

Article

# Bayesian Modeling of Flood Frequency Analysis in Bangladesh Using Hamiltonian Monte Carlo Techniques

Md Ashraful Alam , Craig Farnham and Kazuo Emura \*

Department of Housing and Environmental Design, Graduate School of Human Life Science, Osaka City University, Osaka 558-8585, Japan; alam13ocu@gmail.com (M.A.A.); farnham@life.osaka-cu.ac.jp (C.F.)

\* Correspondence: emura@life.osaka-cu.ac.jp; Tel.: +81-6-6605-2820

Received: 24 May 2018; Accepted: 2 July 2018; Published: 6 July 2018



**Abstract:** In recent years, several Bayesian Markov chain Monte Carlo (MCMC) methods have been proposed in extreme value analysis (EVA) for assessing the flood risk in a certain location. In this study, the Hamiltonian Monte Carlo (HMC) method was employed to obtain the approximations to the posterior marginal distribution of the Generalized Extreme Value (GEV) model by using annual maximum discharges in two major river basins in Bangladesh. As a comparison, the well-known Metropolis-Hasting (MH) algorithm was also applied, but did not converge well and yielded skewness values opposite those of HMC and the statistical characteristic of the data sets. The discharge records of the Ganges and Brahmaputra rivers in Bangladesh for the past 42 years were analyzed. To estimate flood risk, a return level with 95% confidence intervals (CI) has also been calculated. Results show that the shape parameter of each station was greater than zero, which describes the heavy-tailed Fréchet cases of the GEV distributions. One station, Bahadurabad in the Brahmaputra river basin, estimated  $141,387 \text{ m}^3 \cdot \text{s}^{-1}$  with a 95% CI range of [112,636, 170,138] for the 100-year return level, and the 1000-year return level was  $195,018 \text{ m}^3 \cdot \text{s}^{-1}$  with a 95% CI of [122,493, 267,544]. The other station, Hardinge Bridge at the Ganges basin, estimated  $124,134 \text{ m}^3 \cdot \text{s}^{-1}$  with a 95% CI of [108,726, 139,543] for the 100-year return level, and the 1000-year return level was  $170,537 \text{ m}^3 \cdot \text{s}^{-1}$  with a 95% CI of [133,784, 207,289]. As Bangladesh is a flood-prone country, the approach of Bayesian with HMC in EVA can help policy-makers to plan initiatives that could result in preventing damage to both lives and assets.

**Keywords:** Bayesian MCMC; Hamiltonian Monte Carlo; generalized extreme value (GEV); flood frequency analysis; GBM basin

## 1. Introduction

In Bangladesh, major floods are frequent, due to its unique geographic location. About one-quarter to one-third of the country is inundated by overflowing rivers during the monsoon season almost every year. Bangladesh is in the active delta of three major rivers, the Ganges, Brahmaputra and Meghna (GBM). The intensity and time duration of floods in Bangladesh typically depends on the GBM river system. Bangladesh is an agriculture-based economy. Calculating the risk level of river discharge is important for making plans to protect the ecosystem and increase crop and fish production.

Extreme value analysis (EVA) is a branch of probability theory which deals with the stochastic behavior of extreme values of a set of random variables. Typically, EVA is applied for describing a rare event, (e.g., the upper or lower tails of a distribution) [1]. The techniques of EVA are becoming widely applied in various disciplines, including environmental science [2], engineering [3], finance [4], hydroclimatology [5,6] and water resources engineering and management [7]. The fundamental

concepts, techniques and guidelines of the application of the theory of EVA were detailed by Coles [8]. In hydrology, the goal of EVA is to estimate the risk to human beings and environments by extrapolating a certain range of historical observed values, such as for flood frequency analysis (FFA) or rainfall frequency analysis. In particular, EVA often demands estimation of the probability of hydrologic events that are more extreme than any that have already been observed. Here an “extreme event” is defined as the maximum value of a quantity over a given period, for example, the maximum annual discharge in a river basin.

In the application of EVA to flood peaks in at-site and regional settings, the generalized extreme value (GEV) distribution has been widely used [9–11]. The GEV distribution, introduced by Jenkinson, is a three-parameter distribution that associates three extreme value types into a single form [12], and is commonly used for frequency analysis of extreme values. It has been applied in various studies on regional FFA [13–15]. One of the principle goals of FFA is the estimation of the frequency of rare events. For practical use in the design of hydraulic structures, water management strategies and flood insurance, there is great demand for improved accuracy in the estimations yielded by FFA [16,17].

Several techniques have been recommended for parameter estimation in extreme value models. Each technique has its pros and cons. The likelihood-based technique is attractive, but the difficulty is the regularity conditions that are required for the usual asymptotic properties associated with the maximum likelihood estimator to be valid. Subsequently, Bayesian technique, an alternative parameter estimator technique of maximum likelihood estimators, has become popular in recent years [8]. The popularity can be clarified by the rediscovery of the utility of Markov chain Monte Carlo (MCMC) algorithms in the early 1990s [18]. The other reasons for the popularity of Bayesian analysis are the ability to comprise other sources of information through a prior distribution, and the posterior distribution, the output result of a Bayesian analysis, provides a more complete inference than the corresponding maximum likelihood analysis.

MCMC methods are well-known simulation methods, and are widely used for simulating data from an arbitrary distribution. MCMC methods afford a way of simulating from complex distributions by simulating from Markov chains which have the target distributions as their stationary distributions. Besides the general algorithms of MCMC, namely the Gibbs [19], the Metropolis [20,21] and the Metropolis-Hastings [22], for the purpose of estimating the parameters in GEV distributions, Hamiltonian Monte Carlo (HMC) algorithms are more efficient. In addition, the HMC parameter estimation is relatively robust and much faster [23]. In extreme value analysis with GEV models, avoidance of random-walk behavior is one of the major advantages of HMC [24]. Though HMC is much more computationally costly than Metropolis or Gibbs sampling, its proposals are typically much more effective. Therefore, it does not need as many samples to define the posterior distribution.

In determining the hydrologic risk of failure, the return period analysis is widely used. A common use of the return period is to extrapolate the recurrence interval of an event, such as flood peak discharges, droughts, earthquakes, landslides, and others. As an example, one can estimate the magnitude of a flood peak that is expected to occur once in 100 years or any other chosen period [25]. It can be estimated from quantiles of a parametric probability distribution, such as GEV distribution, fitted to the extreme values [26]. A full Bayesian MCMC provides a more exact description of flood risk and the parameter uncertainties with more accurate credible intervals.

The main aim of this article is to calculate a Bayesian regional model for the annual maximum discharge of two major rivers in Bangladesh, as well as flood risk estimation and extrapolation of the return period for the two rivers. For the block maxima or annual maximum discharge, the GEV distribution is used. For describing the Bayesian, the Hamiltonian Monte Carlo is used as the MCMC algorithm.

## 2. Data and Study Area

The area of the GBM river basin, the third largest freshwater outlet to the world's oceans [27], is about 1.7 million km<sup>2</sup> [28], of which 93% of the catchment is outside the geographic boundary of the

country [29]. In the GBM basin, most of the rivers originate from the Himalayas, north of the country, and flow through the country to the Bay of Bengal, south of Bangladesh. The Ganges river originates at the Gangotri glaciers in the Himalayas, passes through Nepal, China and India and enters through the west part of the country, finally ending in the Bay of Bengal at Bangladesh. It is a snowmelt-fed river; its natural flow is controlled by several dams which were constructed by the countries upstream. The dam with the most direct influence on the flow of the Ganges into Bangladesh is the Farakka Barrage in India, located about 16 km upstream from the border. The main objective of the dam is to divert about  $1130 \text{ m}^3 \cdot \text{s}^{-1}$  of water via a canal to the Hooghly River for use in India during the dry season [30]. This reduces the discharge that reaches Bangladesh. The dam was commissioned in May 1975. This has caused a dispute between India and Bangladesh, with some temporary agreements leading to a water use treaty in 1996 [31] which established the discharge share for each country during the dry season from January to May, with 50/50 sharing when the discharge is lowest and a maximum diversion to India of  $1130 \text{ m}^3 \cdot \text{s}^{-1}$ . The treaty does not apply during the monsoon season when water is plentiful, and the diverted water flow is less than 10% of the average discharge during the monsoon period.

Since the Farakka Barrage started operation, the downstream flow has reduced significantly during the dry season months (from December to May). Among the wet months (from June to November), the flow of water also tends to be reduced in the early wet months (June and July) [32]. During June and July, the average monthly maximum peak discharge decreased by 26% and 4% respectively [33]. Also, the sediment movement has been reduced by the low flows. The lack of low-flow transport could be accountable for raising the riverbed levels [34]. However, in August, September and October, the river flow of this basin is dominated by monsoon rainfall, about 70% of rainfall in Bangladesh occurs during this time. Therefore, the Farakka Barrage has less impact on the monsoon-dominated period [32]. According to officials, “all flood gates of Farakka are kept open during the peak rainy season” [35]. Thus, during the period when annual maximum discharge occurs it can be assumed that there is no active, varying regulation of the discharge. However, the more passive effect of the open dam itself will affect the flow.

Pal & Pani [36] compared the Ganges discharge upstream and downstream of the Farakka Barrage. Over the study period (1998–2014) the annual maximum discharge downstream was always greater than upstream, which they conclude is “probably” a result of extra water release from the barrage during strong monsoon flow.

As a measure to counter the reduced flow in the dry season by retaining monsoon rain water for use in the dry season in Bangladesh, another barrage has been planned on the Ganges within the country. However, the project was postponed in 2017, pending further negotiations between India and Bangladesh [37].

The Brahmaputra river begins from a glacier in the Kailash range in Tibet at an elevation of 5300 m above sea level. It is a major transboundary river and the drainage area is around 5,300,000 km<sup>2</sup>. The basin is situated within four different countries: China (50.5%), India (33.6%), Bangladesh (8.1%) and Bhutan (7.8%) [38]. This river enters into the country from the north. China is planning to construct new dams in Tibet that will likely influence the Brahmaputra [39].

The Meghna river is comparatively small and runs through a mountainous region in India before entering Bangladesh. Major characteristics of the GBM river basin are shown in Table 1 (for more details see [40] and references therein).

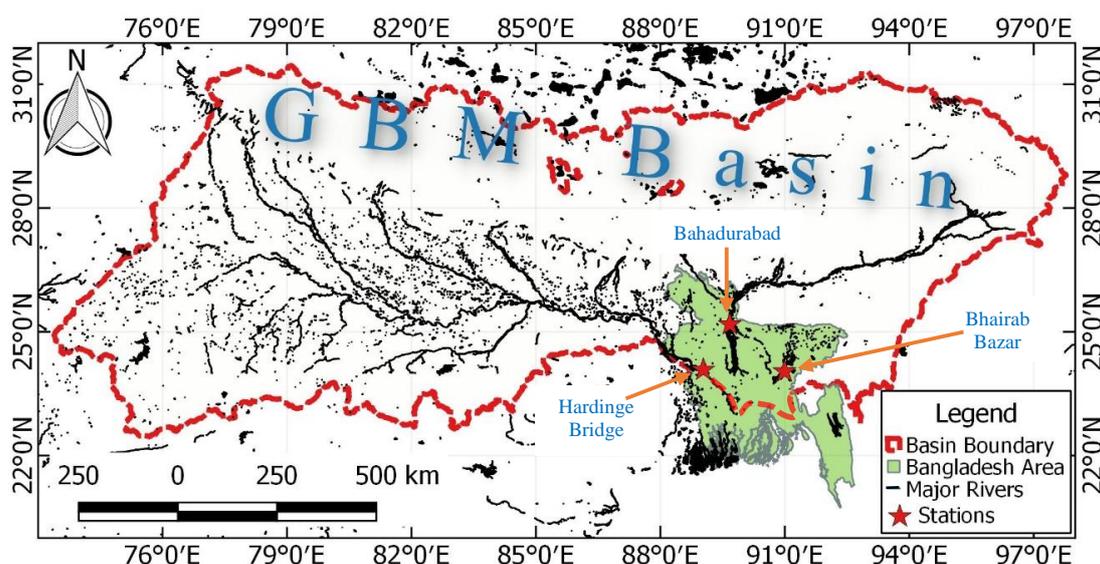
The basins are rich with crop-related biodiversity and ecological diversity. Together they support livelihoods in this area. In the Ganges basin area, the average annual precipitation is about 1500 mm, and in the Brahmaputra basin it is about 2025 mm [41]. This is expected to increase under a warmer climate [42]. Perez and Henebry estimated that from June–October the maximum of accumulated precipitation is over 250 mm month<sup>-1</sup> along the foothills of the Himalayas extending from Himachal Pradesh to Eastern Nepal in the Ganges basin. For the eastern part of the Ganges basin up to 80° E, the authors estimated moderate precipitation varying between 100 and 250 mm month<sup>-1</sup>, while west of

80° E an estimated 100 mm month<sup>-1</sup> accumulates in the same time duration. In the Brahmaputra basin, over 600 mm month<sup>-1</sup> were estimated from Sikkim to Arunachal Pradesh. The northern (above 30° N) part of this basin, over the Tibet Plateau, is mostly dry [43].

**Table 1.** Major characteristics of the Ganges, Brahmaputra and Meghna (GBM) river basins (Sources: [44–46]).

Item	Ganges	Brahmaputra	Meghna	
Basin area (km <sup>2</sup> )	907,000	583,000	65,000	
River length (km)	2000	1800	946	
Total number of dams	75	6	-	
Elevation (m above sea level (asl))	Area below 500 m asl:	72%	20%	75%
	Area above 3000 m asl:	11%	60%	0%
	Lowest	530	3430	2
	Highest	70,868	102,535	19,900
Land use (% area)	Average	11,300	20,000	4600
	Agriculture	68%	19%	27%
	Forest	11%	31%	54%

Bangladesh is situated between latitudes 20°30' N and 26°45' N, and longitude 88°0' E and 92°45' E (Figure 1). The total area is 147,570 km<sup>2</sup>. Bangladesh is a riverine country; the land was formed by the river delta process. So, most of the country (79%) is a floodplain. There are some hilly areas, 12% of the total area, which are situated in the southeast and northeast regions of Bangladesh. The rest of the area (9%) is occupied by four uplifted blocks, which are in the northwest and central parts of the country. The climate of this area is the tropical monsoon type, with a hot summer monsoon and dry season in the winter. The effect of climate on hydrology has many facets. During the summer monsoon time, from June to October, extreme seasonal rainfall occurs. Also, in the Himalayas, the storage of water in the form of snow and ice (in glaciers) provides a large water reservoir that regulates annual water distribution. As most of the rivers originate in the Himalayas, the upper catchment areas are snow-covered and flow through steep mountains. As climatic variability (sea-surface temperature, atmospheric circulation system, El Niño-southern oscillation are well-known climatic factors in this area) occurs, the impact is felt in downstream countries, that is, India and Bangladesh. Also, the seasonal rainfall of upstream countries is a major issue that contributes significantly to the streamflow contribution in the major rivers of Bangladesh. Therefore, the flood peaks are basically determined by basin-wide precipitation [47].



**Figure 1.** The GBM basin with three outlets: Hardinge Bridge in the Ganges, Bahadurabad in the Brahmaputra and Bhairab Bazar in the Meghna river basins.

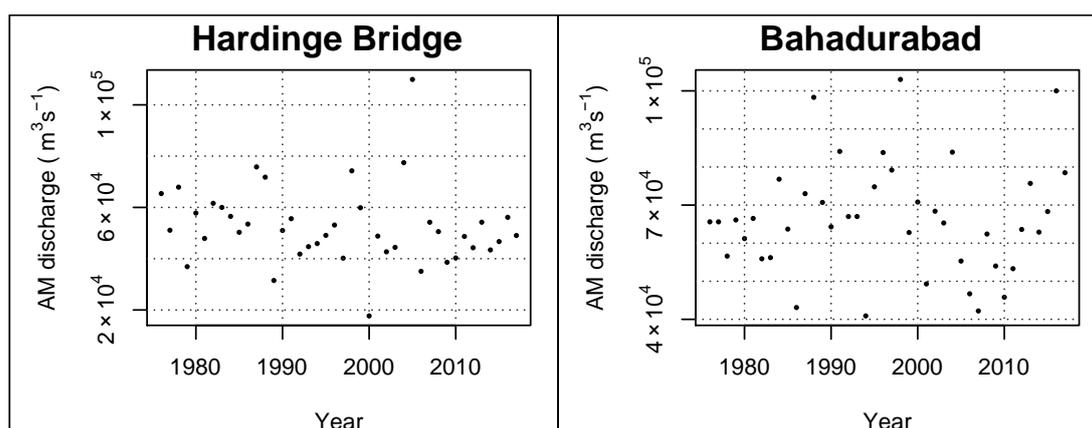
Streamflow data were collected from the Hydrology Division of the Bangladesh Water Development Board (BWDB). There are only 16 stations in Bangladesh, with more than 40 years of river discharge measurement data. Taking the 3 stations closest to the outlets of the 3 rivers of the GBM system should represent the discharge of the GBM system as a whole. The 3 rivers are linked, having a common outlet to the Bay of Bengal. The Brahmaputra and Ganges join to become the Padma to the west of Dhaka. Then the Meghna joins the Padma to the south of Dhaka and flows to the bay. Here, we chose the stations located just upstream of each of these major forks in the GBM system. Station Hardinge Bridge is in the Ganges river. Station Bahadurabad is in the Brahmaputra river, and station Bhairab Bazar is in the Meghna.

In this study, 42 years of data, from 1976–2017, are used. This excludes the period before the start of operations of the Farakka Barrage in 1975. Normally, water discharges are measured weekly by the velocity-area method. The Brahmaputra river is highly braided, so here the discharge measurement was carried out on multiple channels. The Bhairab Bazar station on the Meghna has a large amount of missing data over the time period (over 40%) and could not be useful for this analysis. As the Meghna only contributes about 10% of the total discharge of the GBM basin, this loss of data should not have much effect in an overall assessment of the trends in the GBM system. Thus, only the 2 stations (Hardinge Bridge and Bahadurabad) are used in this analysis. The stations' basic information is presented in Table 2.

**Table 2.** The stations' basic information and description of the data set.

Station Name	River	Latitude	Longitude	Drainage Area (km <sup>2</sup> )	Elevation (m asl)	Observed Data	Missing Values (%)
Hardinge Bridge	Ganges	24.08° N	89.03° E	907,000	10	1976–2017	5
Bahadurabad	Brahmaputra	25.18° N	89.67° E	583,000	22	1976–2017	3
Bhairab Bazar	Meghna	24.04° N	90.99° E	65,000	8	1976–2017	40

For modeling block maxima, here the Annual Maxima (AM), extreme value theory was used. Figure 2 shows the AM of river discharge of the two stations with a time series of 42 years from 1976 to 2017.



**Figure 2.** Annual maximum (AM) discharge data of two major rivers in Bangladesh.

In the 20th century, the country faced major floods in 1951, 1987, 1988, 1998. More recent floods include the years 2004 and 2007. In 1998, 68% of total land of the country was inundated by water. At Bahadurabad station in 1988, the estimated maximum water level was 20.62 m and flowed above the danger level for 27 days. At this station, the theoretical danger level is above 19.5 m. In 1998, the estimated maximum water level was 20.37 m and flowed above the danger level for 66 days. On the other hand, the theoretical danger level of the Hardinge Bridge station is 14.25 m above water level.

In 1988, the estimated maximum level was 14.87 m and flowed above danger level for 23 days. In 1998, it was 15.19 m where water flowed above danger level for 27 days. In recent years, in 2004, 2007, 2015 water flowed above danger level at least once in Bahadurabad station. When the discharge of water flow increases in both river basins during the monsoon period (June to October), the water level increases, causing floods on the both sides of the river.

### 3. Materials and Methods

The GEV distribution is a standard tool for modeling flood peaks by using the annual maximum series (AMS). In this work, a Bayesian approach using the Hamiltonian Monte Carlo (HMC) and Metropolis Hasting (MH) algorithms was applied to estimate the parameters in the GEV model. In addition, extreme flood quantiles (up to 1000-year flood magnitude analysis) were extrapolated for risk estimation.

#### 3.1. The GEV Distribution

The GEV distribution comprises into a single form all three Extreme Value (EV) distributions: Gumbel (EV-I,  $\xi = 0$ ), Fréchet (EV-II,  $\xi > 0$ ), and Weibull (EV-III,  $\xi < 0$ ). The density function  $g(z)$  and cumulative distribution function  $G(z)$  of the GEV distribution can be written as [23], [8]:

$$g(z) = \begin{cases} \frac{1}{\sigma} (1 + \xi \frac{z-\mu}{\sigma})^{-\frac{1}{\xi}-1} \exp\left\{-\left(1 + \xi \frac{z-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right\}, & \xi \neq 0 \\ \frac{1}{\sigma} \exp\left\{-\left(\frac{z-\mu}{\sigma}\right) - \exp\left(-\frac{z-\mu}{\sigma}\right)\right\}, & \xi = 0 \end{cases} \quad (1)$$

$$G(z) = \exp\left\{-\left[1 + \xi \left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \quad (2)$$

The distribution function is defined on the set  $\{z: 1 + \xi (z - \mu)/\sigma > 0\}$ , where the parameters satisfy  $-\infty < \mu < \infty$ ,  $\sigma > 0$  and  $\infty < \xi < \infty$ . The distribution model has three parameters:  $\mu$  is the location parameter,  $\sigma$  is the scale parameter and  $\xi$  is the shape parameter. The value of the shape parameter,  $\xi$  expresses the tail behavior of the distribution. The Gumbel distribution (when  $\xi = 0$ ) has an exponentially decaying tail. The Fréchet distribution has a more slowly decaying tail and the Weibull distribution has an upper limit. The density function of all three forms of GEV distributions are shown in Figure 3 as an example case.

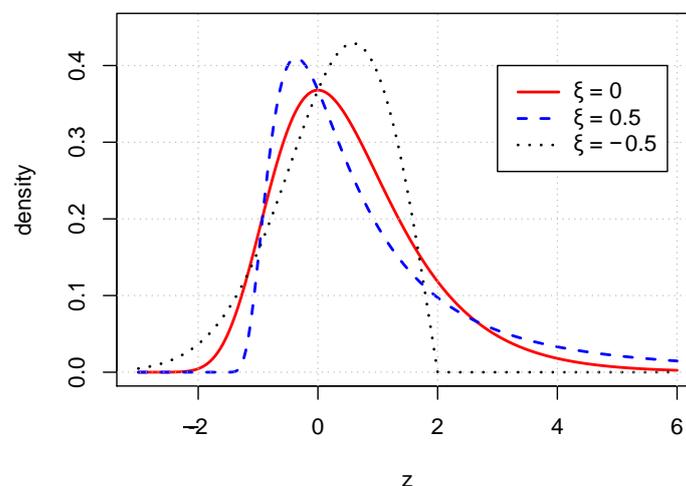


Figure 3. Density functions of the GEV distributions with  $\mu = 0$ ,  $\sigma = 1$ .

Now suppose that the observed data set  $z = (z_1, \dots, z_n)$  are realizations of a random variable with a density from the parametric family  $F = \{f(z; \theta): \theta \in \Theta\}$ . Also, suppose that prior beliefs about  $\theta$  can be expressed by a probability density function  $\pi(\theta)$  with no reference to the observed data. The likelihood for  $\theta$  can be expressed as:

$$\begin{aligned} L(\theta|z) &= f(z|\theta) \\ &= \prod_{i=1}^n f(z_i; \theta) \end{aligned} \tag{3}$$

The prior information and the likelihood are combined using Bayes' theorem to express a posterior distribution for  $\theta$  as follows:

$$\begin{aligned} \pi(\theta|z) &= \frac{\pi(\theta)L(\theta|z)}{f(z)} \\ &= \frac{\pi(\theta)L(\theta|z)}{\int_{\Theta} \pi(\theta)L(\theta|z)d\theta} \end{aligned} \tag{4}$$

So, the distribution notifies the beliefs about  $\theta$  after observing the data and can be expressed as:

$$\begin{aligned} \pi(\theta | z) &\propto \pi(\theta) \times L(\theta | z) \\ \text{i.e., posterior} &\propto \text{prior} \times \text{likelihood} \end{aligned} \tag{5}$$

The function of an EVA is to express the extremes of an observed process to find the probability of extreme events occurring in the future. If an appropriate prior can be specified, there are good reasons to choose Bayesian procedures. The main reason for rejecting the Bayesian procedures is the difficulty in calculating the integral in Equation (4). This problem can be solved by using simulation-based techniques such as MCMC to simulate realizations of the posterior distribution. Parameter estimates of the posterior distribution can then be obtained from the simulated sample.

### 3.2. MCMC Techniques (HMC)

MCMC techniques describe a method of simulating from a complex distribution by simulating from Markov chains which have the target distributions as their stationary distributions. There are many MCMC techniques, including HMC, which was used in this work. HMC was originally proposed by Duane [48] for simulating molecular dynamics under the name of Hybrid Monte Carlo. For an up to date review about the theoretical and practical aspects on HMC, the reader is referred to Neal [24]. The HMC method is used in the context of GEV models in this work.

In a HMC algorithm, a fictitious dynamical system is considered in which auxiliary variables called momentum,  $\mathbf{p} \in \mathbb{R}^N$ , are introduced and the uncertain parameters  $\theta \in \mathbb{R}^N$  in the target distribution  $\pi(\theta)$  are treated as the variables for the displacement. The total energy, also known as the Hamiltonian function, of the fictitious dynamical system is defined by  $H(\theta, \mathbf{p}) = V(\theta) + W(\mathbf{p})$ , where its potential energy is defined by  $V(\theta) = -\ln(P(\theta | D))$  and the kinetic energy is defined by  $W(\mathbf{p}) = \mathbf{p}^T \mathbf{M}^{-1} \mathbf{p} / 2$ , which only depends on  $\mathbf{p}$  and some chosen positive definite "mass" matrix  $\mathbf{M} \in \mathbb{R}^{N \times N}$ . The Hamiltonian method is mentioned as a half-stochastic, hybrid approach that has a Molecular Dynamic trajectory. This property can be convenient when the HMC algorithm is implemented for large and complicated systems [49–52]. This method combines a molecular dynamic (MD) trajectory with a Monte Carlo (MC) acceptance-rejection step, for this reason it is called a hybrid method [51].

Using Hamilton's Equations, the evolution of  $(\theta, \mathbf{p})$  through fictitious time  $t$  can be expressed as:

$$\frac{d\theta}{dt} = \mathbf{M}^{-1} \mathbf{p}(t) \tag{6}$$

$$\frac{d\mathbf{p}}{dt} = -\nabla V(\theta(t)) \tag{7}$$

The Hamiltonian function can be used to define a joint distribution, which can be expressed as:

$$f(\theta, \mathbf{p}) \propto \exp(-\beta_B H(\theta, \mathbf{p})) \tag{8}$$

where  $\beta_B = 1/K_B T$  and  $K_B$  is defined as the Boltzmann constant and  $T$  is temperature. Practically, Equations (6) and (7) have to be solved numerically using some time-stepping algorithm, most commonly-used is a leapfrog algorithm [43]. For time step  $\delta t$ , it works as follows:

$$\mathbf{p}\left(t + \frac{\delta t}{2}\right) = \mathbf{p}(t) - \frac{\delta t}{2} \nabla V(\theta(t)) \quad (9)$$

$$\theta(t + \delta t) = \theta(t) + \delta t \mathbf{M}^{-1} \mathbf{p}\left(t + \frac{\delta t}{2}\right) \quad (10)$$

$$\mathbf{p}(t + \delta t) = \mathbf{p}\left(t + \frac{\delta t}{2}\right) - \frac{\delta t}{2} \nabla V(\theta(t + \delta t)) \quad (11)$$

where  $\nabla V$  is obtained by finite differences as:

$$\frac{\partial V}{\partial \theta_i} = \frac{V(\theta + \Delta h) - V(\theta - \Delta h)}{2h\Delta_i} \quad (12)$$

where,  $\Delta = [\Delta_1, \Delta_2, \dots, \Delta_N]^T$  is defined as the perturbation vector and  $h$  is a scalar that dictates the size of the perturbation of  $\theta$ .

After each iteration of Equations (9)–(11), the output candidate is accepted or rejected according to the Metropolis criterion based on the value of the Hamiltonian  $H(\theta, \mathbf{p})$ . Summarization of the HMC algorithm can be expressed as follows:

1. Initialize  $\theta_0$  to begin the algorithm.
2. Simulate  $\mathbf{p}_0$  such that  $\mathbf{p}_0 \sim N(0, \mathbf{M})$ .
3. Initiate the leapfrog algorithm with  $(\theta, \mathbf{p})$  and evaluate the algorithm for  $L$  time steps to obtain  $(\theta^*, \mathbf{p}^*)$ .
4. Update  $\theta^*$  and obtain new analytic frequencies  $f_i^a$  and then compute  $H(\theta^*, \mathbf{p}^*)$ .
5. Accept  $(\theta^*, \mathbf{p}^*)$  with probability  $\min\{1, \exp(-\beta_B \Delta H)\}$ .
6. Generate a random number,  $r$
7. If the probability in step 5  $> r$ , then let  $\theta$  in the next iteration be  $\theta^*$ , else keep as  $\theta$ .

Repeat the steps 3–7 for  $N_s$  samples.

Here,  $\Delta H = H(\theta^*, \mathbf{p}^*) - H(\theta, \mathbf{p})$ , and the estimated mean of the uncertain parameters can be expressed as:

$$\hat{\theta} = E(\theta) \cong \frac{1}{N_s} \sum_{i=1}^{N_s} \theta^i \quad (13)$$

where the  $E(*)$  defines the mean value of a quantity  $*$ . The variance of the uncertain parameters is expressed as:

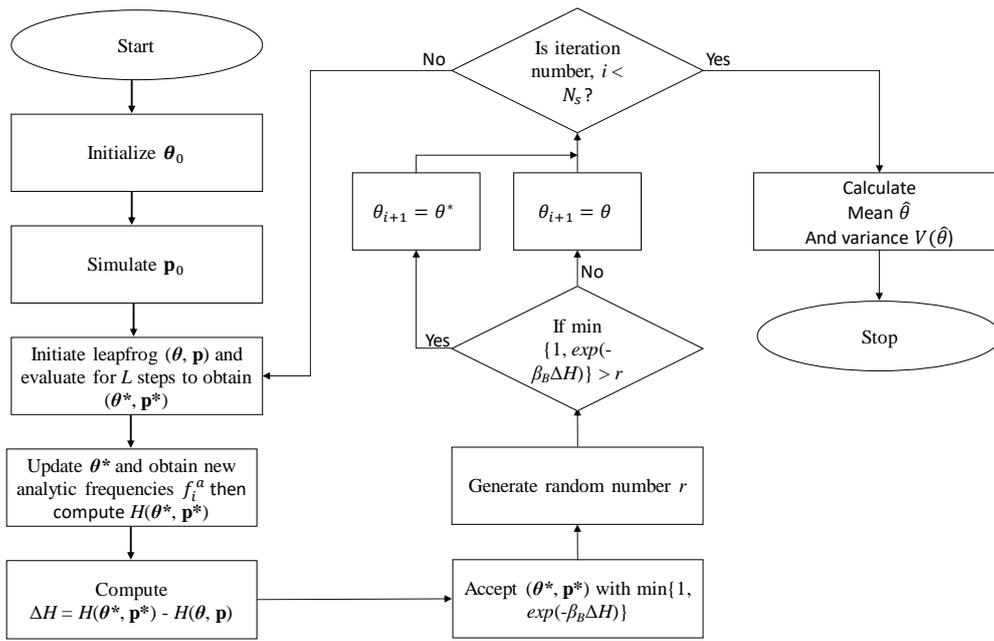
$$V(\hat{\theta}) = E\left((\theta - \hat{\theta})^2\right) \cong \frac{1}{N_s} \sum_{i=1}^{N_s} (\theta^i - \hat{\theta})^2 \quad (14)$$

where the standard error is defined by  $\sigma_\theta = \sqrt{V(\hat{\theta})}$ .

The combination of the Bayesian inference and the HMC algorithm for parameter estimation is presented in Figure 4 as a flowchart.

In this approach, for estimating a Bayesian model using MCMC as the HMC algorithm, an open source programming language called Stan [53,54] was used for solving the mathematical terms.

To create an alternate benchmark to assess the accuracy of the MCMC results, another MCMC method, the MH algorithm, was used to estimate the parameters of the distribution models. An outline of the MH algorithm is given in Appendix A.



**Figure 4.** Flowchart of the estimation of a parameter using the Bayesian inference and the Hamiltonian Monte Carlo (HMC) algorithm.

### 3.3. Inference for Return Levels

Estimates of return levels of the annual maxima are of particular interest in hydrologic extremes, as they give an estimate of the level the process is expected to exceed once, on average, in a given number of years. For the GEV distribution model, the return levels are defined in the following Equation (15):

$$\hat{z}_r = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} [1 - y_r^{-\hat{\xi}}], & \text{for } \hat{\xi} \neq 0, \\ \hat{\mu} - \hat{\sigma} \log y_r, & \text{for } \hat{\xi} = 0, \end{cases} \quad (15)$$

where  $y_r = -\log(1 - 1/r)$ ;  $r$  is the return period (up to 1000-year flood magnitude analysis was estimated in this article). Moreover, by the delta method,

$$\text{Var}(\hat{z}_r) \approx \nabla z_r^T V_c \nabla z_r \quad (16)$$

where  $V_c$  is the variance-covariance matrix of  $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$  and

$$\begin{aligned} \nabla z_r^T &= \left[ \frac{\partial z_r}{\partial \mu}, \frac{\partial z_r}{\partial \sigma}, \frac{\partial z_r}{\partial \xi} \right] \\ &= \left[ 1, -\xi^{-1} (1 - y_r^{-\xi}), \sigma \xi^{-2} (1 - y_r^{-\xi}) - \sigma \xi^{-1} y_r^{-\xi} \log y_r \right] \end{aligned} \quad (17)$$

evaluated at  $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ . In the return level estimation, 95% confidence intervals (CI) were also calculated and presented.

## 4. Result and Discussion

Simulation based techniques, MCMC such as HMC and MH, have provided a way for analyzing the Bayesian techniques of extreme value data for calculating the risk level of flood frequency analysis in the major rivers in Bangladesh. This section shows the estimated parameters of the distribution model and presents the model using diagnostic plots defined as probability plots, quantile plots and density plots. Finally, the return levels of annual peak discharge were calculated, which yielded risk estimation for the subject river basin.

#### 4.1. Statistical Qualities of the Sample Data

The characteristics of the AM samples and best-fit distribution results of each station are shown in Table 3. The stations have similar AM values for mean and standard deviation. In a statistical analysis, skewness is a measure of the asymmetry of the probability distribution of a random variable about its mean. The Hardinge Bridge station is highly positive-skewed and the Bahadurabad station is more symmetric.

In a time series analysis, checking stationarity and homogeneity is essential if the results are to be applied to predict future events. Both parametric and non-parametric methods are used to detect the statistical significance of monotonic trends or non-stationarity.

The Von Neumann (VN) ratio test [55,56] and standard normal homogeneity (SNH) test [57,58] were used in the XLSTAT software package to check the homogeneity of the sample data set. These two test results are shown in Table 3. At the 5% significance level for the data sample size of 42, the critical value is 8.214 for measuring the homogeneity using the SNH test [58]. For both river data sets, the test statistic was less than the critical value. Thus, the sample data can pass as homogeneous. The VN ratio test also passed as homogeneous for both data series at the 5% significance level [56].

In a stationary data series, the statistical qualities are not time-dependent. While non-stationary variables provide misleading and spurious results. The Augmented Dickey-Fuller (ADF) test [59,60], a type of statistical test called a unit root test, was used to check the stationarity of sample data in this work. The ADF statistic is a negative number, and more negative is the stronger rejection of the hypothesis that there is a unit root, and thus the data series has stationarity. The results are shown in Table 3. The ADF statistic value of the stations were less than the critical value of  $-3.544$  at the 5% significance level. Thus, the time series of this sample data pass as stationary. As the analyzed sample data were homogeneous and stationary, flood frequency analysis can be performed.

Also, the goodness-of-fit test statistic results of the commonly used frequency distributions, parametric distributions, in hydrology (see Methodology section in Alam et al. [5]) were also used and the results are presented in Table 3.

**Table 3.** Characteristics of the AM sample data of each station.

Characteristics	Hardinge Bridge Station	Bahadurabad Station
Sample size	42	42
Mean ( $\text{m}^3 \cdot \text{s}^{-1}$ )	52,517	66,490
Standard deviation ( $\text{m}^3 \cdot \text{s}^{-1}$ )	15,079	14,655
Skewness	1.22	0.52
SNH test	2.67 (Pass)	5.07 (Pass)
VN ratio test	1.94 (Pass)	1.50 (Pass)
ADF test	$-5.028$ (Pass)	$-3.648$ (Pass)
First three best-fit distributions by K-S test, with test statistic result in brackets from top to bottom.	GEV (0.067)	LPT3 (0.085)
	Gumbel (0.068)	PT3 (0.088)
	PT3 (0.071)	GEV (0.089)
First three best-fit distributions by A-D test, with test statistic result in brackets from top to bottom.	GEV (0.484)	LPT3 (0.397)
	LPT3 (0.522)	PT3 (0.403)
	Gumbel (0.568)	GEV (0.41)

The GEV distribution yielded the best-fit for Hardinge Bridge station. For Bahadurabad station, the GEV distribution was ranked third according to the test statistic results of goodness-of-fit test of the K-S and A-D. But the difference of the test statistic result of the GEV with the first ranked distribution was comparatively smaller.

#### 4.2. Parameter Estimation of the GEV by the Bayesian MCMC Method

In practical applications, the objective of an EVA is typically an estimate of the probability of future events reaching extreme levels, naturally expressed through a predictive distribution. For example, an appropriate model for the annual maximum  $Z$  of a process is  $Z \sim \text{GEV}(\mu, \sigma, \xi)$ . Estimation of

$\theta = (\mu, \sigma, \xi)$  can be calculated based on the previous observed data set  $z = (z_1, \dots, z_n)$ . Defining a prior distribution is a necessary element of a Bayesian analysis. The prior distribution assumed was a trivariate normal on  $(\mu, \log(\sigma), \xi)$  with mean vector zero and diagonal and diagonal variance covariance matrices. Figure 5 shows the trace plots and posterior densities of the GEV parameters of both stations by using HMC. The HMC algorithm was implemented using the “rstan” package in the “R” programming language.

Typically, two properties are desired in the trace plots: stationarity and good mixing. Stationarity means the path is staying within the posterior distribution. More specifically, the traces all stick around a very stable central tendency and a center of gravity of each dimension of the posterior. In Figure 5, on the left side, experimental trace plots of the parameters have a stable stationarity characteristic. The second property, “good mixing” of a chain means each successive sample within each parameter is not highly correlated with sample before it. Visually, a rapid zig-zag motion of each path is seen, as the trace crosses the posterior distribution without getting mired anyplace. In the experimental trace plots, the second characteristic also shows very well. Two chains were used which are marked by black and red. The number of samples from the chains can be controlled by using the “warmup” and “iter” parameters. In the simulation process, 1000 warmup and 3000 iterations were used. The estimated posterior densities of the parameters are also shown in the same figure on the right side.

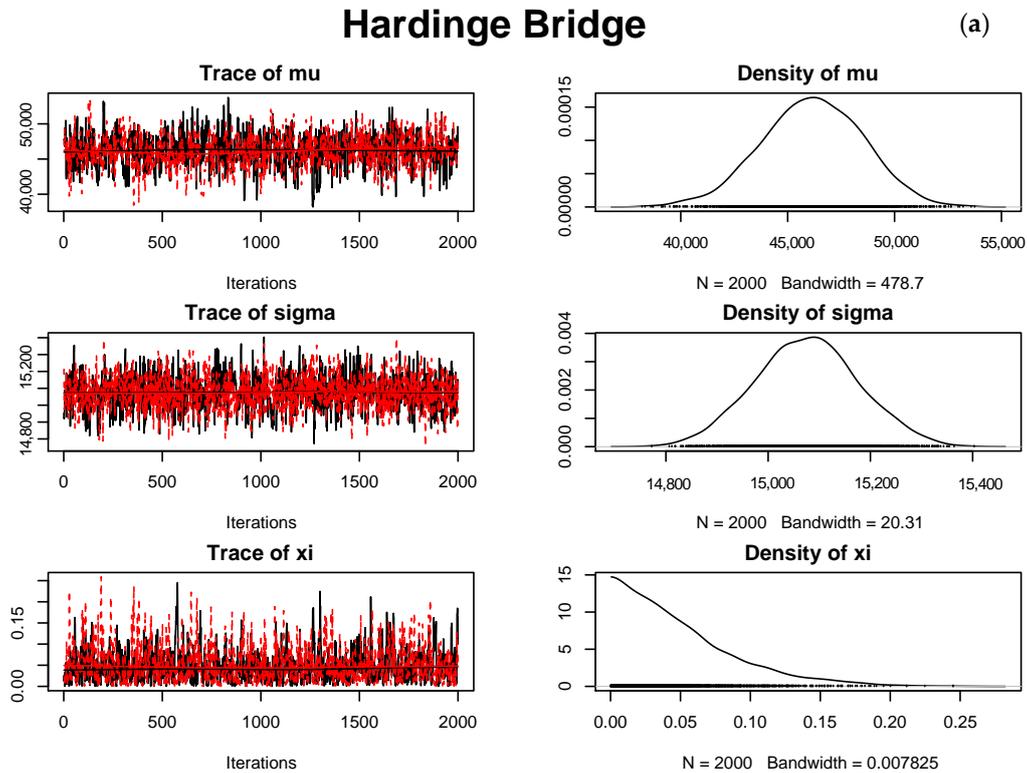
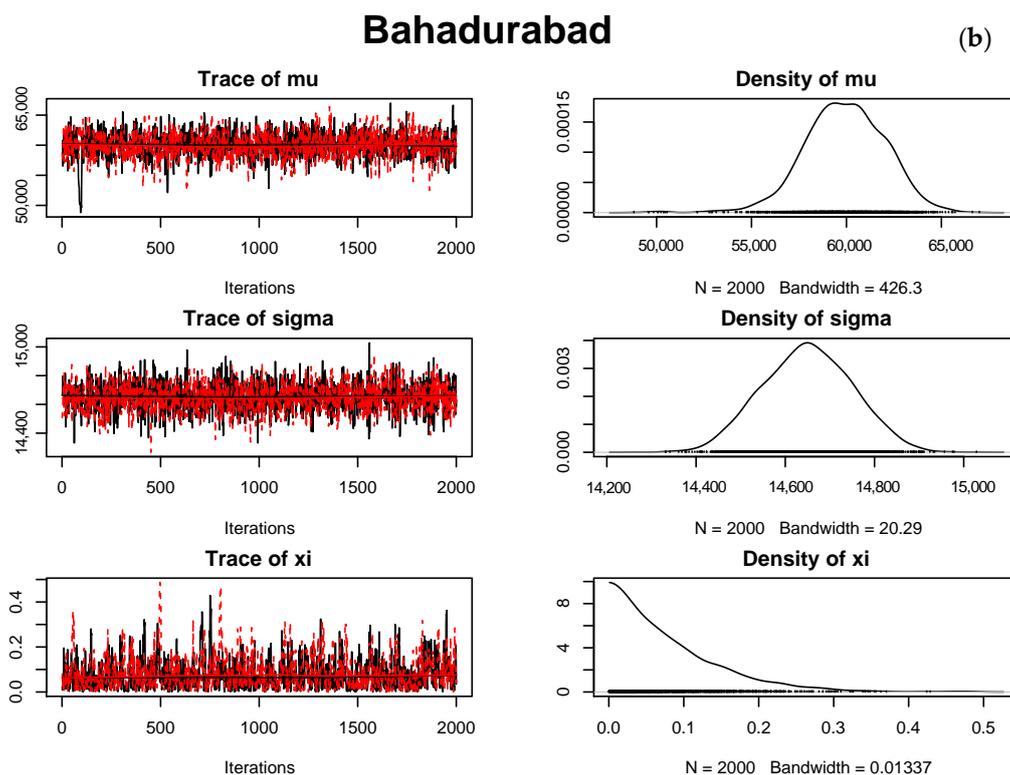


Figure 5. Cont.



**Figure 5.** Trace plots and posterior densities of the GEV parameters. (a) shows the Hardinge Bridge station and (b) shows the Bahadurabad station.

Besides the HMC method, the Metropolis Hasting (MH) algorithm, the most commonly used algorithm in the Bayesian MCMC method, was also applied for a basis of comparison to judge the suitability of the HMC method in this case. The algorithm is discussed in Appendix A. Figures A1 and A2 in Appendix A show the trace plot and posterior densities of the GEV parameters of the two stations where the MH method was applied. Judging from the figures, the parameter  $\mu$  in each station was not converged well. On the other hand, the HMC gave a good convergence. Therefore, the estimated parameters and simulated values of the HMC were considered for the further analysis in this work.

The posterior means of the parameters with 95% confidence intervals for each station are given in Table 4 including the results of both the HMC and MH methods. Also, the effective number of samples, indicated as “n\_eff”, and healthy chain status, indicated as “Rhat” are presented in the same table.

**Table 4.** Posteriors means with 95% credible intervals for the GEV parameters.

Station Name	Parameters	Mean (95% Credible Intervals; Min, Max)		Convergence Diagnostics for HMC Method	
		By HMC Method	By MH Method	n_eff	Rhat ( $\hat{R}$ )
Hardinge Bridge station	Mu	46,157 (41,507, 50,807)	45,948 (45,920, 45,980)	716	1.00
	Sigma	15,076 (14,878, 15,274)	13,353 (10,860, 16,620)	872	1.00
	xi	0.05 (−0.03, 0.13)	−0.068 (−0.177, 0.073)	740	1.00
Bahadurabad station	Mu	59,924 (55,752, 64,095)	60,971 (60,890, 61,010)	853	1.00
	Sigma	14,650 (14,453, 14,847)	13,461 (10,860, 16,940)	1449	1.00
	xi	0.08 (−0.06, 0.22)	−0.12 (−0.29, 0.1)	559	1.00

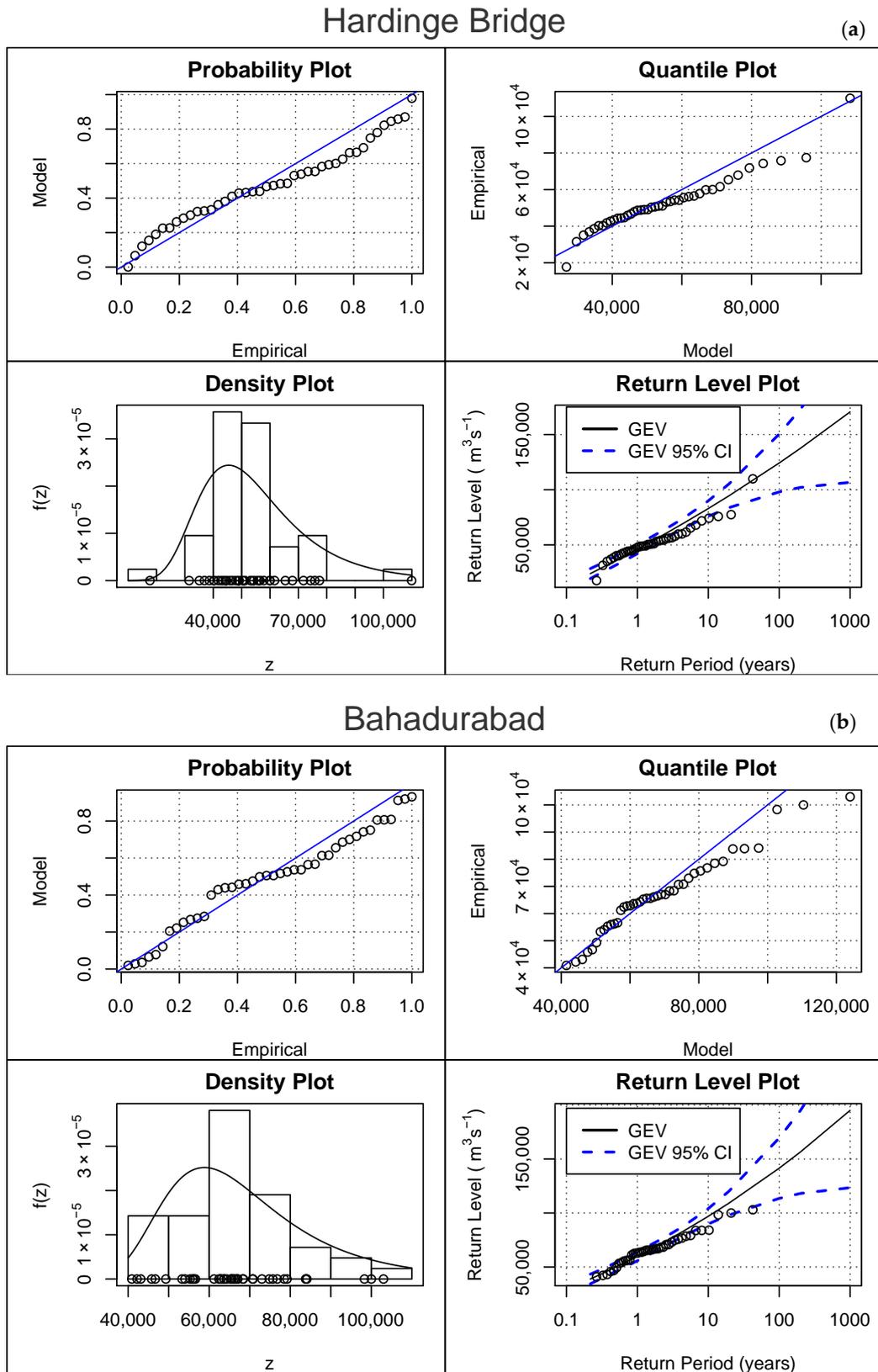
In both stations, the shape parameters found by the HMC method are greater than zero, which yields heavy-tailed Fréchet cases. In Figure 5, it is also obvious that the posterior density of the shape parameter is positive skewed. The location and scale parameters define how dry or wet place is and how much variability there is in the yearly extremes. In Table 4,  $n_{\text{eff}}$  is an important tool to find the convergence diagnostic characteristics. When the effective number of samples,  $n_{\text{eff}}$ , is much lower than the actual number of iterations (3000) minus warmup (1000) in the chains (2 chains), it means the chains are inefficient, but possibly still valid. Warmup samples are used to adjust sampling, and so are not actually part of the target posterior distribution. It does not matter in the result how long the warmup continues. The second convergence diagnostic is the Gelman-Rubin convergence diagnostic,  $R_{\text{hat}}$  ( $\hat{R}$ ). When the output result of  $R_{\text{hat}}$  after convergence is above 1, it indicates that the chain has not yet converged, thus the result based on the  $n_{\text{eff}}$  should not be trusted. In a healthy set of chains,  $R_{\text{hat}}$  should approach 1. In Table 4,  $R_{\text{hat}}$  of all the parameters approached 1.00, and  $n_{\text{eff}}$  were well below the actual number of iterations (3000) minus warmup (1000). The MH method yields the shape parameter as negative for both stations. This is the opposite of both the HMC result and the statistical qualities of both data sets (see Table 3).

#### 4.3. Diagnostic Plots and Return Level

The goal for fitting a statistical model to an observed data set is to reach a conclusion about some aspect of the population from which the data were drawn. Such a conclusion can be drawn onto the fitted model. Therefore, it is essential to check the model fits well. The various diagnostic plots for checking the accuracy of the GEV model fitted to the stream flow data of the two main rivers are shown in Figure 6. The practical application part of the EVA is the calculation of the return period analysis, which yields risk estimations for the event. These are also shown in Figure 6. In this figure the return level plot shows the discharge ( $\text{m}^3 \cdot \text{s}^{-1}$ ) heights of the maximum 1000-year return period with 95% CI. For calculating the 95% CI of the return level it is important to calculate the variance-covariance matrix,  $V_c$ , of  $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ . The approximate variance-covariance matrix, which were calculated from the simulated samples, of the parameter estimates of both stations are:

$$\text{For Hardinge Bridge, } V_c = \begin{bmatrix} 5.6 \times 10^6 & -1946 & -9.798 \\ -1946 & 1.02 \times 10^4 & 0.054 \\ -9.798 & 0.054 & 0.0017 \end{bmatrix}$$

$$\text{For Bahadurabad, } V_c = \begin{bmatrix} 4.5 \times 10^6 & 2707 & -19.76 \\ 2707 & 4.1 \times 10^5 & 0.108 \\ -19.76 & 0.108 & 0.0052 \end{bmatrix}$$



**Figure 6.** Diagnostic plots for the GEV fit and return level plot. The upper panel (a) shows the Hardinge Bridge station and the lower panel (b) shows the Bahadurabad station.

A probability plot is defined as a comparison of the empirical and fitted distribution functions. With ordered block maximum data, the empirical distribution function should lie close to the unit diagonal. Any significant deviations from the linearity are symbolic of the approximation failing in the GEV model. A weakness of the probability plot for extreme value models is that the probability plot provides the least information in the region of most interest. This weakness is avoided by the quantile plot. Both the probability plot and the quantile plot draw the same information expressed on a different scale. However, the observations gained on different scales can be significant, because what looks like a reasonable fit on one scale may be a poor fit on the other scale. In Figure 6, both the probability plot and the quantile plot show the validity of the fitted model. Each set of the plotted points is near-linear. The corresponding density plot in Figure 6 also seems consistent with the histogram of the data. As the shape parameter is positive, the density plots of both stations show a positive skewness. In summary, all three diagnostic plots support the fitted GEV model.

The return level plot,  $\hat{z}_r$  (from Equation (15)) against  $y_r = -\log(1 - 1/r)$  on a logarithmic scale, is particularly convenient for presenting extreme value models. In this process, the tail of the distribution is compressed, as a result return level estimates are easily made for long return periods. Return level heights were calculated by using the value of parameters in the Equation (15) for a  $r$  year return period. The CI of return levels were obtained from Equation (16). For example, the estimated 100-year return level at Bahadurabad was  $141,387 \text{ m}^3 \cdot \text{s}^{-1}$ , and  $\text{Var}(\hat{z}_{100})$  was  $2.15 \times 10^8$ . Hence, a 95% CI was calculated as [112,636, 170,138]. The corresponding estimate for the 1000-year return level at the same station was  $195,018 \text{ m}^3 \cdot \text{s}^{-1}$ , with a 95% CI range of [122,493, 267,544] where  $\text{Var}(\hat{z}_{1000})$  was  $1.37 \times 10^9$ . For Hardinge Bridge, the estimated 100-year return level was  $124,134 \text{ m}^3 \cdot \text{s}^{-1}$ , and  $\text{Var}(\hat{z}_{100})$  was  $6.18 \times 10^7$ . So, the estimated 95% CI was [108,726, 139,543]. The corresponding estimate for the 1000-year return level at Hardinge Bridge was  $170,537 \text{ m}^3 \cdot \text{s}^{-1}$ , with a 95% CI range of [133,784, 207,289] where  $\text{Var}(\hat{z}_{1000})$  was  $3.52 \times 10^8$ . For expressing the uncertainty level, the estimation of CI is important in risk analysis as well as the design purposes. For a certain return period, for example, for a 20-year or 50-year duration, the return level heights with 95% CI can be found from the return level plot in Figure 6.

It must be noted that in the Bahadurabad data there is a single data point ( $109,913 \text{ m}^3 \cdot \text{s}^{-1}$ ) which may seem to be an outlier, as it is over 35% larger than the next highest data point. Although the river level was not as high as many of the smaller AM cases, the velocity was high. This maximum occurred in 2005, when there was no major flood. It is possible this is a result of influence of the Farakka Barrage, even though the operators claim all gates are left open during the rainy season. It is also possible that all AM data points are in part affected by the Farakka Barrage. Comparing AM for 40 years before Farakka was built to 30 years after, Gain and Giupponi found that both floods and droughts happen more frequently in the post-Farakka period [32]. Thus, it may be assumed that the results here could only be valid for the current Farakka Barrage operating period (from 1976) until such time when other major dams are built, possibly in the near future, or if operation policy of the Farakka gates changes.

## 5. Conclusions

This analysis has provided a demonstration of the Bayesian approach to modeling the extreme discharge of two major rivers in Bangladesh. Bangladesh is a flood-prone country because of its geographic characteristics. Knowing the risk level when designing infrastructure makes it possible to reduce the damage by water induced calamities in this area. Block maxima data, each from 42 years of data, was used in a GEV distribution model. There are several techniques to estimate the parameters in extreme value models. Bayesian analysis was used here because it provides a more complete inference than other parameter estimation methods, for example, the maximum likelihood method. For solving the Bayesian approach, a simulation-based technique (MCMC) was selected. As a more efficient, robust and much faster MCMC algorithm, the Hamiltonian Monte Carlo (HMC) algorithm was used. Avoidance of random-walk behavior is a major advantage of this algorithm. Before implementing the sample data in Bayesian MCMC flood frequency analysis, the homogeneity and stationarity was

checked by the provided test. The sample data passed as homogeneous and stationarity. Research into the operations of the dam with the most direct influence on the flow of the Ganges into Bangladesh, the Farakka Barrage, showed that its gates are all left open during the monsoon season. Although the barrage has been shown to yield higher AM than before [32]. There is no active regulation of discharge by that dam during the monsoon.

Estimation of parameters,  $\theta = (\mu, \sigma, \xi)$  were calculated based on the historical observed data set  $z = (z_1, \dots, z_n)$ . In Bayesian analysis for the GEV distribution, the prior distribution is assumed to be a trivariate normal on  $(\mu, \log(\sigma), \xi)$  with mean vector zero and diagonal and diagonal variance covariance matrices. Two important characteristics, stationarity and good mixing, were checked in the trace plots. The resulting traces of the parameters had stable stationarity characteristics, and good mixing. Two chains were used, where the number of samples from the chains can be controlled by using the “warmup” and “iter” parameters. In the simulation process, 1000 warmup and 3000 iterations were used. There results also show the estimated posterior densities of parameters. Besides the graphical diagnostic process, numerical diagnostic methods; the effective number of samples ( $n_{\text{eff}}$ ) and the Gelman-Rubin convergence diagnostic  $\hat{R}$  were also used. Both gave a good convergence result. 95% credible intervals of the parameters were also presented. For both stations, the shape parameters were greater than zero, that shows a heavy-tailed Fréchet case with a positive skewness. The well-known Metropolis-Hasting (MH) algorithm was also used as a comparison with the HMC algorithm to assess the benefit of using HMC. The MH method resulted in poor convergence of the location parameter for each station, also yielding a negative shape parameter, which is opposite the HMC result, as well as opposite the statistical qualities of the data sets. Thus, the HMC is much better suited than the MH in this case.

Various diagnostic plots, such as probability plot, quantile plot and density plot with histogram of the data, were created to judge the accuracy of the GEV model fitted to the stream flow data of the two main rivers which were supported to the fitted GEV model. Return period estimations with 95% CI, which yields risk assessments for the events were also calculated. For the Bahadurabad station (in the Brahmaputra river basin), the estimated 100-year return level was  $141,387 \text{ m}^3 \cdot \text{s}^{-1}$  with a 95% CI range of [112,636, 170,138] and for 1000-year return level was  $195,018 \text{ m}^3 \cdot \text{s}^{-1}$  where 95% CI was [122,493, 267,544]. The other station at Hardinge Bridge (in the Ganges river basin), the estimated 100-year return level was  $124,134 \text{ m}^3 \cdot \text{s}^{-1}$  with 95% CI range of [108,726, 139,543], and for the 1000-year return level was  $170,537 \text{ m}^3 \cdot \text{s}^{-1}$  with a 95% CI range of [133,784, 207,289]. For a certain return period, such as for a 20-year or 50-year duration, it was also estimated with 95% CI. These results (for the Bahadurabad station) may be influenced by the Farakka Barrage. Future construction of dams or a change in Farakka gate control policy would change the hydrology of the river system. As Bangladesh is a flood prone country, this study can help policy-makers to plan initiatives that could result in preventing damage to both lives and assets. Further, the distributions found here could be used in current research on the effects of the Farakka Barrage or by future researchers as a historic baseline of comparison to determine the effects of future dam construction in the GBM system.

**Author Contributions:** M.A.A. and K.E. conceived the research theme and proposed the analysis method. M.A.A. collected the data, performed the analysis and wrote the paper. C.F. contributed to writing the paper and evaluated the results, chose and advised on calculation methods with K.E.

**Funding:** This research was funded by the Osaka City University Graduate School of Human Life Science.

**Acknowledgments:** Thanks to the Bangladesh Water Development Board for providing the river discharge data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix

### Appendix A.1 Metropolis-Hastings (MH) Algorithm

MH is an algorithm that allows sampling from a generic probability distribution which is called the target distribution. It was developed by Metropolis [21] and Hastings [22]. The procedure

is to formulate a transition kernel  $p(\theta, \phi)$  such that the equilibrium distribution of the chain is  $\pi$ . This transition kernel is constructed by two elements- a proposal distribution or arbitrary transition kernel  $q(\theta, \phi)$ , and an acceptance probability  $a(\theta, \phi)$ . The acceptance probability suggested by Hastings [22] is as follows:

$$a(\theta, \phi) = \min \left\{ 1, \frac{\pi(\phi)q(\phi, \theta)}{\pi(\theta)q(\theta, \phi)} \right\} \tag{A1}$$

To obtain a chain with a limiting distribution  $\pi$ , the algorithm can be followed as:

1. Initialize the chain to  $\theta_0$  and counter to  $j = 1$  to begin the algorithm.
2. Simulate a proposal value  $\phi$  using the kernel  $q(\theta_{j-1}, \phi)$ .
3. Find the acceptance probability of the proposed value  $a(\theta_{j-1}, \phi)$ .
4. Accept  $\theta_j = \phi$  with probability  $a(\theta_{j-1}, \phi)$  and take  $\theta_j = \theta_{j-1}$  otherwise.
5. Increase the counter from  $j$  to  $j + 1$  and return to step 2.

The MH algorithm was implemented using the “R2OpenBUGS” package in the “R” programming language.

Appendix A.2 Trace Plots and Posterior Densities Using MH Algorithm

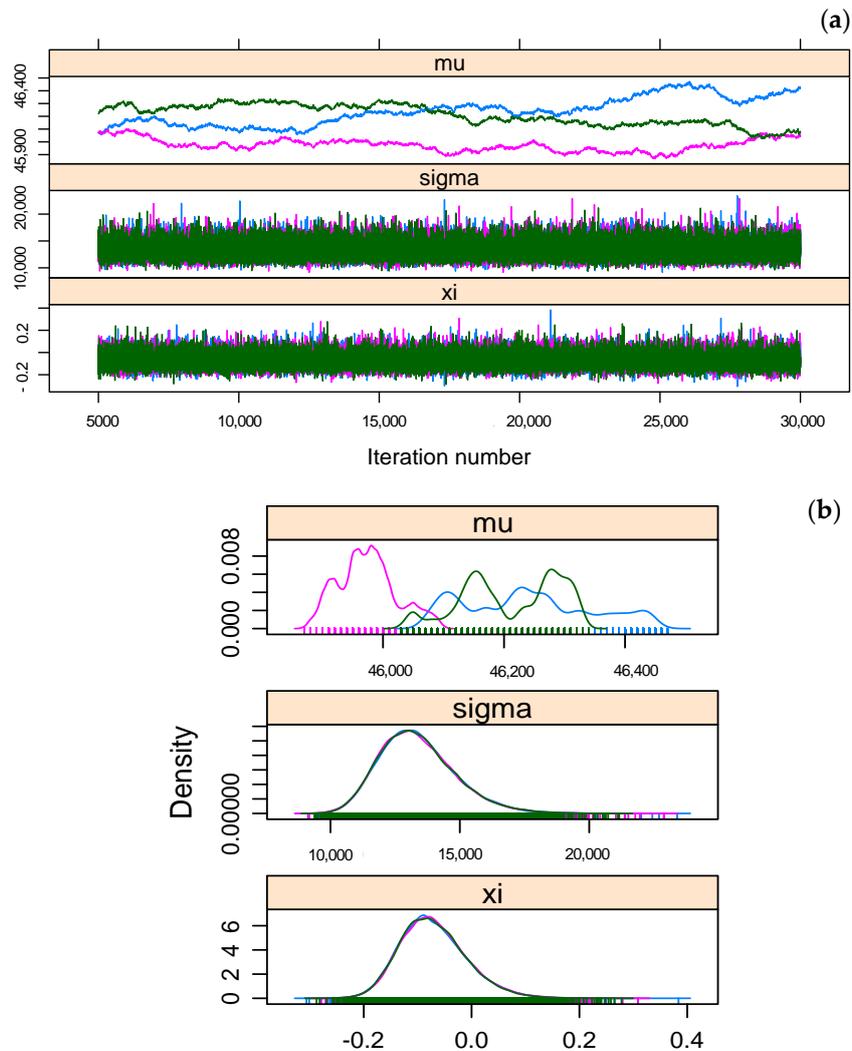
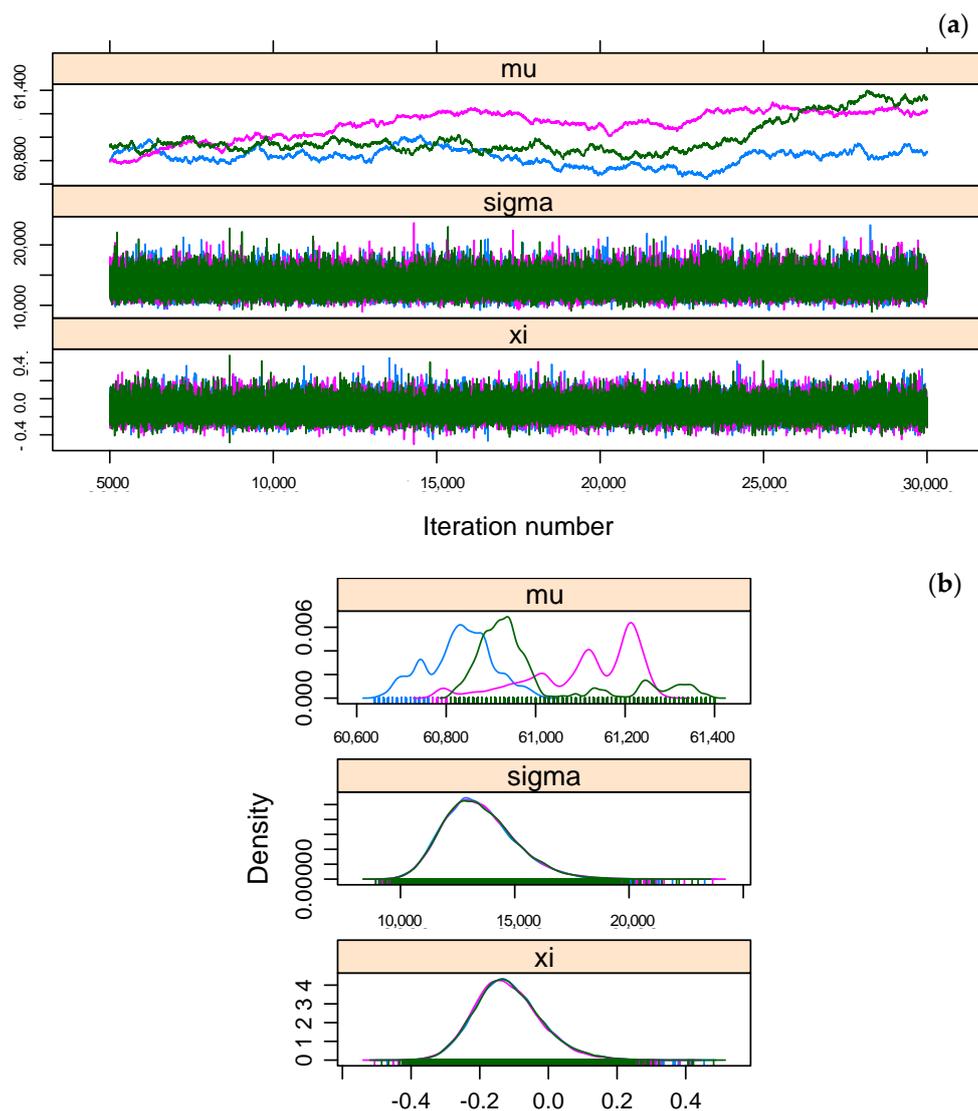


Figure A1. Trace plots (a) and posterior densities (b) of the GEV parameters using MH algorithm of Harding Bridge station.



**Figure A2.** Trace plots (a) and posterior densities (b) of the GEV parameters using MH algorithm of Bahadurabad station.

## References

1. MacDonald, A.; Scarrott, C.J.; Lee, D.; Darlow, B.; Reale, M.; Russell, G. A flexible extreme value mixture model. *Comput. Stat. Data Anal.* **2011**, *55*, 2137–2157. [[CrossRef](#)]
2. Smith, R.L. Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Stat. Sci.* **1989**, *4*, 367–377. [[CrossRef](#)]
3. Castillo, E.; Hadi, A.S.; Balakrishnan, N.; Sarabia, J.-M. *Extreme Value and Related Models with Applications in Engineering and Science*; John Wiley & Sons: New York, NY, USA, 2004; pp. 3–18.
4. Longin, F. *Extreme Events in Finance: A Handbook of Extreme Value Theory and Its Applications*, 1st ed.; John Wiley & Sons: Hoboken, NJ, USA, 2016; ISBN 1-118-65019-0.
5. Alam, M.A.; Emura, K.; Farnham, C.; Yuan, J. Best-Fit Probability Distributions and Return Periods for Maximum Monthly Rainfall in Bangladesh. *Climate* **2018**, *6*, 9. [[CrossRef](#)]
6. Alam, M.A.; Farnham, C.; Emura, K. Best-Fit Probability Models for Maximum Monthly Rainfall in Bangladesh Using Gaussian Mixture Distributions. *Geosciences* **2018**, *8*, 138. [[CrossRef](#)]
7. Katz, R.W.; Parlange, M.B.; Naveau, P. Statistics of extremes in hydrology. *Adv. Water Resour.* **2002**, *25*, 1287–1304. [[CrossRef](#)]

8. Coles, S.; Bawa, J.; Trenner, L.; Dorazio, P. *An Introduction to Statistical Modeling of Extreme Values*, 1st ed.; Springer: London, UK, 2001; Volume 208, ISBN 978-1-84996-874-4.
9. Hosking, J.R.M. Algorithm AS 215: Maximum-Likelihood Estimation of the Parameters of the Generalized Extreme-Value Distribution. *J. R. Stat. Soc.* **1985**, *34*, 301–310. [[CrossRef](#)]
10. Smith, J.A. Estimating the upper tail of flood frequency distributions. *Water Resour. Res.* **1987**, *23*, 1657–1666. [[CrossRef](#)]
11. Hosking, J.R.M.; Wallis, J.R. *Regional Frequency Analysis of Floods in Central Appalachia*; IBM Research Division: Armonk, NY, USA, 1996; p. 17.
12. Jenkinson, A.F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q. J. R. Meteorol. Soc.* **1955**, *81*, 158–171. [[CrossRef](#)]
13. Jingyi, Z.; Hall, M.J. Regional flood frequency analysis for the Gan-Ming River basin in China. *J. Hydrol.* **2004**, *296*, 98–117. [[CrossRef](#)]
14. Kumar, R.; Chatterjee, C. Closure to “Regional Flood Frequency Analysis Using L-Moments for North Brahmaputra Region of India” by Rakesh Kumar and Chandranath Chatterjee. *J. Hydrol. Eng.* **2006**, *11*, 380–382. [[CrossRef](#)]
15. Cunderlik, J.M.; Ouarda, T.B. Regional flood-duration-frequency modeling in the changing environment. *J. Hydrol.* **2006**, *318*, 276–291. [[CrossRef](#)]
16. Browne, M.J.; Hoyt, R.E. The demand for flood insurance: Empirical evidence. *J. Risk Uncertain.* **2000**, *20*, 291–306. [[CrossRef](#)]
17. Carolan, M.S. One step forward, two steps back: Flood management policy in the United States. *Environ. Politics* **2007**, *16*, 36–51. [[CrossRef](#)]
18. Gelfand, A.E.; Smith, A.F. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **1990**, *85*, 398–409. [[CrossRef](#)]
19. Geman, S.; Geman, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. In *Readings in Computer Vision*; Elsevier: Amsterdam, The Nederland, 1987; pp. 564–584.
20. Metropolis, N.; Ulam, S. The monte carlo method. *J. Am. Stat. Assoc.* **1949**, *44*, 335–341. [[CrossRef](#)] [[PubMed](#)]
21. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092. [[CrossRef](#)]
22. Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–109. [[CrossRef](#)]
23. Hartmann, M.; Ehlers, R.S. Bayesian inference for generalized extreme value distributions via Hamiltonian Monte Carlo. *Commun. Stat-Simul. Comput.* **2017**, *46*, 5285–5302. [[CrossRef](#)]
24. Neal, R.M. MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*; Chapman & Hall/CRC: London, UK, 2011; Volume 2, pp. 113–162.
25. Benjamin, J.R.; Cornell, C.A. *Probability, Statistics, and Decision for Civil Engineers*, 1st ed.; McGraw-Hill: New York, NY, USA, 1970; ISBN 978-0-486-78072-6.
26. Wilks, D.S. Comparison of three-parameter probability distributions for representing annual extreme and partial duration precipitation series. *Water Resour. Res.* **1993**, *29*, 3543–3549. [[CrossRef](#)]
27. Chowdhury, M.D.; Ward, N. Hydro-meteorological variability in the greater Ganges-Brahmaputra-Meghna basins. *Int. J. Climatol.* **2004**, *24*, 1495–1508. [[CrossRef](#)]
28. Islam, A.S.; Haque, A.; Bala, S.K. Hydrologic characteristics of floods in Ganges-Brahmaputra-Meghna (GBM) delta. *Nat. Hazards* **2010**, *54*, 797–811. [[CrossRef](#)]
29. Salehin, M.; Haque, A.; Rahman, M.R.; Khan, S.A.; Bala, S.K. Hydrological aspects of 2004 floods in Bangladesh. *J. Hydrol. Meteorol.* **2007**, *4*, 33–44.
30. Farakka Barrage Project Homepage, Farakka Barrage Project. Available online: [www.fbp.gov.in](http://www.fbp.gov.in) (accessed on 23 June 2018).
31. Government of the People’s Republic of Bangladesh. *Treaty between the Government of the Republic of India and the Government of the People’s Republic of Bangladesh on Sharing of the Ganga/Ganges Waters at Farakka*; Government of the People’s Republic of Bangladesh: Dhaka, Bangladesh, 1996.
32. Gain, A.K.; Giupponi, C. Impact of the Farakka Dam on thresholds of the hydrologic flow regime in the Lower Ganges River Basin (Bangladesh). *Water* **2014**, *6*, 2501–2518. [[CrossRef](#)]
33. Mirza, M.M.Q. Hydrological changes in Bangladesh. In *The Ganges Water Diversion: Environmental Effects and Implications*; Springer: New York, NY, USA, 2004; pp. 13–37.

34. Mirza, M.M.Q. Diversion of the Ganges water at Farakka and its effects on salinity in Bangladesh. *Environ. Manag.* **1998**, *22*, 711–722. [[CrossRef](#)]
35. Farakka Propaganda Irks Bangladesh. Available online: <https://www.dhakatribune.com> (accessed on 30 June 2018).
36. Pal, R.; Pani, P. Seasonality, barrage (Farakka) regulated hydrology and flood scenarios of the Ganga River: A study based on MNDWI and simple Gumbel model. *Model. Earth Syst. Environ.* **2016**, *2*, 57. [[CrossRef](#)]
37. Chowdhury, S.I. Delhi keeps Dhaka Waiting. *New Age*. Available online: <http://www.newagebd.net/> (accessed on 25 May 2018).
38. Immerzeel, W. Historical trends and future predictions of climate variability in the Brahmaputra basin. *Int. J. Climatol.* **2008**, *28*, 243–254. [[CrossRef](#)]
39. Bangladesh “very Concerned” over Chaina Building Dams on Brahmaputra. *Times of India*. Available online: <http://timesofindia.indiatimes.com> (accessed on 31 May 2018).
40. Masood, M.; Yeh, P.-F.; Hanasaki, N.; Takeuchi, K. Model study of the impacts of future climate change on the hydrology of Ganges-Brahmaputra-Meghna basin. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 747–770. [[CrossRef](#)]
41. Mirza, M.M.Q.; Warrick, R.A.; Ericksen, N.J.; Kenny, G.J. Are floods getting worse in the Ganges, Brahmaputra and Meghna basins? *Glob. Environ. Chang. Part B Environ. Hazards* **2001**, *3*, 37–48. [[CrossRef](#)]
42. Pervez, M.S.; Henebry, G.M. Projections of the Ganges-Brahmaputra precipitation—Downscaled from GCM predictors. *J. Hydrol.* **2014**, *517*, 120–134. [[CrossRef](#)]
43. Pervez, M.S.; Henebry, G.M. Spatial and seasonal responses of precipitation in the Ganges and Brahmaputra river basins to ENSO and Indian Ocean dipole modes: Implications for flooding and drought. *Nat. Hazards Earth Syst. Sci.* **2015**, *15*, 147. [[CrossRef](#)]
44. Nishat, A.; Faisal, I.M. An assessment of the institutional mechanisms for water negotiations in the Ganges-Brahmaputra-Meghna system. *Int. Negot.* **2000**, *5*, 289–310. [[CrossRef](#)]
45. Lehner, B.; Verdin, K.; Jarvis, A. *Technical Documentation Version 1.0*; USGS Earth Resources Observation and Science: Sioux Falls, SD, USA, 2006.
46. Tateishi, R.; Hoan, N.T.; Kobayashi, T.; Alsaadeh, B.; Tana, G.; Phong, D.X. Production of global land cover data-GLCNMO2008. *J. Geogr. Geol.* **2014**, *6*, 99. [[CrossRef](#)]
47. Chowdhury, M.R.; Sato, Y. Flood monitoring in Bangladesh: Experience from normal and catastrophic floods. *J. Jpn. Assoc. Hydrol. Sci.* **1996**, *26*, 241–252.
48. Duane, S.; Kennedy, A.D.; Pendleton, B.J.; Roweth, D. Hybrid monte Carlo. *Phys. Lett. B* **1987**, *195*, 216–222. [[CrossRef](#)]
49. Boulkaibet, I.; Mthembu, L.; Marwala, T.; Friswell, M.I.; Adhikari, S. Finite element model updating using Hamiltonian Monte Carlo techniques. *Inverse Probl. Sci. Eng.* **2017**, *25*, 1042–1070. [[CrossRef](#)]
50. Cheung, S.H.; Beck, J.L. Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters. *J. Eng. Mech.* **2009**, *135*, 243–255. [[CrossRef](#)]
51. Beskos, A.; Pillai, N.; Roberts, G.; Sanz-Serna, J.-M.; Stuart, A. Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli* **2013**, *19*, 1501–1534. [[CrossRef](#)]
52. Hanson, K.M. Markov Chain Monte Carlo posterior sampling with the Hamiltonian method. In *Medical Imaging 2001: Image Processing*; International Society for Optics and Photonics: Washington, DC, USA, 2001; Volume 4322, pp. 456–468.
53. Carpenter, B.; Gelman, A.; Hoffman, M.D.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker, M.; Guo, J.; Li, P.; Riddell, A. Stan: A probabilistic programming language. *J. Stat. Softw.* **2017**, *76*. [[CrossRef](#)]
54. Stan Modeling Language: User’s Guide and Reference Manual. Version 2.17.0. Stan Development Team, 2017. Available online: <http://mc-stan.org/users/documentation/> (accessed on 5 September 2017).
55. Von Neumann, J. Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Stat.* **1941**, *12*, 367–395. [[CrossRef](#)]
56. Machiwal, D.; Jha, M.K. Comparative evaluation of statistical tests for time series analysis: Application to hydrological time series/Evaluation comparative de tests statistiques pour l’analyse de séries temporelles: Application à des séries temporelles hydrologiques. *Hydrol. Sci. J.* **2008**, *53*, 353–366. [[CrossRef](#)]
57. Alexandersson, H. A homogeneity test applied to precipitation data. *Int. J. Climatol.* **1986**, *6*, 661–675. [[CrossRef](#)]
58. Khaliq, M.N.; Ouarda, T. On the critical values of the standard normal homogeneity test (SNHT). *Int. J. Climatol.* **2007**, *27*, 681–687. [[CrossRef](#)]

59. Said, S.E.; Dickey, D.A. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **1984**, *71*, 599–607. [[CrossRef](#)]
60. Banerjee, A.; Dolado, J.J.; Galbraith, J.W.; Hendry, D. *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*; OUP Catalogue: Oxford, UK, 1993.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).