



Lund University, Faculty of Science, Mathematical Statistics  
Bachelor thesis in Mathematical Statistics

---

**An Extreme Value Approach to Modeling  
Risk of Extreme Rainfall in Bangladesh**

---

*Authors :*  
Emmy Andersson  
Evelina Nilsson

*Supervisor :*  
Nader Tajvidi

May 7, 2018

## Abstract

The extreme value theory has been applied on daily rainfall in the five most exposed areas of Bangladesh between the years 1980-2016 in order to estimate extreme rainfalls for the next 10, 50 and 100 years. These types of computations are necessary for optimising planning and preparations for extreme future rainfalls which can lead to minimising property damage and ultimately saving lives.

Generalised extreme value distribution is fitted to annual maxima according to the block maxima method. In addition, the generalised Pareto distribution is fitted to the daily rainfall according to the Peaks-Over-Threshold method. The different parameters were estimated with a 95 % confidence interval both through the delta- and the profile likelihood methods. Thereafter, the return period is computed according to each model using the same estimated confidence intervals.

## Populärvetenskaplig sammanfattning

Extremvärdesteori har applicerats på daglig nederbörd i Bangladesh, mellan 1980-2016, från fem av de mest utsatta områdena för att beräkna den förväntade nederbörden de kommande 10, 50 och 100 åren. Denna typ av beräkningar är väsentliga för att kunna vidta åtgärder som i sin tur minskar skador på egendom och framförallt räddar liv.

Generaliserad extremvärdesfördelning har använts för att anpassa årligt maxima enligt block maxima-modellen. Även generaliserad Pareto fördelning används för att anpassa den dagliga nederbörden enligt peaks over threshold-metoden. De olika parametrarna skattades med ett 95 % konfidensintervall genom både delta- och profile likelihood-metoden. Därefter beräknades den förväntade nederbörden med båda tidigare nämnda modeller, även dessa med ett 95 % konfidensintervall.

## **Acknowledgement**

We would first like to give a special thank you to our supervisor Nader Tajvidi for all his help through this project. He has been a great support throughout the whole process.

We would also like to thank Nasrine Sultana for providing us with all the necessary data and articles for this research.

# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>                                    | <b>5</b>  |
| <b>2</b> | <b>Theory</b>  | <b>7</b>  |
| 2.1      | Generalized extreme value distribution (GEV) . . . . . | 7         |
| 2.1.1    | Block size selection . . . . .                         | 9         |
| 2.1.2    | Return level . . . . .                                 | 10        |
| 2.1.3    | Inference for GEV distribution . . . . .               | 10        |
| 2.2      | Generalized Pareto distribution (GPD) . . . . .        | 11        |
| 2.2.1    | Threshold selection . . . . .                          | 13        |
| 2.2.2    | Return level . . . . .                                 | 14        |
| 2.2.3    | Inference of GPD . . . . .                             | 16        |
| 2.3      | Model Diagnostic . . . . .                             | 16        |
| 2.3.1    | Probability plot . . . . .                             | 16        |
| 2.3.2    | Quantile plot . . . . .                                | 17        |
| 2.4      | Probable maximum rainfall . . . . .                    | 17        |
| <b>3</b> | <b>Methodology</b>                                     | <b>18</b> |
| 3.1      | Data . . . . .   | 19        |
| 3.2      | Block-size selection for GEV . . . . .                 | 21        |
| 3.3      | Threshold selection for GPD . . . . .                  | 22        |
| <b>4</b> | <b>Rainfall pattern for the different stations</b>     | <b>26</b> |
| 4.1      | Barisal . . . . .                                      | 26        |
| 4.2      | Chandpur . . . . .                                     | 26        |
| 4.3      | Satkhira . . . . .                                     | 26        |
| 4.4      | Madaripur . . . . .                                    | 26        |
| 4.5      | Faridpur . . . . .                                     | 27        |
| <b>5</b> | <b>Results</b>   | <b>28</b> |
| 5.1      | Parameter estimation for extreme rainfall . . . . .    | 28        |
| 5.2      | Return level for extreme rainfall . . . . .            | 31        |
| 5.3      | Probable maximum rainfall . . . . .                    | 34        |
| <b>6</b> | <b>Conclusions</b>                                     | <b>35</b> |
| <b>7</b> | <b>References</b>                                      | <b>37</b> |

# 1 Introduction

Bangladesh is a country in southern Asia that often experiences heavy flooding due to its geographical location as well as its flat and low level landscape. The country is located on a latitude subject to tropical monsoon climates which brings heavy rainfall to the area. Brahmaputra and Ganges are two large rivers running through the country creating large areas of river delta flowing into the Bay of Bengal. The heavy rainfall that can be experienced during just a few hours leads to devastating consequences for the population, making Bangladesh one of the most vulnerable countries to flooding, [6].

Global warming has also a major effect on the amount of rain Bangladesh experiences. According to IPCC AR:5 Bangladesh is expected to have a 5 % increase of rainfall by year 2030, [8]. Between years 1996-2015 there were 13 581 recorded casualties due to extreme weather events in Bangladesh, making it the 6th most exposed country, [7]. The rising temperatures and more frequent flooding is affecting the agriculture leading to scarce food and water resources, which is a big threat to a country which population is rapidly growing, [8].

Due to these circumstances in Bangladesh many have been forced to evacuate and leave their homes, and have become so called climate refugees. There are, however, plausible measures to be taken against these types of consequences that global warming brings involving improvement of irrigation systems, sea defence, flood management, and use of rain water. In order to act on these areas one depends on the forecast of the extensive rainfall and its associated damage, [8].

The extreme value theory has been developed considerably in the last couple of decades and is often used in computations regarding forecasts of flooding, storms, insurance claims, and price fluctuations, [9]. Two approaches exist for practical extreme value analysis; the block maxima and peaks over threshold method. Both methods have been used in this thesis in order to predict rainfall the coming 10, 50 and 100 years in Bangladesh, [10].

All annual maxima, in the block maxima method, and values above a given threshold, in the peaks over threshold method, have been assumed to be independent. This thesis only treats univariate analysis extreme value theory,

that means that all stations have been analyzed separately.

This thesis is structured as follows. Section 2 describes the theory of the different methods used in this analysis. Followed by section 3 which is a description of the procedure and the different trade-offs that was made. Section 4 shortly describes the different rainfall patterns from the 5 different measure stations and in section 5 all av the results and plots are presented. In the last part, section 6, the results are discussed and analyzed.

## 2 Theory

In this section a summary of the univariate extreme value theory is presented. Also a short description of the statistical models that have been applied in this thesis is presented in this section.

### 2.1 Generalized extreme value distribution (GEV)

Let  $X_1, \dots, X_n$  be a sequence of independent random variables having common distribution function  $F$ . Suppose  $M_n$  represents the maximum of a sequence for a given  $n$ ,

$$M_n = \max(X_1, \dots, X_n).$$

If  $n$  is the number of observations in a year, then  $M_n$  corresponds to the annual maximum.

The distribution of  $M_n$  can be derived exactly for all values of  $n$  :

$$\begin{aligned} Pr(M_n \leq z) &= Pr(X_1 \leq z, \dots, X_n \leq z) \\ &= Pr(X_1 \leq z) \times \dots \times Pr(X_n \leq z) = (F(z))^n \end{aligned}$$

Since the distribution function  $F$  is unknown this is not very helpful in practice. Therefore, we proceed by looking at the behavior of  $F^n$  as  $n \rightarrow \infty$ . For any  $z \leq z_+$ , where  $z_+$  is the upper end-point of  $F$ ,  $F^n(z) \rightarrow 0$  as  $n \rightarrow \infty$ , so that the distribution of  $M_n$  degenerates to a point mass on  $z_+$ . To find a non-degenerate distribution let,

$$M_n^* = \frac{M_n - b_n}{a_n}.$$

Further, suppose  $\{a_n > 0\}$  and  $\{b_n\}$  are some constants. Appropriate choices of the  $\{a_n\}$  and  $\{b_n\}$  stabilize the location and scale of  $M_n^*$  as  $n$  increases, avoiding the difficulties that arise with the variable  $M_n$ . We therefore seek



limit distributions for  $M_n^*$ , with appropriate choices of  $\{a_n\}$  and  $\{b_n\}$ , rather than  $M_n$ .

**Theorem 2.1**  
(Coles 2004, [9])

If there exist sequence of constants  $\{a_n > 0\}$  and  $\{b_n \geq 0\}$  such that

$$Pr\left(\frac{M_n - b_n}{a_n} \leq z\right) \rightarrow G(z) \quad \text{as } n \rightarrow \infty$$

where  $G$  is a non-degenerate distribution function, then  $G$  belongs to one of the following families:

Gumbel:

$$G(z) = \exp\{-\exp[-\frac{z-b}{a}]\}, \quad -\infty < z < \infty$$

Fréchet:

$$G(z) = \begin{cases} 0, & z \leq b, \\ \exp\{-(\frac{z-b}{a})^\alpha\}, & z > b; \end{cases}$$

Weibull:

$$G(z) = \begin{cases} \exp\{-[-(\frac{z-b}{a})^\alpha]\}, & z \leq b, \\ 1, & z > b; \end{cases}$$

for some parameters  $a > 0$  and  $b \in R$   $\alpha > 0$ .

Theorem 2.1 states that the normalized sample maxima  $\frac{M_n - b_n}{a_n}$  converges in distribution to a random variable having a distribution which belongs to one of the families Gumbel, Fréchet or Weibull. The common name for these three distributions is extreme value distribution. All families have a location and scale parameter,  $b$  and  $a$ . Only the Fréchet and Weibull families have a shape parameter,  $\alpha$ .

Gumbel, Fréchet and Weibull can be combined into a single family of models having distribution function of the form:

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

**Theorem 2.2**

(Coles 2004, [9])

If there exist sequences of constants  $\{a_n > 0\}$  and  $\{b_n\}$  such that

$$Pr\{\frac{M_n - b_n}{a_n} \leq z\} \rightarrow G(z) \quad \text{as } n \rightarrow \infty$$

for a non-degenerate distribution function  $G$ , then  $G$  is a member of the GEV family

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

defined on  $\{z : 1 + \xi(\frac{z - \mu}{\sigma}) > 0\}$ , where  $-\infty < \mu < \infty, \sigma > 0$  and  $-\infty < \xi < \infty$ .

Suppose we have  $X_1, X_2, \dots, X_n$  observation that are independent identical distributed divided into non-overlapping blocks  $m$  of equal size  $n$ . Then block maximas,  $M_{n,1}, M_{n,2}, \dots, M_{n,m}$ , are the highest value in each block. The distribution of the block maximas can be approximated by the asymptotically GEV distribution defined in Theorem 2.2.

**2.1.1 Block size selection**

When using the block maxima approach in extreme value theory GEV distribution is fitted to the block maximas. The choice of the block size is a trade-off between bias and variance. If the blocks are too small it leads to bias in the estimation. If the block on the other hand is too large there won't be enough block maxima which leads to large estimation variance.

### 2.1.2 Return level

From the fitted generalized extreme value distribution the occurrence of the extreme quantile can be calculated with a certain return level.

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - \{-\log(1-p)\}^{-\xi}] & \text{for } \xi \neq 0 \\ \mu - \sigma \log\{-\log(1-p)\} & \text{for } \xi = 0 \end{cases}$$

where  $z_p$  is the return level connected with the return period  $1/p$ . This means that  $z_p$  is exceeded by the block maximum in any particular time period with probability  $p$ .

### 2.1.3 Inference for GEV distribution

The parameters of GEV distribution can be estimated by using maximum likelihood method which involves maximizing with respect to parameters  $\xi$ ,  $\mu$  and  $\sigma$ .

The difference between the three GEV families, Gumbel, Fréchet and Weibull is the value of the shape parameter,  $\xi$ . When  $\xi$  is less than zero the upper end-point of the distribution can be calculated:  $\mu - \frac{\sigma}{\xi}$ . In the same way the lower end-point can be calculated when  $\xi$  is greater than zero. Maximum likelihood estimate method possess asymptotic behavior when  $\xi > -0.5$ . [4]

MLE of the parameters:

$$\iota(\mu, \sigma, \xi) = m \log \sigma (1 + \frac{1}{\xi}) \sum_{i=1}^m \log [1 + \xi (\frac{z_i - \mu}{\sigma})] - \sum_{i=1}^m [1 + \xi (\frac{z_i - \mu}{\sigma})]^{-1\xi}$$

provided that

$$1 + \xi (\frac{z_i - \mu}{\sigma}) > 0, \quad \text{for } i = 1, \dots, m.$$

GEV distribution

$$\iota(\mu, \sigma) = -m \log \sigma - \sum_{i=1}^m (\frac{z_i - \mu}{\sigma}) - \sum_{i=1}^m \exp\{- (\frac{z_i - \mu}{\sigma})\}$$

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

where  $y_p = -\log(1 - p)$ . Furthermore, by the delta method,

$$\text{Var}(\hat{z}_p) \approx \nabla z_p^T \nabla z_p$$

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

$$\nabla z_p^T = \left[ \frac{\delta z_p}{\delta \mu}, \frac{\delta z_p}{\delta \sigma}, \frac{\delta z_p}{\delta \xi} \right] = [1, -\xi^{-1}(1 - y_p^{-\xi}), \sigma \xi^{-2}(1 - y_p^{-\xi}) - \sigma \xi^{-1} y_p^{-\xi} \log(y_p)]$$

evaluated at  $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ .

## 2.2 Generalized Pareto distribution (GPD)

If an entire time series of daily observations are available peaks over threshold is a more suitable method than the block maxima method. Instead of splitting the data into different blocks the peaks over threshold method is based on fitting generalized Pareto distribution to all values that exceeds a certain threshold  $u$ .

Specifically let  $X_1, X_2, \dots, X_n$  be a sequence of i.i.d random variables, having marginal distribution  $F$ . Denoting an arbitrary term in the  $X_i$  sequence by  $X$ , it follows that a description of the stochastic behavior of extreme events is given by the conditionally probability.

$$\text{Pr}\{X > u + y \mid X > u\} = \frac{1 - F(u + y)}{1 - F(u)}, y > 0.$$

The distribution for the excesses over threshold  $u$  is obtained by using GEV distribution as an approximation to the new distribution GPD.

**Theorem 2.3**  
(Coles 2004, [9])

Let  $X_1, X_2, \dots, X_n$  be a sequence of independent random variables with common distribution function  $F$ , and let  $M_n = \max\{X_1, \dots, X_n\}$ . Denote an arbitrary term in the  $X_i$  sequence by  $X$ , and suppose that  $F$  satisfies  $Pr(\frac{M_n - b_n}{a_n} \leq z) \rightarrow G(z)$  as  $n \rightarrow \infty$ . For large  $n$ ,

$$Pr\{M_n \leq z\} \approx G(z),$$

where

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

For some  $\mu, \sigma > 0$  and  $\xi$ . Then it can be shown that for large enough  $u$ , the distribution function of  $(X - u)$ , conditional on  $X > u$ , is approximately,

$$H(y) = 1 - (1 + \frac{\xi y}{\hat{\sigma}})^{-\frac{1}{\xi}}$$

Defined on  $\{y : y > 0 \text{ and } (1 + \frac{\xi y}{\hat{\sigma}}) > 0\}$ , where  $\hat{\sigma} = \sigma + \xi(u - \mu)$ .

When  $\xi \rightarrow 0$  the generalized Pareto distribution is defined as follows:

$$H(y) = 1 - \exp(-\frac{y}{\hat{\sigma}}), \quad y > 0,$$

Which corresponds to an exponential distribution with parameters  $\frac{1}{\hat{\sigma}}$

Theorem 2.3 describes the connection between GEV distribution and GPD. If the block maximas have approximating distribution  $G$  then threshold excesses have a corresponding approximate distribution within the generalized

Pareto family. The parameters in GPD are uniquely determined by the associated parameters of the GEV distribution. The  $\xi$  parameter is dominant in determining the behavior of the GPD, just as for the GEV distribution. For  $\xi < 0$  the distribution is bounded,  $\xi > 0$  indicates that no upper bound exists and when  $\xi = 0$  the distribution is unbounded.

### 2.2.1 Threshold selection

Just as with block maxima approach, the choice of threshold is a balance between bias and variance. If the chosen threshold is too high it will result in few exceedances and the analysis is unlikely to yield any useful results which will cause high variance. If the threshold instead is too low the larger amount of measurements will be taken into account, leading to bias. By studying mean residual life plots and threshold range plots an appropriate threshold can be selected.

#### Mean residual life plot

The mean residual life plots is based on the mean of GPD. If  $Y$  has a generalized Pareto distribution with parameters  $\sigma$  and  $\xi$ , then

$$E(Y) = \begin{cases} \frac{\sigma}{1-\xi}, & \text{for } \xi < 1 \\ \infty & \text{for } \xi \geq 1 \end{cases}$$

Suppose that that generalized Pareto distribution is valid as a model for excesses of a certain threshold  $u_0$  generated by a series  $X_1, X_2, \dots, X_n$  of which an arbitrary term is denoted by  $X$ . From previous equation we get,  $E(X - u_0 | X > u_0) = \frac{\sigma_{u_0}}{1-\xi}$ , provided  $\xi < 1$  When Pareto distribution is valid for excesses of the threshold  $u_0$ , it is also valid for for all thresholds  $u > u_0$ . For  $u > u_0$ ,

$$E(X - u | X > u) = \frac{\sigma_u}{1-\xi} = \frac{\sigma_{u_0} + \xi u}{1-\xi}$$

$$\left\{ \left( u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) : u < x_{max} \right) \right\},$$

where  $x_{(1)}, \dots, x_{(n_u)}$  consist of the  $n_u$  observations that exceed  $u$ , and  $x_{max}$  is the largest of the  $X_i$ , is termed the mean residual life plot. Above the threshold  $u_0$  at which the generalized Pareto distribution provides a valid approximation to the excesses distribution, the mean residual life plot should be approximately linear above  $u$ .

### Threshold range plot

Threshold range plots can be used as a complement to the mean residual life plot when choosing threshold.

If the generalized Pareto distribution is a reasonable model for excesses of a threshold  $u_0$  then all excesses should also follow the same distribution. However, the value of the generalized Pareto scale parameter for a threshold  $u > u_0$  changes with  $u$  unless  $\xi = 0$ ,

$$\sigma_u = \sigma_{u_0} + \xi(u - u_0).$$

For  $\xi = 0$  the generalized Pareto scale parameter can be reparameterized as follows,

$$\sigma^* = \sigma_u - \xi u.$$

### 2.2.2 Return level

Suppose that a generalized Pareto distribution with parameters  $\sigma$  and  $\xi$  is a suitable model for exceedances of a threshold  $u$  by a variable  $X$ . That is, for  $x > u$ ,

$$Pr\{X > x \mid X > u\} = [1 + \xi(\frac{x - u}{\sigma})]^{-\frac{1}{\xi}}.$$

It follows that

$$Pr\{X > x\} = \zeta_u [1 + \xi(\frac{x - u}{\sigma})]^{-\frac{1}{\xi}},$$

where  $\zeta_u = Pr\{X > u\}$ . Hence, the level  $x_m$  that is exceeded on average once every  $m$  observations is the solution of

$$\zeta_u [1 + \xi (\frac{x_m - u}{\sigma})]^{-\frac{1}{\xi}} = \frac{1}{m}$$

Rearranging,

$$x_m = u + \frac{\sigma}{\xi} [(m\zeta_u)^\xi - 1]$$

Provided  $m$  is sufficiently large to ensure that  $x_m > u$ . This all assumes that  $\xi \neq 0$ . If  $\xi = 0$ , working in the same way with,  $H(y) = 1 - \exp(-\frac{y}{\sigma}, y > 0$ , leads to

$$x_m = u + \sigma \log(m\zeta_u),$$

again provided  $m$  is sufficiently large.

Where  $x_m$  is the  $m$ -observation return level. In general it is often more suitable to present the return level on an annual scale so that the  $N$ -year return level is the level expected to be exceeded once every  $N$  years. Let  $n_y$  be the number of observations per year, then it corresponds to the  $m$ -observation return level, where  $m = Nn_y$ . The  $N$ -year return level is defined by:

$$z_N = u + \frac{\sigma}{\xi} [(Nn_y\zeta_u)^\xi - 1] \quad \text{for } \xi \neq 0$$

$$z_N = u + \sigma \log(Nn_y\zeta_u) \quad \text{for } \xi = 0$$



### 2.2.3 Inference of GPD

Just as for GEV we maximize the log likelihood function to find the estimates of the parameters.

Parameter estimation:

$$\ell(\sigma, \xi) = -k \log(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^k \log\left(1 + \frac{\xi y_i}{\sigma}\right),$$

provided that  $(1 + \sigma_o^{-1}) > 0$  for  $i = 1, \dots, k$ ; otherwise,  $\ell(\sigma, \xi) = -\infty$ . In the case  $\xi = 0$  the log-likelihood is obtained from  $H(y) = 1 - \exp(-\frac{y}{\sigma})$ ,  $y > 0$  as

$$\ell(\sigma) = -k \log(\sigma) - \sigma^{-1} \sum_{i=1}^k y_i.$$

## 2.3 Model Diagnostic

In order to make sure that GEV distribution or GPD are a suitable distribution diagnostics plots are analyzed.

### 2.3.1 Probability plot

A probability plot consist of the points ,

$$\left\{ \left( \hat{F}(x_{(i)}), \frac{i}{n+i} \right) : i = 1, \dots, n \right\}$$

given that an ordered sample  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  is independent observations from an estimated distribution function  $\hat{F}$ .

If all the points of the probability plot lies close to the unit diagonal the  $\hat{F}$  is a reasonable model distribution.

### 2.3.2 Quantile plot

A quantile plot is the inverse of the probability plot and consists of the points

$$\left\{ \left( \hat{F}^{-1} \left( \frac{i}{n+1} \right), x_{(i)} \right) : i = 1, \dots, n \right\}.$$

Given an ordered sample of independent observation  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  with estimated distribution function  $\hat{F}$ .

If  $\hat{F}$  is a reasonable estimate of  $F$ , then the quantile plot should also consist of points close to the unit diagonal.

## 2.4 Probable maximum rainfall

(Rootzén and Tajvidi, 1997, [4]).

Let  $M_T$  be the largest observation during a time period of length  $T$ . Thus  $M_T \leq u$ , or there is at least one excess of  $u$ . The excesses of the level  $u+v$  for  $v > 0$  occur as an Poisson Process with intensity  $\lambda(1 + \frac{\xi v}{\sigma})^{\frac{1}{\xi}}$ . The maximum is smaller than  $u+v$  if and only if the Poisson process has no points in the interval  $[0, T]$ . It follows that:

$$Pr\{M_T \leq u+v\} = exp(-\lambda T(1 + \frac{\xi v}{\sigma})^{\frac{1}{\xi}}) = exp\left\{-\left(1 + \xi \frac{v - ((\lambda T)^\xi - 1)\frac{\sigma}{\xi}}{\sigma(\lambda T)^\xi}\right)_+^{-\frac{1}{\xi}}\right\}.$$

By equating the equation to  $1-p$  and solving for  $v$ , the  $p$ -th upper quantile of the distribution of  $M_t$  is obtained as:

$$x_{T,p} = u + \frac{\sigma}{\xi} \left( \frac{(\lambda T)^\xi}{(-\log(1-p))^\xi} - 1 \right).$$

Where  $x_{T,p}$  is the upper quantile for the risk level  $p$  and a time period of length  $T$ .

### 3 Methodology

Daily rainfall data recorded from 17 different measure stations during the period 1980-2016 was collected from Bangladesh Meteorological Department Climate Division, Agargaon, Dhaka-1207. All 17 stations are located in the southern part of Bangladesh. Based on National plan for disaster management (2010-2015), [13], the five stations that were most exposed to flooding was chosen for this thesis. These stations are Chandpur, Barisal, Faridpur, Madaripur and Satkhira, they are all located in the south-west part of Bangladesh. Station Chandpur was missing data from the whole year 1980. When other stations were compared with Chandpur data from 1981-2016 were used.



Figure 1: Map over Bangladesh, [14]. The red dots represents the 17 stations that we received daily rainfall data from. The five larger dots are the chosen stations.

### 3.1 Data

Daily, monthly maximum and annual maximum rainfall during one day was plotted for all stations. Plots for Station Chandpur is presented in Figure 2 below.

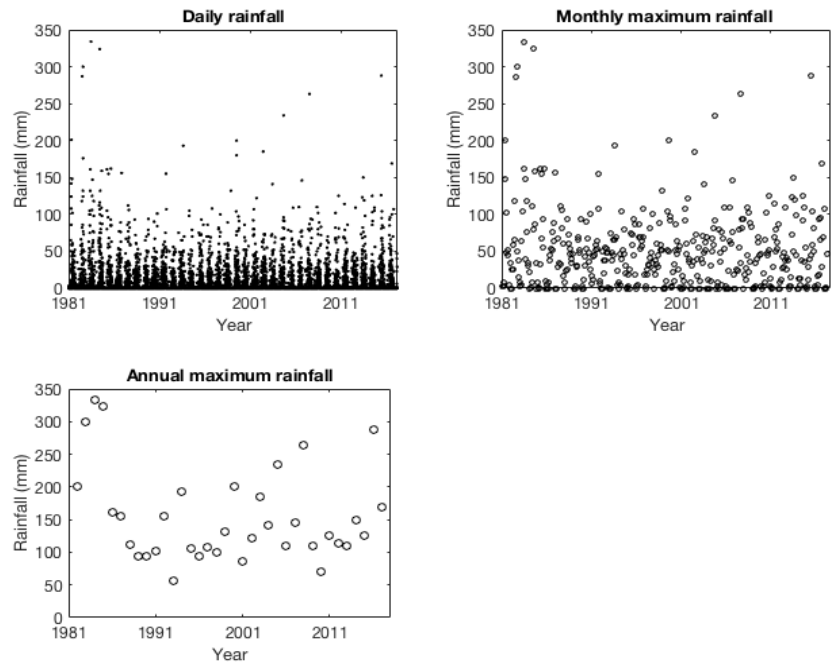


Figure 2: Trend check plots for station Chandpur.

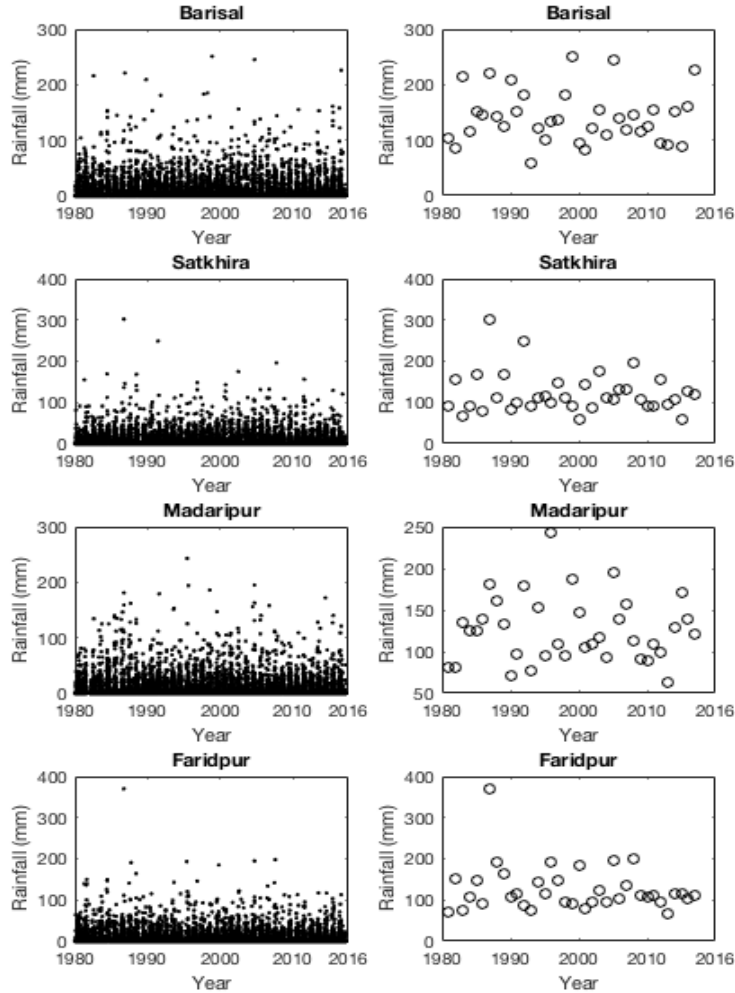


Figure 3: Daily and annual maximum rainfall plots for stations Barisal, Satkhira, Madaripur and Faridpur.

The daily rainfall plot and monthly maximum in Figure 2 shows a clear trend depending on monsoon season. Fig 3 strengthens that it is true for all stations. When studying the annual maximum one can assume that the pattern of variation has stayed constant. This can be checked by analyzing autocorrelation function.

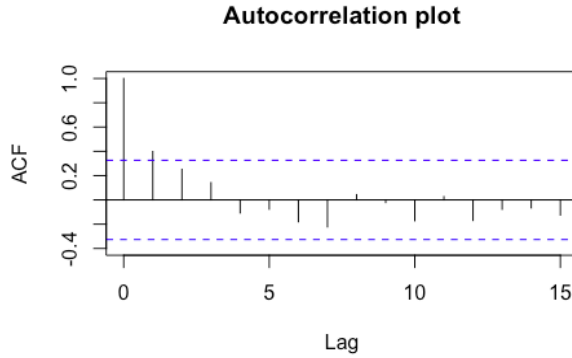


Figure 4: Autocorrelation function for annual maximum rainfall from station Chandpur.

Autocorrelation plots are a tool for checking randomness for time-series data. The x-axis represents the size of the lag between the elements of time-series. The lag 0 is always equal to one because it shows the autocorrelation between itself and each term. The higher value of the spikes the higher autocorrelation for each lag. All spikes that are between the two dashed lines is considered not correlated. The autocorrelation plot for Chandpur is represented in Fig 4. All stations can be assumed to be random time-series according to their autocorrelation plots.

### 3.2 Block-size selection for GEV

The different blocks were split up into size of one year. Every block maxima represents the highest amount of rainfall during one day for all years between 1980-2016. That is 37 block maxima except for station Chandpur which is missing data from 1980 and leaves us with 36 block maxima. Generalized extreme value distribution was fitted to the block maximas. In order to find out whether GEV was a suitable distribution diagnostic plots was analyzed.

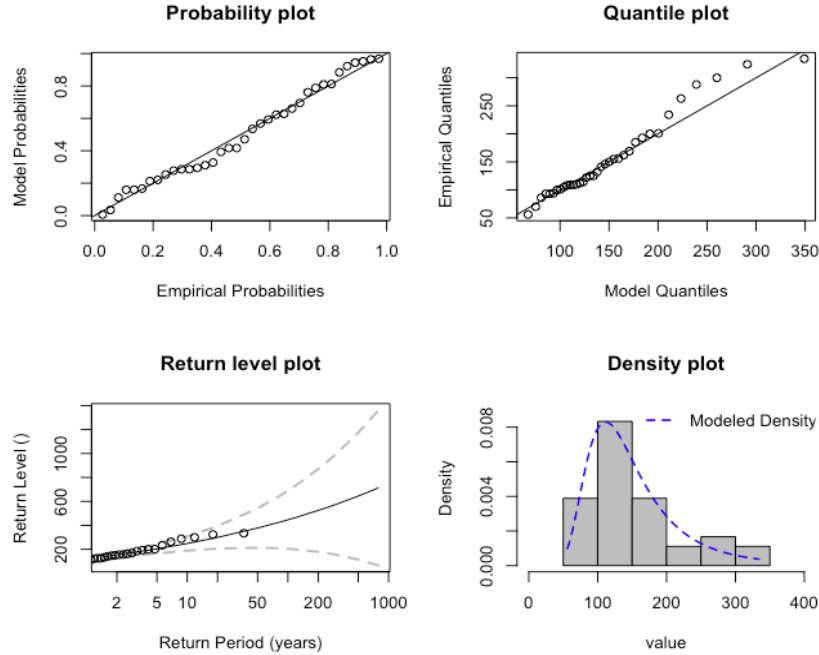


Figure 5: GEV diagnostic for Chandpur with annual maximum block.

As can be seen in Figure 5, the points of the probability plot and quantile plot lies close to the unit diagonal. This implies that generalized extreme value distribution function provides a good fit. The return level plot shows that the empirical return levels match well with those from the fitted distribution function. Finally, the density plot also shows good agreement between the fitted GEV distribution function and the empirical density.

### 3.3 Threshold selection for GPD

Thresholds were chosen based on mean residual life plots, threshold range plots and the number of exceeding values. When threshold was chosen diagnostics plots were analyzed in order to see if GPD was a suitable distribution for the exceeding values. Mean residual life plot, threshold range plots and diagnostic plots for station Chandpur are presented below.

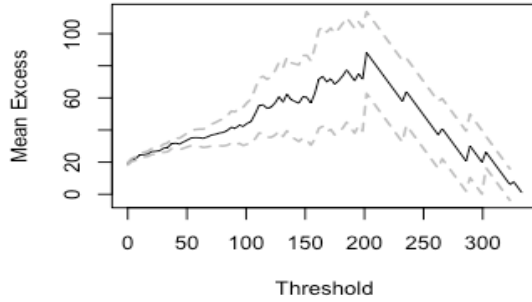


Figure 6: Mean residual life plot for station Chandpur.

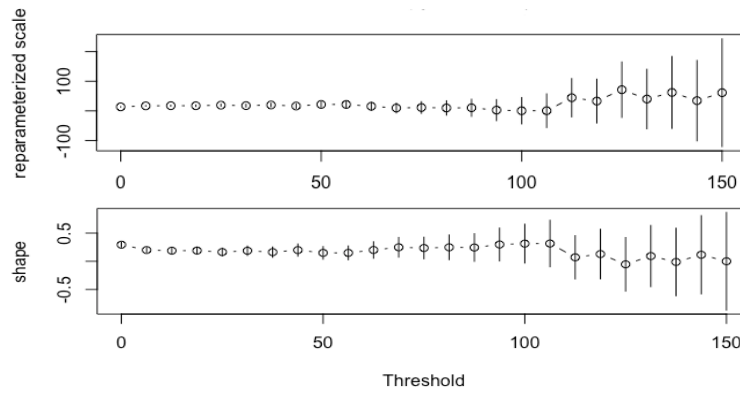


Figure 7: Threshold range plot for station Chandpur.

The threshold for Chandpur was chosen at 70. As can be seen in the mean residual life plot there is some evidence for linearity above  $u = 70$ . One can also see in the threshold range plot that the selected threshold seems reasonable. For threshold 70 there were 181 exceeding values. Generalized Pareto distribution was fitted to all 181 values.



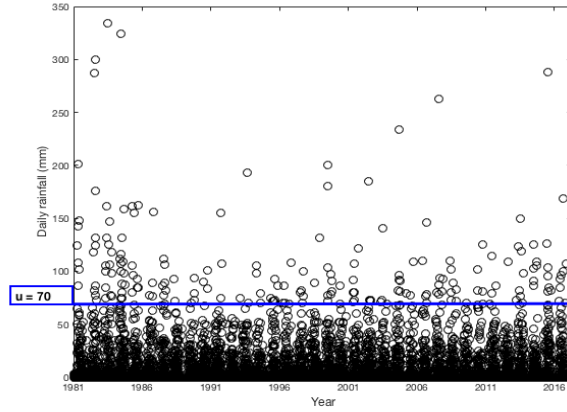


Figure 8: Threshold  $u=70$  for Chandpur.

Same procedure was made for all of the stations. The threshold selection for each station and the number of exceeding values can be found in Table 1 below.

Table 1: Thresholds and number of exceedances for all stations.

| Station   | Threshold | Number of values over threshold |
|-----------|-----------|---------------------------------|
| Barisal   | 70        | 143                             |
| Chandpur  | 70        | 181                             |
| Satkhira  | 60        | 145                             |
| Madaripur | 70        | 165                             |
| Faridpur  | 60        | 179                             |

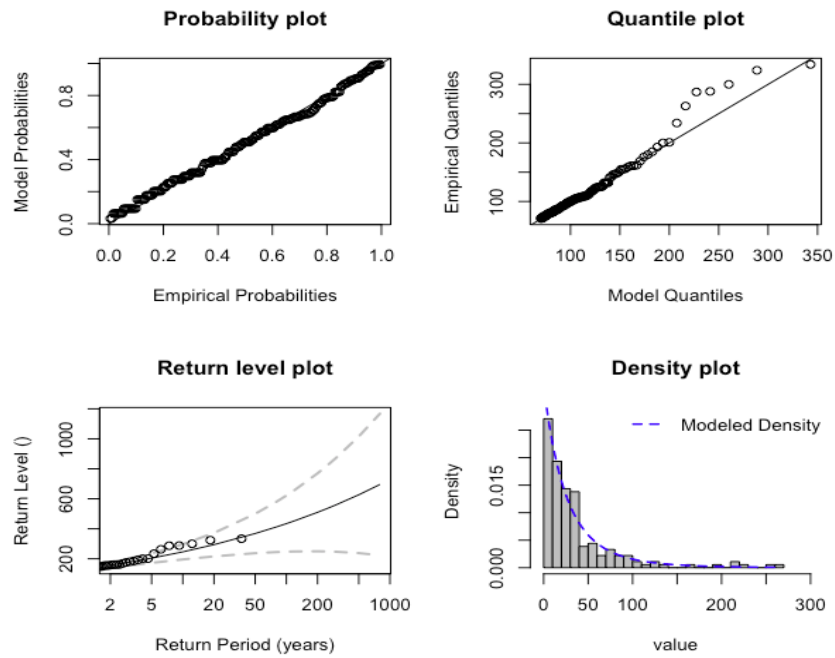


Figure 9: GPD diagnostic for Chandpur with threshold 70 mm.

Just as for GEV distribution diagnostic plots for GPD distribution was made. Figure 8 strengthens that GPD is a reasonable distribution for the exceeding values.

## 4 Rainfall pattern for the different stations

All stations have their wettest period in July and their driest period in December or January which agrees with the monsoon seasons.

### 4.1 Barisal

Barisal is the most southern measure station of all the stations. The maximum rainfall measured at this station was 251 mm which occurred in November 1998. The range of annual maximum during the period was [59,251]. The average annual maximum measured at station Barisal during the period 1980-2017 was 142,001 mm. The average annual rainfall was 2086,004 mm per year.

### 4.2 Chandpur

Chandpur is located in the southern part of Bangladesh and is adjacent to one of the largest rivers, Dakatia, [3]. The maximum rainfall during the measuring period was 334 mm in one day and occurred in June 1983. The lowest annual maximum was 56 mm. The average annual maximum rainfall measured at station Chandpur was 154.582 mm. The average annual rainfall was 2184,103 mm per year which is the highest average annual rainfall of all of the five stations.

### 4.3 Satkhira

Satkhira station is located in southwestern Bangladesh, near the Indian border. The maximum rainfall measured at this station was 302 mm which occurred in September 1986, the same year and month as Faridpurs maximum rainfall. The lowest annual maximum was measured to 58 mm. The average annual maximum measured at station Satkhira during the period 1980-2017 was 122,491 mm, the average annual rainfall was 1710,500 mm per year.

### 4.4 Madaripur

Faridpur is on the north of Madaripur, Barisal is on the south. The range of annual maximum during the period was [63,243], the highest value occurred

in June 1995. The average annual maximum measured at station Madaripur during the period 1980-2017 was 125,844 mm, the average annual rainfall was 1952,102 mm per year.

## **4.5 Faridpur**

The measure station Faridpur is located in the central part of Bangladesh near the capital Dhaka. The maximum rainfall measured at this station was 370 mm which occurred in September 1986 and the range of annual maximum rainfall during the period was [65,370]. The average annual maximum during the period was 126,572 mm and the average annual rainfall was 1812,181 mm per year.

## 5 Results

All parameters for GEV distribution and GPD have been estimated and are presented in Table 2 below. Return levels and confidence intervals for both distribution using both delta method and profile likelihood are presented and briefly explained in this section.

### 5.1 Parameter estimation for extreme rainfall

Maximum likelihood estimation method is used to estimate the parameters of generalized extreme value distribution and generalized Pareto distribution with a 95 % confidence interval. The parameters were estimated using two different methods, delta method and profile likelihood. The two methods gave very similar results for the various confidence intervals. The profile likelihood estimates are presented in the figures and tables below.

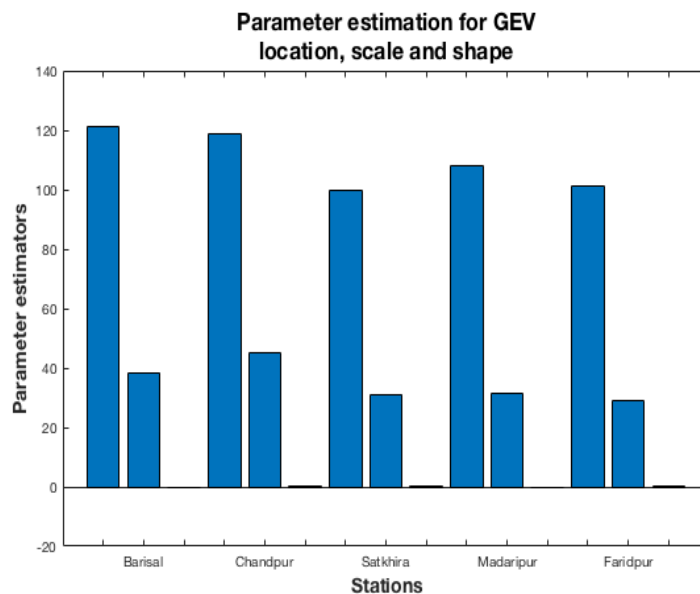


Figure 10: Parameter estimation for GEV distribution. Presented in order location, scale and shape parameters.

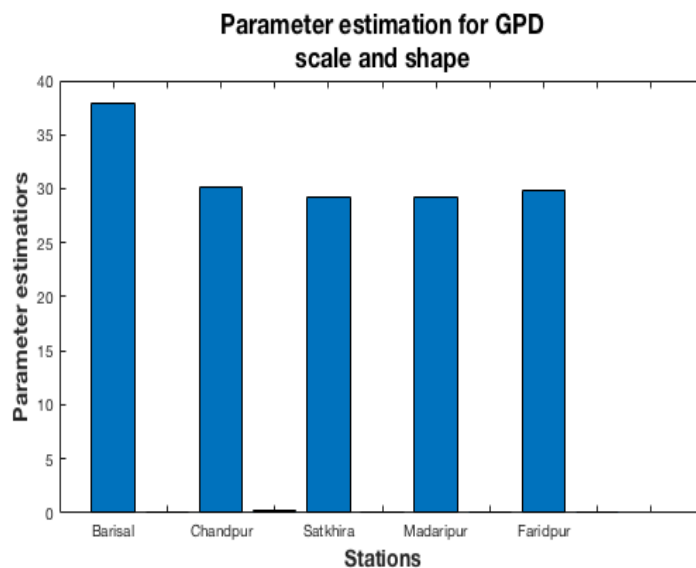


Figure 11: Parameter estimation for GPD. Presented in order scale and shape parameters.

Station Barisal has a negative shape parameter for both GEV distribution and GPD which implies that the distribution belongs to the Weibull family. The confidence interval for station Barisal includes positive values as well so we can not ignore the fact that it might belong to one of the other families. In the case when the maximum likelihood estimate of shape parameter  $\xi < 0$  this implies that the support of the GEV distribution is bounden.

The shape parameters of Madaripur are of opposite sign for GEV and GPD. The confidence intervals have values in both domains. Because of this it might belong to the Weibull family but assumptions like that can not be made.

Station Chandpur the shape parameter is positive and when looking at the GPD confidence interval all values are positive. This strengthens that it belongs to the Fréchet family with no upper end-point.

Faridpur and Satkhira have positive shape parameters but since the confidence intervals includes both negative and positive values any assumptions can not be made.

Table 2: Parameter estimation, standard deviation and a 95 % confidence intervals for GEV and GPD using profile likelihood.

| Station   |          | GEV      |        |        | GPD    |        |
|-----------|----------|----------|--------|--------|--------|--------|
|           |          | Location | Scale  | Shape  | Scale  | Shape  |
| Barisal   | Estimate | 120.987  | 38.389 | -0.038 | 37.864 | -0.018 |
|           | Sd       | 7.185    | 5.205  | 0.134  | 4.596  | 0.088  |
|           | Lower CI | 106.905  | 28.188 | -0.301 | 28.857 | -0.191 |
|           | Upper CI | 135.068  | 48.590 | 0.224  | 46.872 | 0.154  |
| Chandpur  | Estimate | 118.927  | 45.055 | 0.185  | 30.154 | 0.196  |
|           | Sd       | 8.152    | 6.703  | 0.138  | 3.390  | 0.086  |
|           | Lower CI | 102.243  | 31.918 | -0.085 | 25.510 | 0.029  |
|           | Upper CI | 135.610  | 58.192 | 0.455  | 36.798 | 0.364  |
| Satkhira  | Estimate | 99.743   | 31.100 | 0.135  | 29.269 | 0.064  |
|           | Sd       | 5.733    | 4.366  | 0.122  | 3.289  | 0.078  |
|           | Lower CI | 88.508   | 22.542 | -0.104 | 22.822 | -0.090 |
|           | Upper CI | 110.979  | 39.657 | 0.373  | 35.715 | 0.064  |
| Madaripur | Estimate | 108.044  | 31.245 | -0.012 | 29.168 | 0.026  |
|           | Sd       | 5.890    | 4.346  | 0.138  | 3.534  | 0.093  |
|           | Lower CI | 96.500   | 22.728 | -0.283 | 22.242 | -0.156 |
|           | Upper CI | 119.588  | 39.763 | 0.259  | 36.094 | 0.208  |
| Faridpur  | Estimate | 101.455  | 28.883 | 0.233  | 29.880 | 0.065  |
|           | Sd       | 5.405    | 4.362  | 0.139  | 3.022  | 0.068  |
|           | Lower CI | 90.862   | 20.333 | -0.040 | 23.957 | -0.069 |
|           | Upper CI | 112.049  | 37.433 | 0.504  | 35.802 | 0.199  |

## 5.2 Return level for extreme rainfall

Return levels are the estimates of the extreme quantiles of the distribution during a certain return period,  $\frac{1}{p}$ , in this case either 10, 50 or 100 years. The confidence intervals are calculated with a 95 % significance level. Both delta method and profile likelihood have been used and they are both presented in the two tables below. According to the delta method station Chandpur's 100 year return level is 282.833 mm for GEV distribution, this means that the expected rainfall 282.833 is to be exceeded on average once every 100 years.

The estimates of the return levels are very similar for both GEV distribution and GPD. As can be seen in the tables below the profile likelihood method presents confidence intervals with higher values and over all smaller intervals compared to the delta method. Station Chandpur has significantly higher values compared to the other stations when it comes to both delta method and profile likelihood. Although there is a difference both fits presents return levels that are quite similar. Figure 12 shows that the confidence intervals for the profile likelihood approach are quite asymmetric. The confidence intervals implies that the data provides weaker information for the higher return periods.

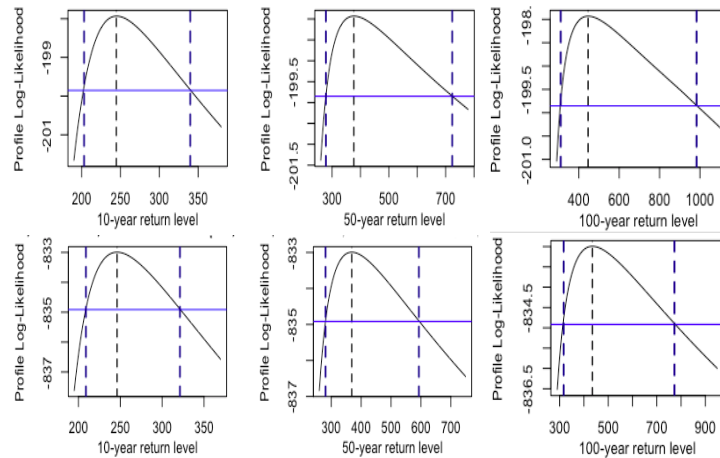


Figure 12: Profile log-likelihood return level plots for the GEV and GPD fit for station Chandpur.



Table 3: Estimated return levels with a 95% confidence intervals for GEV and GPD using profile likelihood.

| Station   |          | GEV     |         |          | GPD     |         |          |
|-----------|----------|---------|---------|----------|---------|---------|----------|
|           |          | 10 year | 50 year | 100 year | 10 year | 50 year | 100 year |
| Barisal   | Estimate | 203.741 | 260.074 | 282.833  | 202.947 | 259.123 | 282.810  |
|           | Lower CI | 180.887 | 220.870 | 233.250  | 182.043 | 228.670 | 234.805  |
|           | Upper CI | 248.880 | 388.853 | 467.763  | 244.719 | 361.900 | 423.308  |
| Chandpur  | Estimate | 244.682 | 376.650 | 445.763  | 246.130 | 368.573 | 434.443  |
|           | Lower CI | 203.043 | 278.128 | 310.111  | 212.746 | 281.731 | 345.332  |
|           | Upper CI | 339.838 | 722.271 | 982.311  | 321.159 | 594.286 | 775.444  |
| Satkhira  | Estimate | 181.484 | 259.352 | 297.846  | 182.337 | 245.243 | 274.399  |
|           | Lower CI | 155.528 | 204.523 | 223.877  | 167.872 | 220.334 | 227.450  |
|           | Upper CI | 236.432 | 444.221 | 586.498  | 223.228 | 351.327 | 424.693  |
| Madaripur | Estimate | 177.404 | 227.114 | 247.831  | 185.662 | 238.594 | 262.088  |
|           | Lower CI | 157.804 | 193.094 | 204.804  | 169.746 | 214.557 | 218.075  |
|           | Upper CI | 217.695 | 352.208 | 434.833  | 226.756 | 344.667 | 411.501  |
| Faridpur  | Estimate | 186.851 | 284.971 | 339.205  | 190.782 | 255.815 | 285.978  |
|           | Lower CI | 157.934 | 214.482 | 237.399  | 174.656 | 226.581 | 233.909  |
|           | Upper CI | 255.299 | 560.256 | 745.462  | 230.827 | 355.131 | 425.180  |

Table 4: Estimated return levels with a 95 % confidence intervals for GEV and GPD using the delta method.

| Station   |          | GEV     |         |          | GPD     |         |          |
|-----------|----------|---------|---------|----------|---------|---------|----------|
|           |          | 10 year | 50 year | 100 year | 10 year | 50 year | 100 year |
| Barisal   | Estimate | 203.741 | 260.074 | 282.833  | 202.947 | 259.123 | 282.810  |
|           | Lower CI | 175.023 | 196.189 | 196.890  | 174.318 | 199.398 | 204.927  |
|           | Upper CI | 232.459 | 323.959 | 368.777  | 231.577 | 318.849 | 360.693  |
| Chandpur  | Estimate | 244.682 | 376.650 | 445.763  | 246.130 | 368.573 | 434.443  |
|           | Lower CI | 187.493 | 213.024 | 204.263  | 195.615 | 237.337 | 247.287  |
|           | Upper CI | 301.872 | 540.276 | 687.264  | 296.646 | 499.809 | 621.560  |
| Satkhira  | Estimate | 181.484 | 259.352 | 297.846  | 182.337 | 245.243 | 274.399  |
|           | Lower CI | 147.344 | 170.572 | 171.024  | 154.304 | 183.261 | 191.376  |
|           | Upper CI | 215.624 | 348.132 | 424.668  | 210.371 | 307.226 | 357.421  |
| Madaripur | Estimate | 177.404 | 227.114 | 247.831  | 185.662 | 238.594 | 262.088  |
|           | Lower CI | 152.582 | 169.747 | 169.567  | 158.605 | 178.018 | 181.487  |
|           | Upper CI | 202.227 | 284.482 | 326.095  | 212.719 | 299.171 | 342.689  |
| Faridpur  | Estimate | 186.851 | 284.971 | 339.205  | 190.782 | 255.815 | 285.978  |
|           | Lower CI | 146.771 | 163.173 | 155.022  | 162.692 | 195.693 | 206.251  |
|           | Upper CI | 226.931 | 406.769 | 523.388  | 218.873 | 315.937 | 365.705  |

### 5.3 Probable maximum rainfall

Return periods for different years and quantiles are presented in the table below. This method presents significantly higher values compared to block maxima or peaks over threshold method. This might be because this approach is based on significantly more measurements. Even in this method station Chandpur presents the highest predicted future rainfall.

Table 5: Estimated quantiles, probable maximum rainfall, in mm for different risk levels and time periods

| <b>Probable maximum rainfall</b> |                         |                 |                 |                  |
|----------------------------------|-------------------------|-----------------|-----------------|------------------|
| <b>Barisal</b>                   | <b>Risk/Time period</b> | <b>10 years</b> | <b>50 years</b> | <b>100 years</b> |
|                                  | <b>10%</b>              | 288.324         | 341.910         | 364.632          |
|                                  | <b>5%</b>               | 312.527         | 365.497         | 387.831          |
|                                  | <b>1%</b>               | 366.162         | 417.567         | 439.241          |
| <b>Chandpur</b>                  | <b>Risk/Time period</b> | <b>10 years</b> | <b>50 years</b> | <b>100 years</b> |
|                                  | <b>10%</b>              | 431.653         | 622.924         | 725.816          |
|                                  | <b>5%</b>               | 509.794         | 730.063         | 848.555          |
|                                  | <b>1%</b>               | 733.344         | 1036.600        | 1199.700         |
| <b>Satkhira</b>                  | <b>Risk/Time period</b> | <b>10 years</b> | <b>50 years</b> | <b>100 years</b> |
|                                  | <b>10%</b>              | 270.678         | 343.195         | 376.806          |
|                                  | <b>5%</b>               | 302.189         | 378.130         | 413.328          |
|                                  | <b>1%</b>               | 379.149         | 463.454         | 502.529          |
| <b>Madaripur</b>                 | <b>Risk/Time period</b> | <b>10 years</b> | <b>50 years</b> | <b>100 years</b> |
|                                  | <b>10%</b>              | 261.150         | 317.331         | 342.265          |
|                                  | <b>5%</b>               | 285.985         | 343.234         | 368.644          |
|                                  | <b>1%</b>               | 343.980         | 403.727         | 430.245          |
| <b>Faridpur</b>                  | <b>Risk/Time period</b> | <b>10 years</b> | <b>50 years</b> | <b>100 years</b> |
|                                  | <b>10%</b>              | 284.628         | 359.874         | 394.766          |
|                                  | <b>5%</b>               | 317.319         | 396.140         | 432.690          |
|                                  | <b>1%</b>               | 397.199         | 484.758         | 525.360          |

## 6 Conclusions

The aim of this thesis was to make an univariate extreme value analysis to calculate the extreme rainfall pattern over the next 10, 50 and 100 years. Daily rainfall data from five different measuring stations over 37 or 36 years was used. All five stations are geographically close to each other located in the south part of Bangladesh. Since the measurement stations only are located in the southern parts of Bangladesh, we can not comment on the estimated rainfall in other parts of the country.

Trend checks were made and daily rainfall plots shows a clear trend for each station depending on monsoon seasons. Autocorrelations function plots were checked and the data was assumed to be independent. No increase in the amount of rainfall can be seen during the measuring period. The most extreme rainfalls are not from the recent period in particular. All stations were compared to each other and a correlation between all of the stations was shown. Although the dependence is not very strong this indicates that the stations had followed the same rainfall pattern during the measuring period.

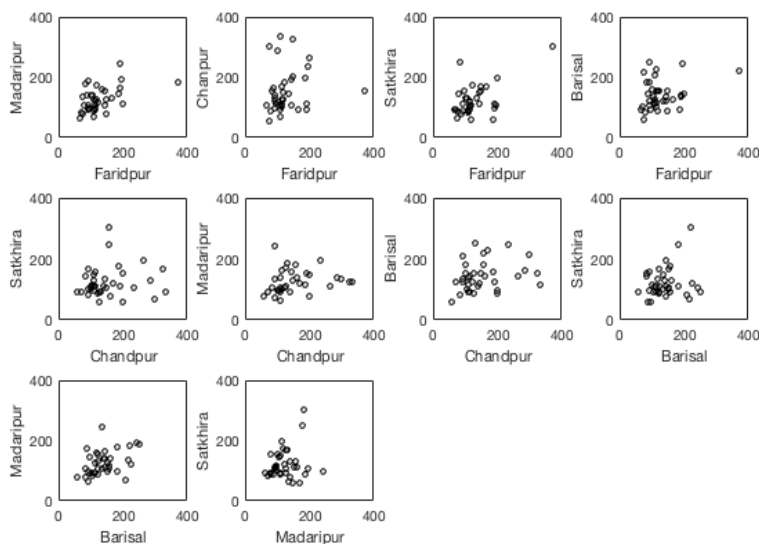


Figure 13: Scatter plots of annual maximum rainfall between the stations.

First block maxima method was used, the blocks were split up into size of years and generalized extreme values distribution was fitted to the annual maxima. The second method used was peaks over threshold. Thresholds were chosen for the different stations and generalized Pareto distribution was applied to all of the exceeding values. The good fit of the two distributions was checked by diagnostic plots.

All parameters, and their confidence intervals, for both distributions were estimated using two different methods, delta and profile likelihood. Return levels with confidence intervals were also calculated for each distribution using both of these methods. As tables 1-3 show both distributions and methods present very similar values for the parameter estimates and the return levels. This strengthens the assumption that both distributions are a reasonable fit.

The calculated return levels for GEV distribution and GPD does indicate an increase in rainfall for most of the stations. Station Faridpur and station Satkhira have a lower expected maximum rainfall in one day the next 100 years than have already occurred during the measuring period. Although, the probable maximum rainfall method presents higher predicted rainfall the coming 10, 50 and 100 years. According to this model the daily rainfall in one day will increase significantly. This method is considered more reliable because of wider range of observation.

The data was first treated in the program Matlab where all matrices were structured and trend checks were made. All stations were missing data from different days during the measuring period. Very few of these missing measurements were during monsoon period. These missed values may have caused some error margin in the calculations. For peaks over threshold method and probable maximum rainfall method clustering must be taken into consideration. It might be a risk that some of the exceeding values are dependent since some of them might be during a short measuring period. All calculation were made in the program R using the package in2extremes.

## 7 References

1. Harmeling, Sven. 2008. 'Weather-Related Loss Events And Their Impacts On Countris In 2007 And In A Long-Term Comparison', Global Climate Risk Index 2009, Germanwatch.
2. Harmeling, Sven and Eckstein, David. 2013. 'Who suffers most from extreme weather events? Weather-related loss events in 2011 and 1992 to 2011', Global Climate Risk Index 2013, Germanwatch.
3. Mahmood, Shakeel Ahmed Ibne. 2012. 'Impact of Climate Change in Bangladesh: The Role of Public Administration and Government's Integrity. Journal of Ecology and the Natural Environment Vol.4(8), 223–240.
4. Rootzen, Holger and Tajvidi, Nader. 1997. Extreme Value Statistics and Wind Storm Losses:A Case Study. *Scandinavian Actuarial Journal*, 70-94.
5. Ferreira, Ana and De Haan, Laurens. 2014. On the block maxima method in extreme value. *University of Lisbon and Erasmus Univ Rotterdam*, 1-4.
6. Binte Murxhed, Sonia, Islam, A.K.M Sailful and Alam Khan, M.Shah. 2011. Impact of climate change on rainfall intensity in Bangladesh. *3<sup>rd</sup> International Conference on Water and Flood Managment*.
7. Kreft, Sönke, Eckstein, David and Melchior, Inga . 2017. Global climate risk index 2017: Who Suffers Most From Extreme Weather Events? Weather-related Loss Events in 2015 and 1996 to 2015 *Germanwatch*
8. The Intergovernmental Panel on Climate Change (IPCC). 2007. *Climate change 2007: Impacts, adaptation and vulnerability*, Ch:10.
9. Coles, Stuart.2004. An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics, Ch: 2-4.
10. S. Cooley, Daniel. 2005. Statistical Analysis of Extremes Motivated by Weather and Climate Studies: Applied and Theoretical Advances. *University of Colorado at Boulder*, 1-7.

11. Plan international. 2017. *Life under water: When monsoon season strikes in Bangladesh.*
12. Sultana, Nasrin. 2015. A Multivariate Extreme Value Approach to Modeling Risk of Flooding and Drought in Bangladesh. *Lund University, Centre for Mathematical Sciences.*
13. National Plan for Disaster Management Government of the People's Republic of Bangladesh 2010-2015. 2010. *Government of the People's Republic of Bangladesh.*
14. Google map. <https://www.google.com/maps/place/Bangladesh/@23.4521973,85.8466443,6z/data=!3m1!4b1!4m5!3m4!1s0x30adaaed80e18ba7:0xf2d28e0c4e1fc6b!8m2!3d23.684994!4d90.356331>, (2018-02-13).