Delft University of Technology

CIE5050

Inferring rare events: The role of the length of observations

Author: M.D. Mascini - 4494644

November 26, 2021

Abstract

Extreme value analyses (EVA) are often used to determine the frequency of extreme events. The length of the available observations is an important aspect when performing EVA. It is generally known that more available data results in better estimates with less uncertainties. The main objective of this research report was to assess what the influence of the length of the observations is when inferring rare events. This was done by first analyzing the sensitivity of inferred return levels from synthetic data from three known distributions. Also, three case studies were analyzed to observe the sensitivity of inferred return levels and return periods from the observations. The results of the analyses were that a larger sample size generally leads to a higher confidence in the estimates of inferred return levels from synthetic data. However, there will always remain some uncertainty associated with the estimates. The confidence in the inferred return levels from observations also generally increase for an increasing sample size. However, this can not always be observed and other aspects can be more dominant than an increasing sample size. No correlation could be observed between the sample size and the inferred return periods. The conclusion of the research was that, while there is a positive correlation between the sample size and the confidence in the estimates, there will always remain some uncertainties. It is therefore important to always communicate the uncertainties associated with estimates.

Contents

List of Figures

List of Tables

1 Introduction

Extreme hydro-meteorological events have occurred throughout history and are a result of the natural variability of the climate. Extreme events, such as flooding, heat waves or droughts, can lead to loss of lives and enormous economical damage. Take for example the 1953 flood disaster in the Netherlands. Extreme sea levels resulted in dike breaching at multiple locations. The lives of 1836 people in the Netherlands were lost and more than 72000 people were evacuated (Rijkswaterstaat, n.d.). Moreover, there has been an increase in the observed hydroclimatological extremes since the industrial revolution (AghaKouchak et al., 2013). Due to this increase and the vulnerability to extreme events, the analysis of extreme events have gained a lot of interest over the years. This has resulted in an increase in research to the mathematical foundation and numerous applications of Extreme Value Analysis (EVA) (Makkonen & Tikanmaki, 2019). Because of the rising sea level and the growing population in flood prone ares, a bigger part of the worlds population will be living in flood-prone area in the next decade. Tellman et al. (2021) found that in the years 2000 to 2018, 255-290 million people were directly affected by floods. They also found that in the coming decade the number of people affected by floods will be ten times higher than in 2000. This is only one example of how the changing climate is making us more vulnerable to extreme events.

An extreme event is generally characterized by its low probability of occurrence. However, an extreme event can also be characterized as an event with a magnitude above a certain threshold. The frequency of extreme events are mostly communicated by assigning return periods to them. The reason frequencies are presented this way is that a return period is more easily interpreted by the public than a small probability of occurrence (Anderson et al., 2013). An event with a probability of occurrence of 1/100 per year happens on average once every 100 years. The return period of such an event is 100 years and for most people this is more easily interpreted than an occurrence probability of 0.01 per year. The magnitude of such an event is called a return level or a quantile. Extreme events can be both extreme high as well as extreme low. In this research report the focus will be on extreme high events.

The analysis of the frequency of extreme events is used in many disciplines such as climatology, hydraulic engineering and hydrology. The objective of frequency analysis of extreme events is to relate the magnitude of the event to their frequency of occurrence by using probability distributions (Alam et al., 2018). By definition, an extreme event has a low probability of occurrence and because of this, few of these events have been recorded. Most of the time we want to protect ourselves against events with a large magnitude that have not been recorded yet in historical data (Coles, 2001). This implies that we are interested in an event with a return period that is larger than the available years of historical observations. Such an extreme event can be estimated from a theoretical distribution function fitted on the available observations. The theoretical distribution can be extrapolated to find estimates for events with high return periods.
When EVA is used to infer extreme events, there is always an uncertainty associated to the estimates. How confident we can be in an inferred event, depends on multiple aspects. One important aspect is the length of the available observations. It is agreed upon in literature that more observations result in better estimates for the theoretical distributions. Generally, the confidence in the estimates increases with an increasing length of observations. A problem when performing EVA is that long timeseries are often not available. It is not uncommon that timeseries have a length of no more than 50 years (Caires, 2011). Therefore, it is important to know how the sample size influences the confidence that we have in inferred events. If it is known how the availability of data might affect the results, this can be taken into account prior to doing such an analysis. Also, it can then be included as information when communicating the results of an analysis. The confidence in the inferred events often not communicated. However, the confidence that we have in the inferred events is a valuable part of the information.

The main objective of this research report is to assess what the effect of the length of observations is when inferring rare events. The effect of the length of observations will be determined by answering the following sub-questions:

1. How sensitive is the inferred return level to the sample size when using synthetic data?
2. How sensitive is the inferred return level to the sample size when using observations?
3. How sensitive is the inferred return period to the sample size when using observations?
4. How can the uncertainty of inferred return periods due to limited observations be communicated to include it as valuable information in decision making processes?

The research limits itself to the Generalized Extreme Value Distribution and the General Pareto Distribution. Another limitation of the research it that it will only include stationary analyses. This means that the data sets (synthetic and observations) do not present a statistical significant trend over time. A final limitation of the research is that only one method for parameter estimation will be used.

The structure of the report is as follows. Chapter 2 consists of the most important concepts and an overview of the available literature that summarizes the most important publications on the topic. Chapter 3 describes the research methods that were used. In chapter 4 the main results of the research are presented. Chapter 5 contains the discussion of the results. In Chapter 6, the conclusion and recommendations are given.

2 Literature review

In this chapter, an overview of the most important concepts and the available literature is presented. In chapter 2.1 the most important distributions used for extreme value analysis will be elaborated. In chapter 2.3, a method for parameter estimation will be discussed. In chapter 2.3, uncertainty in EVA will be elaborated upon and two methods for quantifying uncertainty will be discussed. Chapter 2.4 discusses the most relevant publications regarding the influence of the sample size.

2.1 Distributions for extreme value analysis

The most common distributions used for EVA are the Generalized Extreme Value Distribution (GEV) and the Generalized Pareto Distribution (GPD). While the two are connected, the main difference is the method that is used to sample the dataset. For the GEV distribution the Block Maxima (BM) method is used. Most commonly blocks of 1 year are chosen to obtain annual maxima. For the GPD a Peak Over Threshold (POT) method is used. With POT, extremes above a certain threshold value are selected and used to fit the distribution to (Pisarenko & Rodkin, 2017).

2.1.1 Generalized Extreme Value Distribution

The GEV was first proposed by Jenkinson (1955) and is defined by three parameters, namely the location parameter (μ), the scale parameter (σ) and the shape parameter (ξ). The GEV consists of 3 sub-families that are labelled as type I, II and III but are widely known as the Gumbel, Fréchet and the reversed Weibull distributions (Coles, 2001). The three sub-families differ in the behaviour of the tail. The behaviour of the tail is defined by the shape parameter ξ. The shape parameter is therefore seen as the most important parameter of the distribution. The Gumbel sub-family has a shape parameter $\xi = 0$ (or $\xi \to 0$), Fréchet is characterized by $\xi > 0$ and Weibull is characterized by a shape parameter $\xi < 0$. The probability density function (pdf) of the GEV distribution is given in equation 1. The plotted pdf's for the three types are displayed in figure 1.

$$f_{\mathrm{GEV}}(x;\mu,\sigma,\xi) = \frac{1}{\sigma}\exp\left(-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right)\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi-1}\{z: 1+\xi(z-\mu)/\sigma > 0\} \quad (1)$$

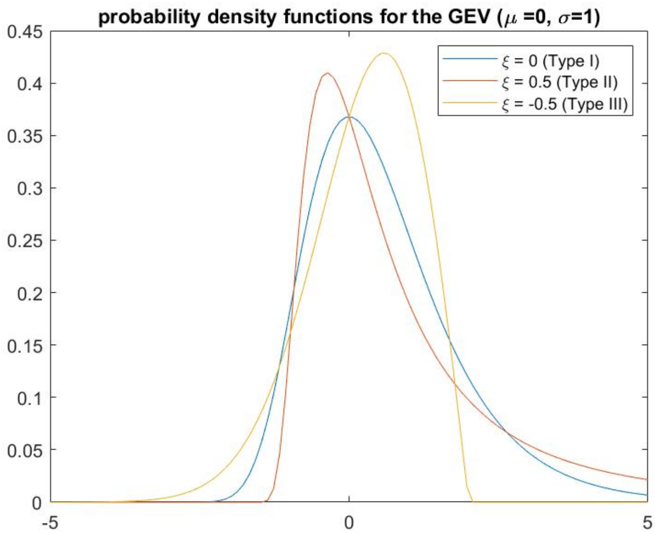


Figure 1: Probability density functions for the three sub-families of the GEV

The figure shows that the Gumbel distribution has infinite bounds on both the upper and the lower tail. The Fréchet distribution is characterized by an finite bound on the lower tail and an infinite bound on the upper tail. The reversed Weibull distribution has an infinite bound on the lower tail and a finite bound on the upper tail.

The GEV distribution is commonly used to model block maxima extremes. A block is a fixed period of time from which the maximum observation is selected. For example, if we consider a block with a period of one year, the obtained sample of maxima is referred to as annual Maxima. The extremes should be independent and they should be described by the same distribution. This concept is known as that the data should be independent and identically distributed (idd).

2.1.2 Generalized Pareto Distribution

The GPD distribution takes as input data the excesses above a certain threshold (u). The GPD for the excesses above the threshold has the same three parameters as the GEV distribution, with the difference that the location parameter is equal to the threshold. Moreover, the shape parameter is equal to the shape parameter of the GEV and it thus determines the behaviour of the upper tail. Since the GPD is used for excesses above a threshold, there is a boundary at the threshold. This means that the probability of occurrence of everything below the threshold is 0. This can be seen in the pdfs for the GPD distributions with threshold $u = 0$ and varying shape parameters in figure 2. In the figure the behaviour of the upper tail for the varying shape parameters can clearly be observed.

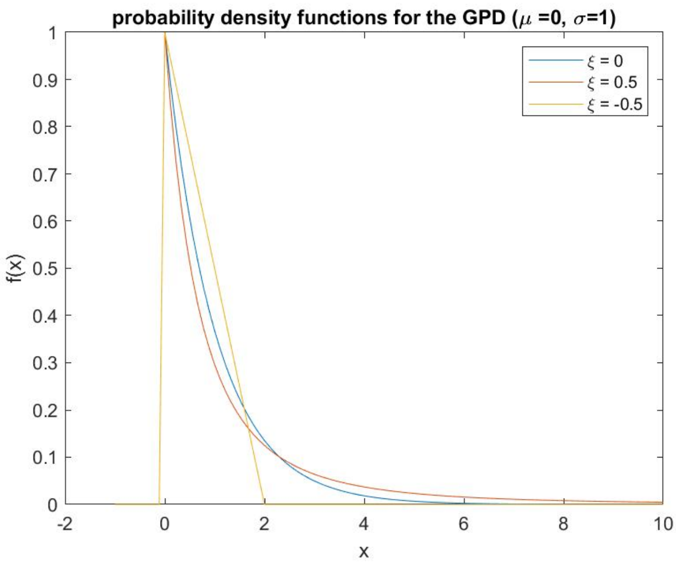


Figure 2: Probability density functions for GPD distribution with varying shape parameters

The choice of the threshold (and also the block size of the associated GEV) is important when fitting an extreme value function. When there is a low threshold, there are more points to which the distribution can be fitted to. However, this can also mean that not all events are truly extreme events and the fitted distribution can be biased. When too high a threshold is selected, there are only a few points to which the distribution can be fitted leading to a large estimation variance. An appropriate threshold can be determined with the help of different techniques. The easiest technique is by creating a Mean Residual Life Plot (MRLP) as proposed by Coles (2001). This can be done by plotting the mean of the excesses above a threshold u as a function of the threshold u. A threshold u_0 from which the MRLP has an approximately linear slope can be considered a valid threshold. Another method to determine the threshold is based on the fact that the excesses over the threshold should be Poisson distributed (Bommier, 2014). For a Poisson process, the mean (μ) and the variance (σ^2) are equal to each other. Meaning that the variance over the mean should be equal to 1. The variance over the mean is called the dispersion index and is given in equation 2. A threshold can be chosen by plotting the DI over the thresholds and selecting a threshold for which the DI is close to 1.

$$DI = \frac{\sigma^2}{\mu} \tag{2}$$

Another graphical method to determine a suitable threshold is to make a parameter stability plot. This method is based on the fact that the estimates for the shape and scale parameters remains constant. If the excesses above a threshold u_0 follow a GPD with parameters ξ and σ, then the excesses above a threshold $u > u_0$ have the same shape and scale parameters (Bommier, 2014). The plot can be made by plotting the shape and scale parameters over the chosen threshold and selecting a threshold for which the parameters are constant. In the research methods, only the MRLP and the DI plot are used for the selection of the threshold.

2.2 Methods for parameter estimation

One of the most commonly applied methods used to infer parameters of the GEV and GPD distributions is the Maximum Likelihood Estimation (MLE) method. MLE is a method to estimate parameters with help of the likelihood function. A likelihood function is a function that relates the probability of observing the available data, given a distribution with parameter Θ. For the GEV distribution this means that $\Theta(x, \xi, \sigma, \mu)$. It measures how much the available data supports the function for all possible parameters. By maximizing the likelihood function, Θ can be found such that observing the data is most probable from the distribution with a parameter Θ. The MLE for a parameter can be seen as a best guess for the parameter, given the available data. Different data from the same true process, might lead to a different estimate of the same parameter. Other methods include graphical methods, the Method of Moments and the L-moments approach (Coles, 2001). More advances methods include the Bayesian inference approach for parameter estimation. However, in this research report we will only focus on the MLE approach for convenience.

2.3 Uncertainty in Extreme Value Analysis

The performance of the fitted distributions and the confidence that we have in the fitted parameters depend on both the method that is used to fit the distribution and the length of the available observations. Most of the time long timeseries are not available (Pinheiro & Grotjahn, 2015). It is therefore important to also include the uncertainty associated with the estimates. This can be done for example with Confidence Intervals (CI). Moreover, it is of importance to include the estimation of the uncertainty for the inferred extremes when communicating the results of an analysis. Otherwise it is not a complete representation of the available knowledge (Coles, 2001) (Benstock & Cegla, 2017). The MLE approach for parameter estimation can preform poorly when there is a small data set available. However, there is no consensus (yet) on a minimum length of observations for reliable results when using the MLE method (Soukissian & Tsalis, 2018). Furthermore, while it is agreed upon in literature that a longer time series results in better estimates with less uncertainties, it is not known how the number of historical observations influences these uncertainties in the estimates exactly. Different methods can be used to calculate the CI of the estimates. Two methods are expanded upon in the following sections.

2.3.1 Bootstrapping

A confidence interval can be computed with the help of a bootstrapping method. With a bootstrapping method, resampled sets are drawn from the data to obtain subsets. For all the subsets the MLE of the parameters can be determined. Functions of the MLE of the parameters (like the return levels) can then be computed. From all the computed return levels, the 95% confidence interval can be determined by selecting the 2.5% and 97.5% percentiles.

2.3.2 Matlab function

Matlab has a function installed that returns the MLE for the parameters. In the function there also is the option to return the 95% confidence intervals for the parameter estimates. These are computed based on the fact that for a large enough n, the MLE of the parameters should be normally distributed (Cai & Hames, 2010) (Coles, 2001). The CI for the shape and the location parameter are computed using a Normal distribution. In the Normal distribution the mean is the MLE of ξ and μ respectively. For the standard deviation, the standard error of the parameters is used. The standard error is computed with the asymptotic covariance matrix of the parameter estimates (MATLAB, 2020). The CI of the scale parameter is computed with the normal approximation for $ln(\sigma)$. The function uses the natural logarithm of the MLE of σ as the mean of the distribution and the standard error divided by the MLE of σ as the standard deviation (MATLAB, 2020).

2.4 The influence of the sample size

(Martins & Stedinger, 2000) and (Katz et al., 2002) found that the MLE method can be a valuable method for parameter estimation due to its simplicity and the reliable results. However, they also found that when the sample size is small (n $\leq$ 25), the method can not converge. This can result in unrealistic estimates for the parameters of the GEV. (Soukissian & Tsalis, 2018) looked into the effects of the length of observations and the methods for parameter estimation when using the GEV distribution with block maxima. They tested 4 different parameter estimation techniques, one of them being the MLE method, and tested the techniques for samples with varying length n with synthetic data. They found that, in general, a larger sample size results in better estimates of the parameters. They also found that generally, the L-moments method performs better than the

MLE method when tested on synthetic data. They used observations of wind speed data at 4 different locations as case studies. The results showed that sample sizes greater than 35 result in stable estimates for return levels in 3 of the 4 cases. The analysis to the method of parameter estimation showed that the L-moments method performed better in the majority of the cases. Finally, they suggested that when using a MLE method, a sample size of at least 35 years is needed for a substantial reduction of the uncertainty.
A research study by Cai & Hames (2010) proposed a new method to determine the minimum sample size required for a valid EVA with the GEV distribution. The method used the fact that generally the MLE of the parameters of the GEV distribution follow a normal distribution when the sample size goes to infinity. They used this to test if the parameters fitted to the data (with varying lengths) indeed follow a normal distribution. They used their proposed method both on synthetic data as well on observations to obtain minimum sample sizes. For the synthetic data they used randomly chosen parameters for the GEV distribution with a negative shape parameter. For that particular case they found that a minimum sample size of 40 was needed. Thereafter, they used 4 different sets of observations with annual maxima (sea levels, wind-speeds, discharges and temperatures). Based on their proposed research method they found that the minimum length of observations to obtain an estimation for the 100 year return period should be between 44 - 72 years for the 4 case studies.

Rutten et al. (2008) proposed a method that uses EVA while including uncertainties to obtain load factors on offshore platforms. They used the GPD distribution with a POT method to do so. The results were that two factors influence the width of the CI the most when inferring extreme return levels with the GPD. These are the number of values and the shape parameter. It should be noted that the number of values in their study is the number of values above the chosen threshold. However, a longer dataset will also result in a higher number of excesses.

Different research has also been executed to overcome the problem of limited availability of observations of extremes. Volpi et al. (2019) proposed a method for parameter estimations that makes use of the entire set of observations prior to the selection of the maxima. Where classical EVA uses only the annual maxima or the peak storm events, they used all the data about events in the timeseries. The main difference from the proposed approach to traditional EVA is that independence of the events is not a necessity. In traditional EVA, events are selected in such a way that they are independent from each other. For example, when a POT approach is used to select the peak precipitation events. When a storm with a duration of multiple days is recorded, only the highest peak of the storm is selected. If multiple peaks from one storm would be selected, this would mean that the events are not independent. In the proposed approach, all events are included. They found that the approach results in more conservative estimates for return periods with a higher uncertainty when compared to annual maxima methods. They found that the method could be a promising one, but that more research is needed before it can be used in practice.

3 Methods

The research was conducted in two parts. The first part consisted of an analysis of synthetic data. The research methods used can be found in section 3.1. The second part consisted of the analyses of three case studies. The methods that were used for the analyses are included in chapter 3.2.

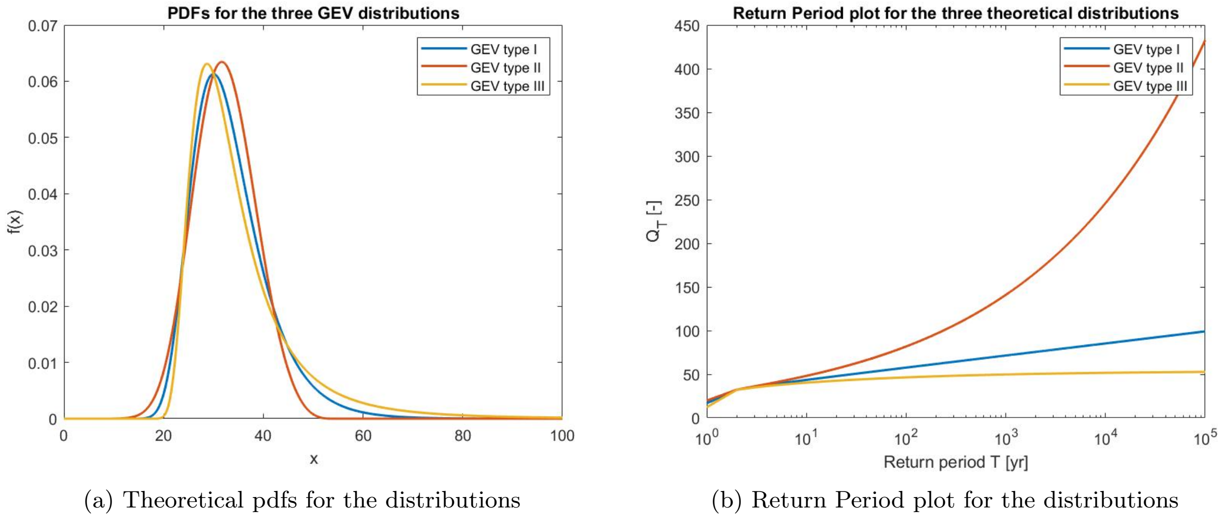
3.1 Synthetic data

In the first part of the research, synthetic data were used to evaluate the sensitivity of the estimated extremes inferred from the GEV distribution to the length of observations. The GEV distribution consists of 3 families, each with a characterizing shape parameter that determines the tail behaviour (see chapter 2.1.1). (Rutten et al., 2008) found that the shape parameter can have an influence on the width of the CI. That is why the analysis was done for each type of the GEV distribution. For each of the families within the GEV, fixed parameters were chosen at random (meaning that the data do not represent a variable). The fixed parameters that were used in the analysis are given in table 1.

Family	Shape ξ	Scale σ	Location μ
Type I	10^{-6}	6	30
Type II	0.25	6	30
Type III	-0.25	6	30

Table 1: Fixed parameters for analysis of synthetic data

The pdfs and a plot of the theoretical return periods of the distributions is given in figure 3. In figure 3b, the difference in tail dependence can clearly be observed. It should be noted that the x-axis is in log-scale. For the type I distribution, the return level keeps increasing for an increasing return period. The type II distribution also shows this behaviour but the increase in return level is much steeper than for the type I. The type III distribution shows asymptotic behaviour for an increasing return period.



(a) Theoretical pdfs for the distributions

(b) Return Period plot for the distributions

Figure 3: Theoretical PDF and Return Period plot for the distributions in table 1

The sensitivity analysis was performed by generating samples of different length from the distributions in table 1. The GEV distribution was fitted to the samples with the MLE method for parameter estimation. The sample lengths started from n=30. The reason for this is that the research in the literature it followed that the MLE method can work properly from n=25-35. With the estimated parameters and four fixed return periods (T = 30, 50, 100 and 1000 years), the corresponding estimated return levels (Q_{30}, Q_{50}, Q_{100}, Q_{1000}) and the 95% confidence bounds for the estimates were determined. The bounds of the 95% confidence interval were determined in two different ways. First, a bootstrapping method was used to generate 1000 samples for which the MLE of the parameters were determined. With the parameters, the return levels were determined. To clarify: for a sample with size n=30, there were 1000 estimates of the return levels. From these 1000 estimated return levels, the 2.5 and 97.5 percentiles were chosen that determined the 95% CI. The seconds method that was used, was a method that came directly from MatLab (see chapter 2.3.2. When fitting a distribution, Matlab returns the MLE as well as the 95% CI values of the parameters. The 95% CI values were used to calculate the 95% confidence intervals of the return levels directly. This method is from heron referred to as the 'MatLab parameters' method. The process is described step by step below.

1. From the in table 1 defined distributions, samples were taken with size n=30:1000. Per sample size, 1000 samples were taken.
2. The GEV distribution was fitted to each sample with the MLE method to obtain the MLE of the parameters. Also the 95% CI of the parameters were determined with the Matlab parameter estimation method.
3. With the 95% CI of the three parameters (obtained with the Matlab parameter method), the 95% CI was determined for the 4 return periods.
4. With the parameters fitted to the 1000 bootstrap (for each sample size), the return levels belonging to the 4 defined return periods were calculated.
5. The $2.5th$, $50th$ and $97.5th$ percentiles were determined from the 1000 estimates for each sample size. The $50th$ percentile was seen as the ML estimate of the return levels and the 2.5^{th} and $97.5th$ percentiles as the bounds of the CI calculated with the bootstrapping method.
6. The results were plotted over the length of observations.

After the analysis to the sensitivity of inferred return levels, another analysis was done. For this analysis, the same distributions as described in table 1 were used. From the literature it followed that both the length of the observations as well as the magnitude of the considered return period have an influence on the width of the CI. The goal of this part of the research was to observe which of the two is more important for the confidence we have in the estimates. Either the number of observations that are used to fit the distribution or the magnitude of the inferred return period. In other words: do we have more confidence in estimating a 50 year period with 50 observations, or in estimating a 100 year return period with 100 observations? To determine this, first the theoretical return levels for the 4 return periods were determined ($Q_{T,30}$, $Q_{T,50}$, $Q_{T,100}$, $Q_{T,1000}$) from the distributions. After that, the following steps were taken.

1. From the in table 1 defined distributions, samples were taken with size n=30, 50, 100, 1000. For each sample length, 1000 samples were taken.
2. The GEV distribution was fitted to each sample to obtain the MLE of the parameters.
3. With the MLE of the parameters the return period associated with the theoretical return levels ($Q_{T,30}$, $Q_{T,50}$, $Q_{T,100}$, $Q_{T,1000}$)) were calculated.
4. From all the estimates, the outliers were removed.
5. A boxplot was made in Matlab of all the estimates to see the deviation of the estimates. In the boxplot, the 25th and 75th percentiles of the results were showed.

Sometimes the MLE can not converge well to estimate the parameters. There can be multiple reasons for why this is the case. The literature showed that when there is a low number of observations, sometimes the MLE method can not converge (Soukissian & Tsalis, 2018). This can lead to large outliers in the distributions. Another reason why the analysis sometimes leads to large outliers is that the distribution has an upper bound. For example, figure 3b showed the return period - return level plot. The type III distribution shows asymptotic behaviour for large return periods. This means that the return period $\rightarrow \infty$ for return levels that are larger than the asymptote. This means that when the estimated distribution is such that it has a lower asymptote than the theoretical distribution, the estimation of the return periods can lead to large outliers. These outliers can be in the order of of $T=10^6$ or even infinity. To not let the outliers influence the results, they were removed.

3.2 Case studies

In the second part of the research, an analysis of three different case studies was done. The goal of part II of the research was to deduct how sensitive the inferred return levels and periods are to the sample size, when using observations. For the analysis, timeseries were selected such that they included an extreme event in the summer of 2021. The sets of observations and the recent extreme event are shortly described below. After a description of the datasets, the method that was used to deduct the sensitivity of the return levels and periods is expanded upon.

3.2.1 Description the case studies and the data

Case study 1: precipitation in New York

The first case study included the daily precipitation events in mm in New York measured by the weather station of LaGuardia airport. The dataset that was used included data from the year 1940 until 2021. The dataset was retrieved from the World Meteorological Organization (WMO) Climate Explorer website. A graphical representation of the the data over the years is displayed in figure 4.

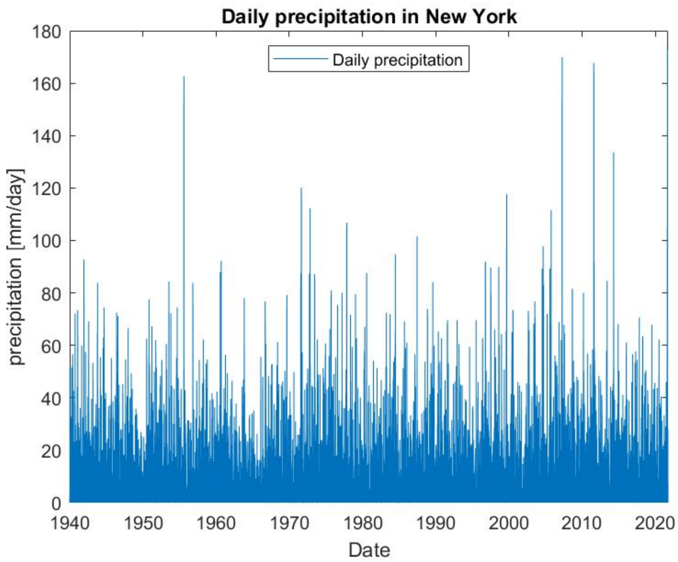


Figure 4: Precipitation in New York over time [mm/day]

On September 1^{st} 2021, hurricane Ida reached New York city. The hurricane came with a lot of rainfall, and hourly and daily records were set. Ida caused damages in the New York area for at least 50 million dollar and 45 people were killed in the region (Intelligencer, 2021). The 6 hour precipitation event was estimated as as an event with a return period of 200 years in the news (Kahn, 2021).

Case study 2: precipitation in Limburg

The second case study included the precipitation events in Limburg in mm/48h. The original data (obtained from the WMO website) was in the units mm/day. This was converted to 48h precipitation events. The dataset that was used contained information from 1960 to 2021 from the weather station Ubachsberg in the Netherlands. A graphical representation of the 48h precipitation in Limburg is included in figure 5.

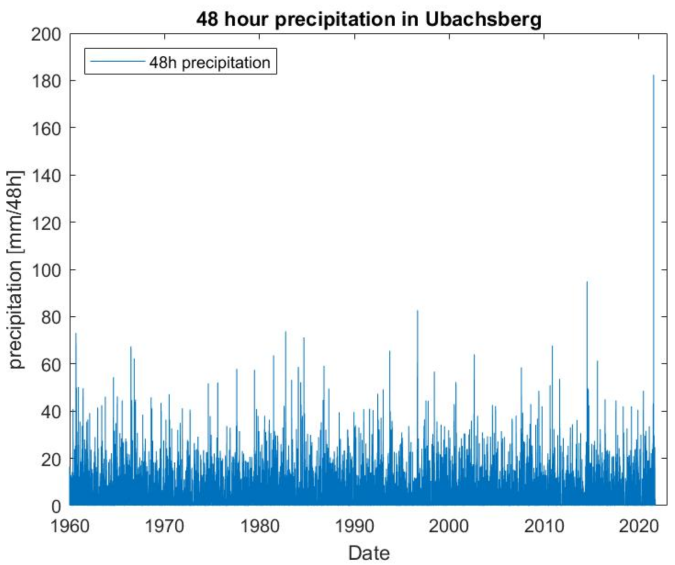


Figure 5: Precipitation in Limburg over time [mm/48h]

In July of 2021, heavy precipitation events took place in the south of the Netherlands, parts of Belgium and in the Western part of Germany. As a result, flooding happened in all three countries and thousands of people were displaced. The 48 hour precipitation recorded at the weather station Ubachsberg was 182 mm. STOWA (2021) analysed the precipitation data from 219 weather stations in the Netherlands in 2019 (so before the event of 2021). They found that such an extreme event has on average a return period of 1000 years. A Mann-Kendall trend test was done for the rainfall events to determine if a statistical significant trend could be observed. The result of the test was that when the 2021 event is included in the observations, a trend can be observed in the data. This means that performing a stationary analysis of the event will result in an underestimation of inferred return periods. With this in mind, the analysis was done.

Case study 3: Temperatures in Sicily The third case study included daily maximum temperatures in Sicily in $^{\circ}C$ measured by the weather station at Sigonella Airport. The dataset that was used contained information from 1960 to 2021, with 3 small gaps. There was no (complete) data for the years 1973, 1986, 1987 and 2001. These gaps were ignored in the analysis. The dataset was retrieved from the WMO Climate Explorer website. A graphical representation of the daily maximum temperatures over time is included in figure 6.

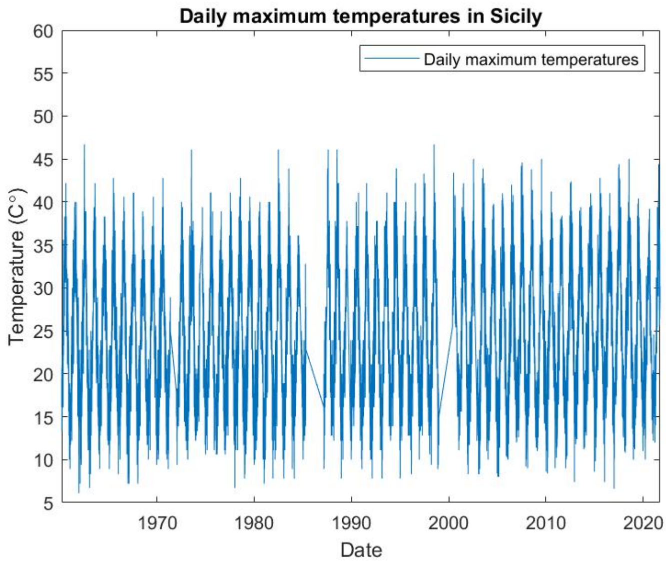


Figure 6: Daily maximum temperatures in Sicily [° C]

On August 11^{th} 2021, a monitoring station near Siracusa (Sicily) recorded a temperature of 48.8 degrees Celsius. The event still needs to be verified by the World Meteorological Organization. If it is confirmed, it will be the highest temperature recorded in Europe (Pianigiani, 2021). No return period was reported on anywhere in the news. Unfortunately, the temperature was also not recorded by the weather station where the dataset was retrieved from. The distance between the weather station used in the analysis and the weather station at Siracusa is approximately 50 km. This was kept in mind when performing the analysis.

3.2.2 The sensitivity of return levels

For all three case studies described above, analyses were done to the sensitivity of the inferred return levels and return periods. Before the analyses could be done, it was checked if a significant trend could be detected in the annual maxima of the data. Since the analyses that were done included stationary analyses only, a trend in the data is not wanted. A trend can affect the results of the analyses. Also, the absence of a statistical significant trend will ensure that the data are independent and identically distributed (idd).

For the analysis of the observations, either a GEV distribution was fitted to annual Maxima or a GPD distribution was fitted with to POT data set. To determine which distribution fitted the data best a QQ-plot was made. The plot compared the empirical and theoretical quantiles. Moreover, a plot was made that compared the empirical cdf (ECDF) (plotted with the Weibull plotting position) to the fitted CDF. The selection of the distribution was made based on visual observations of the QQ-plot and the ECDF/fitted CDF plot.

The analyses consisted of an analyses of different chronological subsets with varying lengths of the datasets. It was assumed that if the chosen distribution was a correct fit for the entire timeseries, this would also be the case for the subsets of the timeseries. For example, the GEV distribution turned out to be an appropriate fit for the the annual Maxima of the precipitation in Limburg for the years 1960-2021. This was checked with a QQ-plot and the CDF - ECDF plot. It was then assumed that the GEV distribution would also be a correct fit for the annual maxima precipitation in Limburg for the years 1960-1989 (a chronological subset of 30 years). After the type of distribution was chosen, the entire dataset was used to determine the MLE of the parameters. A GPD distribution was fitted to the precipitation events in New York with the POT analysis of the events. To select an appropriate threshold a MRLP and a DI plot were made (see section 2.1.2). This was done for subsets of length n=30, 35, 40, ..., 80 years and for the entire dataset (n=82 years). A GEV distribution was fitted to the annual maxima 48h precipitation events in Limburg. A GEV distribution was also fitted to the annual maxima temperatures in Sicily.

After the distributions were chosen and the MLE of the parameters were determined for the entire dataset, a similar sensitivity analysis as in part I of the research was done. In the analysis of the case studies, the CI were only determined with the bootstrapping method. The analysis of the synthetic data showed that the CI of the two methods resulted in a similar course of the graphs. Including both CI in the analysis of the case studies did not provide more information. That is why it was chosen to only calculate the CI with the bootstrapping method. The steps that were taken are described as follows:

1. From the observations, a chronological sample M with length (n=30, 31, ..., N years) was taken. Where N is the total length in years of the dataset.
2. The chosen distribution was fitted to the sample M and the MLE of the parameters were determined.
3. A bootstrapping approach was applied. 1000 random samples with length n=30...N were drawn from a distribution with the MLE parameters of step 2.
4. To the 1000 samples, the GEV distribution was fitted and the MLE of the parameters were determined.
5. With the MLE of the parameters, the return levels Q_{30}, Q_{50}, Q_{100}, Q_{1000} were calculated.
6. The 2.5, 50, and 97.5 percentiles of the return levels were determined from the 1000 samples.
7. The results were plotted over the length of the subset n.

3.2.3 The sensitivity of the return period

After the sensitivity of the 30, 50, 100 and 1000 year return levels were analyzed, an analysis to the sensitivity of the return period was done. For the analysis, the extreme events from summer 2021 were used. The analysis consisted of a year by year analysis that computed the return period associated with the 2021 event. To clarify, an example is given. From the dataset with precipitation in New York, chronological subsets of length $n = 30, 31, ..., 82$ were taken. During hurricane Ida, an event of 173 mm/day was recorded. For all the chronological subsets, the return period associated to this event was determined. The varaition of the estimated return period could be observed over the sample size n. The steps that were taken are described below:

1. From the observations, a chronological sample M with length $(n = 30, 31, ..., Nyears)$ was taken. Where N is the total length in years of the dataset.
2. The chosen distribution (GEV or GPD) was fitted tot the sample M and the MLE of the parameters were determined.
3. A bootstrapping approach was applied. 1000 random samples with length $n = 30...N$ were drawn from a distribution with the MLE parameters determined.
4. To the 1000 samples, the GEV distribution was fitted and the MLE of the parameters were determined
5. The return period to the 2021 extreme event was estimated with the parameters.
6. The 2.5, 50, and 97.5 percentiles of the return periods were determined from the 1000 samples.
7. The results were plotted over the sample size.

4 Results

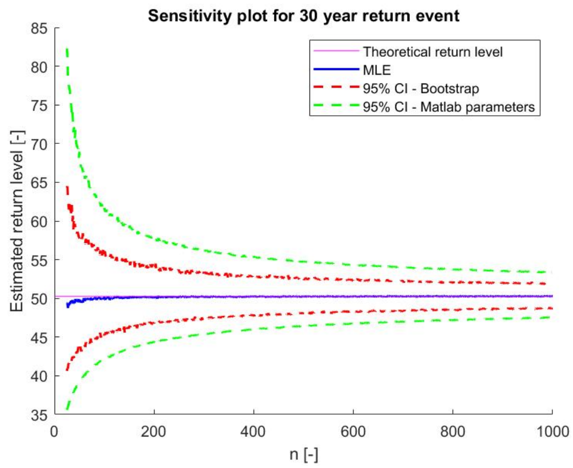
In this chapter the results of the research will be presented. The conducted research consisted of 2 parts. In the first part synthetic data was used to deduct the sensitivity of inferred return levels. The results of the analysis are presented in chapter 4.1. The second part of the research included the analyses of three case studies. The results of the analyses are presented in chapter 4.2.

4.1 Results Synthetic data analysis

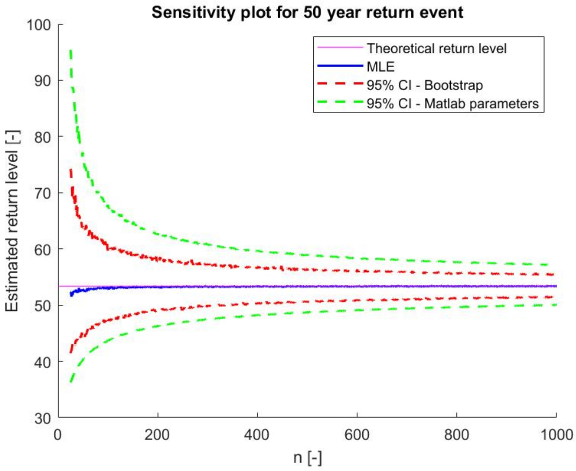
We used synthetic data to assess the sensitivity of the quatification of rare events to the length of observations used. We carried out the sensitivity analysis for the 3 types of GEV families (I, II and III). The parameters used in the analysis are included in table 1.

4.1.1 Results GEV type I - Gumbel

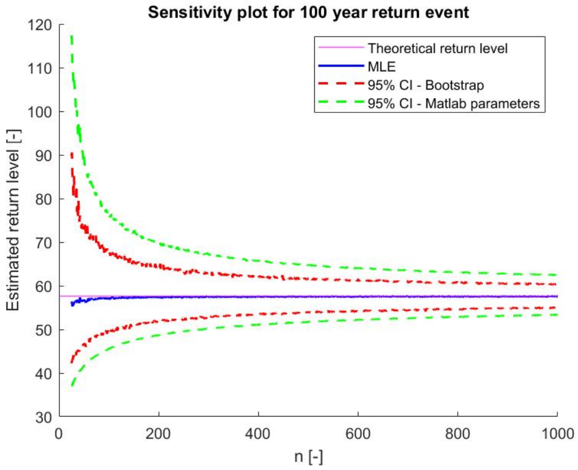
The results of the sensitivity analysis for the type I GEV distribution, i.e., Gumbel distribution, are presented in figure 7. The estimates for the 30, 50, 100 and 1000 year events are included in subplots 7a - 7d. Note that the y-axis of the subplots are not the same scale because the variation of the estimated return levels is large. The CI of the 1000 year event is much wider than that of the 30 year event. To include all the information in the figure, the y-axis of the subplots are not the same. The red and green dashed lines refer to the 95% CI based on the bootstrapping method and the method with the MatLab estimates for the parameters respectively. The blue line refers to the median of the MLE of the return level and is from hereon described as the MLE. The figure shows that the CI calculated with the Matlab parameter methods is wider than the CI calculated with the bootstrapping method. An explanation could be that the MatLab parameter method uses the 2.5% and 97.5% values of all the individual parameters to calculate the return levels of the upper and lower bounds. This means that the calculated 2.5% confidence bound for the return level is calculated by taking the lower 2.5% of all of the parameters. The bootstrapping method first fits the GEV distribution to the samples to calculate the return levels. After that the 2.5% and 97.5% percentiles of all the return levels are selected. The difference with the Matlab method is that for a lot of the bootstrap samples, not all three parameters will be as low (or high) at the same time. This will result in not as low (and not as high) confidence bounds. Hence, it results in a less wide CI.



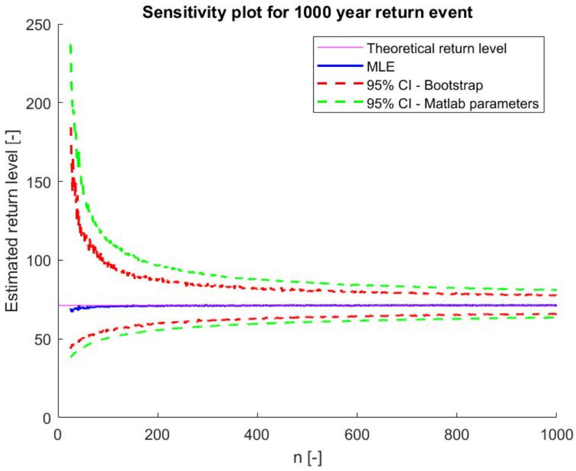
(a) 30 year return event



(b) 50 year return event



(c) 100 year return event



(d) 1000 year return event

Figure 7: The estimated return levels for the type I GEV distribution over the length of observations. The figures include the theoretical return level, the MLE of the return level and the 95% CI computed with a bootstrap method and with Matlab parameters.

The plots show that the MLE of the return levels are generally close to the theoretical value of the return levels. Although they slightly underestimate the return levels evaluated using smaller sample sizes (up to n=50), they can generally be considered a valid estimate. The figures also show that the width of the CI increases, with increasing return period. This can be explained by the behaviour of the distribution. The literature showed that the shape parameter is the parameter that determines the tail dependence. If the shape parameter that is fitted to a sample is slightly higher than the theoretical shape parameter, the tail of the distribution will also be slightly different. If the shape parameter is slightly lower or higher this will not have a huge impact on the estimates for an event with a low return period. However, when it is used to estimate an event with a high return period, it will result in a large deviation from the theoretical return level. A figure that clarifies this is included in figure 8. The figure shows that the CI calculated with the bootstrapping method (shaded areas), increases in width for an increasing return period. It also shows that the CI for a larger sample size is smaller (red shaded area) than for a smaller sample size (blue shaded area). This explains why the CI is wider for the return levels associated to higher return periods.

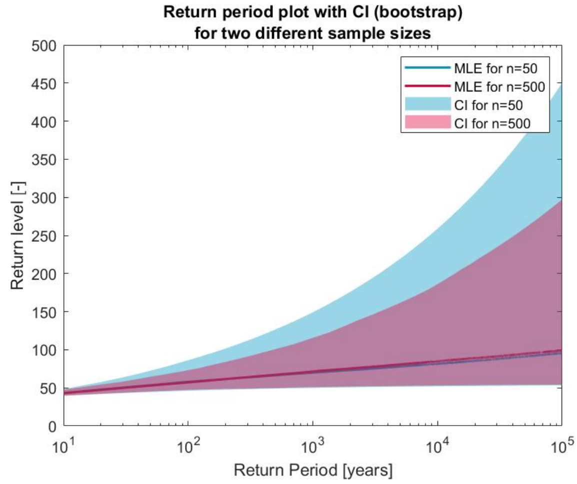
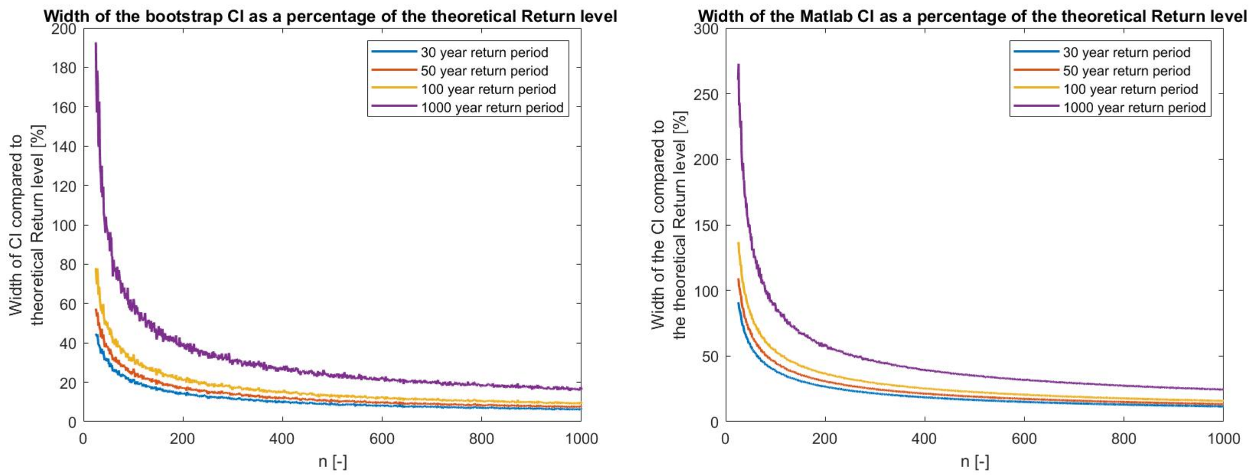


Figure 8: The return period plot for the analysis of the type I (Gumbel) GEV distribution. The estimated return level and the CI calculated with a bootstrapping method are included for 2 sample sizes (n=50, n=500).

Per subplot, we see that the width of the 95% CI decreases with increasing number of observations, which is as expected. Generally, a larger number of observations results in a more certain estimate. However, the width of the CI stays relatively wide, even for a very large sample size. To quantify this, the width of the CI is plotted as a function of the theoreitcal return level in figure 9. Subplot 9a represents the width of the CI obtained with the bootstrap method and 9b the CI obtained with the Matlab parameter estimates.



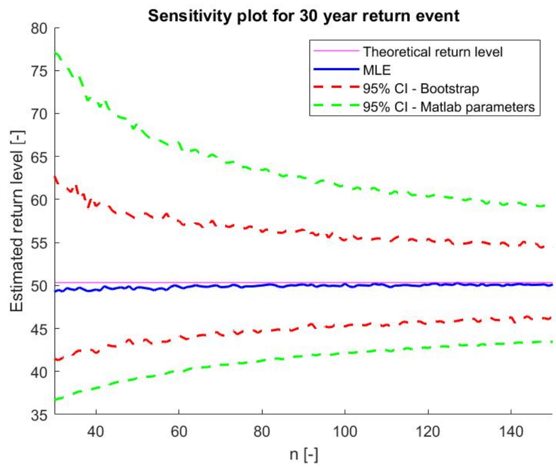
(a) Bootstrap CI

(b) Matlab parameters CI

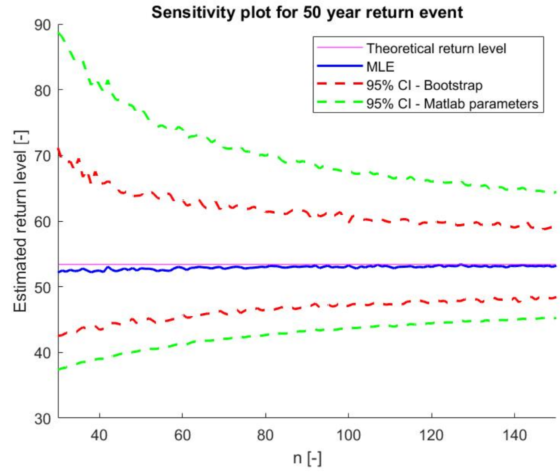
Figure 9: The width of the CI relative to the theoretical return level for the type I GEV distribution. CI obtained with the bootstrap method (9a) and CI obtained with the Matlab parameter method (9b)

Figure 9a shows that even with a sample size of 1000, the width of the CI obtained with the bootstrap method is around 20% of the return level for the estimate of the 1000 year event. The Matlab parameter methods results in an even higher relative width of around 30 % (figure 9b). The plots in figure 9 as well as in figure 7 show that the largest variation in the width of the CI is for low number of observations. Between 30 - 150 observations, the width of the CI decreases fast for increasing observations. Between 150 - 400 observations, the width still decreases, but not as fast. After around 400 observations, the decrease of the CI is minimum (for both methods).

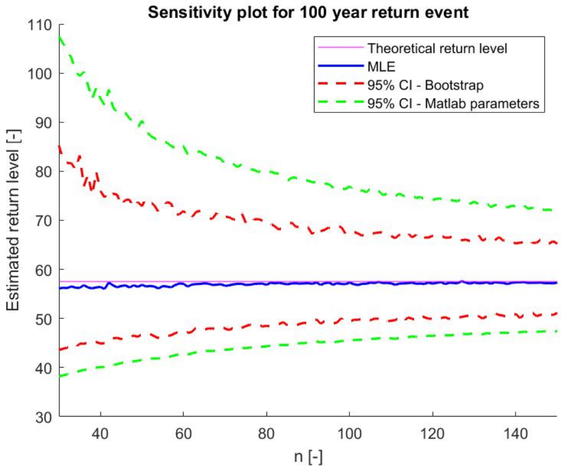
While it is interesting to see how the estimates for the quantiles change with the length of observations for very large sample sizes, in practice there are no records of extremes of 1000 years available. It is therefore more interesting to see how the estimates vary over the sample size for a realistic sample size. Moreover, figures 7 and 9 show that the largest variability of the estimates and the CIs is for sample sizes up to n=150. Therefore, a close up of the results presented in figure 7 is given in figure 10.



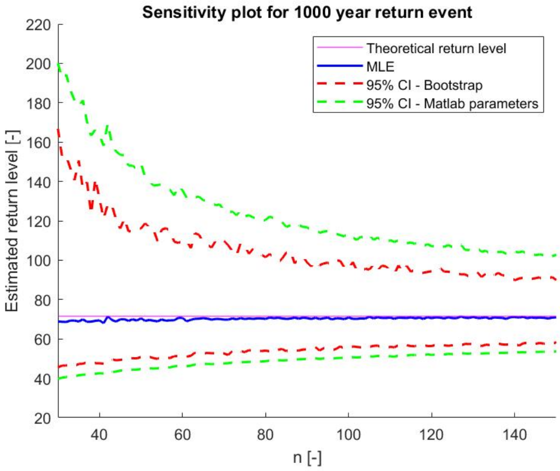
(a) 30 year return event



(b) 50 year return event



(c) 100 year return event



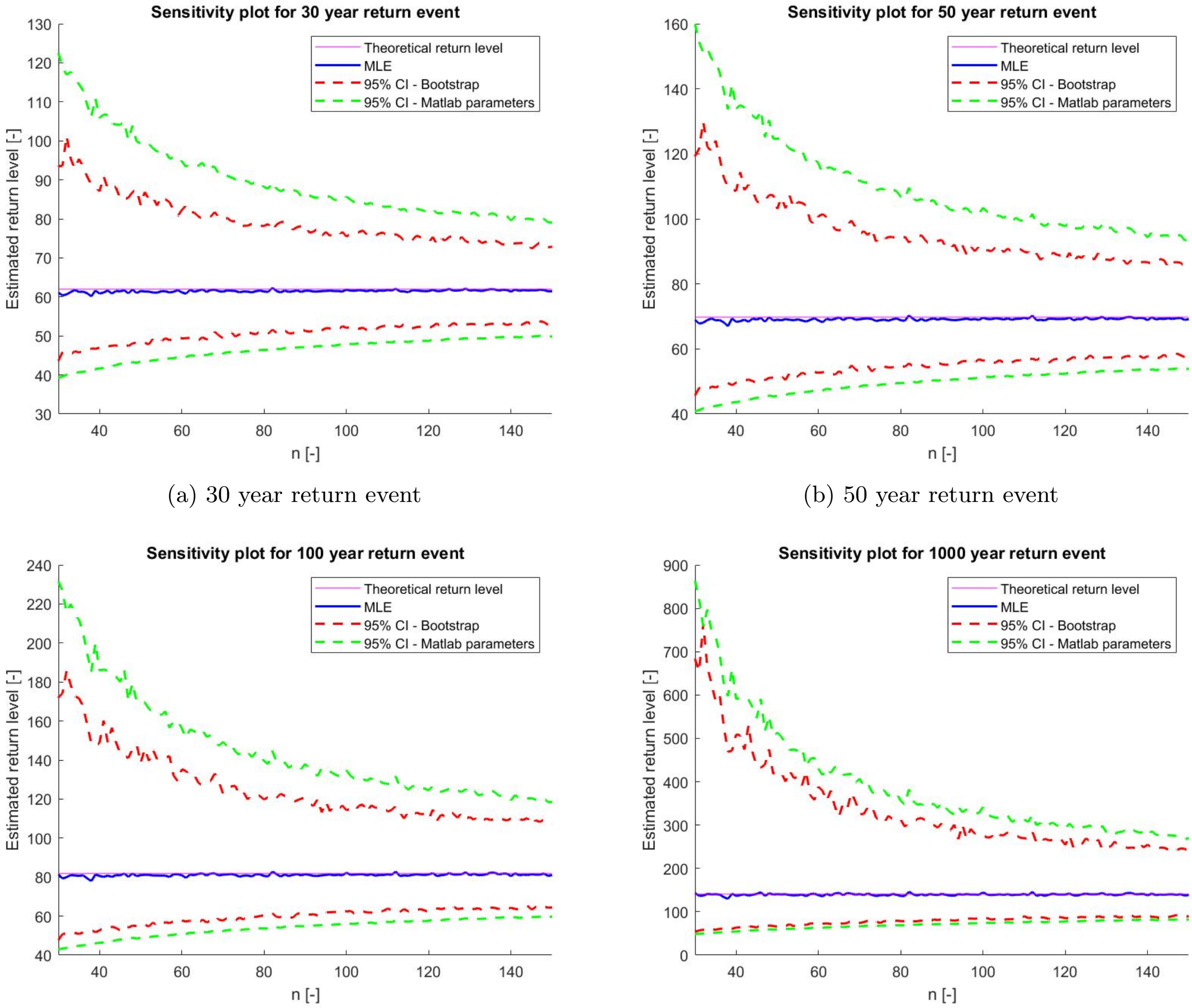
(d) 1000 year return event

Figure 10: A close up of the estimated return levels for the type I GEV distribution over the length of observations. The figures include the theoretical return level, the MLE of the return level and the 95% CI computed with a bootstrap method and with Matlab parameters

The figures show that the MLE of the return levels approach the theoretical return level for an increasing sample size.

4.1.2 Results GEV type II - Fréchet

For the type II GEV distribution the same analysis was done as presented in the previous section. The sensitivity over the length of observations up to n=1000 is included in appendix A. The close up of the analysis is included in figure 11.



(a) 30 year return event

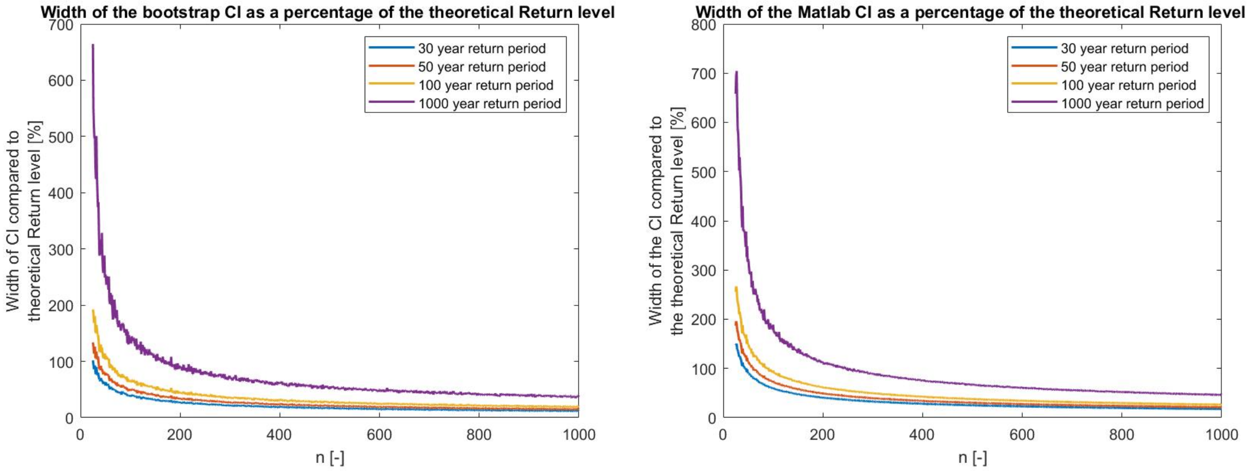
(b) 50 year return event

(c) 100 year return event

(d) 1000 year return event

Figure 11: A close up of the estimated return levels for the type II GEV distribution over the length of observations. The figures include the theoretical return level, the MLE of the return level and the 95% CI computed with a bootstrap method and with Matlab parameters

The figures show that for the type II distribution 50^{th} percentile of the MLE for the return levels is very close to the theoretical return levels. It can also be observed that the confidence bounds are wide when compared to plots for the type I GEV presented in figure 10. This can also be observed in figure 12, which shows the width of the CIs plotted relative to the theoretical return level. This can be explained by the shape parameter that is characterizing for a type II GEV distribution. The shape parameter was $= 0.25$ (see table 1). In figure 3b, the return period - return level plot was displayed. In that graph the steep gradient of the graph for the type II GEV distribution can be observed for an increasing return period. Since $\xi > 0$, there is no boundary at the upper tail, and the return levels will keep increasing for an increasing return period. Therefore, a small variation in the shape parameter, will result in a large variation in the estimated return level. The course of the graph of the type I GEV distribution in figure 3b, is much less steep. A variation of the shape parameter will result in a smaller variation in the calculated return level. This can explain the wide CIs that can be observed in figure 11 when compared to figure 10. What is also noticeable is that the upper and lower bound of the CIs (for both methods) show asymmetric behaviour. This can also be explained by the non-linear response of the estimated return levels to the shape parameters. A slightly larger shape parameter will result in a relatively larger variation from the theoretical return level than a slightly lower shape parameter.



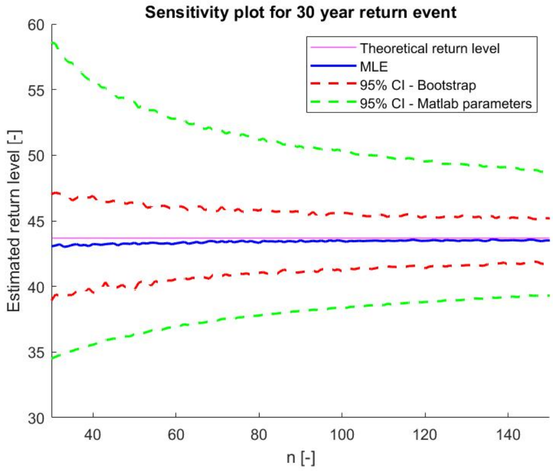
(a) Bootstrap CI

(b) Matlab parameters CI

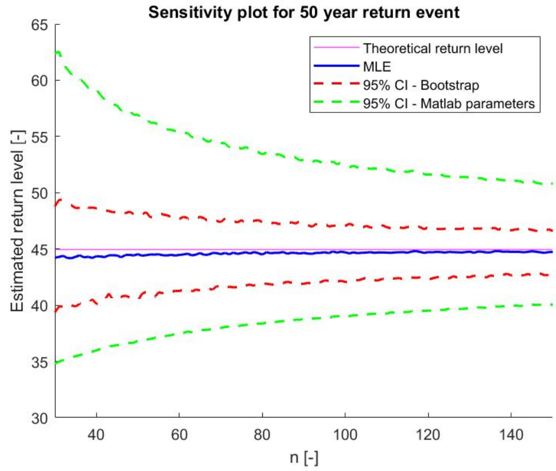
Figure 12: The width of the CI relative to the theoretical return level for the type II GEV distribution. CI obtained with the bootstrap method (12a) and CI obtained with the Matlab parameter method (12b)

4.1.3 Results GEV type III - Reversed Weibull

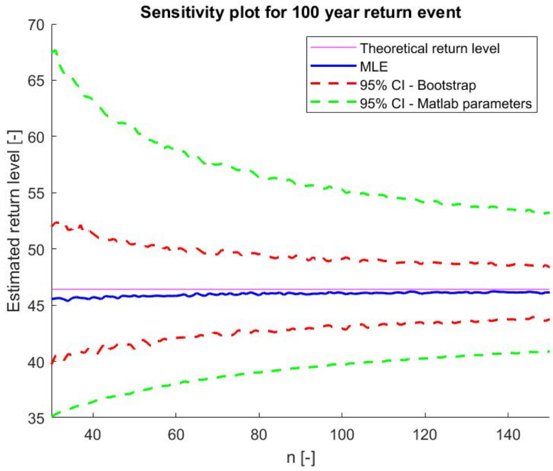
The results from the sensitivity analysis for the type III GEV distribution up to a sample size of n=1000 are included in appendix A. A close up of the results up to a length of n=150 is included in figure 13.



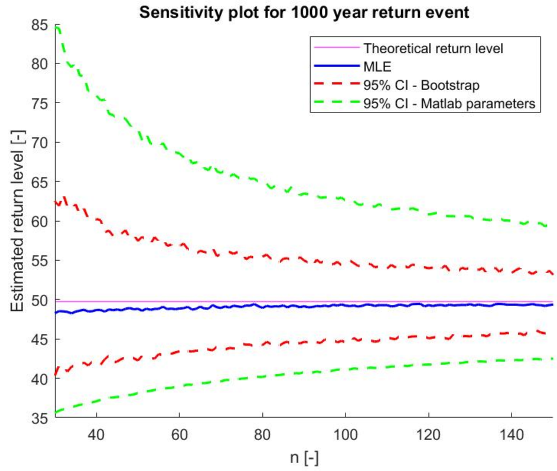
(a) 30 year return event



(b) 50 year return event



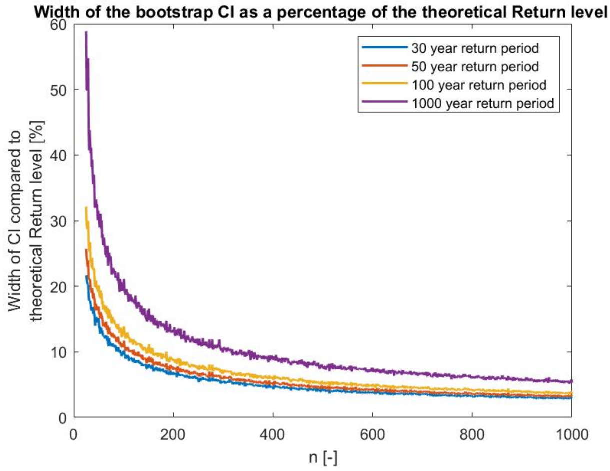
(c) 100 year return event



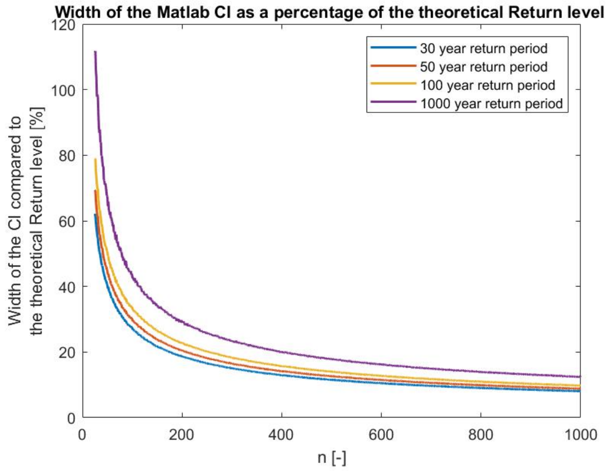
(d) 1000 year return event

Figure 13: A close up of the estimated return levels for the type III GEV distribution over the length of observations. The figures include the theoretical return level, the MLE of the return level and the 95% CI computed with a bootstrap method and with Matlab parameters

The figures show that the CIs (with both methods) are less wide than for the type II and type I distributions. For clarity, the width of the CIs are also plotted as a function of the theoretical return level in figure 14. The subplots show that the confidence in the estimates is better than for the type I and type II distributions. This can be explained by the negative shape parameter that is characterizing for the type III GEV distribution. Figure 3b (showing the theoretical return level over the return periods), shows that the type III GEV distribution has an asymmptote at the upper tail. A small deviation in the shape parameter will create a similar graph with values close to that of the theoretical distribution. A direct result is that the confidence interval is relatively small. The upper and lower bound of the CI show appear to have a near symmetric course.



(a) Bootstrap CI



(b) Matlab parameters CI

Figure 14: The width of the CI relative to the theoretical return level for the type III GEV distribution. CI obtained with the bootstrap method (a) and CI obtained with the Matlab parameter method (b)

4.1.4 Comparing the different types of the GEV distribution

In the sections above, the results from the analysis of the three types of GEV distributions were already briefly discussed. The main results were that the confidence in the estimates of the return levels increases with increasing sample size, up to a sample size of roughly 400. After that, the confidence increases almost nothing when the sample size is increased. The analysis showed that the confidence in the estimated return levels can be calculated in different ways. In the analysis the bootstrapping method and the Matlab parameter estimation method were used. The CI calculated with the Matlab parameter method resulted in a wider CI than the bootstrap method for all results. An explanation for this can be that the Matlab parameter approach uses the lower and upper bound of all the parameters, while the bootstrap approach uses the lower and upper bound of the calculated return levels. Since we have seen that the two approaches behave in a same manner, doing more analyses with both methods does not provide any more information. That is why in the case studies only the bootstrapping method was used to calculate the CI.

The results also pointed out the differences in the behaviour of the three distributions. To clarify this, a return period - return level plot was made with all three distributions in it. A fixed sample size of n=400 was used for the calculations and the CI were determined with the bootstrapping method. The plot is displayed in figure 15. This figure clearly shows the difference in behaviour of the three distributions and their CI. The figure proves that for the same sample size, the type II GEV distribution results in the widest CI. The asymmetric behaviour of the upper and lower bound of the CI can be observed in the graph. This can most clearly be observed for the type II GEV distribution. The graph clearly shows how the asymmetry of the lower and upper bound of the CI increases for an increasing return period. The figure also shows the asymptotic behaviour of the type III distribution. Because of this behaviour, the type III distribution also results in the smallest CI.

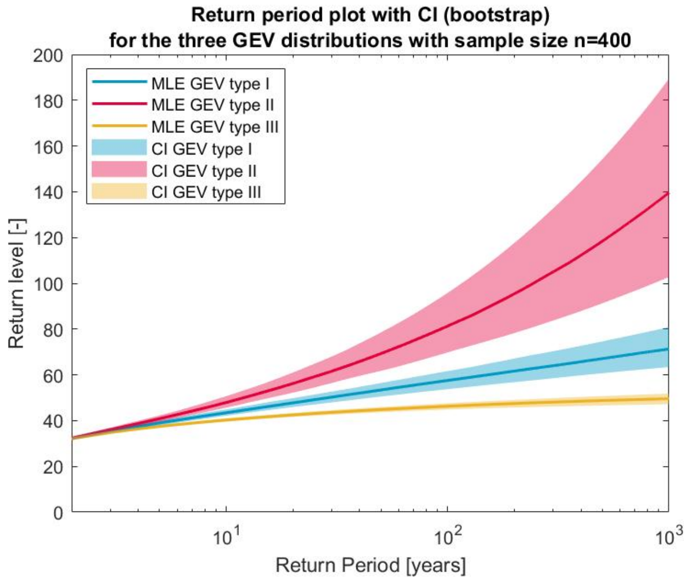


Figure 15: Return period plot for the three distributions that were used in the analysis for a fixed length of the observations of n=400. The CIs in the figure were obtained with the bootstrapping method.

4.1.5 Length of observations versus the inferred return period

This section included the results of the analysis where the influence of the sample size was compared to the influence of the magnitude of the return event. The results of the analysis for the type I GEV distribution are included in figure 16. The results of the type II and III GEV distributions are included in figure 17 and 18 respectively. On the x-axis, the return level (Q_X) and the number of observations are presented. For example: for the estimation of the return period for return level Q_{30}, 30 observations were used. The theoretical return period of that quantile is T=30. In the figures, the blue lines represent the 25^{th} and 75^{th} percentiles and the red line represents the 50^{th} percentile. The values within the blue bounds are considered the 50% CI. This CI is determined by MatLab directly from the bootstrapping results. Note that the y-axis for Q_{1000} has a different scale than the y-axis for the other quantiles. This is done solely so the results could be easier interpreted.

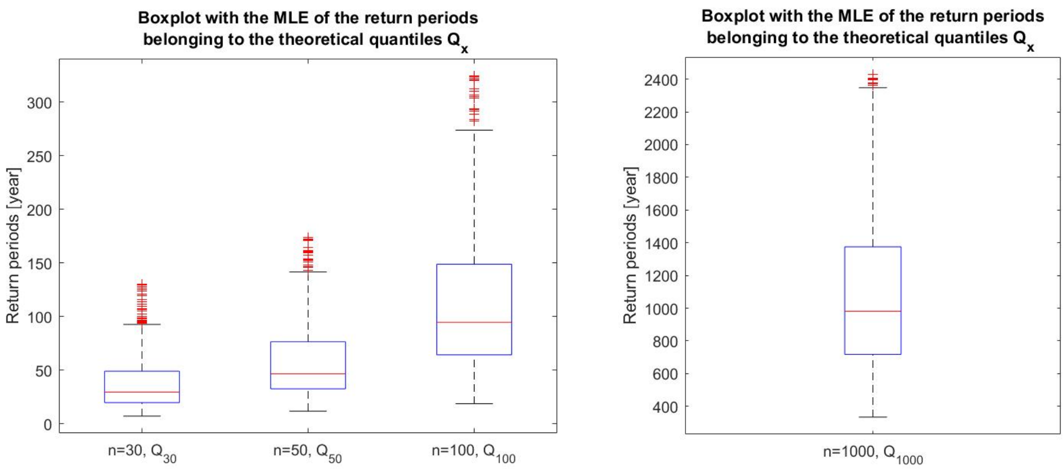


Figure 16: Estimated return period for the theoretical quantiles for the type I GEV distribution

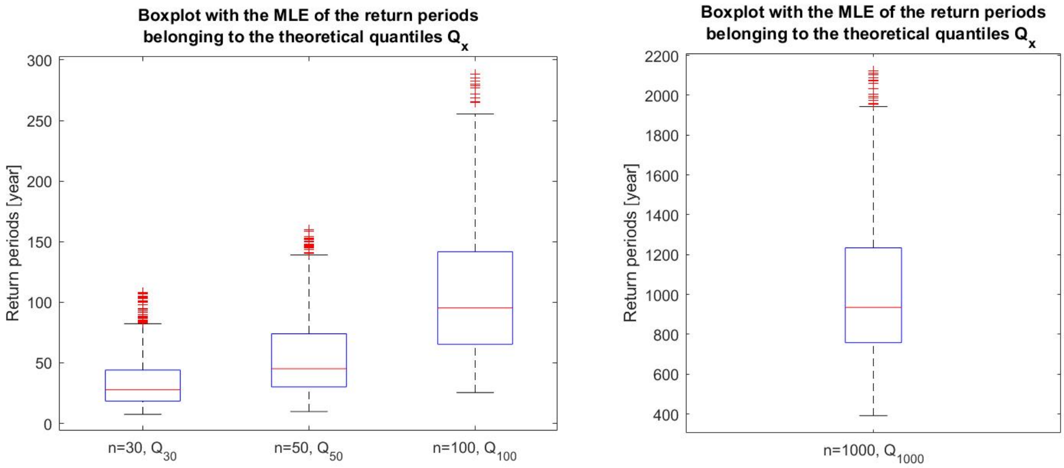


Figure 17: Estimated return period for the theoretical quantiles for the type II GEV distribution.

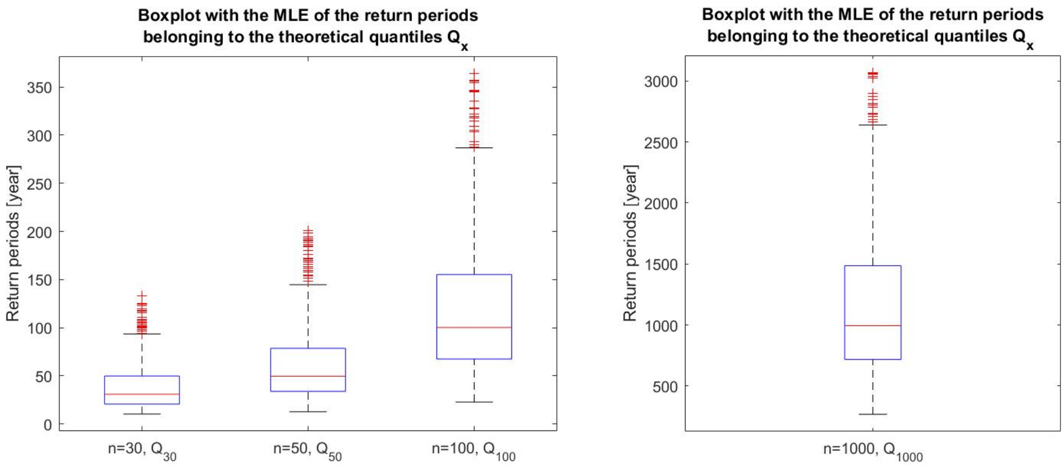


Figure 18: Estimated return period for the theoretical quantiles for the type II GEV distribution.

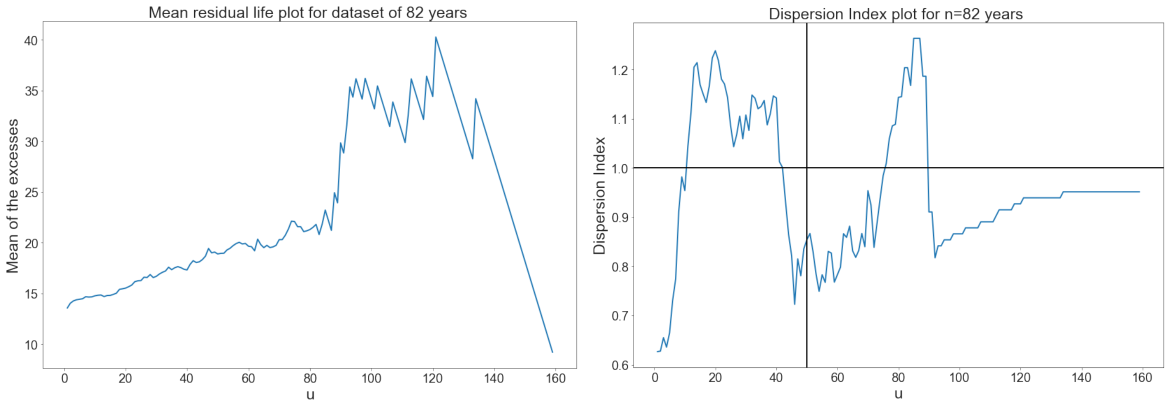
The figures for all three types of the GEV distribution show, that there is a higher confidence in the lower return periods that were calculated with a smaller sample size. The higher the return period and sample size, the wider the CI is. This suggests that the influence of the increasing return level (and associated return period) is dominant over the increasing sample size.

The figures also show that the CI of the type II GEV distribution is smaller than those for the type I and III GEV. This suggests that we have higher confidence in the return periods estimated from the type II distribution. However, we saw in the previously presented results that the confidence in the estimated return levels is lower for the type II GEV distribution. This can be explained with the graph in figure 15. The figure shows that the course of the graph for a type II GEV distribution is steeper when compared to the type I and II GEV. This means that for a small deviation in the return level (on the y-axis), the deviation in the return period will also be small (as when compared to the type I and III). In this analysis that means that when estimating the return periods with the parameters, the estimations will differ less from the theoretical return period when the type II distribution is used. The results of the type III GEV distribution have the widest CI. This can be explained by the asymptotic behaviour of the distribution. A small deviation in the return period, will result in a large deviation in the return period.

4.2 Results case study I: Precipitation in New York

In this chapter the results of case study 1: Precipitation in New York are presented. A description of the dataset that was used for the analysis is included in section 3.2.1. A Mann-Kendall trend test was done to detect if there was a trend in the data. The result of the test was that there was no trend of statistical significance in the data.

The distribution that best fitted the daily precipitation events in New York was the GPD. The threshold level was chosen with the help of the mean residual life plot (figure 19a) and the Dispersion Index (DI) plot (figure 19b). A fixed minimum distance of 1 day was chosen between consecutive peaks. This was done so only the highest peak of a storm was selected. Figure 19a shows that there is a somewhat linear slope in the plot between a threshold u=20 and u=60. Figure 19b shows the dispersion index. The figure shows that DI varies over the thresholds. The values are however reasonably close to 1. A threshold of 50 mm/day was selected.



(a) The mean residual life plot for the daily precipitation events in New York.

(b) Dispersion Index plot for the daily precipitation events in New York

Figure 19: Figures used for the graphical threshold selection.

In the analysis, chronological subsets of the entire dataset were used. The subsets varied from a length of 30 years to a length of 82 years (the entire dataset). The validity of the chosen threshold had to be checked for these subsets. It was chosen to check the validity of the threshold for subsets with length 30, 35, ..., 80 and 82 years. The MRLP and the DI plot for these susbsets are included in appendix B. The plots show that the assumption of a fixed threshold of 50 mm/day is valid for the included subsets.
After the threshold was chosen, the POT method was used to select the peak events. A graphical representation of all the precipitation events in New York and the events above the threshold of u=50mm/day is shown in figure 20.

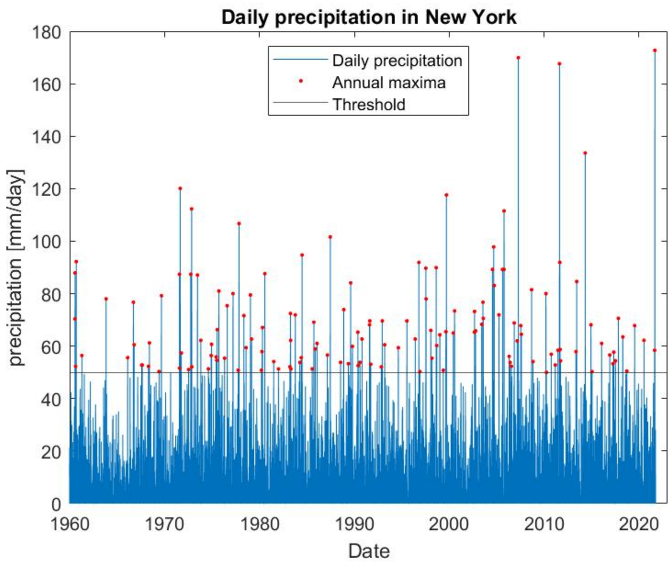
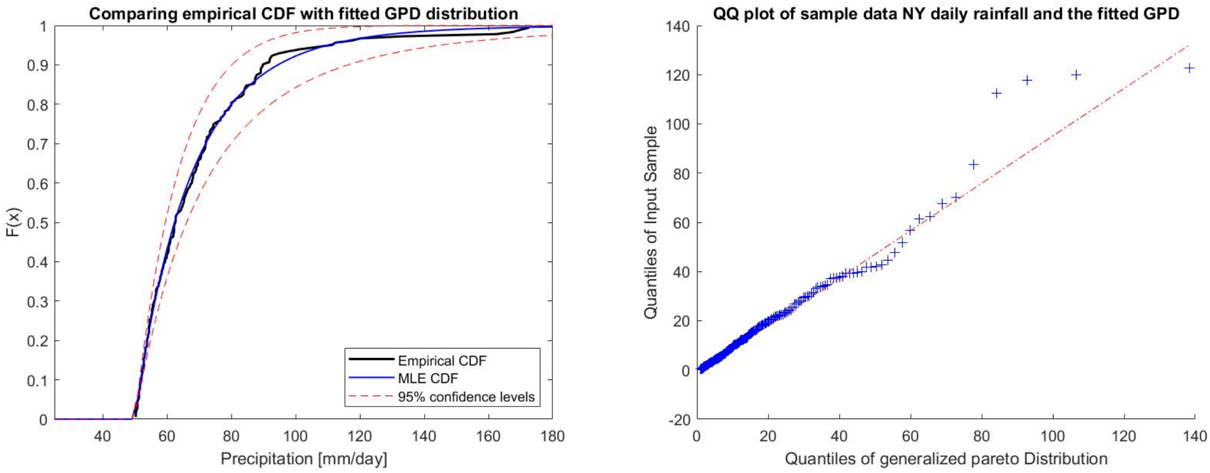


Figure 20: A graphical representation of the precipitation events over time in New York [mm/day]. The peak events above the threshold (u=50 mm/day) are indicated in red.

The chosen threshold resulted in a set of 182 peak events (2.2 peaks per year) to which the GPD was fitted. The MLE of the parameters of the GPD (fitted to the entire dataset) were $\xi = 0.098$ and $\sigma = 17.23$. Figure 21 shows a figure of the empirical CDF and the CDF of the fitted distribution as well as a QQ plot. The figures show that the distribution gives a reasonable fit. However, especially for the upper quantiles, the distribution varies from the observations. Since the GPD gave a better fit than the GEV distribution, the GPD fit was favoured and the distribution was used for the analysis.



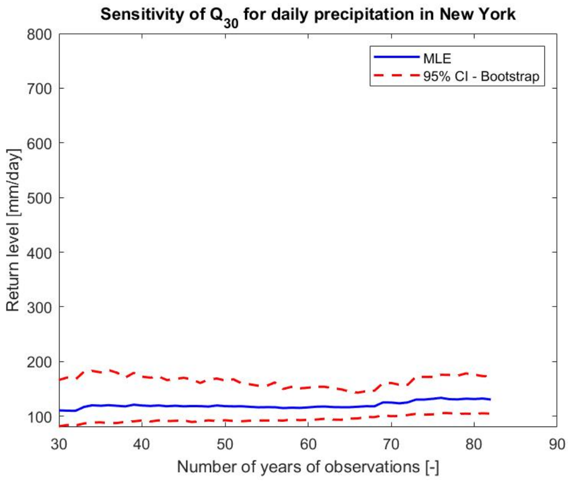
(a) The comparison between the empirical CDF and theoretical CDF of 82 years of observations.

(b) QQ plot of the sample data and the fitted GPD distribution for 82 years of observations.

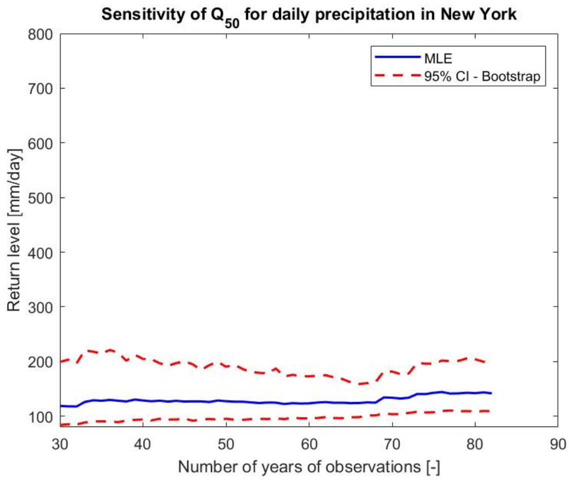
Figure 21: Figures for testing the goodness of fit of the fitted GPD distribution to the sample data of the precipitation events in New York.

4.2.1 Sensitivity of the estimated return levels

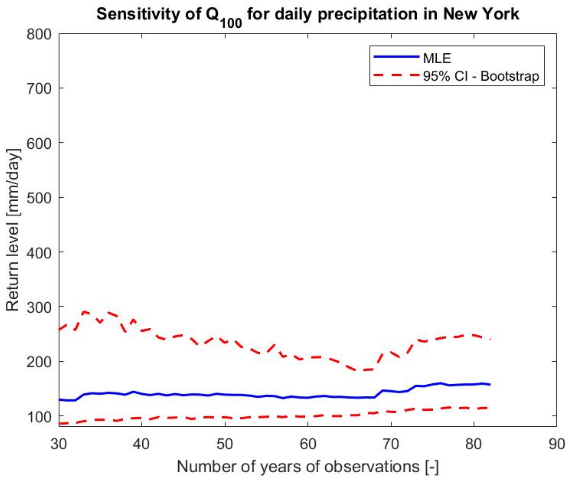
After the type of distribution and the threshold were chosen, a year by year analysis was done. In this analysis the 30, 50, 100 and 1000 year return levels were determined from the data. The analysis started at 30 years of observations and ended at 82 years of observations. During the analysis, the threshold limit of $u_0 = 50mm/day$ was kept constant.
The results of estimated return levels are displayed in figure 22. The bootstrapping method was used to determine the CI. To observe the variability of the width of the CI clearly, the scale of the y-axis is kept constant. The figures show that the confidence bound increases for increasing return periods. This was as expected since the same was observed in the analysis of the synthetic data. The figures show that overall there seems to be a decrease in the width of the CI over the number of years of observations.



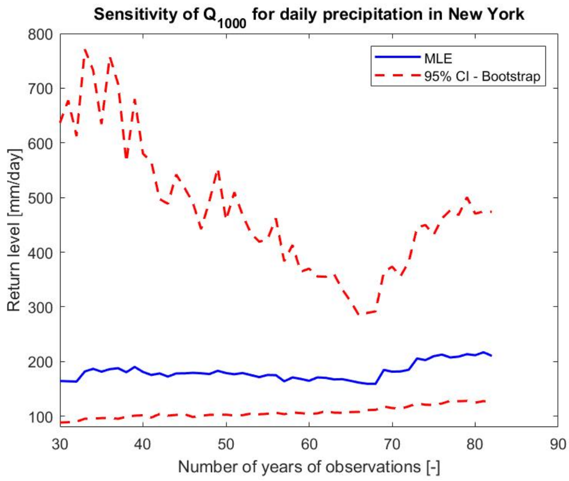
(a) 30 year return event



(b) 50 year return event



(c) 100 year return event



(d) 1000 year return event

Figure 22: The estimated return levels of the precipitation in New York for the 30, 50, 100 and 1000 year return events over the length of observations. The associated 95% CI (in red) is calculated with a bootstrapping method.

To better observe the variability of the width of the CI over the years of observations, a plot was made of the width of the CI relative to the 50th percentile of the MLE. The plot is included in figure 23. The figure shows the decrease in the width in the CI for the four return periods as the years of observations increase. However, the confidence in the estimates remains low, even for many years of observations. We see that after 82 years of observations the CI for the 30 year event is around 50% of the estimated return level. Graph 22a shows that the MLE values of the 30 year return periods is around between 100 mm/day. A CI that is 50%, means that we have 95% certainty that the 30 year return event is somewhere between 50 and 150 mm/day. This is a very wide range. The widths of the CI for higher return periods are even higher. It can therefor be argued if the fitted distribution gives valuable information regarding the events.

There can be two possible explanations for the low confidence in the estimates. The first one is that the length of the observations is not long enough. A clear decrease in the width of the CI can be observed, and it could be argued that a longer set of observations will result in better estimates. However, figure 23 also shows that the width of the CI does not decrease anymore after 65 observations. Another explanation for the low confidence is that the GPD fit (or the threshold selected) turned out to not be an appropriate fit for the data. Observing the QQ-plot and the CDF/ECDF plot in figure 21, this is probably what happened.

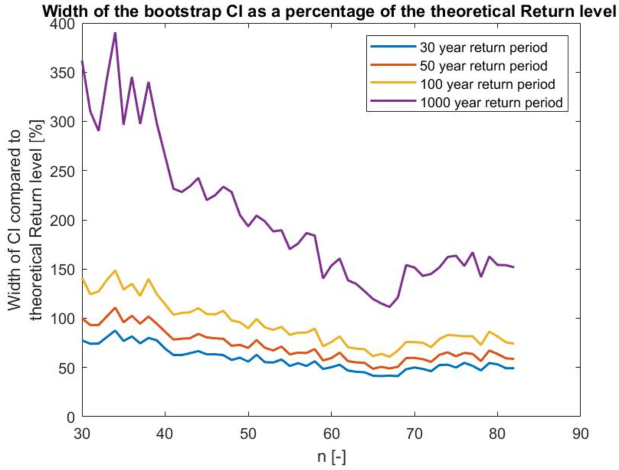
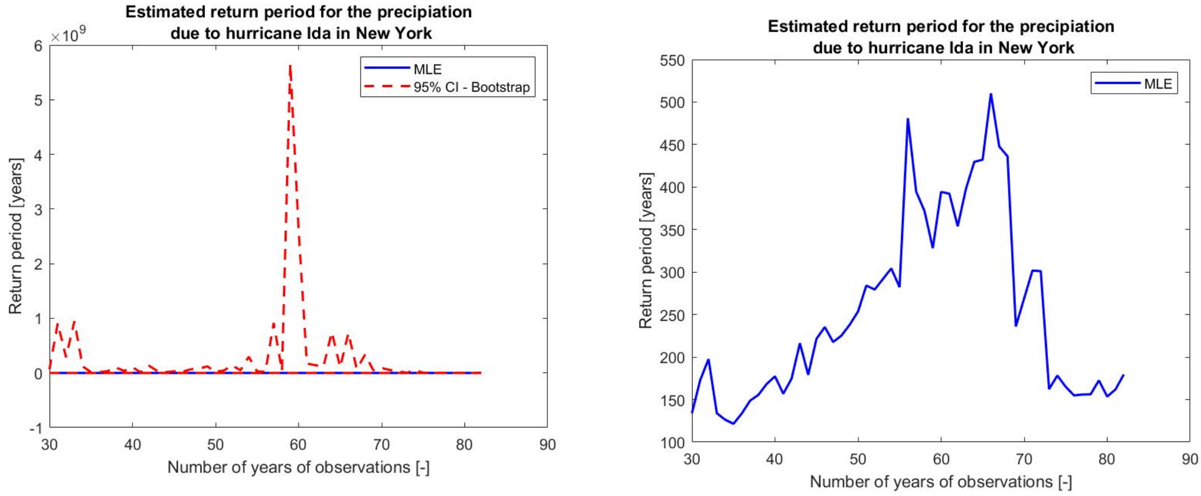


Figure 23: The width of the CI as a function of the MLE of the return levels over the years of observations for the precipitation in New York.

4.2.2 Variability of the return period for Hurricane Ida

An analysis was done to the sensitivity of the inferred return period of the precipitation recorded in New York during hurricane Ida. With a chronological year by year analysis, the MLE of the return period associated to the event was determined for every subset taken from the entire dataset. The CI was also determined with the bootstrapping method. The results of the analysis are displayed in figure 24. Figure 24a presents the MLE of the return period together with the calculated CI. The figure shows that the upper bound of the CI is at some points extremely high (around 10^9 years). An explanation for this is that the event is around the upper bound of the distribution. When this is the case, a small deviation in the parameters will result in a large deviation in the estimated return period. This is something that was also observed in chapter 4.1.5. The result is that the confidence we have in the estimates is very low. With this kept in mind, the results were plotted once more without the CI. This is shown in the graph in figure 24b. The figure shows that there is a large variability over the number of observations. After 82 years of observations, a return period of 162 years is found. The event was recorded in the newspaper as an event with a return period of 200 years. However, this recorded return period was for the 6 hour precipitation event, while in this analysis the 24 hour precipitation event was used. This difference in temporal scale makes it difficult (and not completely correct) to compare the found return periods directly. However, the found return periods are relatively close together. The fact remains that the found estimate from the analysis comes with a very large uncertainty.



(a) The estimated return period for the precipitation during hurricane Ida in New York over the number of years of observations. The CI is obtained with the bootstrapping method and displayed in red.

(b) The estimated return period for the precipitation during hurricane Ida in New York over the number of years of observations.

Figure 24: Results of the sensitivity analysis of the inferred return period of the precipitation event in New York in September 2021.

4.3 Results Case study 2: Precipitation in Limburg

A Mann-Kendall trend test was done to detect if there was a trend in the data. The result of the test was that no statistical significant trend that could be detected when using a dataset that excluded the event of 2021. If the dataset including the 2021 event was used, a trend could be detected. Apparently, the 2021 event was so high, that when the event is included in the time series a trend could be detected. We then decided to remove the 2021 event for the following analysis. Hence, a data set of the 48h precipitation up to the end of year 2020 was used. A graphical representation of the events and the annual maxima is given in figure 25.

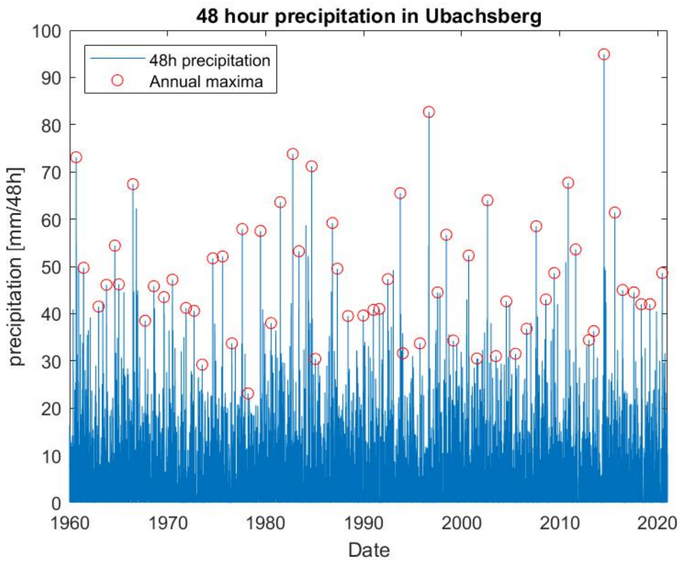
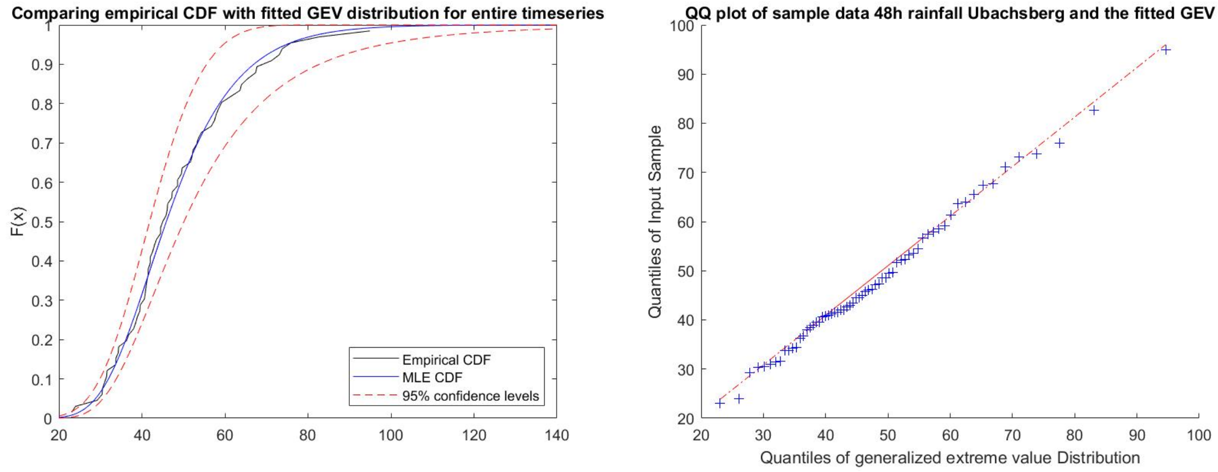


Figure 25: A graphical representation of the precipitation events over time in Ubachsberg, Limburg [mm/48h]. The annual maxima are indicated in red.

The distribution that best fitted the precipitation events in Limburg was the GEV distribution. This was determined with the help of a QQ-plot and a plot that compared the CDF of the theoretical distribution to the empirical CDF (the ECDF). The MLE of the parameters for the entire dataset were $\xi = 0.1364$, $\sigma = 11.95$ and $\mu = 41.23$. This means that it can be categorized as a type II GEV distribution. The QQ-plot is presented in figure 26b and the comparison between the CDF to the ECDF is presented in figure 26b.



(a) The comparison between the empirical CDF and theoretical CDF for the entire dataset.

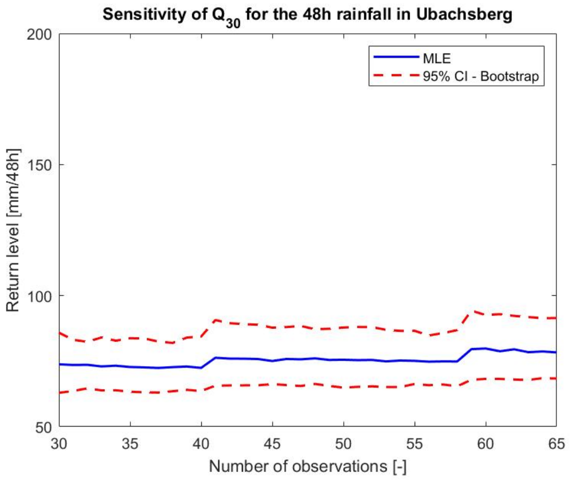
(b) QQ-plot of the sample data and the fitted GPD distribution for the entire dataset.

Figure 26: Graphical presentation of the goodness of fit of the annual maxima precipitation in Limburg and the fitted GEV distribution.

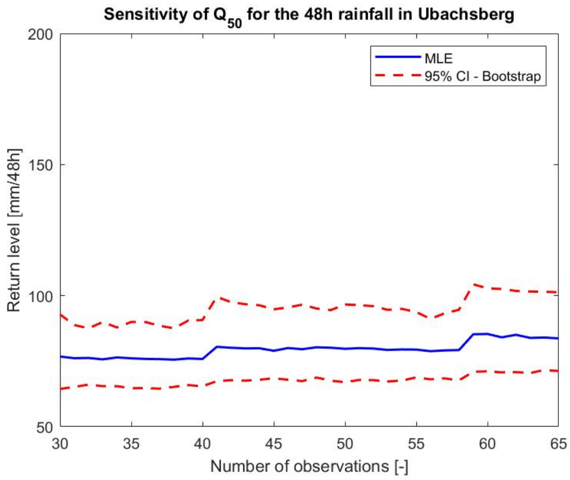
The figures show that the GEV distribution fits the data well and it was therefore determined that the distribution could be used for the analysis.

4.3.1 Sensitivity of the estimated return levels

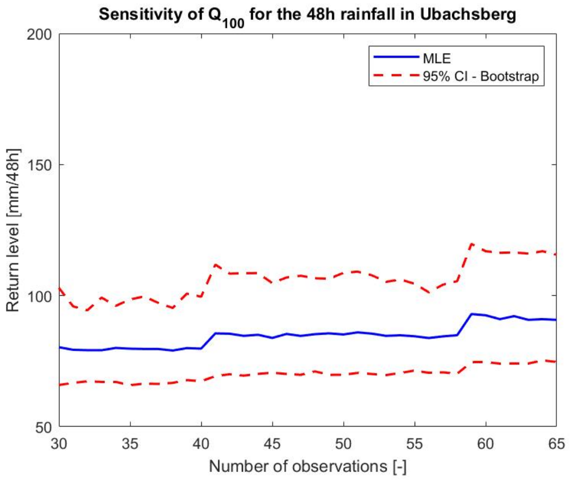
As described in chapter 3.2.2, a chronological year by year analysis was done to observe the sensitivity to the length of observation. For each subset the GEV distribution was fitted and the parameters were estimated. With the bootstrap method, the 50th percentile and the 95% CI were determined. The results of the sensitivity analysis are included in figure 27.



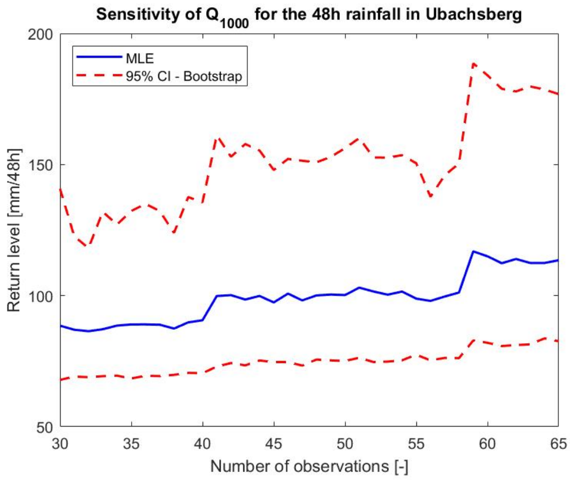
(a) 30 year return event



(b) 50 year return event



(c) 100 year return event



(d) 1000 year return event

Figure 27: The estimated return levels of the precipitation in Limburg for the 30, 50, 100 and 1000 year return events over the length of observations. The associated 95% CI (in red) is calculated with a bootstrapping method.

The figures show that there is no decrease in the width of the CI over the length of observations. There even seems to be a small increase in the width of the CI. In figure 28, the width of the CI is plotted as a function of the 50th percentile estimate of the return levels. In this figure also the increase in the width of the CI can be observed. The figure shows that the width of the CI has small variations over the number of observations, but overall remains constant. The subplots in figure 27 show that the estimates for the four quantiles increase slightly over the length of observations. However, there are two clear points in time that the distribution changes. This is around n=41 and n=58. These correspond to the years 2001 and 2018. When observing the graph in figure 25, there is not a very clear irregularity that can be observed in the annual maxima. A possible explanation could be that there is a trend in the events but only for smaller sample sizes. The performed Mann-Kendall trend test was only done for the entire length of the dataset. A possible explanation for the 'jumps', is that there is a small trend. It remains uncertain why the width of the CI does not decrease in size for an increasing length of observations. This was expected after the analysis in section 4.1. Also, from the figures in figure 26, it followed that the GEV distribution was an appropriate fit.

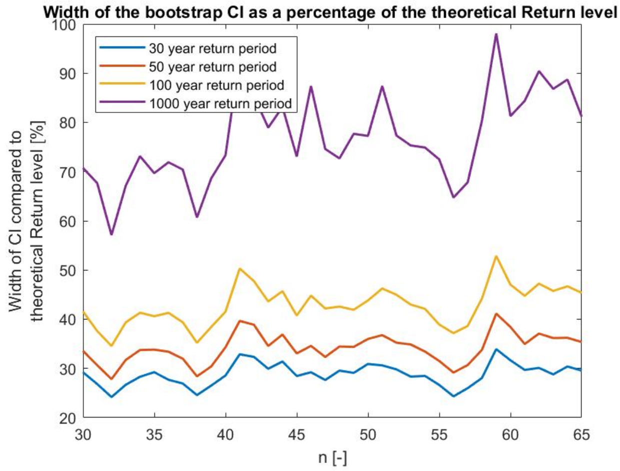


Figure 28: The width of the CI as a function of the MLE of the return levels over the years of observations for the precipitation in Limburg.

4.3.2 Variability of the return period of the 2021 precipitation in Limburg

The results of the analysis of the return period of the 48 precipitation in Limburg from July 2021 are presented in this chapter. The event is shown in green in figure 29. The figure shows that the event is much higher than any other 48h event recorded by the weather station of Ubachsberg.

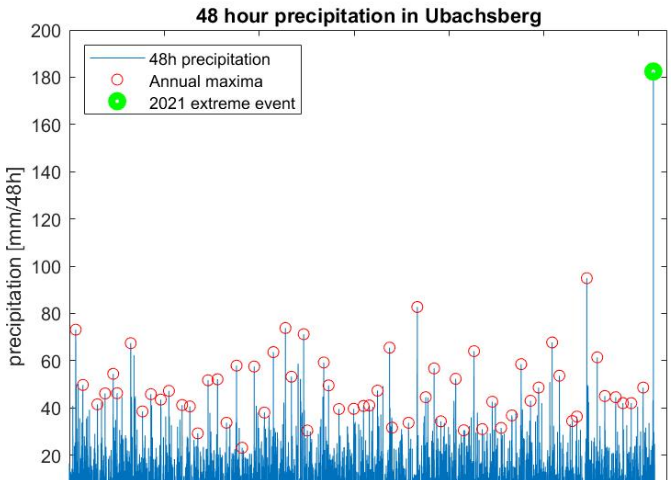
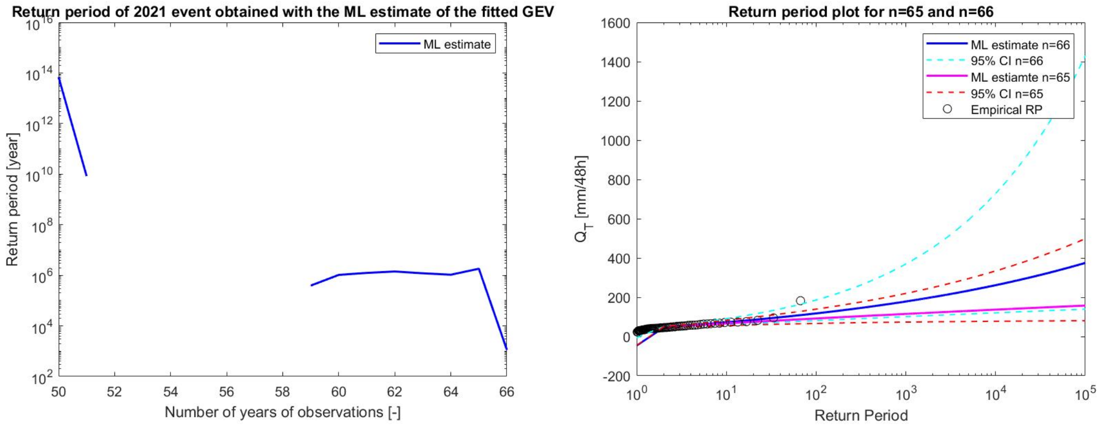


Figure 29: A graphical representation of the precipitation events over time in Ubachsberg, Limburg [mm/48h]. The annual maxima are indicated in red and the 2021 event is indicated in green.

The results of the analysis to the sensitivity of the estimated return period are shown in figure 30a. In this figure, the calculated CI is not included. The reason for this is, like for the case of the precipitation in New York, that the upper bound of the CI is extremely high and goes to infinity for most sample sizes. That is why it was chosen to only present the MLE of the return period.



(a) Estimation of the return period with the ML estimate for the parameters

(b) Return period plot for a dataset excluding the 2021 event (n=65) and including the event (n=66)

Figure 30: Figures for 2021 event Limburg

The figure shows that the estimates for the return period are extremely high. There is also a large gap in the figure. For those points the the calculated return period was infinite. This indicates that the found event is above the asymptote of the distribution, and no return period can be found for the event. When the event is included in the analysis (the last point in the graph, number of observations = 66), there is a drop in the return period to roughly 1000 years. Considering that the point before there is in the order of 10^5 (see figure 30a), this suggests a big change in the distribution. The ML estimate for the distribution and the 95% CI (with the bootstrapping method) were used to plot the return period graph. This was done for the cases n=65 (dataset without 2021 event) and n=66 (dataset including 2021 event). The graphs are shown in figure 30b. From this figure it can clearly be seen why the return period for the 2021 event is so high when calculated with the dataset of length n=65. The event was 182 mm precipitation in 48 hours. In purple, the graph is given for the distribution that was fitted to a dataset without this event. It appears that the distribution has an upper bound, which would suggest it is a type III GEV distribution. The MLE for the shape parameter for this distribution is -0.02, indicating indeed a type III GEV distribution. However, when we observe the graph for the distribution that was fitted to a dataset including the event (figure 30b in blue), we see that the graph is still increasing, even at very high return periods. This would indicate a type I or II distribution. The distribution has a MLE for the shape parameter of 0.16, indicating a type II GEV distribution. The results suggests, that by adding the 1 extreme event, the distribution shifts. It should be noted, that the Mann-Kendall trend test showed that a trend could be observed in the data when the 2021 event was included. This means that the sample shows nonstationarity and that a stationary analysis might not be appropriate. This should be considered when interpreting the results presented above.

4.4 Results Case study 3: Temperatures in Sicily

A Mann-Kendall trend test was done to detect if there was a trend in the data. The result of the test was that there was no trend that could be detected in the annual maxima. A graphical representation of the daily maximum temperatures over time and the annual maxima are presented in figure 31.

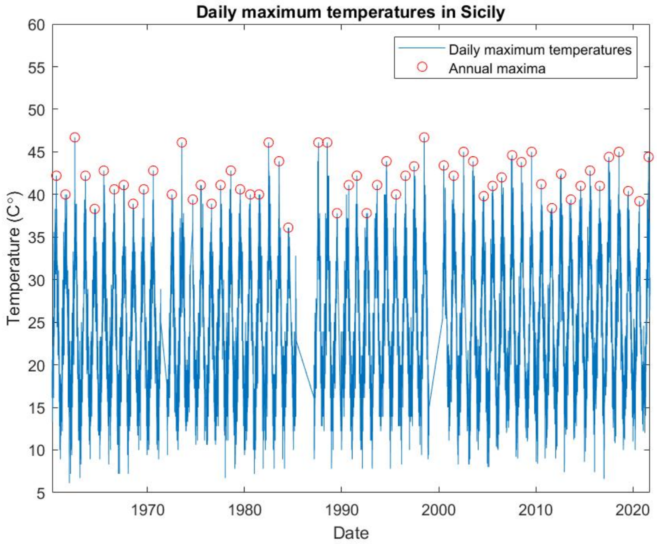
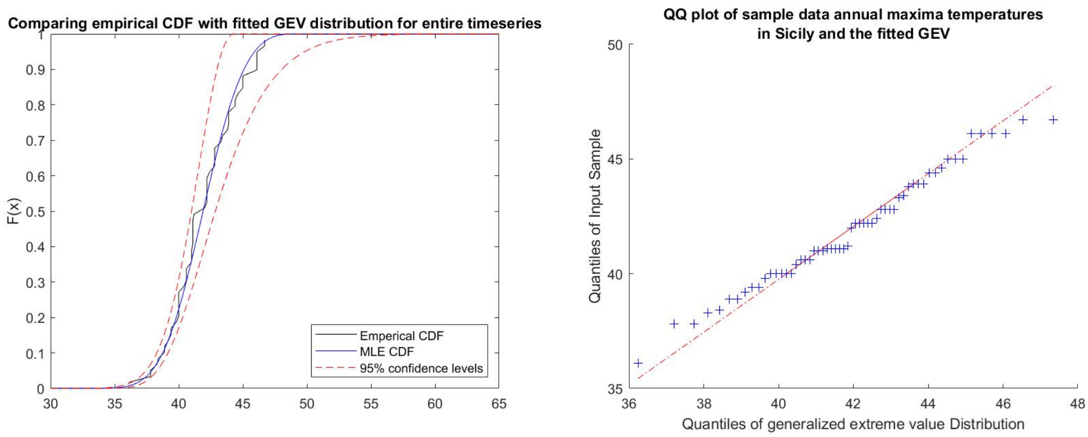


Figure 31: A graphical representation of the daily maximum temperatures over time at the Sigonella airport, Sicily [° C]. The annual maxima are indicated in red.

The distribution that best fitted the annual maxima temperatures in Sicily was the GEV distribution. This was determined with the help of a QQ-plot and a plot that compared the CDF of the theoretical distribution to the ECDF. The MLE of the parameters for the entire dataset were $\xi = -0.29$, $\sigma = 2.44$ and $\mu = 41.04$. This means that it can be categorized as a type III GEV distribution. The QQ-plot is presented in figure 32b and the comparison between the CDF and the ECDF is presented in figure 32a.



(a) The comparison between the empirical CDF and theoretical CDF for the entire dataset

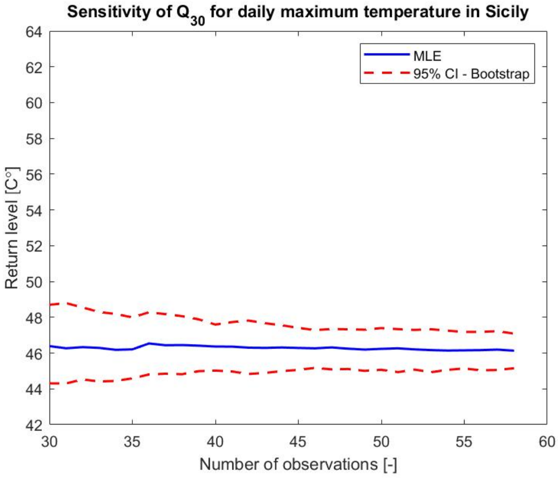
(b) QQ-plot of the sample data and the fitted GPDdistribution for 82 years of observations.

Figure 32: Graphical tests for the goodness of fit of the annual maxima temperatures and the fitted GEV distribution.

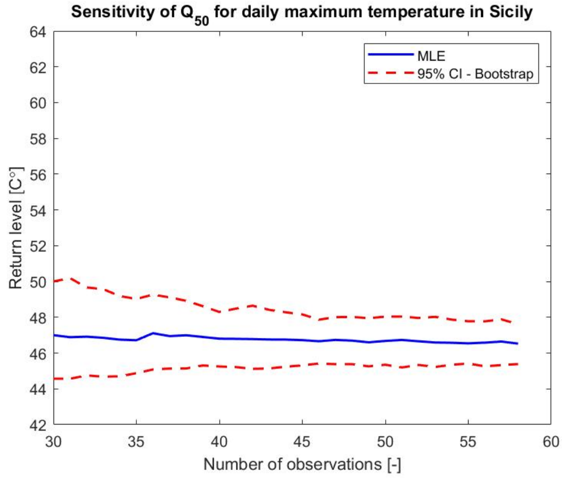
The figures show that the quantiles of the fitted GEV distribution match the empirical quantiles well, but not perfect. In the most upper quantile the fitted GEV distribution gives a slight overestimation and in the lower quantiles it results in a slight underestimation. Nonetheless, the fit is good enough to use the model for the analysis.

4.4.1 Sensitivity of the estimated return levels

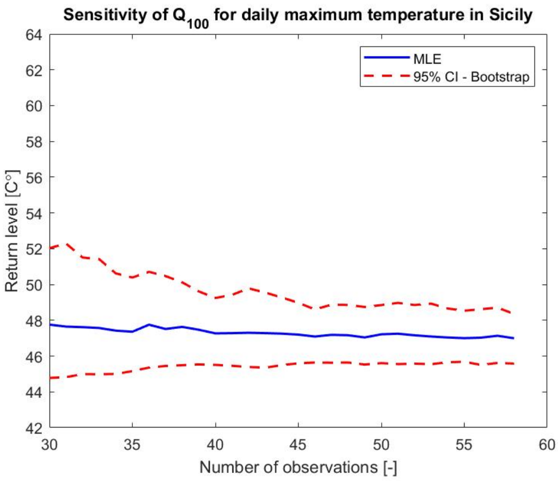
The results of the analysis to the sensitivity of the inferred return periods are displayed in figure 33. The figures show the 50th percentile MLE of the return levels as well as the calculated 95% CI. The subplots are displayed with the same scale in the y-axis. The figures show that the MLE return levels remain more or less constant over the length of observations. Moreover, they show that the width of the CI decreases with an increasing length of observation. To clarify this, the width of the CI was plotted as a function of the 50th percentile of the MLE. The graph is included in figure 34.



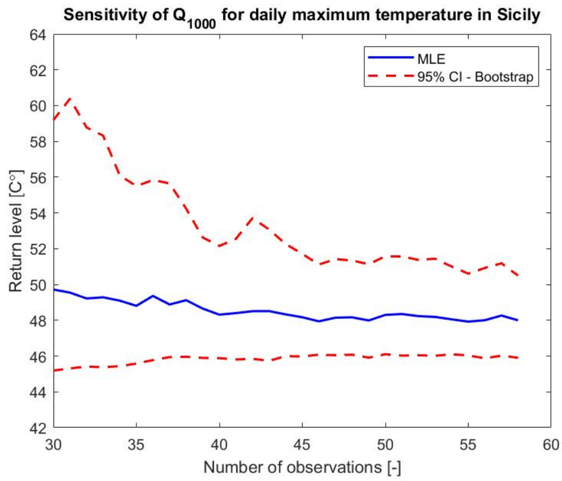
(a) 30 year return event



(b) 50 year return event



(c) 100 year return event



(d) 1000 year return event

Figure 33: The estimated return levels of the temperatures in Sicily for the 30, 50, 100 and 1000 year return events over the length of observations. The associated 95% CI (in red) is calculated with a bootstrapping method.

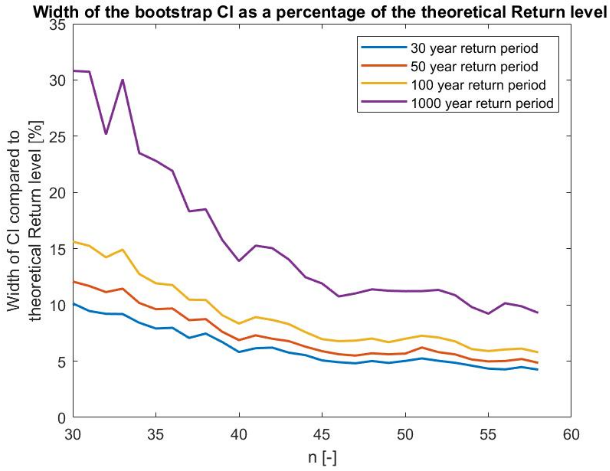


Figure 34: The width of the CI as a function of the MLE of the return levels over the years of observations for the temperatures in Sicily.

Figure 34 show that there is indeed a decrease in the width of the CI over the numbers of observations. The width of the confidence interval after 58 years of observations is around 5% of the 50th quantile for the 30-, 50- and 100 year return events. For the 1000 year return event this is around 10%.

4.4.2 Variability of the return period of the 2021 event in Sicily

In this section the results of the analysis of the sensitivity of the return period for the 2021 event in Sicily are included. For the analysis the (unconfirmed) temperature of 48.8 degrees was used. This temperature was not recorded by the weather station where the rest of the data was from. However, the two locations were close to each other (roughly 50 km apart). The return period of the temperature of 48.8 degrees Celsius was estimated by the chronological subsets of the dataset with length N=30,31,...58. The results are presented in figure 35. The figure present the MLE of the estimated return period. Again, the upper bound of the CI went to infinity and is therefore not included in the graph.

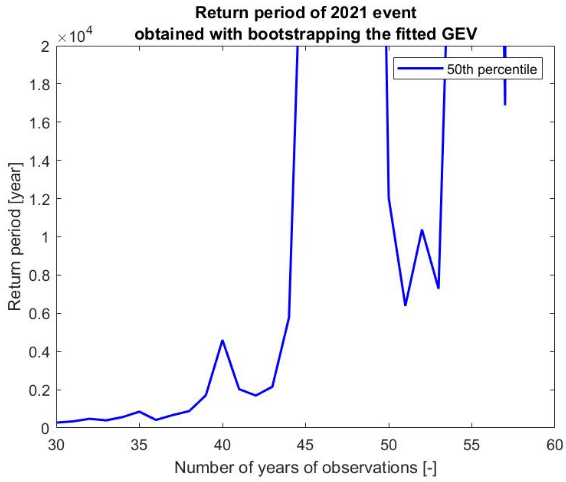


Figure 35: Estimation of the return period of 2021 event with two methods

The figures show that the return period is very sensitive to the subset that is used to estimate the parameters. While at the beginning of the figure the return period estimated for the event would be around 1000 years, when longer time series are used the estimate for the return period are around $2.6 \cdot 10^5$ years at n=47, and goes to infinity at n=46 and n=55. The reason that this happens is that we are trying to find a return period for an event that is larger than the upper bound of the distribution. Therefor, the analysis does not provide very valuable information.

5 Discussion

In this chapter a discussion of the results is presented. For consistency, the discussion is split up in two parts.

5.1 Discussion of the analysis of synthetic data

The analysis of the synthetic data showed that the estimates for the inferred return levels get closer to the theoretical value with an increasing sample size. The results also showed that an increase in the sample size generally results in a higher confidence in the estimated return levels. However, increasing the sample size above 400 samples, did not result in a significantly higher confidence. The results showed that for small sample sizes (n<150), the confidence in the estimates increases fast for an increasing sample size.

The analysis showed that two other factors have a large influence on the confidence in the return levels. The type of GEV distribution that was used in the analysis showed to have a large influence on the confidence in the estimates. This can be attributed to the shape parameter that was used. The GEV II (Fréchet) results in more uncertainty when inferring extreme return levels, than the GEV I (Gumbel) and III (reversed Weibull). This means that a shape parameter larger than 0 resulted in the lowest confidence in the estimates. A negative shape parameter resulted in the highest confidence in the estimates. Another factor of influence is the magnitude of the return event that is inferred. Estimating a return level associated with a higher return period, results in a lower confidence in the estimate. The results showed that this effect is larger than the effect of an increasing sample size.

5.2 Discussion of the analysis of the case studies

The results of the three case studies showed that when observations are used, the results are not as easily interpreted as when synthetic data is used. The three case studies and the results are briefly discussed below.

The case study of the precipitation in New York showed a decrease of the width of the CI over the length of observations. Moreover, the MLE for the return levels were more or less stable over the length of observations. This would suggest that the results of the analysis can be used. However, when the confidence in the estimates is included as information, it is clear that there is a very large uncertainty associated to the estimated return level. This is a very clear example of how too much weight can be given to an estimate, if not all the available information is communicated. The low confidence in the estimates can be explained by the relatively poor fit of the GPD distribution to the upper quantile of the data. The analysis of the return period of the 2021 event showed that the return period is very sensitive to the subsample of the dataset that was used. No clear pattern could be observed over the length of observations. It appears that other factors are more dominant than the length of observations when determining the return period.

The case study of the precipitation in Limburg showed no clear decrease in the width of the CI over the length of observations. This was not as expected, since the results from the synthetic data and the results from the first case study did show this. It is noteworthy that the case study of the the precipitation in Limburg, was the case where the chosen distribution appeared to give the best representation of the data. It is therefor not clear what caused that the influence of the length of observations could not be observed. Apparently, other factors were more dominant over the influence of the length of observations. The results of the analysis to the sensitivity of the return period showed that the magnitude of the event was so high that a return period could not be found for almost all subsets. However, the trend test already showed that a trend could be detected when the event was included in the timeseries. It is therefor not completely unexpected that the a return period could not be assigned. A trend in the data suggests a change in distribution. It is interesting that in the news a return period of 1000 years was reported (in 2019), while the found return period from this research is in the order of 10^5 years. That is a very large difference. In the analysis by (STOWA, 2021), data from 219 weather stations was used. Their different approach led to a very different result. This is another example of how sensitive the results of EVA can be.

The case study of the temperatures in Sicily showed a decrease in the width of the CI over the length of observations. The results also show that the relative width of the CI is relatively low for the 30- and 50-year events (around 5%). Putting this in perspective, this means that we can be 95% sure that the 30 year return event is between 44.5 - 49.35 $°C$. Naturally, it depends on the goal of the analysis if this is sufficient. However, generally speaking, this uncertainty is still large.

The analysis of the inferred return periods for all three case studies showed that the inlfuence of the sample size could not be observed. The estimated return periods all prove to be very sensitive. However, in all three case studies, events were chosen with a very high magnitude. The result was that the events were so high that they were around or above the upper bound of the distributions. Meaning that a return period of infinity was assigned to them. In hindsight it would have been better to choose events that were not that extreme, to observe the influence of the length of observations on the estimated return periods and the uncertainty associated.

6 Conclusion and recommendations

This chapter presents the main conclusions of the research. They are included in section 6.1. The recommendations are included in section 6.2.

6.1 Conclusion

The objective of this research report was to assess what the effect of the length of observations is when inferring rare events. From the research in this report, the following can be concluded.

- When inferring the return levels of extreme events from synthetic data, a larger sample size generally leads to a higher confidence in the estimates. However, the width of the CI will not reduce to zero for sample sizes up to 1000 samples. The increase in the confidence after a sample size of 400 is negligible. It can therefore be concluded that the confidence in the inferred return levels is sensitive for the sample size, up to a sample size of n=400.
- Besides the sample size, two other aspects have a large influence on the confidence in the estimates return levels. These are the shape parameter and the magnitude of the return event that is inferred.
- When inferring return levels of extreme events from observations, an increase in the sample size results most of the time in an increase of the confidence in the estimated return level. However, other factors can play a more dominant role.
- There is a large uncertainty associated when return periods of extreme events are inferred from observations. No clear pattern can be observed over the length of observations. The inferred return period proves to be very sensitive to the sample that it is inferred from.
- It is case dependent how many observations are needed for estimates that can be considered reliable. It is therefore hard to include information about uncertainty due to the available sample size, because it is not directly clear where the uncertainty comes from.

The presented research in this report shows that the length of the observations generally has a positive correlation with the confidence that we have in the inferred return levels. However, there will always remain some uncertainty, even with a very long length of observations. Moreover, it is case dependent how many observations are needed to obtain reliable results. When using observations, other factors can be more dominant in determining the uncertainty than the length of observations. It is therefore always important to report the uncertainty associated with estimates for inferred return levels and return periods.

6.2 Recommendations

This section includes the recommendations. The recommendations are split up into recommendations for when performing EVA and recommendations for follow up research.

6.2.1 Recommendations when performing EVA

The analyses presented in this report have shown that inferring extreme events comes with a lot of uncertainty. It can be concluded that even for very long lengths of observations, there will still be an uncertainty in the estimates. Moreover, it can not be determined upfront how the sample size will influence the confidence of the estimates. This is an important aspect for the following reason. While the analysis and inference of extreme events is often done by scientists, measures that anticipate these events are made by decision- and policy makers. There is a knowledge gap between scientists and decision makers. While the scientist know the methods and the sources of uncertainty of their analysis, decision makers do not have that information. When nothing is communicated by the scientists with respect to the confidence, policy makers might accept the estimations as a truth. The analyses in this research report have shown that when using EVA, there is not one truth. There are only estimates that go hand in hand with uncertainties. Therefore, it is strongly advised to communicate the confidence in the estimates very well. The proposed method to do this is by calculating and communicating a CI. To include all the information, it is advised to also communicate the used methods and the sample size.

6.3 Recommendations for follow up research

Over the past years, nonstationary analysis have gained a lot of interest. Due to climate change, extreme events are getting more extreme and occur more frequent. This results in a shift in the distributions representing the events over time. Follow up research could include an analysis of the effect of the length of observations on distributions where a trend can be observed in the data.

References

AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., & Sorooshian, S. (2013). *Extremes in a changing climate* (1st ed.). Springer, Dordrecht. doi: https://doi.org/10.1007/978-94-007-4479-0

Alam, M. A., Emura, K., Farnham, C., & Yuan, J. (2018). Best-fit probability distributions and return periods for maximum monthly rainfall in bangladesh. *Climate*, *6*(1). Retrieved from https://www.mdpi.com/2225-1154/6/1/9 doi: 10.3390/cli6010009

Anderson, M., Bengtsson, L., Cruise, J. F., Kothyari, U. C., Serrano, S. E., Stephenson, D., & Strupczewski, W. (2013). Extremes in a changing climate. *Water Science and Technology Library*, *65*.

Benstock, D., & Cegla, F. (2017, 01). Extreme value analysis (eva) of inspection data and its uncertainties. *NDT E International*, *87*. doi: 10.1016/j.ndteint.2017.01.008

Bommier, E. (2014). *Peaks-over-threshold modelling of environmental data* (Unpublished doctoral dissertation).

Cai, Y., & Hames, D. (2010). Minimum sample size determination for generalized extreme value distribution. *Communications in Statistics - Simulation and Computation*, *40*, 87-98. doi: 10.1080/03610918.2010.530368

Caires, S. (2011). *Extreme value analysis: Wave data* (Tech. Rep.). World Meteorological Organization.

Coles, S. (2001). *An introduction to statistical modeling of extreme values.* Springer series in statistics.

Intelligencer. (2021). *Ida death toll rises to 45 in new york area.* Retrieved 22/10/2021, from https://nymag.com/intelligencer/2021/09/ida-aftermath-floods-new-york-city-updates.html

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, *81*, 158-171. doi: https://doi.org/10.1002/qj.49708134804

Kahn, B. (2021). *New york grinds to a halt as record rain floods city, kills at least 9.* Retrieved 22/10/2021, from https://gizmodo.com/record-rainfall-turned-new-york-into-a-lake-1847604964

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, *25*(8), 1287-1304. Retrieved from https://www.sciencedirect.com/science/article/pii/S0309170802000568 doi: https://doi.org/10.1016/S0309-1708(02)00056-8

Makkonen, L., & Tikanmaki, M. (2019). An improved method of extreme value analysis. *Journal of Hydrology X*, *2*, 1–6. doi: {https://doi.org/10.1016/j.hydroa.2018.100012}

Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, *36*(3), 737-744. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/1999WR900330 doi: https://doi.org/10.1029/1999WR900330

MATLAB. (2020). *version 9.8.0.1451342 (r2020a).* Natick, Massachusetts: The MathWorks Inc.

Pianigiani, G. (2021). *Sicily registers record-high temperature as heat wave sweeps italian island.* Retrieved 22/10/2021, from https://www.nytimes.com/2021/08/12/world/europe/sicily-record-high-temperature-119-degrees.html

Pinheiro, M., & Grotjahn, R. (2015). An introduction to extreme value statistic. Retrieved from http://grotjahn.ucdavis.edu/EWEs/extremes_primer_v9_22_15.pdf

Pisarenko, V., & Rodkin, M. (2017). The estimation of probability of extreme events for small samples. *Pure and Applied Geophysics*, *174*, 1547-1560. doi: https://doi.org/10.1007/s00024-017-1495-0

Rijkswaterstaat. (n.d.). *The flood of 1953.* Retrieved 02/11/21, from https://www.rijkswaterstaat.nl/en/water/water-safety/the-flood-of-1953

Rutten, J. G., van Gelder, P. H. A. J. M., Ewans, K. C., & Efthymiou, M. (2008). Uncertainties in extreme value analysis and their effect on load factors. *ASME 2004 23rd International Conference on Offshore Mechanics and Arctic Engineering*, 163-170. doi: https://doi.org/10.1115/OMAE2004-51163

Soukissian, T. H., & Tsalis, C. (2018). Effects of parameter estimation method and sample size in metocean design conditions. *Ocean Engineering*, *169*, 19-37. Retrieved from https://www.sciencedirect.com/science/article/pii/S0029801818317517 doi: https://doi.org/10.1016/j.oceaneng.2018.09.017

STOWA. (2021). *Hoe extreem was neerslag in limburg?* Retrieved 19/10/21, from https://www.stowa.nl/nieuws/hoe-extreem-was-neerslag-limburg

Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., ... A., S. D. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, *596*, 80-86. doi: https://doi.org/10.1038/s41586-021-03695-w

Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F., & Koutsoyiannis, D. (2019). Save hydrological observations! return period estimation without data decimation. *Journal of Hydrology*, *571*, 782-792. Retrieved from https://www.sciencedirect.com/science/article/pii/S002216941930174X doi: https://doi.org/10.1016/j.jhydrol.2019.02.017

A Results sensitivity analysis Part I

In the figure 36 the sensitivity plots are given for the type II GEV distribution.

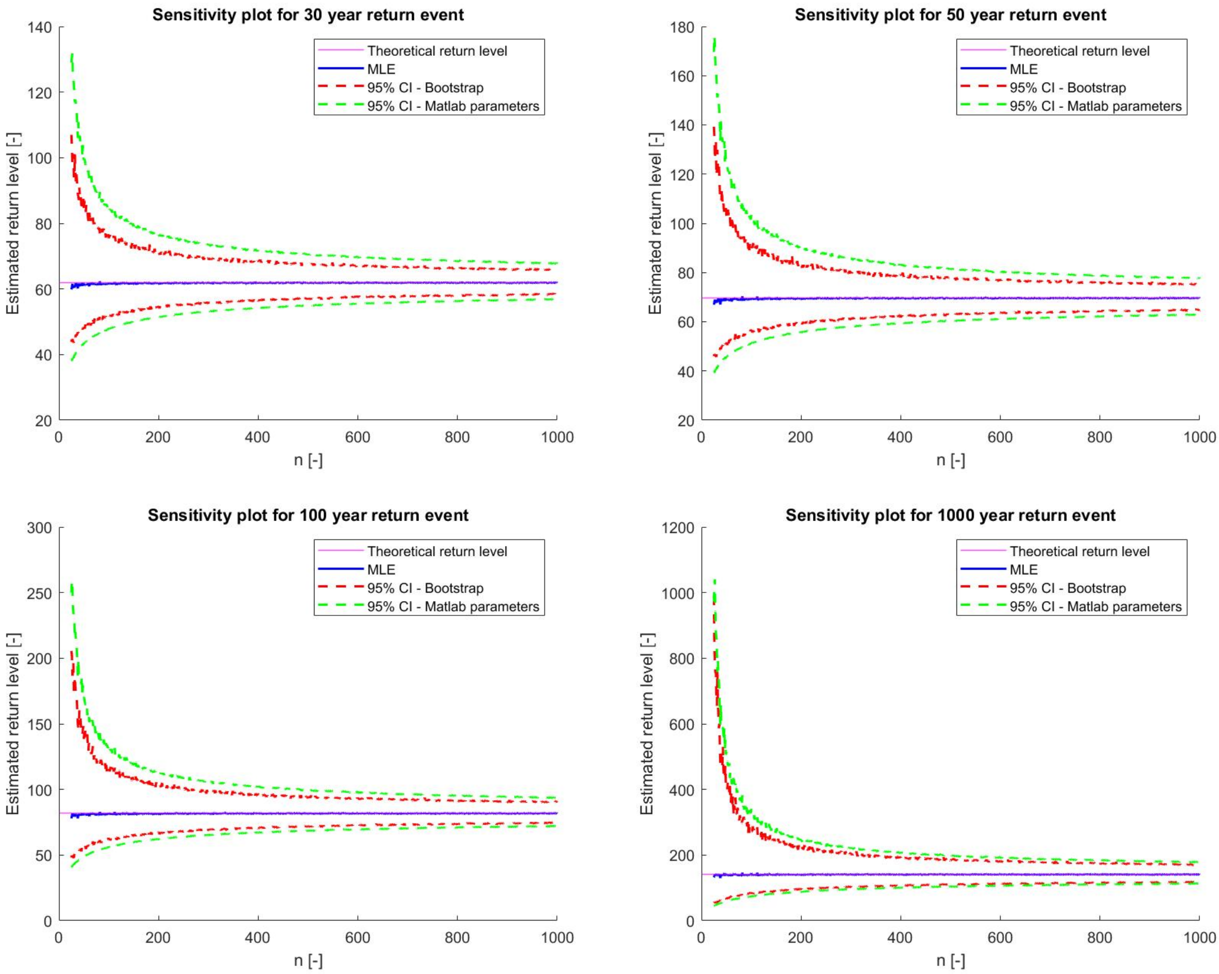


Figure 36: Results of the sensitivity analysis for type II GEV distribution

In figure 37 the results of the sensitivity analysis are given for the type III GEV distribution.

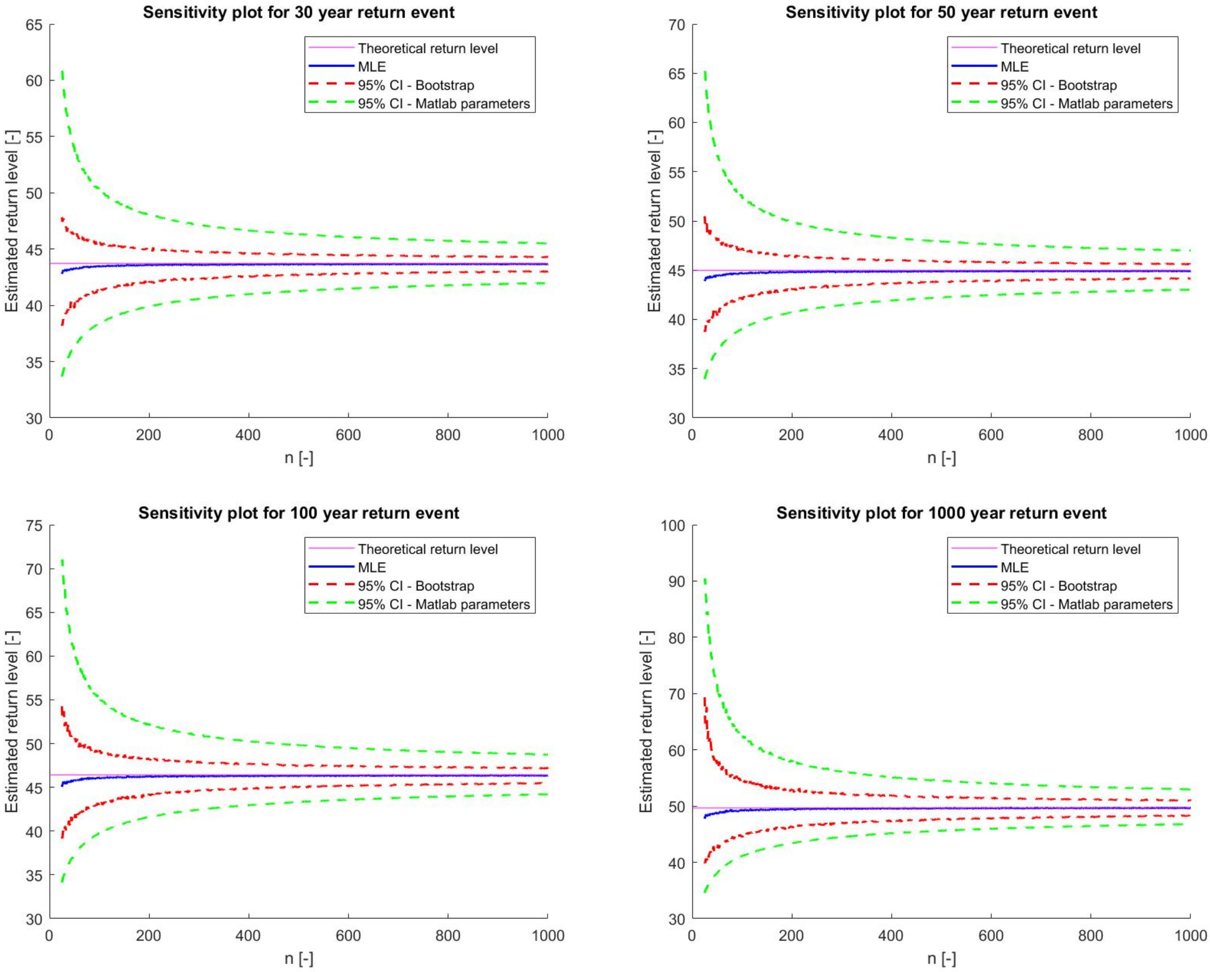


Figure 37: Results of the sensitivity analysis for type III GEV distribution

B Determining Threshold for GPD NY Precipitation

In figure 38, 39 and 40, the Mean residual life plots and the Dispersion Index (DI) plots for subsets with length n=30, 35, ..., 75 and the entire dataset are presented. The plots show that there is a drop in the Dispersion Index after $u = 40$. From the plots it can be concluded that at $u = 50$, the DI is around 1. Also, a linear slope can be assumed between $u = 30$ and $u = 60$ for almost all the subsets. For the subsets with a length of 55 and 60 years, the linearity in $u = 50$ is not very clear. It can also be argued that the graph curves around that point.

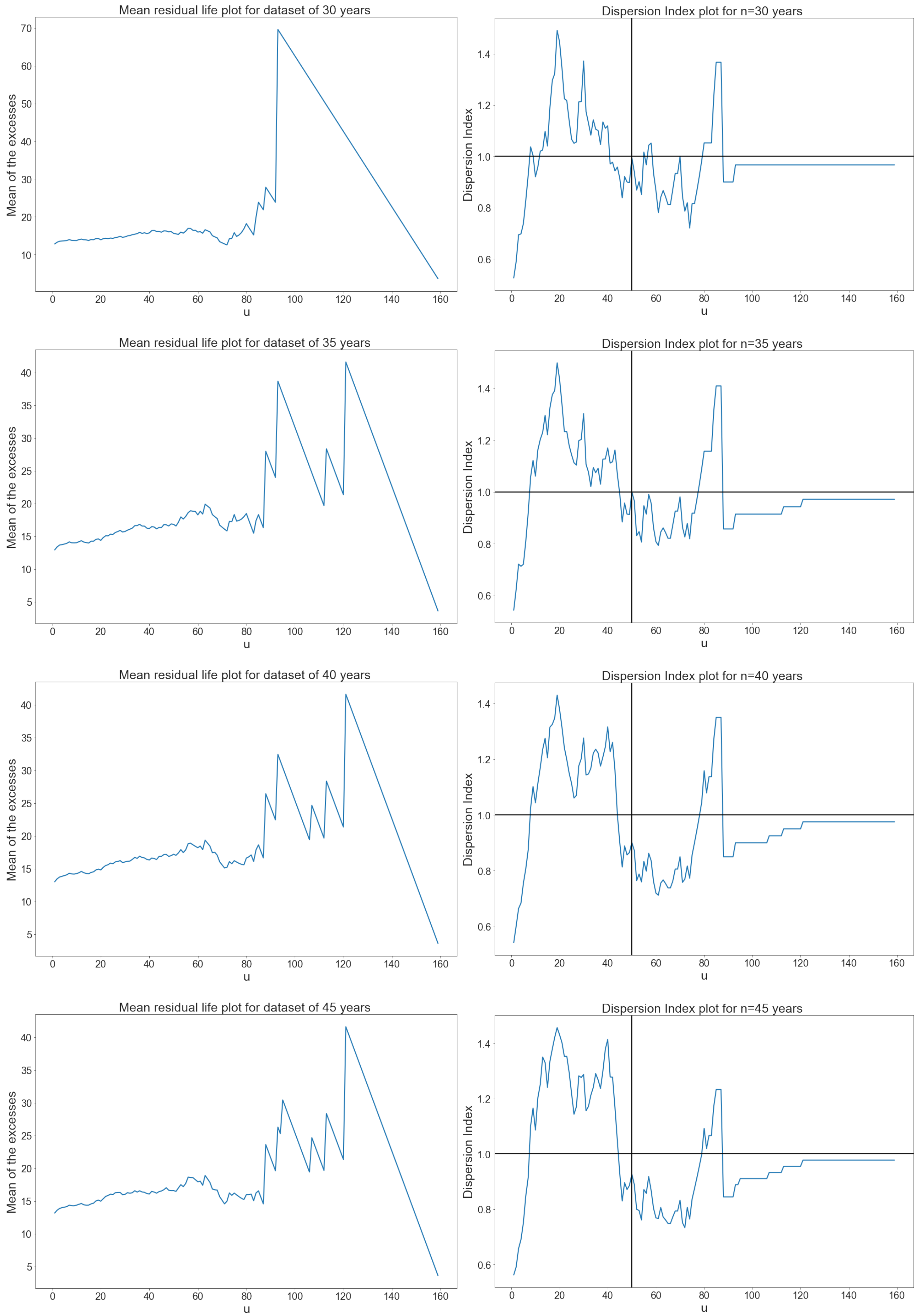


Figure 38: Mean Residual Life Plot and Dispersion Index plot for subsets of n years of the entire dataset

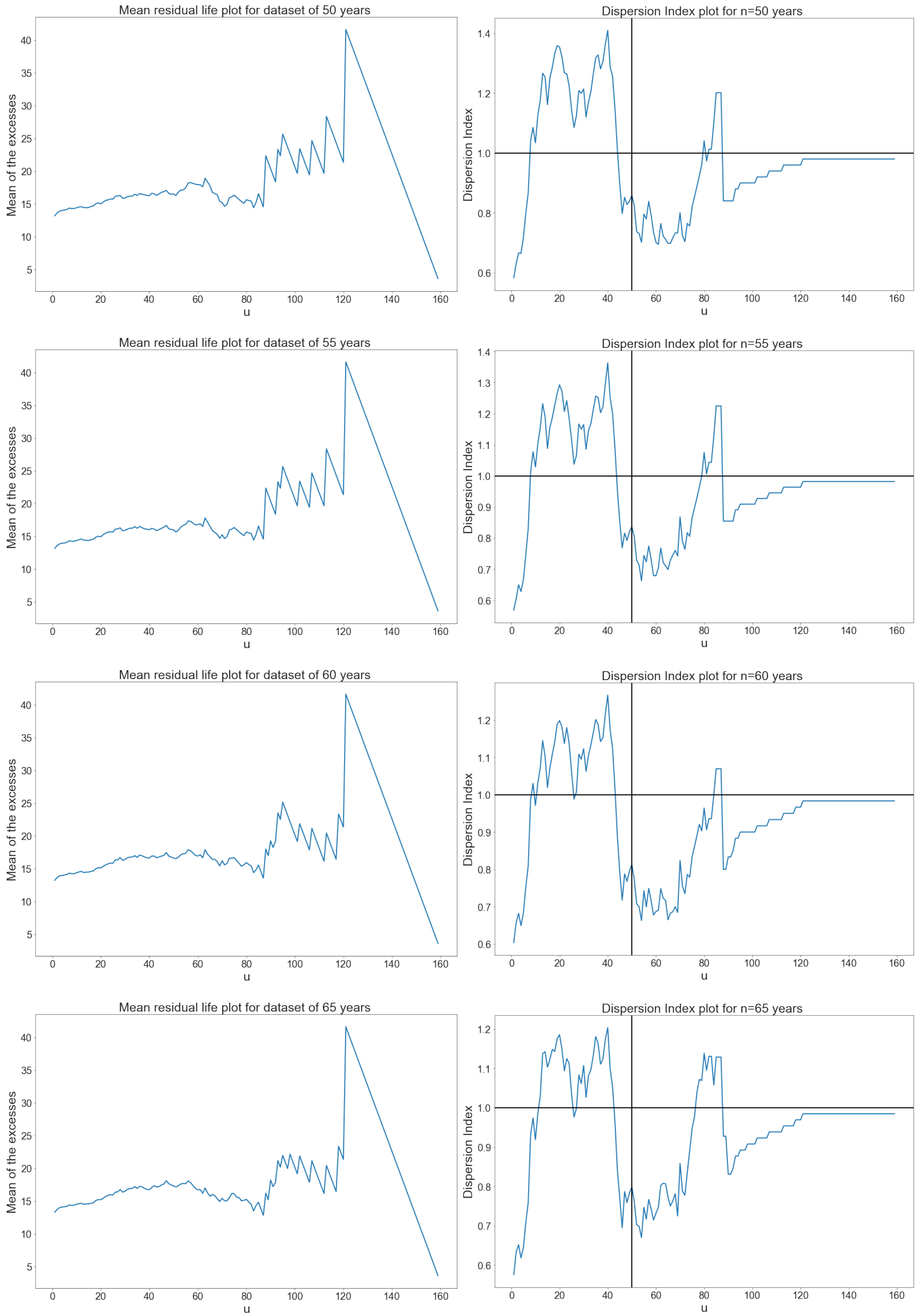


Figure 39: Mean Residual Life Plot and Dispersion Index plot for subsets of n years of the entire dataset

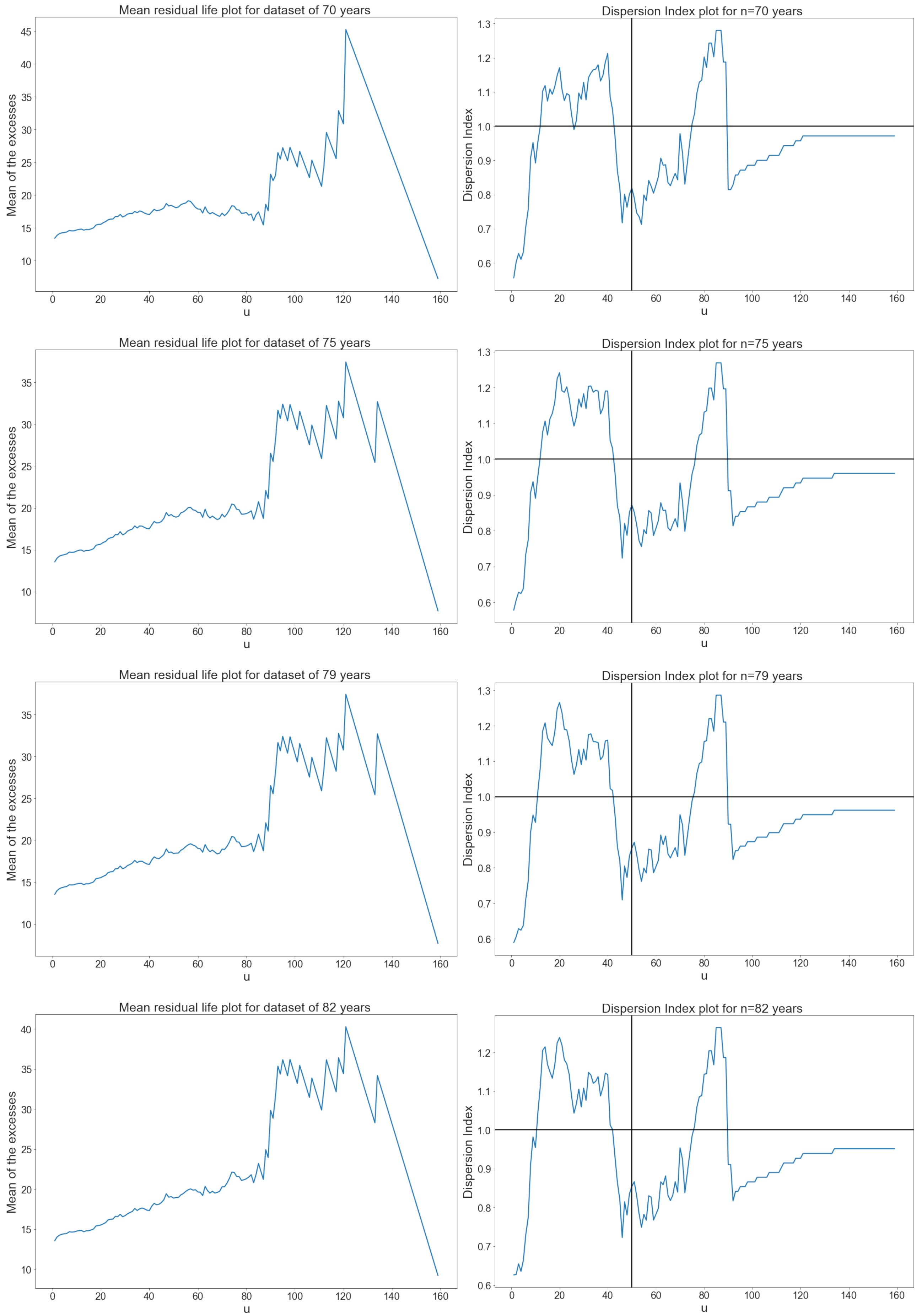


Figure 40: Mean Residual Life Plot and Dispersion Index plot for subsets of n years of the entire dataset