0.003)	0.711 (0.003)	0.4336 (0.0040)
	CI	(2.465, 3.051)	(0.376, 0.898)	(0.007, 0.780)	(2.53, 3.13)	(0.48, 1.12)	(0.14, 0.81)
Pathardhi	Estimates	5.659 (0.335)	1.702 (0.264)	-0.518 (0.131)	5.69 (0.008)	1.65 (0.007)	-0.4666 (0.0034)
	CI	(4.913, 6.406)	(1.113, 2.291)	(-0.810, -0.226)	(4.87, 6.16)	(1.35, 2.47)	(-0.77, -0.17)
Rajim	Estimates	4.498 (0.278)	1.545 (0.213)	-0.239 (0.163)	4.66 (0.006)	1.50 (0.005)	-0.21 (0.0035)
	CI	(4.078, 5.317)	(1.069, 2.021)	(-0.604, 0.124)	4.14, 5.18)	(1.22, 2.05)	(-0.47, 0.13)
Seorinarayan	Estimates	9.635 (0.437)	2.185 (0.323)	-0.391 (0.114)	9.81 (0.010)	2.25 (0.008)	-0.3121 (0.0032)
	CI	(8.861, 10.809)	(1.563, 3.006)	(-0.646,-0.135)	(8.78, 10.7)	(1.79, 3.28)	(-0.58, -0.01)
Sigma	Estimates	8.840 (0.369)	2.160 (0.279)	0.391 (0.116)	8.850 (0.008)	2.213 (0.005)	-0.3426 (0.0025)
	CI	(8.017, 9.66)	(1.637, 2.88)	(-0.651, -0.13)	(8.14, 9.55)	(1.83, 2.79)	(-0.52, -0.11)
Sundargarh	Estimates	6.147 (0.163)	0.790 (0.122)	-0.183 (0.157)	6.432 (0.004)	0.864 (0.003)	-0.1607 (0.0034)
	CI	(6.083, 6.812)	(0.617, 1.163)	(-0.533, 0.167)	(6.13, 6.78)	(0.74, 1.29)	(-0.47, 0.13)
Alipilngal	Estimates	10.589 (0.199)	2.418 (0.189)	-0.958 (0.215)	10.65 (0.008)	2.30 (0.008)	-0.93 (0.0026)
	CI	(10.14, 11.032)	(1.974, 2.861)	(-1.438, -0.478)	(9.63, 11)	(1.89, 3.29)	(-1.14, -0.65)

Table 4. Expected return periods and return levels for 11 sites

Site name	Estimation techniques	Expected return periods						
		20	50	100	200	250	500	1000
		Expected return levels (in metre)						
Dharamjaigarh	MLE	10.196	12.242	14.082	16.225	16.989	19.619	22.691
	Bayesian	10.746	12.962	14.942	17.237	18.052	20.848	24.095
Kesinga	MLE	12.245	12.625	12.830	12.985	13.026	13.133	13.214
	Bayesian	12.342	12.765	12.998	13.178	13.226	13.354	13.453
Kotni	MLE	11.849	12.233	12.444	12.604	12.647	12.759	12.845
	Bayesian	12.284	12.794	13.086	13.318	13.382	13.554	13.691
Manendragarh	MLE	7.772	8.957	9.904	10.903	11.236	12.309	13.443
	Bayesian	7.834	9.080	10.090	11.165	11.526	12.697	13.947
Mohana	MLE	8.381	11.594	14.906	19.249	20.917	27.141	35.316
	Bayesian	7.974	10.946	14.000	17.992	19.523	25.222	32.688
Patahrdhi	MLE	8.487	8.662	8.748	8.807	8.822	8.859	8.885
	Bayesian	8.626	8.838	8.946	9.023	9.043	9.094	9.130
Rajim	MLE	8.554	9.072	9.391	9.660	9.737	9.952	10.135
	Bayesian	8.589	9.161	9.521	9.831	9.921	10.176	10.397
Seorinarayan	MLE	14.358	14.762	14.981	15.147	15.191	15.307	15.395
	Bayesian	14.820	15.375	15.696	15.955	16.026	16.220	16.375
Simga	MLE	13.313	13.712	13.928	14.092	14.136	14.251	14.338
	Bayesian	13.557	14.036	14.308	14.521	14.579	14.734	14.856
Sundargarh	MLE	8.887	9.268	9.513	9.728	9.792	9.973	10.133
	Bayesian	8.891	9.296	9.563	9.800	9.871	10.076	10.258
Alipingal	MLE	13.047	13.086	13.099	13.106	13.107	13.110	13.111
	Bayesian	13.051	13.092	13.107	13.114	13.116	13.119	13.121

CONCLUSIONS

Bayesian estimates are higher than ML estimates at all sites. For all sites, the standard errors of parameters estimated using MLE are larger than those of Bayesian estimates. The confidence intervals for MLE-estimated parameters are broader than credible intervals for Bayesian-estimated values. Non-exceedance probability is used to calculate expected return levels for various return periods. Return levels computed using the Bayesian technique is higher than those predicted using the MLE approach. If return levels high, we may take precautions right away. The, study concludes that Bayesian approach has improved results by allowing inclusion of uncertainties through priors. The outcome of Bayesian analysis gives better information than MLE.

ACKNOWLEDGEMENT

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

REFERENCES

Alam MA, Farnham C and Emura K 2019. Bayesian inference for extreme value flood frequency analysis in Bangladesh using Hamiltonian Monte Carlo techniques. In *MATEC Web of Conferences* **276**: 04006.

Chandra R, Saha U and Mujumdar PP 2015. Model and parameter uncertainty in IDF relationships under climate change. *Advances in Water Resources* **79**: 127-139.

Dombry C 2015. Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli* **21**(1): 420-436.

Ferreira A and De Haan L 2015. On the block maxima method in extreme value theory: PWM estimators. *The Annals of Statistics* **43**(1): 276-298.

Fisher RA and Tippett LHC 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.

Ghosh JK, Delampady M and Samanta T 2006. *An introduction to Bayesian analysis: theory and methods* Vol. 725. New York: Springer.

Maposa D, Cochran JJ, Lesaoana M and Sigauke C 2014. Estimating high quantiles of extreme flood heights in the lower Limpopo River basin of Mozambique using model based Bayesian approach. *Natural Hazards and Earth System Sciences Discussions* **2**(8): 5401-5425.

Nagesh S and 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.

Received 12 October, 2022; Accepted 24 January, 2023