Analysis of trends in extreme rainfall
A regional approach

Martin Roth
The work described in this thesis has been carried out at EURANDOM at the Eindhoven University of Technology and at the Royal Netherlands Meteorological Institute, within the Knowledge for Climate project.

© March 2015, Martin Roth

The pictures on the cover are all taken on July 28, 2014 in different parts of the Netherlands:

UL: Deventer © 2014 Marten Eekhof (DRTV)
UR: Utrecht © 2014 Kees Wenekendonk (www.keeswenekendonk.nl)
LR: Alphen aan den Rijn (www.nos.nl)
LL: Amsterdam (www.nos.nl)

The picture on page 2 is taken on August 2, 2014 in Nijmegen © 2014 Klaas-Jan Grafe.

A catalogue record is available from the Eindhoven University of Technology Library.

Printed by Wöhrmann Print Service, Zutphen, The Netherlands.
Analysis of trends in extreme rainfall
A regional approach

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus, prof.dr.ir. C.J. van Duijn, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op donderdag 12 maart 2015 om 16:00 uur

door

Martin Roth

geboren te Regensburg, Duitsland.
Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

vorzitter: prof.dr. E.H.L. Aarts
1e promotor: prof.dr. J.H. van Zanten (UvA)
2e promotor: prof.dr.ir G. Jongbloed (TUD)
copromotor: dr. T.A. Buishand (KNMI)
leden: prof.dr. E. van den Heuvel
       prof.dr. R.W. van der Hofstad
       prof.dr. H. Rootzén (Chalmers University of Technology)
adviseur: dr. A.M.G. Klein Tank (KNMI)
The research for the present thesis was carried out over the past four years in the nice surroundings of EURANDOM and KNMI. A substantial part of the necessary discussions took also place in Delft. This thesis could never have been completed without the support of many people, to whom I want to express my gratitude.

First of all, I want to say a special thank you to my supervisors Adri Buishand and Geurt Jongbloed. It is owing to your ideas and your eagerness to make things work, despite all challenges, that I was able to finish this thesis. Adri, I want to thank you for letting me profit from your tremendous knowledge of precipitation extremes. You provided a protected working zone for me and you were always available for me, although the past years have been difficult for you. Geurt, your openess is overwhelming and your optimism is really infectious and helped me over some difficult phases. You encouraged me to make my own decisions and stand by them. Together with my further supervisors Harry van Zanten and Albert Klein Tank we had many fruitful discussions. Dear Harry and dear Albert, thank you very much for your insightful remarks and for challenging me to dig deeper. Thank you also for keeping me balanced between being too theoretical and being too applied.

My special thanks go also to Emile Aarts, Edwin van den Heuvel, Remco van der Hofstad, and Holger Rootzén for serving on the doctoral committee.

I enjoyed a lot the time with my great office mates Robert and Alexander at KNMI and Stella and Carlo at EURANDOM. Thank you very much for all the discussions and coffees we had together.

Our lunch and koffie clubje at KNMI with Andrew, Bernadette, Carine, Christiana, Else, Hermine, Gerard, Ine, Jules, Rudmer, and Theo, to name a few, I will miss a lot. I will not forget the many lunches in the park and the nice evenings with
stampot or frietjes and the many let's go’tjes! Thank you for the great time.

At EURANDOM I got to know a lot of inspiring people from all of Europe. To Allesandro, Botond, Carlo, Enrico, Florence, Remco, Robert, Paulo, Peter, Sander, Stella, and Tim thank you very much for making me feel at home in the Netherlands right away from the start and for being such splendid companions!

My parents and my sister have gone with me through all phases of my life. Without your constant love and support in every possible way I would never have come to this point. Thank you so much!

Finally, I want to thank my wife Susanne. You have always encouraged me to do the things that interest me. Your optimism and enthusiasm always kept me on track. I am very curious what future will bring to us. This thesis is dedicated to you and our Rumpeli.
# Contents

1 Introduction 
  1.1 Motivation 
  1.2 Extreme Value Theory 
  1.3 Regional Frequency Analysis 
  1.4 Non-stationarity 
  1.5 Rainfall data 
  1.6 Outline 

2 Regional non-stationary POT model 
  2.1 Introduction 
  2.2 Model description 
    2.2.1 Stationary climate 
    2.2.2 Non-stationary climate 
  2.3 Estimation of the model parameters 
    2.3.1 Threshold 
    2.3.2 Excess distribution 
  2.4 Model selection for the excess distribution 
  2.5 Application to precipitation data 
    2.5.1 Event selection and trend in the threshold 
    2.5.2 Threshold excesses 
  2.6 Conclusion 

3 Projections of precipitation extremes 
  3.1 Introduction

vii
3.2 Methods
3.2.1 Introduction to the POT model
3.2.2 Temporal dependence and declustering
3.2.3 Index Flood approach
3.2.4 Determination of the threshold
3.2.5 Estimating the excess distribution
3.2.6 Bias correction
3.3 Data
3.3.1 Region and precipitation data
3.3.2 Temperature as covariate
3.4 Results and discussion
3.4.1 Temporal dependence and declustering
3.4.2 Trend in the threshold
3.4.3 Trend in the excess distribution
3.4.4 Return levels
3.5 Conclusion

4 Regional threshold selection
4.1 Introduction
4.2 Preliminaries
4.3 Univariate threshold selection approaches
4.3.1 Approaches based on visual inspection
4.3.2 Approaches based on GoF tests
4.4 Threshold selection in the regional POT model
4.4.1 Approaches based on visual inspection
4.4.2 Approaches based on GoF tests
4.5 Simulation study
4.5.1 Marginal model
4.5.2 Spatially independent data
4.5.3 Spatially dependent data
4.6 Application to rainfall data
4.7 Conclusion

5 Trends in moderate rainfall extremes
5.1 Introduction
5.2 Data
5.3 Methods
5.3.1 Mann–Kendall test
5.3.2 Monotone Quantile Regression
5.4 Results
Chapter 1

Introduction

However big floods get, there will always be a bigger one coming; so says one theory of extremes, and experience suggests it is true (Gumbel 1958, quoting President’s Water Commission, 1950).

1.1 Motivation

Extreme rainfall events pose a constant threat to society. Large scale floodings, after extended periods of heavy precipitation, cause disruption and loss of property and can lead to loss of life. Also sub-daily rainfalls can cause severe damages if they exceed the capacity of drainage systems. Fig. 1.1 shows a street flooding event in Nijmegen on August 2, 2014 and the cover of the present thesis shows pictures of the flooding event of July 28, 2014.

The damages caused by two heavy rainfall events in the Netherlands in September and October 1998 alone amounted to roughly 500 million euros (Smits et al. 2004). These events revived the discussion about which protection measures to take facing such hazards. For instance, individuals are advised not to store high value items in the basement of their houses, if possible. Society as a whole has to agree upon guidelines, e.g. design norms, for the dimensioning of both water drainage systems and flood protection structures. For this purpose often design rainfalls are developed, that have to be tolerated by the infrastructure. These design rainfalls are usually based on a total rainfall amount and a temporal profile of the event (e.g. Svensson and Jones 2010). Considering too low design rainfalls can lead to damage as a con-
CHAPTER 1. INTRODUCTION

Figure 1.1: Street flooding in Nijmegen caused by a heavy rain storm on August 2, 2014 (while writing this introduction in 500 m distance).

sequence of failing infrastructure. On the other hand, too high design levels result in increased building costs (e.g. Papalexiou and Koutsoyiannis 2013). In order to achieve a balance between these two, accurate information about the occurrence of rainfall extremes is necessary.

The definition of rainfall extremes always depends on the regional climate (Hartmann et al. 2013, Box 2.4), for example, a rainfall event considered extreme in the Netherlands might be quite common for the monsoon season in India. A widely applied concept in the literature of extremes is given by the return period, i.e. the expected waiting time between two consecutive events exceeding a high level. The corresponding level for an $n$-year return period is termed the $n$-year return level and the annual risk equals $1/n$. Prominent examples of the usage of this concept in the Netherlands are the designs of river and sea dikes, which are based on the 1250- and 10000-year return period, respectively. More examples can be found in the Dutch Nationaal Bestuursakkoord Water (http://www.helpdeskwater.nl/onderwerpen/wetgeving-beleid/nationaal/01280/nationaal/).

The concept of return periods relies on the stationarity assumption (e.g. Rootzén and Katz 2013). The stationarity assumption implies that a variable has a time-invariant (or 1-year-periodic) distribution that can be estimated from measurements (e.g. Milly et al. 2008). In this setting the notions $n$-year return period and $1/n$-annual risk are exchangeable. However, it occurs that people misinterpret the meaning of
the term *return period* and are surprised that the chance of the 100-year return level being exceeded at least once (twice/thrice) in a century is 63% (26%, 7.5%). These calculations can be done using a simple binomial model. Moreover, there is evidence that the frequency of extreme rainfall events might increase as a consequence of climate change. Then, the concept of return periods does not apply anymore (Rootzén and Katz 2013). Klein Tank and Lenderink (2009) reported a 10% increase in the precipitation thresholds, which are exceeded on average once per year over the Netherlands since 1950. Moreover, the number of days with at least 10 mm precipitation in the winter season and 20 mm precipitation in the summer season have increased in most parts of the Netherlands (Sluijter 2011; Van den Hurk et al. 2014).

This thesis focuses on the trends in extreme rainfall over the Netherlands. Accurate estimation of trends in extreme rainfall is necessary to evaluate possible implications of climate change. For instance, the failure risk of hydrological infrastructure might substantially increase, owing to intensifying rainfall extremes.

In the remainder of this introduction some basic concepts of Extreme Value Theory (EVT) and Regional Frequency Analysis (RFA) are presented. These concepts are, under the stationarity assumption, widely applied in hydrology and climatology, and represent the basis of many existing design norms. After that, a short overview of existing approaches to include non-stationarity in the analysis of extremes is given, followed by a few introductory comments on the used rainfall data. Then, the outline of the thesis is given.

### 1.2 Extreme Value Theory

**Why EVT?** The idea of EVT is to provide statistical models for events of rarely observed or unprecedented magnitude. For the analysis of extremes of a given data record, an appropriate subset of the data – representing the extremes – should be chosen. There are some applications where a distribution is fitted to all data; an example is the use of the Weibull distribution to estimate extreme wind speeds (Conradsen et al. 1984). Perrin et al. (2006) criticized this method and argued that the Block Maxima (BM) approach should be used instead. In that approach only the maxima of equally-sized disjoint blocks of the data are used (e.g. annual maxima). Indeed, it seems widely accepted now that the bulk of the data is generally not representative of the extremes and should hence be dismissed in the analysis of extremes, see e.g. Reiss and Thomas (2007). A different approach to the selection of extremes is given by the Peaks-Over-Threshold (POT) approach, where only the exceedances of a high threshold are considered. In this section we introduce the two basic theorems underlying the BM and the POT approach.
CHAPTER 1. INTRODUCTION

History of EVT  The development of EVT began relatively late compared to that of classical statistics. Gumbel (1958) attributed this partly to the centrality of the Gaussian distribution in statistics, as the fundamental theorems of EVT are not related to it in a simple way. While some considerations on extremes are dated back to Nicolaus Bernoulli in 1709, the development of EVT started with L. von Bortkiewicz. In 1898 he published his book The Law of Small Numbers, introducing the Poisson distribution for modeling rare events. A famous example from this book is the good fit of the Poisson distribution to the number of Prussian soldiers killed each year by horse-kick. Bortkiewicz (1922) is the first systematic approach to study the range of random samples drawn from the normal distribution, introducing the concept of the distribution of the largest value (Kotz and Nadarajah 2000). With this concept, the fundamental question of EVT can be stated in a manner resembling the central limit problem (Embrechts et al. 1997):

What are the possible (non-degenerate) limit distributions of the largest value when properly centered and normalized?

After 1922 EVT became an active research field. Dodd (1923) began studying extremes from distributions different from the Gaussian, whereas Tippett (1925) carried out an extensive study of the extremes from the normal distribution and tabulated the mean of the largest value for various sample sizes up to 1000. The first limiting distribution of the largest value was provided by Fréchet (1927) and already in the next year Fisher and Tippett (1928) showed that there exist only three possible limiting distributions.

Theorem 1 (Three types theorem) Let \((X_n)\) be a sequence of independent and identically distributed (i.i.d.) random variables. Define by \(M_n := \max(X_1, \ldots, X_n)\). If there exist norming constants \(b_n > 0, a_n \in \mathbb{R}\) and some non-degenerate distribution function \(H\) such that

\[
\frac{M_n - a_n}{b_n} \xrightarrow{d} H,
\]

then \(H\) belongs to one of the following three Extreme Value (EV) distribution functions:

Type I (Gumbel): \(\Lambda(x) = \exp(-e^{-x}), x \in \mathbb{R}\).

Type II (Fréchet): \(\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0 \end{cases}, \alpha > 0\).

Type III (Weibull): \(\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^{-\alpha}), & x \leq 0 \\ 0, & x > 0 \end{cases}, \alpha < 0\).

A sketch of the proof of this theorem can be found, for instance, in Embrechts et al. (1997).
1.2. Extreme Value Theory

The Generalized Extreme Value (GEV) distribution was first obtained by R. von Mises in 1936, combining the three EV distributions (compare Mises 1954). However, it is often attributed to Jenkinson (1955) instead. The cumulative distribution function (CDF) of the GEV distribution, with shape parameter $\xi$, scale parameter $\sigma$, and location parameter $\mu$, is given by

$$H_{\xi, \sigma, \mu}(x) = \begin{cases} \exp \left\{ - \left[ 1 + \frac{1}{\xi} \left( \frac{x-\mu}{\sigma} \right) \right]^{-1/\xi} \right\}, & \xi \neq 0, \\ \exp \left\{ - \exp \left( - \frac{x-\mu}{\sigma} \right) \right\}, & \xi = 0, \end{cases}$$

where $x$ must satisfy $1 + \xi (x - \mu) / \sigma > 0$. For $\xi > 0$, $\xi < 0$, and $\xi = 0$ we have the Fréchet, Reversed Weibull, and Gumbel distribution, respectively. Moreover, R. von Mises identified some simple, sufficient conditions to determine if the parent distribution $F$ lies in the maximum domain of attraction of the EV distribution $H_{\xi}$, i.e. that there exist norming constants $b_n > 0$ and $a_n \in \mathbb{R}$ such that Eq. (1.1) holds. This is indicated as $F \in \text{MDA}(H_{\xi})$. Gnedenko (1943) fully characterized the maximum domain of attraction for the three EV distributions through necessary and sufficient conditions on the parent distributions.

The following example provides an insight of the general theory. Suppose a sequence $(X_n)$ of independent standard exponential random variables, i.e. $X_i \sim F$ with

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-x}, & x > 0. \end{cases} \quad (1.2)$$

Using $a_n = \ln(n)$ and $b_n = 1$ we obtain

$$P \left( \frac{M_n - a_n}{b_n} \leq x \right) = F^n(x + \ln(n)) = \left( 1 - e^{-x} \right)^n \xrightarrow{n \to \infty} e^{-e^{-x}}, \quad (1.3)$$

which shows that the maximum of standard exponential random variables, properly centered, converges to the Gumbel distribution. The (centered) distribution of $M_n$ for $n = 5, 20, 50$ and the limiting Gumbel distribution are shown in Fig. 1.2. One can see the good approximation by the Gumbel distribution, even for small $n$.

The name patron of the Gumbel distribution, Emil Julius Gumbel, was born into a wealthy family in Munich on July 18, 1891 and graduated in 1913 in mathematics and actuarial science. A year later he was awarded a doctorate in economics with highest honors (Brenner 1990). After World War One, which he joined voluntarily in the beginning, he became one of the leading pacifists and anti-Nazi activists in Germany. Though, he faced increasing hostility from conservative colleagues he habilitated in Heidelberg in 1924. Gumbel documented the numerous political murders and politicized justice in the Weimar Republic (Brenner 1990). In 1932 he lost
his professorship in Heidelberg, due to intense attacks led by the Nazi student organiza-
tions. After emigrating first to France and then to the USA, he became the most
he advocates to fit a EV distribution to the BM of the data. This approach is nowadays
widely applied.

**Transition from BM to POT**  De Haan (1970) extended ideas by J. Karamata on
regular variation and obtained a unifying approach for the determination of the
maximum domain of attraction for all limit distributions, simplifying the work by
Gnedenko (1943). Using the regular variation results, Balkema and De Haan (1974)
derived the second fundamental theorem of EVT, proving an asymptotic result for
the POT approach. Pickands (1975) derived an equivalent result using upper order
statistics.

**Theorem 2 (Pickands–Balkema–De Haan)**  Let \((X_n)\) be a sequence of i.i.d. random
variables (satisfying weak regularity conditions). Then,

\[
P(X - u \leq y \mid X > u) \xrightarrow[n \to x^*]{} G_{\xi,\sigma}(y),
\]

where \(x^*\) is the right endpoint of the parent distribution and \(G_{\xi,\sigma}(y)\) is the Generalized
Pareto Distribution (GPD) with cdf

\[
G_{\xi,\sigma}(y) = \begin{cases} 
1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}, & \xi \neq 0, \\
1 - \exp\left(-\frac{y}{\sigma}\right), & \xi = 0,
\end{cases}
\]
where $\xi$ is the shape parameter and $\sigma$ the scale parameter. The shape parameter $\xi$ is the same as for the GEV distribution.

A proof of the theorem can be found, for instance, in De Haan and Ferreira (2006).

The motivation for the POT approach is twofold. On the one hand, Pickands (1975) noted that the consideration of fixed block sizes – and hence sometimes small maximum values – introduces a small bias in the BM approach, which can be reduced by increasing thresholds in the POT approach. On the other hand, the use of multiple extreme values from one block might lead to a decrease of the statistical uncertainty.

A difference between the BM and the POT approach is that for the latter also the occurrences of the peaks have to be modeled. The most commonly used model for the occurrence process is the Poisson process, with intensity proportional to the expected number of threshold exceedances per year. This practice was introduced in hydrology by Shane and Lynn (1964) and theoretically justified by Leadbetter (1991).

The selection of the threshold has to be made carefully. One criterion for the threshold, which follows from the Poisson process modeling, is that the number of threshold exceedances per year should be approximately Poisson distributed, closing the circle to the work of L. von Bortkiewicz.

The Battle of EV distributions The limit behavior of the three EV distributions varies substantially. Still, practitioners in the second half of the 20th century used mostly the Gumbel distribution. Koutsoyiannis (2004a) presented several reasons for this persistent use of the Gumbel distribution, e.g. the accuracy in the estimated parameters is higher if the shape parameter is fixed and many parent distributions used in hydrology (and climatology) belong to the maximum domain of attraction of the Gumbel distribution. However, because of the potential severe underestimation of the return levels, Koutsoyiannis argued that the Fréchet distribution should be used instead for the estimation of (sub-) daily rainfall (compare also Buishand 1991). He supported his claim by a field study of daily rainfall over different parts of the world (Koutsoyiannis 2004b).

Papalexiou and Koutsoyiannis (2013) updated this work on the battle of extreme value distributions using a much larger data basis. Moreover, they showed that the distribution of the maxima from a generalized Gamma distribution that describes the distribution of wet-day rainfall well seemingly does not converge to the Gumbel distribution. Although the generalized Gamma distribution is in the domain of attraction of a Gumbel distribution, the distribution of the maxima turned out to be better approximated by the Fréchet distribution. This is attributed to the extremely slow rate of convergence to the Gumbel distribution.

Serinaldi and Kilsby (2014) came to similar conclusions using the POT approach
with increasing thresholds, but also show that for decreasing thresholds the Weibull distribution is more adequate. Considering fixed high thresholds and increasing sample size, heavy-tailed distributions prevail again. A difficulty with heavy-tailed distributions is that estimates of high quantiles are unstable. This highlights the need for accurate parameter estimation. A method to reduce the estimation uncertainty is presented in the following section.

1.3 Regional Frequency Analysis

There are currently about 325 stations in the Netherlands, where rainfall is measured one time a day (Fig. 1.3 shows the evolution of the number of the stations since 1850). Being a relatively small country with little orographic differences it seems plausible that the rainfall series from these stations resemble each other.

RFA provides methods to exploit similarities between multiple data series. Cunnane (1988) and Buishand (1991) reviewed several of these methods. The oldest and simplest is the station-year method. For this method, all rainfall series are combined into one long record. However, in the presence of positive dependence between the different series, the estimates of large quantiles of the EV distribution can suffer substantial bias (Buishand 1984a). Moreover, it cannot be expected that the station-year data come from the same distribution. Dalrymple (1960) introduced an important modification of the station-year method, which tackles the latter issue by standardiz-

Figure 1.3: Temporal evolution of the number of rain gauges from 1850 to 2014 (after Buishand et al. (2013)).
1.3. Regional Frequency Analysis

ing the data with a site-specific index variable, usually the mean annual maximum. From the standardized data a common quantile function, also known as regional growth curve, is obtained for the whole region. The individual $R$-year return level is then the product of the index-variable and the $R$-year return level from the regional growth curve.

The UK Flood Studies Report (Natural Environment Research Council 1975) has a strong regional component, adopting an index-variable approach. The measurement stations were grouped according to their 5-year return level, which served also as the site-specific index variable. Common growth curves were derived for each group. In this approach regions do not necessarily have to be spatial contiguous. However, most studies cluster stations according to geographical or administrative regions (Stewart et al. 1999).

An alternative to the station-year approach is the averaging of at-site estimates. Spatial dependence hardly influences the bias of the parameter estimates, but might have substantial effect on their variance (Stedinger 1983; Hosking and Wallis 1988). Because at-site estimates of the shape parameter are based on relatively short records, estimators with small bias for small sample sizes, such as probability weighted moments (PWM) and L-Moments estimators, are needed. These have also relatively small variance for EV distributions commonly encountered in the analysis of daily rainfall. Hosking and Wallis (1997) provided an excellent reference of RFA based on L-Moments, establishing it as quasi standard.

Another approach to RFA is Maximum Likelihood (ML) estimation under the independence assumption (Buishand 1991). Through extensions of the information matrix the effect of spatial dependence on the uncertainty can be taken into account (Smith 1990b; Varin et al. 2011). With the increased need to take non-stationarity into account, this approach becomes especially appealing, owing to the flexible likelihood concept.

While most of the regional approaches are based on BM data, Madsen et al. (1997b) and Madsen and Rosbjerg (1997) presented a combination of the index-variable approach with POT data. They concluded that the POT approach is superior to the BM approach with respect to the accuracy of the $R$-year return level, in particular for positive shape parameters $\xi$. However, there are also some problems in the regional POT approach, such as the extraction of the extreme events and the quantification of the spatial dependence (Stewart et al. 1999). A recent overview of countries, which use RFA methods in the determination of design rainfalls is given in Svensson and Jones (2010).
CHAPTER 1. INTRODUCTION

1.4 Non-stationarity

The consensus in the climatological community now is that the characteristics of extreme precipitation may change in the future (IPCC 2013). Wrongly assuming stationarity might lead to underestimation with potentially severe consequences (Wigley 2009).

Different descriptive indices of climate extremes have been developed for the monitoring of climate change, compare e.g. Klein Tank et al. (2009). Several publications made use of these indices for the study of changes in extreme precipitation, among others Klein Tank and Können (2003), Moberg and Jones (2005), Alexander et al. (2006), and Turco and Llasat (2011). However, these climate indices are not suitable to determine trends for events with small annual risk (Klein Tank et al. 2009). Therefore, EVT has to be used.

The simplest and still widely used approach is to consider two distinctive periods and then compute the quantity of interest for both periods, assuming that the data in the considered period itself is stationary, compare e.g. Frei et al. (2006) for BM data and Kyselý and Beranová (2009) for POT data. In that case also the classical RFA methods can be applied (e.g. Fowler et al. 2005; Ekström et al. 2005). A different approach to non-stationary EVT is based on the inclusion of time or other covariates representative of the climate change signal. This approach also allows the comparison of different covariates for the inclusion of non-stationarity, e.g. by the AIC criterion. In the BM approach the inclusion is achieved by modeling the parameters of the GEV dependent on the considered covariate, e.g. a location parameter with linear trend in time:

\[ \mu_t = \mu_0 + \mu_1 \cdot t. \]

A few examples of this approach are Coles (2001), Kharin and Zwiers (2005), El Adlouni et al. (2007), and Tramblay et al. (2012).

The POT approach offers two possibilities, either to consider a fixed threshold and a non-homogeneous Poisson process for the occurrence model (e.g. Chavez-Demoulin and Davison 2005) or a time-varying threshold (e.g. Brown et al. 2008; Kyselý et al. 2010), which can be determined, for instance, via quantile regression. The latter approach has the advantage that peaks are collected over the whole record (compare also Tramblay et al. 2013). As for the BM approach, the scale and shape parameters of the GPD can then be modeled depending on covariates. In both approaches the shape parameter is often considered to be constant (Kyselý 2010).

Hanel et al. (2009) combined a non-stationary BM approach with RFA methods, using a new index-variable, namely the at-site location parameter. This study served as basis for the development of the regional, non-stationary POT approach presented in Chapter 2.
Most of the studies mentioned above consider linear or log-linear models for the dependence on the covariate. Sometimes this approach appears not flexible enough. Therefore, Hall and Tajvidi (2000) and Davison and Ramesh (2000) used local likelihood estimation and Chavez-Demoulin and Davison (2005) used spline functions to model the non-linear dependence on the covariate.

1.5 Rainfall data

The most common source of rainfall data are rain gauges, so the analysis of extreme rainfall is traditionally focused on gauge data. Although climate models do not yet simulate precipitation extremes very reliably, they provide useful information on the potential future changes in extremes, see Klein Tank et al. (2009). Thus, the analysis of precipitation extremes in climate model data and the comparison of these extremes with observed precipitation extremes has become more and more important in the last two decades. Gridded data sets are favorable for this comparison, in particular when the observations and the climate model simulation data have the same grid (Haylock et al. 2008). Gridding changes the characteristics of extreme precipitation, because the most extreme values are averaged with less extreme ones. The change in characteristics depends on the resolution of the grid. Moreover, the effect is stronger for regions with sparse station density and in the summer season, where extremes are often of local, convective nature.

In the future radar and satellite data may provide an alternative, see Overeem et al. (2008). They provide good spatial coverage, but the available records are too short to enable the analysis of trends. Moreover, there are substantial biases in the raw remote sensing data. Therefore, high quality rain gauge data are needed for the calibration.

1.6 Outline

This thesis is organized as follows. In Chapter 2 a regional POT model, capable of dealing with non-stationarity through the inclusion of covariates, is presented. Maximum independence likelihood, a special case of composite likelihood, is used for the estimation of the model parameters. Model selection via adapted versions of the Akaike information criterion and likelihood ratio tests that account for spatial dependence are discussed. The model is applied to gridded, daily rainfall over the Netherlands for the winter seasons 1950–2010.

The thesis proceeds in Chapter 3 with an application of this approach to climate model data from the KNMI Regional Climate Model RACMO2, driven by two different General Circulation Models. A novel graphical approach to threshold selection
taking advantage of the regional setting is introduced. Moreover, a new approach for bias correction of the climate model projections of rainfall extremes is discussed. It is shown how the regional, non-stationary POT approach reduces the uncertainty compared to at-site estimation or standard regional approaches.

In Chapter 4 the regional setting is further exploited for the purpose of the threshold selection. A simulation study is carried out to show the benefits of the proposed approach. For this simulation study we introduce a smooth, differentiable hybrid Weibull–GPD distribution. The approach is also applied to station rainfall data from the Vallei en Veluwe water board.

Chapter 5 focuses on the trend in rainfall observations over the Netherlands for the period 1910–2009. For such long periods it is necessary to allow more flexibility than a linear change over time in the trend estimation. Therefore, the trend in high quantiles is estimated under the monotonicity constraint, which allows more forms of increasing extremes than only linear ones. The approach is also adapted for multiple sites, and can be used to estimate the trend in the threshold for the model presented in Chapter 2.

Chapter 6 outlines an approach to estimate the scale parameter of the GPD under the monotonicity constraint.
Regional non-stationary POT model

Abstract. Regional frequency analysis is often used to reduce the uncertainty in the estimation of distribution parameters and quantiles. In this paper a regional Peaks-Over-Threshold (POT) model is introduced that can be used to analyze precipitation extremes in a changing climate. We use a temporally varying threshold, which is determined by quantile regression for each site separately. The marginal distributions of the excesses are described by Generalized Pareto Distributions (GPDs). The parameters of these distributions may vary over time and their spatial variation is modeled by the Index Flood (IF) approach. We consider different models for the temporal dependence of the GPD parameters. Parameter estimation is based on the framework of composite likelihood. Composite likelihood ratio tests that account for spatial dependence are used to test the significance of temporal trends in the model parameters and to test the IF model. We apply the method to gridded, observed daily precipitation data from the Netherlands for the winter season. A general increase of the threshold is observed, especially along the west coast and northern parts of the country. Moreover, there is no indication that the ratio between the GPD scale parameter and the threshold has changed over time, which implies that the scale parameter increases by the same percentage as the threshold. These positive trends lead to an increase of rare extremes of on average 22% over the country during the observed period.

CHAPTER 2. REGIONAL NON-STATIONARY POT MODEL

2.1 Introduction

Design values for infrastructure are often based on characteristics of extreme precipitation. These characteristics may have changed over time owing to climate change, see e.g. Klein Tank and Können (2003) and Milly et al. (2008), which contradicts the stationarity assumption, that is usually made in hydrologic and hydraulic design. Wrongly assuming stationarity generally leads to systematic errors in design values and might have a considerable impact on the risk of failure of hydraulic structures, as shown by Wigley (2009). Climate scientists have analyzed trends in moderate extremes, that occur once or several times per year, based on annual indices. Examples are the empirical annual 90% quantile of the precipitation amounts on wet days or the 1-day or 5-day maximum precipitation amount in each year, see e.g. Klein Tank and Können (2003) and Turco and Llasat (2011).

In this chapter we focus on rare extremes which occur less frequently than once per year. These are often assessed by Extreme Value (EV) models, that are fitted to Block Maxima (BM), e.g. the largest value in a year or season. Considering only BM discards useful data in the case of multiple (independent) extremes in a block, see e.g. Madsen et al. (1997a), Lang et al. (1999) and Kyselý et al. (2010). We follow a common alternative method to analyze extremes by considering all values that exceed a certain high threshold, which is known as POT modeling, see e.g. Coles (2001). A potential advantage of POT modeling is the possibility to include more data in the analysis than in the BM approach, which may reduce the estimation variance. A reduction of the estimation variance may also be achieved by analyzing the $r$-largest maxima in a block, see Coles (2001) and Zhang et al. (2004). However, this method has the potential drawback of using values that may not represent extremes, e.g. in a dry year.

To account for the temporal trend in the distribution, one or several parameters of the EV model are customarily selected to depend on a temporal covariate. One of the first references regarding this implementation of non-stationarity is Smith (1986), who considers the $r$-largest values of the Venice sea level for each year, which exhibit a clear trend. From then on many non-stationary models have been used, see e.g. Kharin and Zwiers (2005) and El Adlouni et al. (2007) for BM data and Brown et al. (2008), Kyselý et al. (2010), and Beguería et al. (2010) for POT data.

Because of the rarity of the extremes, the parameters in the EV models and in particular large quantiles of the precipitation amounts have wide confidence intervals. To reduce the uncertainty in the estimates the use of data sets over a long period and/or Regional Frequency Analysis (RFA), have been recommended (e.g. Hosking and Wallis 1997). Long time series are rare and sometimes not available for a certain region. However, often relatively short records are available for many sites in the region. The idea behind RFA is to exploit the similarities between the sites in that
region, so that all data in the region can be used to obtain quantile estimates for a particular site. RFA approaches to estimate properties of extremes in a stationary climate have been used quite often with BM data, see for an overview Cunnane (1988) and Hosking and Wallis (1997) and more recently Svensson and Jones (2010), but rarely with POT data (Madsen et al. 1997b). For non-stationary extremes only very few studies considered an RFA approach, among them Westra and Sisson (2011), who use a max-stable process model for BM, and Hanel et al. (2009) who apply an IF approach also to BM data. The IF approach is a popular method in RFA. It assumes that the distributions of the extreme precipitation amounts are identical after scaling with a site-specific factor (the index variable). We mimic the approach by Hanel et al. (2009) and develop a POT model, with time-varying parameters, that satisfies the IF assumption. The threshold varies linearly over time and is determined for each site separately. The distribution of the excesses of the threshold is modeled by the GPD. Its parameters may vary linearly over time and their spatial variation is determined by the IF assumption. We apply this model in a case study to gridded observed daily precipitation data for the winter period over the Netherlands.

In section 2.2 we describe the proposed model. We explain the basic methods used to deal with high quantile estimation in the case of stationary data with emphasis on the POT approach. After that, we present our model for the non-stationary climate. In section 2.3 we outline the estimation procedure. The choice between different models is addressed in section 2.4 and in section 2.5 the application of the model to observed daily precipitation data in the Netherlands is discussed.

2.2 Model description

The data we describe with our model consist of measurements at $S$ sites over a period of $T$ time points. The data can be represented in an $S \times T$ space-time matrix

$$x := (x_s(t))_{s \in S, t \in T},$$

where $x_s(t)$ is the value at site $s$ and time $t$, $S := \{1, ..., S\}$ and $T := \{1, ..., T\}$.

In POT modeling exceedances over a high threshold $u_s(t)$ are considered, $s \in S$, $t \in T$. This threshold is generally site-specific and may depend on time. In the case of temporal clustering of the exceedances the largest value in a cluster (peak) is considered only. These peaks will then generally be approximately independent. We assume that the $x_s(t)$ have been declustered and we define $y_s(t)$ as the difference between the daily value at site $s$ and time $t$ and the corresponding value of the threshold, i.e.

$$y_s(t) := x_s(t) - u_s(t),$$
and \( y \) is defined analogously to \( x \). The excesses are the non-negative part of \( y \). Note, that owing to the declustering \( y_s(t) \) is only non-negative if there is a peak. By \( \tilde{T} \) we denote the subset of days where at least one threshold excess occurs, i.e.

\[
\tilde{T} := \{ t \in T \mid \text{there exists an } s \in S \text{ with } y_s(t) \geq 0 \}.
\]

### 2.2.1 Stationary climate

#### Site specific approach

The BM approach for a stationary climate relies on the three types theorem for maxima of independent and identically distributed (i.i.d.) random variables. This theorem allows, under certain regularity conditions, to approximate the distribution of the BM by an EV distribution, see e.g. Embrechts et al. 1997. The three types of EV distributions can be summarized in the Generalized Extreme Value (GEV) family with cumulative distribution function (cdf)

\[
H_{\xi^*, \sigma^*, \mu^*}(x) = \begin{cases} 
\exp \left\{ - \left[ 1 + \xi^* \left( \frac{x - \mu^*}{\sigma^*} \right) \right]^{-1/\xi^*} \right\}, & \xi^* \neq 0, \\
\exp \left\{ - \exp \left( - \frac{x - \mu^*}{\sigma^*} \right) \right\}, & \xi^* = 0,
\end{cases}
\]

for \( 1 + \xi^*(x - \mu^*)/\sigma^* > 0 \), where \( \mu^*, \sigma^* \) and \( \xi^* \) are the location, scale and shape parameter. \( \xi^* > 0 \) corresponds to the Fréchet family, \( \xi^* < 0 \) to the Weibull family and \( \xi^* = 0 \) to the Gumbel family.

When we consider the POT approach rather than the BM approach, we have to model the process of exceedance times and the distribution of the excesses separately. In a stationary climate the threshold \( u \) is constant and the times of exceedance are usually modeled by a homogeneous Poisson process. This implies, that the mean number \( \lambda \) of exceedances in a block (e.g. year or a particular season) is constant over time.

The Pickands–Balkema–De Haan theorem states that the distribution of i.i.d. excesses can be approximated by a GPD, if the threshold \( u \) is sufficiently high and certain regularity conditions hold, see e.g. Reiss and Thomas 2007:

\[
P(Y \leq y | Y \geq 0) = G_{\xi, \sigma}(y) = \begin{cases} 
1 - \left( 1 + \frac{\xi y}{\sigma} \right)^{-1/\xi}, & \xi \neq 0, \\
1 - \exp \left( - \frac{y}{\xi} \right), & \xi = 0,
\end{cases}
\]

for \( y \geq 0 \) if \( \xi \geq 0 \) and \( 0 \leq y \leq -\sigma/\xi \) if \( \xi < 0 \), where \( \sigma \) and \( \xi \) are the scale and the shape parameter. For \( \xi = 0 \) the GPD reduces to the exponential distribution.
2.2. MODEL DESCRIPTION

We are interested in the level \( r(\alpha) \) which is exceeded on average \( \alpha \) times in a block (\( \alpha < \lambda \)). Since there are on average \( \lambda \) peaks in a block, the probability that an arbitrary peak exceeds \( r(\alpha) \) equals \( \alpha / \lambda \), i.e. the excess level \( r(\alpha) - u \) is not exceeded with probability \( 1 - \alpha / \lambda \). Therefore, to obtain \( r(\alpha) \) we add the threshold to the \((1 - \alpha / \lambda)\)-quantile of the excess distribution:

\[
r(\alpha) = u + G^{-1}_{\xi, \sigma}(1 - \alpha / \lambda) = \begin{cases} 
  u - \xi [1 - (\frac{\lambda}{\alpha})^\xi], & \xi \neq 0, \\
  u + \sigma \ln(\frac{\lambda}{\alpha}), & \xi = 0.
\end{cases}
\]  

(2.1)

We will sometimes denote the level \( r(1/m) \) as the \( m \)-year return level to make the comparison with studies for a stationary climate easier. For seasonal data a return period of \( m \) years means, that we expect on average \( 1/m \) excesses per year in the specific season.

If one assumes that the exceedance times originate from a homogeneous Poisson process and the excesses are independent and follow a GPD, it can be shown that the following relationship between the parameters of the GEV and the GPD holds (Buishand 1989; Wang 1991; Madsen et al. 1997b):

\[
\mu^* = \begin{cases} 
  u - \xi (1 - \lambda^\xi), & \xi \neq 0, \\
  u + \sigma \ln(\lambda), & \xi = 0,
\end{cases} \quad \sigma^* = \sigma \lambda^\xi, \quad \xi^* = \xi.
\]  

(2.2)

Note, that the derived GEV distribution is defined only for BM greater than \( u \).

Regional approach

The IF method was originally developed for annual maxima of river discharges by Dalrymple (1960). It assumes that the annual maxima at different sites, after being scaled by a site-specific factor, the ‘index variable’, have a common distribution (e.g. Dalrymple 1960; Hosking and Wallis 1997; Robinson and Sivapalan 1997; Svensson and Jones 2010):

\[
P \left( \frac{M_s}{\eta_s} \leq x \right) = \phi(x) \quad \forall s \in S,
\]  

(2.3)

where \( M_s \) represents a typical block maximum at site \( s \), \( \eta_s \) is the index variable at site \( s \) for \( s \in S \) and the common cdf \( \phi \) does not depend on the site \( s \). From Eq. (2.3) we see, that the site-specific quantile function can be written in the following product form:

\[
q_\tau(s) := Q_{M_s}(\tau) = \eta_s \phi^{-1}(\tau),
\]  

(2.4)
where $Q_{M_s}$ is the quantile function of $M_s$ and $\tau$ is the non-exceedance probability.

Because of using more data than those from the site of interest alone, the IF can provide quantile estimates, which are superior to at-site estimates, even if spatial homogeneity is not entirely achieved after scaling (Cunnane 1988). The IF approach was developed for river discharges but can be applied, whenever multiple samples of similar data are available, see Hosking and Wallis (1997). In particular, for precipitation data the IF assumption has often been used in combination with the GEV family, see e.g. Hosking and Wallis (1997) and Fowler et al. (2005) and Hanel et al. (2009). To further enhance the usage of the available data, Madsen and Rosbjerg (1997) propose the combination of the IF assumption with the POT approach.

A natural analogue of Eq. (2.3) in the POT setting is that the site-specific exceedances, properly scaled by their index variables, have a common distribution. More formally:

$$P\left(\frac{X_s}{\eta_s} \leq x | X_s \geq u_s\right) = \psi(x) \quad \forall s \in S,$$

where the random variable $X_s$ represents the values at site $s$, $\eta_s$ is the site-dependent scaling factor (index variable) and $\psi$ does not depend on site $s$. Note that because $\psi(u_s/\eta_s) = 0$, $\forall s \in S$ and because $\psi$ has a density with mass immediately to the right of $u_s/\eta_s$, it follows that $u_s/\eta_s$ has to be the lower endpoint of the support of $\psi$ for every $s \in S$, i.e.

$$\frac{u_i}{\eta_i} = \frac{u_j}{\eta_j} \quad \forall i, j \in S.$$  

(2.6)

This can be only true if the index variable is a multiple of the threshold, i.e. $\eta_s = cu_s$, $\forall s \in S$, for some positive constant $c$. Without loss of generality we can take $c = 1$. This choice of $\eta_s$ also satisfies the IF equation for the excesses, i.e.,

$$P\left(\frac{Y_s}{\eta_s} \leq y | Y_s \geq 0\right) = \tilde{\psi}(y) \quad \forall s \in S,$$

where $Y_s = X_s - u_s$ and $\tilde{\psi}(y) := \psi(y + 1)$ is independent of site $s$.

A natural choice for a site-specific threshold is a high empirical quantile of the at-site data (see also Smith 1989a). An important consequence of this choice is that the mean number of exceedances per block $\lambda_s$ will be approximately constant over the region, i.e.

$$\lambda_s \equiv \lambda.$$

Under the previous assumptions the distribution of the scaled excesses has the following form:

$$P\left(\frac{Y_s}{u_s} \leq y | Y_s \geq 0\right) = G_{\xi_{\eta_s}, \sigma_{\eta_s}}(y).$$

(2.8)
2.2. Model Description

Eq. (2.7) then implies, that we have the following restrictions on the parameters of the GPD

\[ \frac{\sigma_s}{u_s} \equiv \gamma, \quad \xi_s \equiv \xi, \quad \forall s \in S, \]  

(2.9)
i.e., a common dispersion coefficient \( \gamma \) and a common shape parameter \( \xi \).

We would like to obtain an IF model in the BM setting, if we transfer the parameters from the IF model in the POT setting, using Eq. (2.2). If the BM follow a GEV distribution, it can be shown that the IF assumption is satisfied if the dispersion coefficient \( \gamma^*_s := \sigma^*_s / \mu^*_s \) and the shape parameter \( \xi^*_s \) of the GEV distribution are constant over the region, see e.g. Hanel et al. 2009, i.e.

\[ \gamma^*_s \equiv \gamma^*, \quad \xi^*_s \equiv \xi^*, \quad \forall s \in S. \]  

(2.10)

If we transform the conditions (2.9) according to Eq. (2.2) and use that \( \lambda \) is constant over the region, we obtain the following conditions on the GEV distribution parameters:

\[ \xi^*_s \equiv \xi, \]  

(2.11)

\[ \gamma^*_s = \begin{cases} \frac{\lambda^* \xi}{(1 - \frac{\xi}{\lambda})(1 - \gamma \xi)}, & \xi \neq 0, \\ \frac{1}{1 - \gamma \ln(\lambda)}, & \xi = 0. \end{cases} \]  

(2.12)

That is the conditions in (2.10) are fulfilled.

Summarizing we have developed an IF model with only one spatially varying parameter, the threshold \( u_s \) and the other parameters \( \xi, \gamma, \lambda \) constant over the region. Note that we choose \( \lambda \) to be constant in the first place and therefore obtain a site-specific threshold. This is different from the model proposed by Madsen and Rosbjerg (1997), where \( u_s \) is a priori fixed and only the shape parameter \( \xi \) is constant over the region, whereas \( \sigma \) and \( \lambda \) vary over the region, which violates Eq. (2.2). Moreover, their model is only an IF model for the excesses, whereas our model is an IF model for both the exceedances and the excesses, Eqs. (2.5) and (2.7), respectively.

We get the following GPD model for the exceedances:

\[ P \left( Y_s \leq y | Y_s \geq 0 \right) = G_{\xi, \gamma u_s}(y). \]  

(2.13)

Now we can rewrite Eq. (2.1) for the \( 1/\alpha \)-year return level at site \( s \) as

\[ r_s(\alpha) = \begin{cases} u_s \left( 1 - \frac{\gamma}{\xi} \left[ 1 - \left( \frac{\lambda}{\alpha} \right)^\xi \right] \right), & \xi \neq 0, \\ u_s \left( 1 + \gamma \ln(\lambda/\alpha) \right), & \xi = 0. \end{cases} \]  

(2.14)
As in Eq. (2.4), we see the factorization in a site-specific index variable and a site independent common quantile function.

### 2.2.2 Non-stationary climate

There is no general theory for the distribution of extremes of non-stationary data. Approaches to account for long term trends in extremes are usually pragmatic extensions of the EV models for stationary data (Coles 2001). The classical way to incorporate this non-stationarity in the POT approach, is to keep the threshold constant and model the changing exceedance frequency by an non-homogeneous Poisson process and the excesses by a GPD with time dependent parameters (Smith 1989b; Coles 2001; Yiou et al. 2006; Bengtsson and Nilsson 2007).

We follow a different route by considering a time dependent threshold, see e.g. Yee and Stephenson (2007) and Coelho et al. (2008) and Kyselý et al. (2010), such that the process of exceedances can be approximated by a homogeneous Poisson process. We evaluate this approximation by a number of tests (see section 2.5). A natural way to determine the time varying threshold is quantile regression, which can be described as a way to identify the temporal evolution of a given quantile in a smooth parametric way, see e.g. Koenker (2005) and Friederichs (2010) and Kyselý et al. (2010). Quantile regression is further discussed in section 2.3.1. When we take a time dependent high quantile, estimated by quantile regression, instead of a constant quantile, we can assume that $\lambda$ is constant over space and time. The time dependent GPD is used to describe the excesses of the time varying threshold.

Hanel et al. (2009) generalize the IF assumption to the non-stationary BM setting. Following them we generalize (2.5) in a similar way, which means that, after scaling by a time dependent index variable, for every time point the site-specific cdfs are constant over the region, i.e. $\forall s \in S, \forall t \in T$

$$P \left( \frac{X_s(t)}{\eta_s(t)} \leq x | X_s(t) \geq u_s(t) \right) = \psi_t(x), \quad (2.15)$$

where $\psi_t$ is independent of the site $s$. As in the stationary case we take the threshold as the index variable:

$$\eta_s(t) = u_s(t).$$

Now we can generalize (2.9) in view of (2.15) to

$$\xi_s(t) \equiv \xi(t), \quad \frac{\sigma_s(t)}{u_s(t)} \equiv \gamma(t), \quad (2.16)$$
2.3. ESTIMATION OF THE MODEL PARAMETERS

and Eq. (2.14) can be generalized to the non-stationary setting:

\[
\begin{cases}
    u_s(t) \left(1 - \gamma(t) \left[1 - \left(\frac{\lambda}{\alpha}\right)^{\xi(t)}\right]\right), & \xi(t) \neq 0, \\
    u_s(t) \left(1 + \gamma(t) \ln\left(\frac{\lambda}{\alpha}\right)\right), & \xi(t) = 0.
\end{cases}
\]

(2.17)

As in the stationary case, we can see the factorization into a time and site dependent index variable and a quantile function, which depends on time only.

2.3 Estimation of the model parameters

We have chosen the threshold as a time dependent high quantile. For the estimation of this quantile we use quantile regression, which is outlined in section 2.3.1. Section 2.3.2 illustrates the composite likelihood framework for estimating the time-dependent parameters of the excess distribution.

2.3.1 Threshold

Quantile regression relies on the fact that a sample quantile can be viewed as a solution of an optimization problem, which can be computed efficiently using linear programming, as shown in Koenker and Bassett (1978). For a fixed site \( s \in S \), we can obtain the \( \tau \)-th sample quantile of the observations \( x_s = (x_{s,1}, \ldots, x_{s,T}) \) as

\[
\begin{equation}
\tag{2.18}
\arg \min_{\beta \in \mathbb{R}} \sum_{t=1}^{T} \rho_{\tau}(x_{s,t} - \beta),
\end{equation}
\]

where

\[
\rho_{\tau}(v) = \begin{cases}
    v(\tau - 1), & v < 0, \\
    v^{\tau}, & v \geq 0.
\end{cases}
\]

The value of \( \gamma \) for which the objective function in Eq. (2.18) is minimal is the estimator. Estimators of this type are known as M-estimators. The M-estimator approach provides an alternative for the computation of quantiles, which are often computed using a sorting approach. Its main advantage comes to light when considering quantile regression.

In linear quantile regression it is assumed that the \( \tau \)-th conditional quantile function for given covariates \( z \) has a linear structure, i.e.

\[
\tag{2.19}
Q_{x_s} (\tau|z) = z' \beta(\tau),
\]
e.g. a linear trend in time would be given by
\[
Q_{x_s}(\tau|t) = \beta_0(\tau) + t \cdot \beta_1(\tau).
\]

In view of (2.18) Koenker and Bassett (1978) propose
\[
\arg \min_{\beta_0, \beta_1 \in \mathbb{R}} \sum_{t=1}^{T} \rho_{\tau}(x_{s,t} - \beta_0 - t\beta_1)
\]
as estimator for \( \beta(\tau) \). For details of the transformation of this optimization problem into a linear program, see Koenker (2005).

Note that the threshold is determined for each site separately and that, given the linear quantile function (2.19) holds, we have the following relationship between the mean number of exceedances per block \( \lambda \) and \( \tau \),
\[
(1 - \tau) \cdot T / \#N_B = \lambda,
\]
where \( \#N_B \) is the number of blocks.

### 2.3.2 Excess distribution

Maximum Likelihood (ML) estimation is a common approach to estimate the parameters in a statistical model. The ML framework has attractive asymptotic properties. Moreover, it is very flexible, e.g. it is convenient to incorporate covariates. For these reasons several authors recommend it for the estimation of extreme quantiles, especially when trends occur, see e.g. Coles (2001).

To estimate the regional parameters \( \gamma \) and \( \xi \) of the excess distribution all peaks across the study area are considered simultaneously. For the application of the ML method, the full likelihood function, over all times and sites, is needed. This function is difficult to describe, owing to the spatial dependence and large dimensionality. Moreover, maximization of this function would be virtually impossible. However, if one is interested in the marginal distributions only rather than multivariate extremes, a simplified likelihood for the estimation of the parameters may be used, but standard errors and test procedures have to be adjusted for spatial dependence. In RFA approaches the parameters have sometimes been estimated by the so called independence likelihood, i.e. the likelihood is maximized under the artificial working assumption of spatial independence, see e.g. Moore (1987), Smith (1989a), Buishand (1991), Cooley et al. (2007) and Hanel et al. (2009). Though this method provides asymptotically unbiased parameter estimates, the spatial dependence leads to an increase of the variance of the estimates compared to the independent case. Especially in the earlier papers the adjustment of the error estimation for spatial dependence
2.3. ESTIMATION OF THE MODEL PARAMETERS

was not made. Smith (1990b) is probably the first reference where standard errors and likelihood ratio tests were adjusted for spatial dependence in an RFA approach. His approach is a special case of the composite likelihood method, see Varin et al. (2011) for an extensive overview. Recently it has been applied by Blanchet and Lehn-\v{z}ing (2010) to annual maximum snow depths over Switzerland and by Van de Vyver (2012) to annual extremes of precipitation in Belgium.

In the non-stationary IF model, the parameters $\gamma$ and $\xi$ of the excess distribution depend on time. We postulate a certain structure for these parameters, e.g.

$$\gamma(t) = \gamma_1 + \gamma_2 \cdot (t - \bar{t}), \quad \xi(t) = \xi_1,$$

where $\bar{t}$ is the mean of the time points, so that $\gamma_1$ is the average of $\gamma(t)$ over $t$. Let $\theta = (\gamma_1, \gamma_2, \xi_1)'$ be the vector of parameters, that has to be estimated. The independence likelihood is then given by:

$$L_I(\theta; y) = \prod_{t \in T} \prod_{s \in S, y_s(t) \geq 0} \frac{1}{\gamma(t) u_s(t)} \cdot \left[ 1 + \frac{\xi(t) y_s(t)}{\gamma(t) u_s(t)} \right]^{-1/(\xi(t)-1)},$$

where the condition on $y_s(t) \geq 0$ reflects that we only consider peaks over the threshold. Note, that by the choice of the quantile, the threshold has been fixed beforehand.

Corresponding to the Maximum Likelihood Estimator (MLE) the Maximum Independence Likelihood Estimator (MILE) is the parameter $\hat{\theta}_I$ which maximizes $L_I(\theta, y)$, respectively the independence log likelihood

$$\ell_I(\theta; y) = -\sum_{t \in T} \sum_{s \in S, y_s(t) \geq 0} \left[ \ln(\gamma(t) u_s(t)) + \frac{1 + \xi(t)}{\xi(t)} \ln(1 + \frac{\xi(t) y_s(t)}{\gamma(t) u_s(t)}) \right]. \quad (2.20)$$

We have to optimize this function, with respect to the elements of $\theta$. This can be done using the Broyden–Fletcher–Goldfarb–Shanno method as implemented in the optim-function of GNU R (R Core Team 2014).

For testing the adequacy of the IF model, it is necessary to consider models with a spatially dependent dispersion coefficient, e.g. $\gamma_s(t) = \gamma_s$ and $\xi_s(t) = \xi$. The independence log likelihood for this model is obtained by replacing $\gamma(t)$ by $\gamma_s$ and $\xi(t)$ by $\xi$ in Eq. (2.20). The direct optimization of this likelihood with respect to the $(S + 1)$ parameters is in the case of a large number of sites computationally very demanding. Therefore we exploit the structure of the independence likelihood by using a profile likelihood approach. In the example above we can split, for a given
shape parameter, the optimization over an $S$-dimensional space into $S$ optimization problems in one dimension, i.e. the maximization of the log likelihood for the excesses at site $s$ with respect to $\gamma_s$. This is usually much faster. If one does this on a grid of potential values for the shape parameter one can see the structure of the profile likelihood. Moreover, we can construct a convergent procedure, leading to the estimator for the shape parameter. We recommend as initial value for this procedure the mean of the estimated shape parameters $\hat{\gamma}_s$ of a site-specific model. Another problem with the direct optimization might be the existence of local maxima in the likelihood; with the proposed approach we did not experience any problems with this issue.

The MILE $\hat{\theta}_I$ is asymptotically normal, see e.g. Varin et al. (2011):

$$\sqrt{\# \hat{T}} (\hat{\theta}_I - \theta) \xrightarrow{d} N \left(0, G^{-1}(\theta)\right),$$

where $\# \hat{T}$ is the number of days with one or more threshold exceedances and $G(\theta)$ is the Godambe information:

$$G(\theta) = H(\theta) J^{-1}(\theta) H(\theta), \quad (2.21)$$

where $H(\theta)$ is minus the expected Hessian of $\ell_I$ at $\theta$, also referred to as sensitivity matrix, and $J(\theta)$ is the variability matrix, i.e. the covariance matrix of the score $u(\theta; y) = \nabla_\theta \ell_I(\theta; y)$. In the case of spatial independence, we have $H(\theta) = J(\theta)$ and the Godambe information reduces to the Fisher information, i.e. $G(\theta) = H(\theta)$. Here $H(\theta)$ is estimated as its observed value at $\hat{\theta}_I$, and $J$ as

$$\hat{J} = \frac{1}{\# \hat{T}} \sum_{t \in T} \left. \left. u(\hat{\theta}_I; y(t)) \right| u(\hat{\theta}_I; y(t)) \right|,'$$

where $y(t) = (y_1(t), \ldots, y_S(t))'$ and $u(\hat{\theta}_I; y(t))$ is the contribution of day $t$ to $u(\hat{\theta}_I; y)$. The latter estimate makes use of the fact that the excesses on different days are independent, see e.g. Smith (1990b), Chandler and Bate (2007) and Varin et al. (2011). An estimate $\hat{G}(\theta)$ of the Godambe information is obtained by plugging in the estimates $\hat{H}$ and $\hat{J}$ in Eq. (2.21). This estimate $\hat{G}(\theta)$ is used to assess the uncertainty of the parameters (and quantiles) of the excess distribution, see section 2.4

### 2.4 Model selection for the excess distribution

In this section we describe the methods used to investigate the temporal behavior of the dispersion coefficient and the shape parameter as well as the adequacy of the IF
2.4. MODEL SELECTION FOR THE EXCESS DISTRIBUTION

Information criteria are used as an indication of the suitability of a specific model. Varin et al. (2011) present composite likelihood adaptations of the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which are defined in the usual way

\[
AIC = -2\ell_I(\hat{\theta}_I; y) + 2 \dim(\theta),
\]

\[
BIC = -2\ell_I(\hat{\theta}_I; y) + \ln(\#\tilde{T}) \dim(\theta),
\]

where \( \dim(\theta) \) is an effective number of parameters:

\[
\dim(\theta) = \text{tr} \left( \hat{H}(\theta) \hat{G}(\theta)^{-1} \right).
\]

Moreover, we will test our assumptions using nested models. This means, that we consider subsets \( M_0 \) of the full model \( M_1 \) by constraining \( q \) components of the parameter vector \( \theta \). For instance, we may partition \( \theta = (\psi, \phi)' \) such that the \( q \)-dimensional component \( \psi \) is zero under \( M_0 \). To test this hypothesis, we use the independent likelihood ratio statistic, which is a special case of a composite likelihood ratio (CLR) statistic (Chandler and Bate 2007; Varin et al. 2011):

\[
W = 2 \left[ \ell_I(\hat{\theta}_{M_1}; y) - \ell_I(\hat{\theta}_{M_0}; y) \right], \tag{2.22}
\]

where \( \hat{\theta}_{M_1}, (\hat{\theta}_{M_0}) \) denotes the MILE under model \( M_1 (M_0) \). Varin et al. (2011) present the following asymptotic result for \( W \) under the null hypothesis

\[
W \xrightarrow{d} \sum_{j=1}^{q} \lambda_j Z_j^2, \tag{2.23}
\]

where the \( Z_j \) are independent, standard normal variates and \( \lambda_1, \ldots, \lambda_q \) are the eigenvalues of

\[
(G_{M_1}^{-1})_{\psi} \left( (H_{M_1}^{-1})_{\psi} \right)^{-1}.
\]

Here \( (G_{M_1}^{-1})_{\psi} \) denotes the submatrix of the inverse Godambe information for the full model \( M_1 \) pertaining to the parameter vector \( \psi \) and \( (H_{M_1}^{-1})_{\psi} \) is defined analogously.

In order to obtain the information criteria and the asymptotic distribution of \( W \) under the null hypothesis, we need to estimate the Godambe information, which is difficult when the number of parameters is large. Hence it is not feasible to examine the appropriateness of the IF assumption for regions with many sites, based on the Godambe information.

One possibility to obtain \( p \)-values for the test statistic \( W \), without estimating the
Godambe information, is to apply a bootstrap procedure, see e.g. Varin et al. 2011. We follow Hanel et al. 2009 and use a semiparametric bootstrap approach, to take the dependence structure into account, without explicitly modeling this. The challenge is to produce bootstrap samples according to the null hypothesis, which exhibit approximately the same spatial dependence structure as the original data set. We assume that the underlying spatial dependence is not changing over time, i.e. only the marginal distributions are changing. One could think of a constant copula generating the dependence structure, i.e. for fixed $t$

$$P(Y_s(t) \leq y_s \ \forall s) = C(G_{1,t}(y_1), \ldots, G_{S,t}(y_S)),$$

where $G_{s,t} = G_{\sigma_s(t),\xi_s(t)}$, and $C$ is a copula, for details on copulas see e.g. Nelsen 2006. We generate the bootstrap samples in three steps. In the first step we transform the sample of the excesses $y_s(t)$ into a sample that follows approximately the standard exponential distribution

$$z_s(t) = \begin{cases} \frac{1}{\hat{\xi}_s(t)} \ln(1 + \hat{\xi}_s(t) y_s(t)), & \hat{\xi}_s(t) \neq 0, \\ \frac{y_s(t)}{\hat{\sigma}_s(t)^2}, & \hat{\xi}_s(t) = 0, \end{cases} \quad (2.24)$$

where $\hat{\sigma}_s(t)$ and $\hat{\xi}_s(t)$ are the estimated scale and shape parameters under the full model $M_1$. In the second step, we sample with replacement from $z_s(t)$ monthly blocks of the whole spatial domain to obtain a new sample $\tilde{z}_s(t)$ with approximately standard exponential margins and the same spatial dependence structure as that of $z_s(t)$. In the third step we use the estimated scale and shape parameter under the null hypothesis, denoted as $\hat{\sigma}_s(t)$ and $\hat{\xi}_s(t)$, respectively, to transform the sample $\tilde{z}_s(t)$ to a bootstrap sample of the excesses

$$\tilde{y}_s(t) = \begin{cases} \frac{\hat{\sigma}_s(t)}{\hat{\xi}_s(t)} \left[ \exp(\hat{\xi}_s(t) \tilde{z}_s(t)) - 1 \right], & \hat{\xi}_s(t) \neq 0, \\ \tilde{z}_s(t) \hat{\sigma}_s(t), & \hat{\xi}_s(t) = 0. \end{cases} \quad (2.25)$$

The $\tilde{y}_s(t)$ follow approximately the GPD model $M_0$ and mimic the spatial dependence structure of the original excesses.

From a number of Monte Carlo experiments, Kyselý (2007) and Kyselý (2010) concluded that the (non-parametric) bootstrap generally resulted in too narrow confidence intervals for large quantiles of the distributions, that are commonly used to describe the distribution of precipitation extremes. This has been attributed to the skewness of the estimators of the model parameters in the case of small and moderate sample sizes. This objection might be weakened, when using RFA methods,
2.5 Application to precipitation data

We applied the regional peaks-over-threshold method to observed precipitation data from the Netherlands. We used the daily, gridded E-OBS data (version 5.0), which were made available by the European funded project ENSEMBLES (Haylock et al. 2008). We consider winter (DJF) precipitation for 25 km × 25 km grid squares centered in the Netherlands, for the period December 1, 1950 to February 28, 2010. In total we have 69 grid boxes and 60 winter seasons of daily measurements for each grid box.

The Netherlands has a maritime climate with relatively mild and humid winters. Fig. 2.1 shows the mean over the considered period of the largest daily precipitation value in winter (winter maximum) for each grid box. The spatial variation in Fig. 2.1 is small, 80% of the values lie between 18.2 and 20.4 mm. Previous studies propose to view the Netherlands as a homogeneous region to which the IF assumption applies, see e.g. Overeem et al. (2008) and Hanel et al. (2009).

Figure 2.1: Mean of the winter maxima in mm.
2.5.1 Event selection and trend in the threshold

Daily precipitation in the winter season exhibits some temporal dependence, also at high levels. The relation between the GEV and GPD parameters (Eq. (2.2)) relies on the independence assumption as does the estimation of the variability matrix $J$, therefore, it is necessary to select a subset of independent events. This is usually achieved by specifying a minimum separation time $t_{sep}$ between exceedances over the threshold. The separation time is determined by the temporal dependence in the data at high levels. This temporal dependence is weak for daily precipitation and separation times of one or two days seem to be sufficient, see e.g. Coles (1993) and Kyselý and Beranová (2009). In this chapter a separation time of one day was chosen and the original data were declustered rather than the exceedances. For every $s \in S$ and $t \in T$ we replace $x_t(s)$ by zero, if it is not a local maximum, i.e. if $x_{t-1}(s)$ or $x_{t+1}(s)$ is larger than $x_t(s)$. For example, we obtain from the sequence $(0, 3, 4, 5, 0)$ the new sequence $(0, 0, 0, 5, 0)$ but the sequence $(0, 3, 0, 5, 0)$ remains unchanged. It is clear, that also the excesses obtained from the declustered data are separated at least by one day. However, while the method handles temporal dependence for each site separately, it does not handle situations where the peak at grid box A occurs a day later than on site B. In most cases these will be two separate rainfall events. The main advantage of this declustering algorithm is that the expected number of exceedances per block $\lambda$ will be approximately constant, which is a basic assumption of our model. This follows since we determine the threshold for the new data via quantile regression, as described in section 2.3.1.

We choose the threshold to be the 96% linear regression quantile of the declustered data. Hence, we expect on average 3.61 exceedances per grid box and winter season. The actual sample size varies between 216 and 218 exceedances per grid box, which corresponds to 3.6-3.63 exceedances per grid box and season (the small differences are caused by ties in the daily precipitation amounts). In total we observe 777 days with at least one exceedance. In Fig. 2.2 we show the cumulative sum of the number of exceedances for the successive winter seasons for the grid box around De Bilt with 95% pointwise tolerance intervals for a homogeneous Poisson process, with intensity $\lambda = 3.61$ and conditioned on the total number of exceedances (see Lang et al. 1999, for the derivation of the tolerance interval). The figure is typical for the whole region and indicates, that a homogeneous Poisson process might be appropriate to describe the occurrence times of the exceedances. Moreover, we test the hypothesis that the number of exceedances in a winter season follow a Poisson distribution for each grid box individually. The dispersion index (DI) test exploits the fact, that the variance and the mean of the Poisson distribution are the same, see Cunnane (1979) for details. The ratio of the variance and the mean is sensitive to the separation time between exceedances: the ratio tends to be larger than one if
2.5. APPLICATION TO PRECIPITATION DATA

Figure 2.2: Cumulative sum of the exceedances (solid) for the grid box around De Bilt together with the theoretical mean (dashed) and 95% pointwise tolerance intervals.

the separation time is too short. The Poisson assumption is rejected at the 5% significance level in two of the 69 grid boxes, which is in good agreement with the expected number of rejected grid boxes under the Poisson assumption. If the exceedance times come from a homogeneous Poisson process, these should be distributed uniformly on any time interval, conditioned on the number of exceedances, which can be tested by the Kolmogorov–Smirnov (KS) test, see e.g. Cox and Lewis (1966). This test does not reject uniformity in any grid box. The $p$-values of the Kolmogorov-Smirnov test on uniformity of the event times are shown in Fig. 2.3.

Fig. 2.4 shows for each grid box the mean of the threshold for the 1950–2010 period. The trend in the threshold for the 1950–2010 period is positive over the whole domain, see Fig. 2.5, but is relatively small in the southeastern part of the country and large (up to 0.68 mm per decade) in the west and northern parts, where it is significant at the 5% level. In Fig. 2.6 one can observe, that the temporal evolution of the thresholds is similar for each of the selected grid boxes. The findings are consistent with Buishand et al. (2013) who found a significant positive trend in the mean precipitation for the winter half year (October – March) in the Netherlands during the period 1951 – 2009, but did not detect a clear spatial gradient in the trend of mean winter precipitation.
CHAPTER 2. REGIONAL NON-STATIONARY POT MODEL

Figure 2.3: $p$-values of the KS test on uniformity of the event times.

Figure 2.4: Mean of the threshold for the 1950–2010 period in mm.
2.5. APPLICATION TO PRECIPITATION DATA

Figure 2.5: Trend in the threshold for the 1950–2010 period in mm per decade. The deltas indicate that the observed trend is significant at the 5% level.

Figure 2.6: Left: Temporal evolution of the threshold for six selected grid boxes in mm. Solid lines indicate, that the observed trend is significant at the 5% level and dashed lines symbolize non-significant trends. Right: Location of the corresponding grid boxes. The brown dot denotes the grid box around De Bilt.
2.5.2 Threshold excesses

We consider four different models for the excess distribution, three based on the IF approach, $A – A''$ in Table 2.1, and one with a spatially varying dispersion coefficient and constant shape parameter, model $B$.

Table 2.1: Overview of models used.

<table>
<thead>
<tr>
<th>Model</th>
<th>dispersion</th>
<th>shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$\gamma$</td>
<td>$\xi$</td>
</tr>
<tr>
<td>$A'$</td>
<td>$\gamma_1 + \gamma_2 * (t - \bar{t})$</td>
<td>$\xi$</td>
</tr>
<tr>
<td>$A''$</td>
<td>$\gamma$</td>
<td>$\xi_1 + \xi_2 * (t - \bar{t})$</td>
</tr>
<tr>
<td>$B$</td>
<td>$\gamma_1, \ldots, \gamma_S$</td>
<td>$\xi$</td>
</tr>
</tbody>
</table>

In a first step we want to infer which of the IF models is the best to describe the data. For a first indication the information criteria are computed, as outlined in section 2.4, for each of the three IF models, see Table 2.2. We see from both the composite AIC and the composite BIC, that the incorporation of a trend in the dispersion coefficient $\gamma$ (model $A''$) does not result in a better model. According to the AIC model $A''$, which has a (linear) trend in the shape parameter, is selected. This contrasts with the selection of the simplest model $A$ by the BIC. The effective number of parameters $\dim(\theta)$ is much larger than in the traditional AIC or BIC criteria, i.e. for model $A$ we observe that $\dim(\theta)$ equals 70.5 instead of two for the independent case and for model $A'$ and $A''$ we obtain values for $\dim(\theta)$ of 95.5 and 89.3, respectively, instead of three. This large difference is owing to the strong spatial dependence in the data.

Table 2.2: Information criteria for the IF models. The lowest AIC and BIC values are printed in bold.

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>78387.28</td>
<td>78715.59</td>
</tr>
<tr>
<td>$A'$</td>
<td>78435.60</td>
<td>78880.41</td>
</tr>
<tr>
<td>$A''$</td>
<td>78333.28</td>
<td>78748.95</td>
</tr>
</tbody>
</table>

The shape parameter is crucial for the estimation of very high quantiles. Model $A$ estimates the shape parameter to be 0.03, i.e. just in the Fréchet domain. Model $A''$ estimates a large drop in the shape parameter from 0.10 to -0.09, which would mean a change from the Fréchet family to the Weibull family. In order to gain more insight in the temporal behavior of the shape parameter, we compute the shape parameter for overlapping 20-year subsamples of the data, using model $A$, which has no trend in the model parameters. It appears that a large part of the negative trend in the
shape parameter in model $A''$ is owing to one specific event, namely the extreme rainfall of December 3, 1960, compare also Buishand (1984b) and Van den Brink and Können (2011), resulting in a large drop of the 20-year window estimates in the year 1971, as observed in Fig. 2.7.

![Figure 2.7: Evolution of the shape parameter $\xi$ over time (dotted - model $A$, dashed - model $A''$, solid line with dots - 20 year window estimates for model $A$).](image)

The quantile estimates, obtained from model $A$, are increasing due to the positive trend in the threshold. Because in this model there are no trends in the dispersion coefficient and the shape parameter, the trend in the quantile is proportional to the trend in threshold. The average increase over the country is 22% for the entire period 1950–2010. In contrast to the previous model, the trends in the quantiles from model $A''$ depend on the return period. While the 2-year return level is still increasing due to the positive trend in the threshold, we have that the 25-year return level is decreasing due to the negative trend in the common shape parameter. The 5-year return level is approximately constant, see Fig. 2.8. An interpretation of this is much more complex, than for the quantile estimates, stemming from model $A$.

When we carry out the composite likelihood ratio test, it turns out that neither the trend in the dispersion coefficient nor the trend in the shape parameter are significant, although the $p$-values are quite different for these trends, see Table 2.3. We can also see from Table 2.3, that the bootstrap procedure gives similar results as the use of the asymptotic result in Eq. (2.23). We want to stress, that the regional approach is more likely to detect a trend, if there is one. This can be deduced e.g. by comparing the standard errors of the regional approach with those obtained for the same model in the at-site analysis. For model $A'$ the standard error of the dispersion
CHAPTER 2. REGIONAL NON-STATIONARY POT MODEL

Figure 2.8: Trends of different return levels of daily precipitation for model $A''$ (dashed – 2 year, solid – 5 year, dotted – 25 year).

trend estimate is 35% smaller and for model $A''$ the standard error of the shape trend estimate is 37% smaller than for the at-site approach.

Table 2.3: $p$-values of the CLR-test against model $A$ (2500 samples).

<table>
<thead>
<tr>
<th>Model</th>
<th>asymptotic</th>
<th>bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A'$</td>
<td>82.9%</td>
<td>81.3%</td>
</tr>
<tr>
<td>$A''$</td>
<td>26.7%</td>
<td>12.2%</td>
</tr>
</tbody>
</table>

In the second step we want to test the IF assumption. Therefore we compute the composite likelihood ratio test for the full model $B$ and the nested model $A$. As earlier explained, we cannot estimate the Godambe information well for model $B$. Hence, we proceed only with the bootstrap procedure. We obtain a $p$-value of 0.103 for 2500 bootstrap samples. This means, that the IF assumption does not have to be rejected. Note, because of the large difference in the number of parameters between model $B$ and model $A$, the composite likelihood ratio test will not have much power due to the great number of alternatives. This can be considered as an intrinsic problem, when comparing regional models with site dependent parameters.

In the following we focus on model $A$, i.e. no trend in the dispersion coefficient and shape parameter. Before looking closer to return levels and their associated uncertainty, we investigate the adequacy of the model and the spatial dependence using the seasonal maxima of the standard exponential residuals $z_s(t)$ defined by Eq.
(2.24). The maximum for grid box \( s \) in season \( j \) is denoted as \( m_{s,j} \). From Eq. (2.2) it follows that the \( m_{s,j} \) are approximately Gumbel distributed with location parameter \( \ln(\lambda) \) and scale parameter 1, censored at zero. We have to censor the distribution to account for seasons without any exceedance. The empirical distribution of \( m_{s,j} \) can be compared with the theoretical distribution in a Gumbel plot. Fig. 2.9 shows the spatial mean of the empirical distributions. Apart from the outlier at the largest rank, which is due to the December 1960 event, there is a good correspondence between the averaged observed distribution and the postulated Gumbel distribution. Analogous to Dales and Reed (1989), see also Reed and Stewart (1994), the degree of spatial dependence is determined by comparing the distribution of the seasonal spatial maximum, i.e. the largest seasonal maximum over the region in a season

\[
m_{j} := \max(m_{1,j}, \ldots, m_{S,j}),
\]

with the distribution of the seasonal maximum of an individual grid box. Based on this comparison an effective (spatial) sample size, i.e. number of independent grid boxes, can be computed, which is a measure of joint tail dependence. In the case of fully spatially dependent data, the \( m_{j} \) follow the same Gumbel distribution as the maxima \( m_{s,j} \) at an individual grid box. However, in the absence of spatial dependence the location parameter increases to \( \ln(\lambda S) \). The empirical distribution of the \( m_{j} \), as shown in Fig. 2.9, indicates that the data are neither fully spatially dependent nor independent. To determine the effective sample size \( S_e \) we fit a Gumbel distribution to the \( m_{j} \), keeping the scale parameter fixed at 1. The location parameter \( \hat{\mu} \) of the fit is then equivalent to \( \ln(\lambda S_e) \). Hence, we obtain \( S_e = \exp(\hat{\mu}) / \lambda \), which results in an effective sample size of almost 4 for the data in Fig. 2.9.

The influence of the spatial dependence on the uncertainty of the parameters is, however, directly related to the variability matrix \( J \) in Eq. (2.21) and not to \( S_e \). Fig. 2.10 compares for a particular grid box the estimated return levels of the excess distribution based on the site-specific approach with those obtained from the RFA. Pointwise confidence bands for the return levels based on the delta method using the asymptotic normality of the MILE are also given. The quantile estimates for the two methods are quite similar, but the IF approach reduces the uncertainty in the estimation by 37.5%. We can see in Fig. 2.10 that, much tighter confidence bands are obtained, if no adjustment for spatial dependence is made. Fig. 2.11 visualizes for one grid box the temporal evolution of the threshold, based on at-site estimation, and the 25-year return level, based on the at-site estimation and the IF approach. Additionally, the figure shows confidence bands for the threshold and the 25-year return level. The confidence band for the threshold is based on the two-sided 95% confidence interval using a Huber sandwich estimate, as implemented in the quantreg package (option \( nid \)) (Koenker 2005). The confidence band for the 25-year return
Figure 2.9: Adequacy of the regional GPD model and spatial dependence (blue – spatial mean of the empirical distributions of the seasonal maximum $m_{s,j}$ of the exponential residuals at an individual grid box, red – empirical distribution of their spatial maximum $\tilde{m}_{j}$, black – Gumbel distribution with location parameter $\ln(\lambda)$ and scale parameter 1 (censored at 0), dotted – Gumbel distribution for $S$ independent grid boxes, dashed – Gumbel distribution for 4 independent grid boxes).

Level is computed by adding the lower (upper) limits of the 95% confidence band for the threshold to the lower (upper) limits of the 95% confidence band for the 25-year return level of the excesses (shown in Fig. 2.10). This confidence band has a coverage probability of at least 90%, which can be seen by simple probability arguments. We can see, that the uncertainty in the threshold is small compared to the uncertainty in the return level.
2.5. APPLICATION TO PRECIPITATION DATA

Figure 2.10: Estimated return levels of the excesses (solid lines) with 95% pointwise confidence bands (dashed lines) for the year 1980 at the grid box around De Bilt (black – site-specific, red – IF, blue – no correction of the standard error for spatial dependence).

Figure 2.11: Estimated threshold with 95% pointwise confidence band (black) and 25-year return level based on the at-site estimation (blue) and the IF approach (red), together with pointwise confidence bands, for the grid box around De Bilt.
2.6 Conclusion

A regional approach for non-stationary Peaks-Over-Threshold data has been developed. The threshold is chosen to be a large quantile that varies over time, which is also taken as the index variable. The peaks exceeding the threshold are described by Generalized Pareto Distributions. The IF assumption implies that the ratio of the scale parameter to the threshold and the shape parameter are constant over the region but may vary over time. A consequence of this is that the ratio between different return levels is constant over the region.

The approach was applied to gridded, observed daily precipitation data from the Netherlands for the winter season. A linear increase in the threshold was found, which is significant in the western and northern parts of the country. An apparent trend in the shape parameter was observed, which turned out to be mainly due to one exceptional event. This trend was not significant at the 10%-level and was therefore disregarded. There was no evidence of a change in the ratio of the GPD scale parameter to the threshold, which means that the increase in the threshold is accompanied by an increase in the scale parameter. Therefore, we conclude that rare extremes are increasing proportionally to the increase in the threshold.

Although the uncertainty in the estimation of the excess distribution was considerably reduced by the IF approach, the remaining uncertainty is still substantial. This is owing to the spatial dependence. The uncertainty could be possibly further reduced by considering longer records or by extending the region. For instance, one could think of including the neighboring part of North Germany in the analysis. However, the presented model may not necessarily apply to larger regions, owing to the homogeneity constraints. In particular the different trends in the index variable indicate, that one should be very careful with extending the region. Moreover, one should keep in mind, that the effective sample size may grow slowly with the size of the region owing to the spatial dependence.

The validity of the bootstrap might be questionable and should be assessed by a Monte Carlo experiment, which includes the spatial dependence. However, this is for POT data much more computationally demanding than for BM.

Acknowledgments

We thank three anonymous reviewers for their helpful comments on an earlier version of this chapter. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://eca.knmi.nl).
Projections of precipitation extremes

Abstract. Projections of extreme precipitation are of great importance, considering the potential severe impacts on society. In this chapter a recently developed regional, non-stationary Peaks-Over-Threshold (POT) approach is applied to two transient simulations of the RACMO2 Regional Climate Model (RCM) for the period 1950-2100, driven by two different General Circulation Models (GCMs). The regional approach reduces the estimation uncertainty compared to at-site approaches. The selection of a threshold for the POT model is tackled from a new perspective, taking advantage of the regional setting. Further, a regional quantile regression model using a common relative trend in the threshold is introduced. A considerable bias in the extreme return levels is found with respect to gridded observations. This bias is corrected for by adjusting the parameters in the POT model. In summer a significant increase in extreme precipitation over the study area is found for both RACMO2 simulations, mainly as a result of an increase of the variability of the excesses over the threshold. However, the magnitude of this trend in extreme summer precipitation depends on the driving GCM. In winter an increase in extreme precipitation corresponding with an increase in mean precipitation is found for both simulations. This trend is due to an increase of the threshold.

3.1 Introduction

Information on extreme precipitation is crucial for various societal sectors, e.g. for the design of sewage and drainage systems, roads and tunnels, farming, and the insurance industry. Consensus is growing that the characteristics of extreme precipitation may alter owing to climate change. In order to project the change in extreme precipitation, climate model data have been analyzed and compared to observations. Often this evaluation is done in terms of index values, such as the empirical annual 90% quantile of the precipitation amounts for each year or the 1-day or 5-day maximum precipitation amount in a year, see e.g. Klein Tank and Können (2003) and Turco and Llasat (2011). This approach shows the evolution of precipitation extremes over time. However, the indices have mostly a return period of not more than one year, which is of minor importance for the planning of hydraulic infrastructure, that usually has to withstand events with much longer return periods. To estimate the changes in these rare events, Extreme Value (EV) distributions have been fitted to the extremes for two subsets of the data representing current (e.g. 1980-2010) and future (e.g. 2070-2100) climate, assuming stationarity within the time slices, see e.g. Fowler et al. (2005), Ekström et al. (2005), and Kyselý and Beranová (2009). Considering only two time slices does not give a picture of the evolution of the extremes, which is e.g. necessary if one is interested in the risk of failure of a hydraulic structure during its expected lifetime. Moreover, the selection of the time slices introduces additional uncertainty. A small shift of the time slices may have large influence on the estimated change. As an alternative, EV distributions with time-dependent parameters, which allow the consideration of the full time period, have been used, see e.g. Coles (2001), El Adlouni et al. (2007), Sugahara et al. (2009), Kyselý et al. (2010), and Beguería et al. (2010), and Tramblay et al. (2013).

The estimation of changes in rare extremes is subject to large uncertainty. A general way to reduce the estimation uncertainty is Regional Frequency Analysis (RFA), where the similarities between different sites in a region are exploited (Hosking and Wallis 1997). RFA is mostly applied to (annual) Block Maxima (BM). An alternative to BM is to consider all POT, which is often preferable, owing to the more efficient use of the data.

A regional POT model, combining the RFA approach and POT data, which can be used to analyze precipitation extremes in a changing climate, was developed in Chapter 2. In this model a temporally varying threshold, which is determined by quantile regression, is used to account for changes in the frequency of precipitation extremes. The marginal distributions of the excesses are described by Generalized Pareto Distributions (GPDs), with parameters, that may vary over time and their spatial variation is modeled by the Index Flood (IF) approach. For a detailed introduction to the IF method, see Hosking and Wallis (1997).
3.2 METHODS

The selection of the threshold is a crucial step in the application of the POT approach. However, there is still no standard procedure for this, and usually one relies on visual tools. Among these the plotting of the change of the estimated GPD shape parameter or the Mean Excess (ME) of the exceedances against the height of the threshold is popular. Unfortunately these plots rarely give clear indications of which quantile should be used for the threshold. In this chapter, the individual plots are averaged over the region in order to make the desired constant or linear structure more apparent.

Daily precipitation from two simulations of a RCM driven by different GCM and from gridded observational data is analyzed for the Netherlands and north-western Germany. Instead of linking the POT model parameters to time, a temperature-based covariate is used to include the evolution of climate, see also Hanel et al. (2009), and Van Oldenborgh et al. (2009). Bias correction of the return levels from the RCM simulations is discussed. In addition to the changes in return levels, we present a risk-based design level, that was recently introduced by Rootzén and Katz (2013).

Section 3.2 outlines the methods and section 3.3 introduces the data used. Results and discussion are given in section 3.4, followed by the conclusion in section 3.5.

3.2 Methods

3.2.1 Introduction to the POT model

To study the extremes of independent and identically distributed random variables \(X_i\), one can consider the excesses \(Y_i = X_i - u\) over a (high) threshold \(u\). The Pickands–Balkema–De Haan theorem states that the distribution of the excesses \(Y\), conditioned on \(Y \geq 0\), can be approximated by a GPD, if the threshold \(u\) is sufficiently high and certain regularity conditions hold, see e.g. Reiss and Thomas (2007):

\[
P(Y \leq y | Y \geq 0) = G_{\xi, \sigma}(y) = \begin{cases} 
1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}, & \xi \neq 0, \\
1 - \exp\left(-\frac{y}{\sigma}\right), & \xi = 0,
\end{cases}
\]

for \(y \geq 0\) if \(\xi \geq 0\) and \(0 \leq y \leq -\sigma/\xi\) if \(\xi < 0\), where \(\sigma\) and \(\xi\) are the scale and the shape parameter respectively. For \(\xi = 0\) the GPD reduces to the exponential distribution. The independence requirement can be weakened (e.g. Leadbetter et al. 1983). In the case of short-range dependence the GPD approximation applies if one considers declustered excesses, i.e. the excesses of the local maxima (peaks) in a cluster of exceedances only. Several studies have considered the GPD also for non-stationary data, using temporally varying parameters, see for recent examples Sugahara et al. (2009), Kyselý et al. (2010), and Beguería et al. (2010).
3.2.2 Temporal dependence and declustering

Daily precipitation exhibits temporal dependence, also at high levels. This dependence is generally stronger in winter than in summer (due to the convective nature of most extremes in summer). As mentioned in section 3.2.1 one can account for this temporal dependence by considering declustered excesses. This is usually achieved by specifying a minimum separation time $t_{\text{sep}}$ between exceedances over the threshold, where $t_{\text{sep}}$ is determined by the temporal dependence in the data at high levels. Here, we follow this approach but decluster all data and not the excesses only.

Let $x_s(t)$ be the rainfall at site $s \in \{1, \ldots, S\}$ and day $t \in \{1, \ldots, T\}$. To determine $t_{\text{sep}}$ we compute first the 95% quantile for each site $s$ and calculate the number of clusters of length $n \geq 2$. A cluster of length $n$ is defined as $n$ consecutive exceedances of the 95% quantile. The number of clusters decreases usually very fast with the length $n$. The separation time $t_{\text{sep}}$ is set to $n$, if the number of clusters of length $n + 2$ is sufficiently low. After this initial step we obtain the declustered data by replacing $x_s(t)$ with zero, if it is not a maximum in the subset $x_s(t - t_{\text{sep}}), \ldots, x_s(t), \ldots, x_s(t + t_{\text{sep}})$. From this it is clear, that also the excesses obtained from the declustered data are separated by at least $t_{\text{sep}}$ days.

3.2.3 Index Flood approach

In Chapter 2 a regional approach for multi-site, non-stationary POT rainfall data

$$y_s(t) = x_s(t) - u_s(t),$$

where $u_s(t)$ is a suitable threshold value for site $s$ and day $t$, has been introduced. The approach is based on the IF assumption, i.e. that the non-stationary POT data have, after scaling by a time and site dependent index variable (or index rainfall) $\eta_s(t)$, a common excess distribution. If the site-specific excess distributions are GPD with shape parameter $\xi_s(t)$ and scale parameter $\sigma_s(t)$, then we have for the scaled excesses:

$$P\left(\frac{Y_s(t)}{\eta_s(t)} \leq y | Y_s(t) > 0\right) = G_{\xi_s(t),\gamma(t)}(y),$$

(3.1)

with $Y_s(t)$ the excess at site $s$ and day $t$ and $\gamma(t) = \sigma_s(t)/\eta_s(t)$ a dimensionless dispersion coefficient. The IF assumption thus implies that this coefficient and the shape parameter are constant over the region of interest. In Chapter 2 the threshold $u_s(t)$ was used as index variable:

$$\eta_s(t) = u_s(t).$$
3.2. METHODS

The mean number $\lambda$ of the excesses over $u_s(t)$ in this approach is constant over time and space, which was achieved by using quantile regression to determine $u_s(t)$. With Eq. (3.1) we can compute for each site $s$ and day $t$ the value $r_{s,t}(\alpha)$, that is exceeded on average $\alpha$ times in a season:

$$ r_{s,t}(\alpha) = \begin{cases} 
  u_s(t) \left( 1 - \frac{\gamma(t)}{\xi(t)} \left[ 1 - \left( \frac{\lambda}{\alpha} \right) \xi(t) \right] \right), & \xi(t) \neq 0, \\
  u_s(t) \left( 1 + \gamma(t) \ln \left( \frac{\lambda}{\alpha} \right) \right), & \xi(t) = 0. 
\end{cases} \tag{3.2} $$

In analogy with a stationary setting, the quantity $r_{s,t}(\alpha)$ is termed the $1/\alpha$-year return level, although $1/\alpha$ no longer gives the expected waiting time between exceedances of $r_{s,t}(\alpha)$.

3.2.4 Determination of the threshold

The non-stationary threshold is estimated via quantile regression. However, we have to select an appropriate quantile, i.e. the value of the threshold has to be high enough to justify the GPD assumption.

Quantile selection

The Threshold Stability (TS) plot and the ME plot (also referred to as mean residual life plot) are widely used graphical tools for the selection of the threshold in the POT analysis. The TS plot is based on the fact, that once the GPD model holds at grid point $s$ for some threshold $u_0$ it holds for every threshold $u_s \geq u_0$. In particular we have that the associated shape parameter is the same for $u_s$ and $u_0$. This property is called threshold stability. The threshold stability can be exploited by estimating the shape parameter for a range of high thresholds and plotting

$$ (u_s, \hat{\xi}_s(u_s))_{u_s \geq u_0} $$

where $\hat{\xi}_s(u_s)$ denotes the estimated shape parameter for threshold $u_s$. If the GPD model holds for $u_0$ the graph should be (approximately) constant for $u_s \geq u_0$. However, owing to the decreasing number of excesses above higher thresholds, the plot becomes unstable and the constant behavior is difficult to see. By combining different diagnostics one hopes for a better picture of the threshold to be used. In this chapter we, therefore, consider also the ME plot.

The ME plot relies on a similar consideration and utilizes the ME function

$$ e_s(u) := E[X_s - u | X_s > u] $$
where $X_s$ represents the daily rainfall at site $s$. The empirical version of the ME function is given by

$$
\hat{e}_s(u) = \frac{\sum_{t=1}^{T} (x_s(t) - u) I_{(u, \infty)}(x_s(t))}{\sum_{t=1}^{T} I_{(u, \infty)}(x_s(t))},
$$

where $I_A$ is the indicator function for set $A$, i.e. $I_A(x) = 1$ if $x \in A$ and otherwise zero. If the GPD model holds, $e_s(u)$ is linear in the threshold $u$, see e.g. Embrechts et al. (1997), and $\hat{e}_s(u)$ becomes approximately linear. However, as described in greater detail in Ghosh and Resnick (2010), there are some problems associated with the use of the ME plot. These are in particular that the ME function is only well defined for $\xi < 1$, and – as for the TS plot – the empirical ME function becomes unstable for high values of the threshold, see also the comments of Dr. Kimber in the discussion of Davison and Smith (1990). In our specific application the condition $\xi < 1$ is no restriction, as even the highest estimates of $\xi$ for daily extreme precipitation are far below 1, see e.g. Martins and Stedinger (2000), and Papalexiou and Koutsoyiannis (2013).

We want to select $\tau_0 \in [0, 1)$, such that the GPD model is valid above the $\tau_0$ quantile at every grid point. We denote by $q_s(\tau) := F_{X_s}^{-1}(\tau)$ the $\tau$ quantile of the data at site $s$, by $\hat{\xi}_s(\tau)$ the corresponding estimate of the shape parameter and by $\hat{e}_s(\tau)$ the corresponding value of the empirical ME function. We assume, there exists for each $s$ a $\tau_0^s$, such that the GPD model holds for all excesses $X_s - q_s(\tau_0^s)$. Then we define $\tau_0 := \max_s \tau_0^s$ and it is clear that the GPD model holds for $q_s(\tau)$ for every grid point $s \in \{1, \ldots, S\}$ and $\tau > \tau_0$, i.e. $\hat{\xi}_s(\tau)$ should be approximately constant and $\hat{e}_s(\tau)$ should be approximately linear for $\tau > \tau_0$. Concerning the instabilities of the TS and ME plots, the constant (respectively linear) behavior might not be apparent for every single grid point $s$. Therefore, we propose to consider spatial averaging of the TS (respectively ME) plot to reveal the underlying structure better. The spatially averaged TS plot is given by

$$
\left(\bar{q}(\tau), \bar{\xi}(\tau)\right)_{\tau \in [0,1)}, \quad (3.3)
$$

where $\bar{q}(\tau) = S^{-1} \sum_{s=1}^{S} q_s(\tau)$ and $\bar{\xi}(\tau) = S^{-1} \sum_{s=1}^{S} \hat{\xi}_s(\tau)$. The spatially averaged ME plot is given by

$$
\left(\bar{q}(\tau), \bar{e}(\tau)\right)_{\tau \in [0,1)}, \quad (3.4)
$$

where $\bar{e} = S^{-1} \sum_{s=1}^{S} \hat{e}_s(\tau)$. Note that while the spatially averaged TS plot works if the shape parameter is site-specific, the spatially averaged ME plot requires a common
3.2. METHODS

shape parameter, see also section 4.4. The strength of the spatially averaged plot lies in the increased detection probability of non-constant (respectively non-linear) behavior, when the threshold is too low.

Quantile regression

Quantile regression relies on the fact that a sample quantile can be viewed as the solution of an optimization problem, which can be computed efficiently using linear programming, as shown in Koenker and Bassett (1978). For a fixed site \( s \in \{1, \ldots, S\} \), we can obtain the \( \tau \)-th sample quantile of the declustered observations \( x_s = (x_s(1), \ldots, x_s(T)) \) as

\[
\arg \min_{\beta \in \mathbb{R}} \sum_{t=1}^{T} \rho_{\tau}(x_s(t) - \beta),
\]

where

\[
\rho_{\tau}(v) = \begin{cases} 
  v(\tau - 1), & v < 0, \\
  v\tau, & v \geq 0.
\end{cases}
\]

This can be easily generalized, by replacing \( \beta \) by a suitable regression model, e.g. in order to obtain for each site a linear trend in the temporal covariate \( z(t) \), we determine

\[
\arg \min_{\beta_0^s, \beta_1^s \in \mathbb{R}} \sum_{s=1}^{S} \sum_{t=1}^{T} \rho_{\tau}(x_s(t) - \beta_0^s - \beta_1^s \cdot z(t)),
\]

which can be done for each site separately. It may be useful to assume a common relative trend over the region, compare Hanel et al. (2009) who use a common relative trend in the location parameter of the Generalized Extreme Value (GEV) distribution. Then, we have to determine

\[
\arg \min_{\beta_0^0, \beta_1^1 \in \mathbb{R}} \sum_{s=1}^{S} \sum_{t=1}^{T} \rho_{\tau}\left(x_s(t) - \beta_0^s [1 + \beta_1^1 \cdot z(t)]\right),
\]

where \( \beta_0^s \) is a site specific niveau component and \( \beta_1^1 \) a common relative trend. This common relative trend can be found by profile quantile regression, i.e. we compute

\[
\sum_{s=1}^{S} \min_{\beta_0^s} \sum_{t=1}^{T} \rho_{\tau}\left(x_s(t) - \beta_0^s [1 + \beta_1^1 \cdot z(t)]\right)
\]

over a grid of possible values of \( \beta_1^1 \) and select the one, that minimizes the sum on the right side. The \( \beta_0^s \) are determined for each fixed \( \beta_1^1 \) as the solutions of the indepen-
dent minimization problems

\[
\min_{\beta^0 \in \mathbb{R}} \sum_{t=1}^{T} \rho_t \left( \frac{x_s(t)}{1 + \beta^1 \cdot z(t)} - \beta^0 \right).
\]

We want to test, whether \( \beta^1 \) differs significantly from zero. However, the distribution of \( \beta^1 \) under the null hypothesis, i.e. that there is no trend, is not known. Therefore, we calculate \( p \)-values by a block-wise bootstrap approach (compare e.g. Douglas et al. 2000). Seasonal blocks over the whole spatial domain are sampled with replacement, therefore the newly created data should have approximately the same spatial dependence structure as the original data but no trend component. Calculating \( \beta^1 \) multiple times for these bootstrap data delivers an approximate distribution of \( \beta^1 \) under the null hypothesis.

### 3.2.5 Estimating the excess distribution

For a specific model of the GPD parameters, e.g.

\[
\gamma(t) = \gamma_0 + \gamma_1 \cdot z(t), \quad \xi(t) = \xi_0, \tag{3.8}
\]

and fixed threshold \((u_s(t))_{s \in \{1, \ldots, S\}, t \in \{1, \ldots, T\}\)} we estimate the vector of parameters \( \theta = (\gamma_0, \gamma_1, \xi_0)' \) by maximizing the following function:

\[
\ell_1(\theta; y) = -\sum_{t=1}^{T} \sum_{s=1}^{S} \left[ \ln(\gamma(t)u_s(t)) + \frac{1 + \xi(t)/\xi(t)}{\xi(t)} \ln(1 + \frac{\xi(t)y_s(t)}{\gamma(t)u_s(t)}) \right]. \tag{3.9}
\]

This is the so-called independence log likelihood, i.e. the log likelihood that would be obtained, if peaks at different sites were independent of each other, see also Moore (1987), Smith (1989a), Buishand (1991), and Cooley et al. (2007), and Hanel et al. (2009). We rely on this simplified likelihood function, because the estimation of the full likelihood function would be virtually impossible, due to spatial dependence and large dimensionality, compare also Thibaud et al. (2013). This method provides asymptotically unbiased parameter estimates, but the spatial dependence in the data results in a (highly) increased variance of the estimates compared to the variance that would be obtained for independent data. Therefore, Smith (1990b) suggested to adjust the standard errors and likelihood ratio tests in a way that is now generalized in the composite likelihood framework, see Varin et al. (2011) for an extensive overview. For applications of this approach, see e.g. Blanchet and Lehning (2010) for annual maximum snow depths over Switzerland and Van de Vyver (2012) for annual extremes of precipitation in Belgium. We refer to Chapter 2 for the details of
the estimation of the GPD parameters.

### 3.2.6 Bias correction

Climate models represent the current status of knowledge about the climate system but are imperfect still. Systematic differences occur between climate model data and observations, which translate into biases in derived quantities like return levels of daily precipitation. There are different ways to correct for these biases. In the discussion below, we will use the following notation. By superscript ‘Mod’ we denote all entities, such as parameters and return level, that are derived from the climate model data and by superscript ‘Obs’ those from the observations. Further, we denote by $T_C$ the end of the overlapping or control period.

A simple way to adjust the simulated return levels, in the spirit of quantile matching, is to apply a return period specific change factor:

$$
\tilde{r}_{s,t}(\alpha) = r_{s,t}^{\text{Mod}}(\alpha) \cdot \frac{\sum_{i=1}^{T_C} r_{s,i}^{\text{Obs}}(\alpha)}{\sum_{i=1}^{T_C} r_{s,i}^{\text{Mod}}(\alpha)}. \tag{3.10}
$$

A drawback of this approach is that in general the adjusted quantiles will no longer follow a GPD. It is even possible that they do not increase monotonically with increasing return period. This can be overcome if we adjust the GPD parameters and the threshold instead, i.e.

$$
\tilde{\xi}(t) = \xi^{\text{Mod}}(t) - \frac{1}{T_C} \sum_{i=1}^{T_C} (\xi^{\text{Obs}}(i) - \xi^{\text{Mod}}(i)),
$$

$$
\tilde{\gamma}(t) = \gamma^{\text{Mod}}(t) \cdot \frac{\sum_{i=1}^{T_C} \gamma^{\text{Obs}}(i)}{\sum_{i=1}^{T_C} \gamma^{\text{Mod}}(i)}, \tag{3.11}
$$

$$
\tilde{u}_s(t) = u_s^{\text{Mod}}(t) \cdot \frac{\sum_{i=1}^{T_C} u_s^{\text{Obs}}(i)}{\sum_{i=1}^{T_C} u_s^{\text{Mod}}(i)}.
$$

For the shape parameter an additive adjustment was chosen owing to the small and possibly negative values of this parameter. Using the adjusted parameters in Eq. (3.2), we get an estimate of $r_{s,t}(\alpha)$. The proposed correction yields, that the temporal mean of the adjusted return levels over the control period approximates the temporal mean of the estimated return levels for the observational data. The approach can be generalized by taking not only the biases in the means of the threshold and the GPD parameters into account but also the biases in their trends. However, one should be very careful doing so as it might lead to implausible effects owing to the many parameters involved.
CHAPTER 3. PROJECTIONS OF PRECIPITATION EXTREMES

The uncertainty in the projected return levels, due to the unknown GPD parameters, can be assessed by the following procedure. First we generate bootstrap samples of the estimated GPD parameters for both the observations and the climate model data, e.g. by the method used in Chapter 2, based on exponential residuals of the excesses. From these bootstrap samples a bootstrap sample of the adjusted return level is obtained, by applying correction scheme (3.11) to the estimated shape and dispersion parameters in the bootstrap samples from the climate model data and observations. Note that for long return periods the uncertainty of the threshold can be neglected, but this cannot be done for short return periods. Then one must generate bootstrap samples from all days in the season of interest rather than from days with threshold exceedances only.

3.3 Data

3.3.1 Region and precipitation data

The spatial domain of the study includes the Netherlands and a part of northwestern Germany, roughly northwest of the line Cologne–Hamburg. The region is relatively flat with hills up to 300 m in the extreme south and south-east. The averaged elevation is for about 60% of the grid boxes lower than 25 m and only in less than 5% of the grid boxes higher than 100 m. Fig. 3.1 shows the observed mean annual precipitation totals for the considered grid points, which range from about 650 mm to 850 mm, without a clear spatial pattern. Precipitation data have been provided on a $0.22^\circ \times 0.22^\circ$ rotated pole grid and a total of 158 grid points falls into the specified region.

For this region we consider two transient simulations from the RACMO2 RCM of the Royal Netherlands Meteorological Institute (Van Meijgaard et al. 2008) from 1950 to 2100. The first one was driven by the ECHAM5 GCM, developed at the Max Planck Institute for Meteorology in Hamburg (Roeckner et al. 2003), and the second was driven by the GCM Model for Interdisciplinary Research on Climate (MIROC) (K-1 Model Developers 2004). In the following these two simulations are denoted as R/ECHAM5 and R/MIROC respectively. We selected these simulations for the present case study, because their precipitation projections are quite different, although both simulations are based on the SRES A1B emission scenario (IPCC 2000).

We compare the climate model data with gridded observations from the E-OBS data set (Haylock et al. 2008), version 6.0, which is available on the same grid. With 3.1 stations on average per grid box (in total 492) the station density in this part of the E-OBS domain is quite high. Therefore, the data are considered to be adequate for comparison with gridded RCM data. Table 3.1 contrasts the daily regional mean
winter (December - February) and summer (June - August) precipitation for the climate model data with those for the observations during the control period 1950–2010. For the climate model data also the means for the future period 2071–2100 and the associated change are given. The R/ECHAM5 simulation shows a clear increase in winter and decrease in summer, which are both significant at the 1% level according to a two sample \( t \)-test based on the seasonal totals. For R/MIROC the changes in mean winter and summer precipitation are small. It is noteworthy, that the climate simulations do not preserve the seasonal cycle: for the observations the mean precipitation is higher in summer than in winter, but for the climate model data the opposite is found.

### 3.3.2 Temperature as covariate

The enhanced greenhouse gas effect is anticipated to be small or not existent in the first decades of the simulation and increasing by the end of the 20th century. This is inconsistent with a simple linear trend over time for the threshold and GPD parameters. Instead of applying more complicated relationships with time, leading to increased estimation uncertainty, a covariate that is considered representative of the enhanced greenhouse gas effect is used.

With rising temperatures the water holding capacity of the atmosphere increases.
Table 3.1: Seasonal mean precipitation in mm / day.

<table>
<thead>
<tr>
<th>Season</th>
<th>Data</th>
<th>1950 - 2011</th>
<th>2071-2100</th>
<th>Change (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter</td>
<td>E-OBS</td>
<td>2.08</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>R/ECHAM5</td>
<td>2.98</td>
<td>3.55</td>
<td>19.1</td>
</tr>
<tr>
<td></td>
<td>R/MIROC</td>
<td>2.48</td>
<td>2.52</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>E-OBS</td>
<td>2.46</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Summer</td>
<td>R/ECHAM5</td>
<td>2.70</td>
<td>2.27</td>
<td>-15.9</td>
</tr>
<tr>
<td></td>
<td>R/MIROC</td>
<td>2.10</td>
<td>2.17</td>
<td>3.3</td>
</tr>
</tbody>
</table>

As extreme precipitation strongly depends on the available precipitable water, compare e.g. Lenderink and Van Meijgaard (2008), temperature is a natural covariate for the non-stationary POT approach. Hanel et al. (2009) used the seasonal global temperature anomaly, from the driving GCM, as covariate. However, a prominent property of the standard time covariate is that trends over time are easily comparable between different models. In order to maintain this property, and in view of the common driving scenario, we consider a common, regional temperature covariate.

First we determine for each year the seasonal mean temperature in the R/ECHAM5 and R/MIROC data, aggregated over all grid boxes in the RACMO2 domain with at least 50% land coverage. These seasonal mean temperatures were averaged for the two climate simulations and the anomalies were then computed.
3.4. RESULTS AND DISCUSSION

with respect to the control period. The temperature anomalies are decomposed into a short-range dependent component and an increasing trend component using monotone regression (Wu et al. 2001). For an introduction to monotone regression, see e.g. Robertson et al. (1988). In this chapter the monotone median algorithm of Koenker and Ng (2005) is used, and the fitted values constitute the common covariate $z(t)$. Fig. 3.2 shows $z(t)$ for the winter and the summer season. Note the breaks in winter around 1960 and 2020, and in summer around 1990 and 2055. These are owing to the piecewise linear fit in the regression setting and do not have a specific physical meaning. The fitted curves are assumed to be representative of the enhanced greenhouse effect, as projected in the two climate simulations.

Over the control period the observed trend in temperature is larger than the (averaged) trend of the climate simulations in this region. Van Oldenborgh et al. (2009) explored several explanations for the stronger observed than modeled warming in western Europe, the most important being a stronger trend to westerly circulation in winter and a stronger trend towards more short-wave radiation in summer than simulated by the climate models. The pronounced trend in observed temperature is accounted for in an alternative covariate $z_{E,OBS}(t)$, based on a monotone regression fit to the observed seasonal temperature anomalies.

3.4 Results and discussion

3.4.1 Temporal dependence and declustering

Prior to the analysis of the extremes in the precipitation data, we apply the declustering scheme, outlined in section 3.2.2. In winter we observe large numbers of clusters of lengths 2 and 3 of the exceedances of the 95% quantile, both in the climate model data and the observations, indicating a considerable amount of temporal dependence. As the number of clusters of length 4 is less than one per decade, we choose in winter a separation time of 2 days. Consistent with the literature, compare e.g. Kyselý and Beranová (2009), the temporal dependence in summer is much weaker, manifesting in a small number of clusters of length 3. Therefore, a separation time of 1 day is taken as sufficient for the summer season.

3.4.2 Trend in the threshold

Quantile selection

We illustrate the quantile selection approach proposed in section 3.2.4, using the winter data from the R/MIROC simulation. From the spatially averaged TS and ME plots (Fig. 3.3 and Fig. 3.4), it can be seen that the 95% quantile is too low for the GPD
model to hold, i.e. the spatially averaged estimates of the shape parameter to the right of the spatially averaged 95% quantile in the TS plot are neither constant nor is the ME plot linear. The 97.5% quantile seems to be high enough for the GPD model to hold. Similar pictures for the winter data from the ECHAM5 driven simulation and the observations were obtained. Therefore, we took the 97.5% quantile as threshold for all winter data. For the summer data it was necessary to reject also the 97.5% quantile and the 98.5% quantile was used as threshold, i.e. on average we consider 1.38 excesses per grid point and season. Note, that this is a considerably higher quantile than in many other studies (Friederichs 2010; Kyselý et al. 2010; Halmstad et al. 2012).

![Figure 3.3: Spatially averaged TS plot for R/MIROC, winter season. The dashed vertical lines mark the spatially averaged quantile for $\tau = 0.95, 0.975$ and 0.985.](image)

**Quantile regression**

For each grid point and season, we fitted the simple linear quantile regression model in Eq. (3.6) to the observed precipitation data. As predictor for the 97.5% quantile in winter (respectively the 98.5% quantile in summer) we used the regional temperature anomaly $z(t)$. This was repeated for the climate model data both for the control period and the full period. The resulting local trends (i.e. slope parameters $\beta_1$) are shown in Fig. 3.5 for winter and Fig. 3.6 for summer. In winter the observed data show rather large positive slopes, which are not found for the climate model data in the control period. The large increase in precipitation extremes is consistent with the observed increase in 10-day precipitation maxima in this part of Europe. Van Haren
et al. (2013a) attribute part of the latter trend to the change in circulation, which explains also partly the relatively large increase in mean winter temperature. In summer, the climate model output exhibits more negative slopes than the observations. As the considered region is bordered by the North Sea, this might be attributed to a lower than observed trend in sea surface temperature in this coastal area in the climate models (compare Van Haren et al. 2013b).

It is striking that for both climate simulations the trends obtained for the control period do not resemble those obtained for the full period. In particular, the spread of the estimated trends over the region is much smaller for the full period than for the control period, consistent with the decreased estimation uncertainty for the full period. Moreover, there is no clear spatial pattern in the trends for the full period nor for the control period. Therefore, we assume in the following a common relative trend over the region, i.e. we fit the common relative trend model, given in Eq. (3.7), to the daily precipitation data. Table 3.2 shows, that the climate simulations give quite different trends for the control period, in summer even with opposite sign, but for the full period the trends are much closer. This is consistent with the results for the local trends. The values for the observations are quite large. This is partly caused by the fact that the temperature covariate $z(t)$ underestimates the temperature trend in the observations (section 3.3.2). If we replace the covariate $z(t)$ by $z_{E-OBS}(t)$ we obtain an increase of 3.7% per degree Celsius change for the observed winter data and 2.8% in summer, resembling the estimates from the climate simulations for the full period.

We computed $p$-values for the common trend by the block-wise bootstrap, out-
Table 3.2: Change in the threshold (% per degree warming; bold script indicates significance at the 5% level), i.e. $\beta^1$ in the model with a common relative trend, Eq. (3.7).

<table>
<thead>
<tr>
<th>Season</th>
<th>Data</th>
<th>1950-2011</th>
<th>1950-2100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E-OBS</td>
<td>10.6</td>
<td>-</td>
</tr>
<tr>
<td>Winter</td>
<td>R/ECHAM5</td>
<td>3.8</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>R/MIROC</td>
<td>9.4</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>E-OBS</td>
<td>7.4</td>
<td>-</td>
</tr>
<tr>
<td>Summer</td>
<td>R/ECHAM5</td>
<td>-6.3</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>R/MIROC</td>
<td>8.8</td>
<td>4.6</td>
</tr>
</tbody>
</table>

lined in section 3.2.4. It turns out that the trends are significant only for the full period in winter for both models and in summer only for R/MIROC, which is consistent with the significance of the local trends, compare the number of pluses in Fig. 3.5 and Fig. 3.6. Therefore, we are considering in the following temporally varying thresholds only for those data sets, for the other data sets the thresholds are taken to be constant.
3.4. RESULTS AND DISCUSSION

Figure 3.5: Change in the threshold (mm per degree warming; upside deltas indicate positive trends significant at the 5% level) for the winter season, i.e. $\beta^1_2$ in Eq. (3.6).
Figure 3.6: Same as Fig. 3.5 but for the summer season (downside deltas indicate significant, negative trends).
3.4. RESULTS AND DISCUSSION

3.4.3 Trend in the excess distribution

For the GPD parameters it was assumed that the dispersion coefficient varies linearly with the temperature covariate and that the shape parameter is constant, see Eq. (3.8). The significance of the trend in the dispersion coefficient was determined by the composite likelihood ratio test, which is an extension of the classical likelihood ratio test, that takes the spatial dependence into account (compare Varin et al. 2011, and Chapter 2). The results are given in Table 3.3. Just as for the threshold, the control period does not provide a clear picture and the strong positive trend in the dispersion coefficient of the winter observation data vanishes, if the observed regional temperature anomalies $z_{\text{E-OBS}}$ are used as covariate. However, for the whole period both climate simulations agree reasonably well in each season. In winter, where a significant trend in the threshold was found for both models, the trend in the dispersion coefficient is negligible. In summer the dispersion coefficient is significantly increasing for R/ECHAM5 as well as for R/MIROC. Hanel and Buishand (2011) reported a similar trend in the dispersion coefficient of the 1-day summer precipitation maxima in this region for an ensemble of 15 transient RCM simulations.

Table 3.3: Slope parameter of the linear regression model for the GPD dispersion coefficient, i.e. $\gamma_1$ in Eq. (3.8), for each season (bold - significant at 5% level).

<table>
<thead>
<tr>
<th>Season</th>
<th>Data</th>
<th>1950-2011</th>
<th>1950-2100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter</td>
<td>E-OBS</td>
<td>0.0905</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>R/ECHAM5</td>
<td>-0.0093</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>R/MIROC</td>
<td>0.0324</td>
<td>0.0086</td>
</tr>
<tr>
<td>E-OBS</td>
<td></td>
<td>0.0085</td>
<td>-</td>
</tr>
<tr>
<td>Summer</td>
<td>R/ECHAM5</td>
<td>-0.0470</td>
<td>0.0258</td>
</tr>
<tr>
<td></td>
<td>R/MIROC</td>
<td>0.0247</td>
<td>0.0400</td>
</tr>
</tbody>
</table>

The need to include a trend in the dispersion coefficient can be demonstrated with the standard exponential residuals of the model fit, compare Eq. (2.24). Fig. 3.7 shows these transformed residuals for the R/MIROC summer season obtained using the model without and with trend in the dispersion coefficient. Whereas the former exhibit a highly significant linear trend, the linear trend in the latter is close to zero and there is no indication for other trends.

Table 3.4 shows for both climate simulations and the observations the shape parameter $\xi$, the temporally averaged dispersion coefficient $\gamma$, and the temporally and regionally averaged threshold $u$. While the shape parameters and the threshold values from the climate simulations are comparable to the observed ones, the dispersion
coefficients from the climate model data are too low in winter and too high in summer. Note that the shape parameter for the E-OBS summer season almost equals the constant 0.15, found by Koutsoyiannis (2004b) for annual BM of daily precipitation in different parts of the world.

Table 3.4: Mean threshold $u$ (mm), dispersion coefficient $\gamma$, and shape parameter $\xi$ for both seasons.

<table>
<thead>
<tr>
<th>Season</th>
<th>Data</th>
<th>1950 - 2011</th>
<th>1950-2100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$u$</td>
<td>$\gamma$</td>
<td>$\xi$</td>
</tr>
<tr>
<td>Winter</td>
<td>E-OBS</td>
<td>11.8</td>
<td>0.417</td>
</tr>
<tr>
<td></td>
<td>R/ECHAM5</td>
<td>14.0</td>
<td>0.303</td>
</tr>
<tr>
<td></td>
<td>R/MIROC</td>
<td>11.5</td>
<td>0.348</td>
</tr>
<tr>
<td>Summer</td>
<td>E-OBS</td>
<td>17.9</td>
<td>0.321</td>
</tr>
<tr>
<td></td>
<td>R/ECHAM5</td>
<td>17.5</td>
<td>0.413</td>
</tr>
<tr>
<td></td>
<td>R/MIROC</td>
<td>15.9</td>
<td>0.506</td>
</tr>
</tbody>
</table>
3.4. RESULTS AND DISCUSSION

3.4.4 Return levels

For the selected threshold and the estimated parameters we can compute a time-dependent return level using Eq. (3.2). As an alternative to the non-stationary POT approach, we consider a ‘time slice’ approach. Therefore, we estimate for each 30-year window of the data (i.e. 1950–1979, 1951–1980, . . .) a common shape parameter and common dispersion coefficient as if the data were stationary. Then the return levels are computed based on these estimates. For the MIROC driven simulation Fig. 3.8 shows, for both approaches, the 50-year return level of daily precipitation in the summer season at the grid point closest to De Bilt, see Fig. 2.6. The confidence bands are obtained using the asymptotic normality of the Maximum Independence Likelihood Estimator (MILE) of the GPD parameters (Davison 2003; Varin et al. 2011). This ignores the uncertainty in the threshold, which is small for the 50-year return level compared to the uncertainty due to the GPD parameters. Overall, the figure shows a good agreement between both methods. However, the 95% confidence band for the non-stationary POT approach is considerably narrower than that for the time-slice approach, owing to the increased number of data points used for the estimation. In fact, the relative standard error reduces from about 8.7% for the moving window approach to about 4.5% for the non-stationary approach. Moreover, a monotone trend

Figure 3.8: 50-year summer return level, r(0.02), at the grid point closest to De Bilt for R/MIROC with 95% confidence bands. The solid blue line (respectively blue band) is based on regional 30-year time slices and the solid red line (respectively red band) is based on the non-stationary approach. The gray shadow indicates the 95% confidence band based on the non-stationary at-site model.
in the return level is more plausible than the irregular pattern of the trend for the time-slice approach, where slightly different selections of the windows can produce quite different estimates of the change, e.g. it matters a lot if the period 2063–2092 \( r(0.02) = 85 \text{ mm} \) or the period 2071–2100 \( r(0.02) = 99 \text{ mm} \) is taken as future period. This is even more delicate because the control period can be chosen in different ways too. Fig. 3.8 also shows the confidence band if only at-site data are used to estimate the parameters of the non-stationary model. These are four times wider than those obtained by the regional estimation approach.

The 50-year return level of the 1-day summer maximum precipitation near De Bilt from the MIROC driven simulation in Fig. 3.8 is significantly larger than the estimate of 52 mm from the observations. Therefore, a bias correction is needed. Bias correction can be very sensitive to trends, e.g. when they are close to zero but of different signs. Thus, we take only significant trends in the threshold and GPD parameters into account. Note that the proposed correction scheme, see section 3.2.6, assumes that the bias is constant over time. Maraun (2012) finds this justified for seasonal precipitation sums over most of Europe and in particular over the study area considered here. A small experiment, splitting the historical period in two subperiods, shows that bias-corrected return levels are preferable to the uncorrected model output, see Appendix A.2.

Fig. 3.9 shows the bias-corrected 50-year return level for the summer season for both the R/ECHAM5 and R/MIROC data, based on the adjusted threshold and GPD parameters. The results for winter are shown in Fig. 3.10. As the trend in the control period is negligible for both models and seasons, the bias correction removes almost the whole bias, which would not be the case if the trend in the observations was significant. We see that in summer R/MIROC projects a much stronger increase (45\%) than R/ECHAM5 (15\%), while the situation in winter is the opposite. Then, R/MIROC projects a 11\% increase in extreme precipitation, and R/ECHAM5 a 22\% increase. The latter is about the same as the trend in the mean, see Table 3.1. In winter similar changes in mean and extreme precipitation have been reported for other RCM simulations in parts of Europe (Frei et al. 2006; Kyselý et al. 2011; Hanel and Buishand 2012). However, in summer the trend in mean precipitation does not give a good indication of the trend in extremes.

In Fig. 3.10 one can see that the width of the confidence band for the return level is the same for both climate model simulations and that it does not vary over time. This is owing to the bias-correction scheme. As there is no trend in the GPD parameters of the climate model data in the winter season, the bias correction always yields the values of the (bootstrapped) GPD parameters from the observations. Thus, the width of the confidence band is determined by the variance of these parameter estimates. This implies that only longer observation records can reduce the uncertainty. However, in summer the trend in the dispersion coefficient is significant for the cli-
3.4. RESULTS AND DISCUSSION

Figure 3.9: Bias-corrected 50-year summer return level, $r(0.02)$, at the grid box closest to De Bilt for R/MIROC (blue) and R/ECHAM5 (red) with 95% confidence bands obtained by the bootstrap procedure, outlined in section 3.2.6. The constant 50-year return level of the observations is given as reference (dashed line).

Figure 3.10: Same as Fig. 3.9 but for the winter season.

mate model data. Including this trend in the POT model increases the width of the confidence band towards the end of the 21st century, as shown in Fig. 3.9.

Return levels of precipitation are often the basis of hydrologic design. In a stationary climate the risk that a system fails within its expected life time, i.e. the en-
counter probability, is directly related to the return period, e.g. the probability that a 50-year return level is exceeded at least once during 50 years is 64%. This holds no longer for a non-stationary climate. Rootzén and Katz (2013) recently proposed a risk-based design level, that can be used both for a stationary and a non-stationary climate. The design life level for a specified period gives the value, that is exceeded with probability $p$ during the design life period. Table 3.5 shows the 10% and 5% design life level for De Bilt, computed for two different 50-year design life periods, using the bias-corrected GPD parameters for the summer season. For both climate simulations a significant increase is found for the future period 2016-2065. The 95% confidence interval for the change in design life level is based on the same bootstrap samples as those used to generate the confidence band for the 50-year return level in Fig. 3.9. The relative changes for $p = 0.10$ and $p = 0.05$ are almost identical. The change in the design life level for the ECHAM5 driven simulation is half the size of the change in the MIROC driven simulation, similar to the difference in the change in the summer 50-year return level, see Fig. 3.9. However, the change in the design life level can be compared only roughly with that in the return level, as it takes the whole design life period into account rather than only two points in time. A caveat is, that the 5% design life level corresponds to a very rare rainfall amount. In a stationary climate the associated return period to this design life level is 1000 (exactly 975) years, which is on the edge of reasonable extrapolation. In general, for a stationary climate, the return period corresponding to the $p$ design life level for a design life period of $n$ years, follows from

$$
R = \frac{1}{1 - \sqrt{1 - p}} \approx \frac{1}{1 - (1 - \frac{p}{n})} = \frac{n}{p}.
$$

\begin{table}[h]
\centering
\begin{tabular}{lllll}
\hline
$p$ & Data & 1950 -1999 & 2016 - 2065 & Change (%) \\
\hline
& E-OBS & 82 & - & - \\
0.10 & R/ECHAM5 & 82 & 92 & 12.9 (4.3, 22.6) \\
& R/MIROC & 81 & 104 & 27.9 (19.8, 36.7) \\
& E-OBS & 95 & - & - \\
0.05 & R/ECHAM5 & 94 & 106 & 13.4 (4.4, 23.5) \\
& R/MIROC & 93 & 120 & 28.7 (20.0, 37.6) \\
\hline
\end{tabular}
\caption{Design life level (mm) for a 50-year life time at De Bilt. The values for R/ECHAM5 and R/MIROC are based on the bias-corrected GPD parameters for the summer season. The values in parentheses denote the 95% confidence interval for the percentage change as obtained by the bootstrap procedure, outlined in section 3.2.6.}
\end{table}
3.5 Conclusion

In this chapter we applied the regional non-stationary POT model, introduced in Chapter 2, to precipitation extremes of two transient climate simulations for the period 1950-2100, conducted with the RCM RACMO2, driven by the GCMs ECHAM5 and MIROC. The simulated 1-day summer and winter precipitation extremes in the Netherlands and north-western Germany were compared with those in the gridded observation data set E-OBS for the period 1950-2011.

Visual inspection of the spatially averaged TS plot and the spatially averaged ME plot leads to a rejection of the 95% quantile as threshold, which is often used in the literature for precipitation data. It may be useful to develop regional Goodness of Fit (GoF) tests to make the selection of the threshold more objectively. A regional quantile regression model using a common relative trend for the threshold was used to smooth the large spatial scattering of the local trends in the threshold for the control period. The absence of a spatial pattern in the trends for both the control and full period justifies this model. The non-stationary approach leads to return levels, that are consistent with those obtained by a 30-year moving window approach, but exhibit less uncertainty. For the considered 50-year return level the uncertainty is reduced by a factor of two, compared to the moving window approach. The simulated return levels exhibit a considerable positive bias, which was corrected for by adjusting the parameters of the POT model. The uncertainty in the adjusted return levels is then strongly governed by the variance of the estimated GPD parameters from the observations. Therefore, accurate estimates of the GPD parameters, based on high quality observed precipitation records, are needed to project future extremes.

For the winter season, the two climate model simulations project a significant trend in the threshold. This is in line with a positive trend in the location parameter of the GEV distribution fitted to the 5-day winter precipitation maxima in this region for an ensemble of 15 transient RCM simulations (Hanel and Buishand 2011). The difference between the R/ECHAM5 and R/MIROC simulations does not give the full range of possible future projections. In summer the differences between the climate model data are larger. The ECHAM5 driven simulation projects a 15% increase of the 50-year return level, based on a significant increase of the dispersion coefficient, which corresponds well to the increase in the dispersion coefficient of the GEV distribution found in Hanel and Buishand (2011). The MIROC driven simulation projects additionally a significant increase of the threshold, resulting in a 45% increase of the 50-year return level.
Acknowledgments

The R/ECHAM5 simulation and the E-OBS data set were partially funded by the EU FP6 Integrated Project ENSEMBLES (Contract number 505539). The R/MIROC simulation was kindly made available by G. Lenderink and E. van Meijgaard. We are grateful to two anonymous reviewers for their helpful comments on an earlier version of this chapter.
Regional threshold selection

Abstract. A hurdle in the Peaks-Over-Threshold (POT) approach for analyzing extreme values is the selection of the threshold. A method is developed to reduce this obstacle in the presence of multiple, similar data samples. This is the case, for instance, in many environmental applications. The idea is to combine univariate threshold selection methods into a regional method. Regionalized versions of the Threshold Stability (TS) and the Mean Excess (ME) plot are presented as graphical tools for threshold selection. Moreover, quantitative approaches based on the bootstrap distribution of the spatially averaged Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) test statistics are introduced. It is demonstrated that the proposed regional method leads to an increased sensitivity for too low thresholds, compared to methods that do not take into account the regional information. The approach can be used for a wide range of univariate threshold selection methods. We test the methods using simulated data and present an application to rainfall data from the Dutch water board Vallei en Veluwe.

CHAPTER 4. REGIONAL THRESHOLD SELECTION

4.1 Introduction

Extreme events, like financial shocks, storms or flooding, often affect society disastrously. In order to take the right measures for their regulation, e.g. by building a dyke of appropriate height, gaining inference about these extremes is crucial. In particular it might be important to predict which level of future extremes is to be expected. This can be analyzed by Extreme Value Theory (EVT), which allows extrapolation beyond the range of observed data, based on asymptotically justified models (Coles 2001; Wadsworth and Tawn 2012).

Two common approaches for the study of extremes of independent and identically distributed (i.i.d.) random variables \( X(t) \) \((t = 1, \ldots, T)\) are the Block Maxima (BM) and the Peaks-Over-Threshold (POT) approach, see e.g. Coles (2001). The BM approach considers the maxima of disjoint, equally-sized blocks of the data, which are then usually modeled by a Generalized Extreme Value (GEV) distribution. This is justified by the three types theorem, see e.g. Embrechts et al. (1997). The POT approach on the other hand considers all excesses \( Y(t) = X(t) - u \) over a previously chosen (high) threshold \( u \). For \( x \geq u \) we have, writing \( y = x - u \),

\[
P(X > x) = P(X > y + u) = P(X > u) \cdot P(X > y + u | X > u) = P(X > u) \cdot P(Y > y | Y > 0).
\]

The Pickands–Balkema–De Haan theorem states that the conditional distribution of the excesses \( Y \), can be approximated by a Generalized Pareto Distribution (GPD), if the threshold \( u \) is sufficiently high and certain regularity conditions hold, see e.g. Reiss and Thomas (2007):

\[
P(Y \leq y | Y \geq 0) \approx G_{\xi, \sigma}(y) = \begin{cases} 
1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}, & \xi \neq 0, \\
1 - \exp\left(-\frac{y}{\sigma}\right), & \xi = 0,
\end{cases}
\]

for \( y \geq 0 \) if \( \xi \geq 0 \) and \( 0 \leq y \leq -\sigma/\xi \) if \( \xi < 0 \), where \( \sigma \) and \( \xi \) are the scale and the shape parameter. For \( \xi = 0 \) the GPD reduces to an exponential distribution.

From a theoretical point of view, the POT approach is often preferred over the BM approach, owing to the possibility to consider multiple extremes in a block, and the corresponding reduction in estimation uncertainty. However, in practice where i.i.d. data are the exception, the POT approach is still less common than the BM approach. A reason for this is that the choice of the block size in the BM approach is often quite natural, e.g. a year in many environmental applications, and BM are likely to be independent. For the POT approach in contrast, there is still no standard procedure for the selection of the sufficiently high threshold \( u \), although it is of crucial importance (Scarrott and MacDonald 2012). Moreover, one has to ensure that the
peaks are independent, which is often achieved by declustering methods.

Threshold selection constitutes a trade-off situation between bias and variance. If the selected threshold is too low, the GPD approximation is poor. This causes bias in the estimated return level. However, with increasing thresholds the variance in the estimated GPD parameters increases, due to the small number of observations effectively used.

The aim of the present chapter is to guide threshold selection in the presence of multiple similar data samples. In this context Regional Frequency Analysis (RFA) has been applied to reduce the variance in the parameter estimates. One commonly used variant of RFA is the Index Flood (IF) method, which assumes that all site-specific distributions of the quantity of interest, i.e. maxima or exceedances, are the same after scaling by an index variable (Hosking and Wallis 1997). The IF method has been applied mostly in combination with BM (Fowler et al. 2005; Hanel et al. 2009; Svensson and Jones 2010), but there are several studies in combination with POT data (Madsen and Rosbjerg 1997; Weiss and Bernardara 2013, compare also Chapter 2 and 3) as well. However, the use of multiple similar sites for the threshold selection itself is mentioned for the first time in Chapter 3.

In section 4.2 we explain the underlying assumptions of the regional POT model. In section 4.3 we present a number of the available univariate threshold selection procedures. Based on these, new regional threshold selection approaches are introduced in section 4.4. These approaches are tested in a simulation setting in section 4.5, using a new way to construct smooth mixture densities with GPD tail. In section 4.6 we apply the proposed methods to high quality precipitation measurements over the area of the Vallei en Veluwe water board in the Netherlands. The chapter ends with the conclusion in section 4.7.

4.2 Preliminaries

We represent the spatio-temporal data by the matrix \((X_s(t))_{s,t}\), where \(s \in \{1, \ldots, S\}\) denotes the site and \(t \in \{1, \ldots, T\}\) the time. For each site \(s\) the data \(X_s(t)\), \(1 \leq t \leq T\) are distributed according to an unknown marginal cumulative distribution function (cdf) \(F_s\). We assume that for all sites \(s\) and for some small common exceedance probability \(\xi^0 = 1 - \tau^0\), the distribution of the excesses over the associated quantile \(u^0_s = F_s^{-1}(\tau^0)\) can be approximated by a GPD with scale parameter \(\sigma_s\) and shape parameter \(\xi_s\), i.e.

\[
P \left( X_s - u^0_s \leq y \mid X_s \geq u^0_s \right) \approx G_{\xi_s,\sigma_s}(y), \quad \forall s \in \{1, \ldots, S\}. \tag{H1} \]
Moreover, the following assumptions will be used regularly:

\[ \xi_s \equiv \xi \quad \text{(H2)} \]
\[ \frac{\sigma_s}{\mu_s} \equiv \gamma \quad \text{(H3)} \]

where \( \gamma \) denotes the dispersion coefficient. The assumption (H1) ensures that about the same amount of data is used for each location, which is always fulfilled in RFA of BM data. It is often used implicitly, whenever multiple POT time series are considered, see e.g. Ribatet et al. (2007) and Kyselý et al. (2010), but not used for threshold selection itself.

The variance of the estimated shape parameter is in general very large, therefore assumption (H2) is often used in RFA, see e.g. Hosking and Wallis (1997). Hanel et al. (2009) introduce an IF model for BM based on the corresponding assumptions (H2,H3) for the GEV distributions, i.e. a constant shape parameter and dispersion coefficient over the region. This was the basis for the IF model for POT data relying on assumptions (H1,H2,H3), introduced in Chapter 2. This model will be used also in the simulation setting, yet most of the methods developed in the following rely only on assumption (H1).

Spatial data usually exhibit dependence. Therefore, spatial dependence should be taken into account for the quantitative approaches to threshold selection. A simplifying assumption is that the dependence structure does not change over time, i.e. each row of the matrix \((X_s(t))_{s,t}\) has the same dependency structure. To strictly formulate this we rely on the notion of copulas, which are multivariate distributions with standard uniform margins. Sklar’s theorem, see e.g. Nelsen (2006), ensures, that every multivariate distribution function can be disentangled into its marginal distributions and a copula. Therefore, we assume, that there exists a copula \(C\), such that \(\forall t \in \{1, \ldots, T\}\):

\[
P(X_1(t) \leq x_1, \ldots, X_S(t) \leq x_S) = C(F_1(x_1), \ldots, F_S(x_S)).
\] (4.1)

### 4.3 Univariate threshold selection approaches

Scarrott and MacDonald 2012 provide a review of existing threshold selection methods, here we focus only on two types of commonly used approaches.

#### 4.3.1 Approaches based on visual inspection

The TS plot and the ME plot (also referred to as mean residual life plot) are widely used graphical tools for the selection of the threshold in POT analyses, see e.g. Em-
4.3. **Univariate Threshold Selection Approaches**

Brechts et al. (1997) and Coles (2001). The TS plot is based on the fact that once the GPD model holds at grid point \( s \) for some threshold \( u_s^0 \), it holds for every threshold \( u \geq u_s^0 \). In particular we have that the associated shape parameter \( \hat{\xi}_s(u) \) is the same for \( u \) and \( u_s^0 \). This property is called threshold stability. The threshold stability can be exploited by estimating the shape parameter for a range of high thresholds and plotting \( (u, \hat{\xi}_s(u))_{u \geq 0} \),

where \( \hat{\xi}_s(u) \) denotes the estimated shape parameter based on the excesses over threshold \( u \). If the GPD model holds for \( u_s^0 \) the graph should be (approximately) constant for \( u \geq u_s^0 \). However, owing to the decreasing number of excesses above higher thresholds, the plot becomes unstable and the constant behavior is difficult to see. In the following we will use a slightly modified version of the TS plot, i.e.

\[
(\tau, \hat{\xi}_s(\tau))_{\tau \in [0,1]},
\]

where \( \hat{\xi}_s(\tau) \) is a short notation for \( \hat{\xi}_s(\hat{F}_s^{-1}(\tau)) \) and \( \hat{F}_s^{-1}(\tau) \) denotes the empirical \( \tau \)-quantile.

In order to support proper selection of the threshold, it is recommended to also consider other available diagnostic plots. The ME plot relies on a similar consideration but utilizes the ME function

\[
e_s(u) := \frac{1}{1 - \hat{F}_s(u)} \int_u^\infty (x - u) d\hat{F}_s(x).
\]

The empirical version of the ME function is given by

\[
\hat{e}_s(u) = \frac{\sum_{t=1}^T (x_s(t) - u) \mathbb{I}_{(u, \infty)}(x_s(t))}{\sum_{t=1}^T \mathbb{I}_{(u, \infty)}(x_s(t))},
\]

where \( \mathbb{I}_A \) is the indicator function for set \( A \), i.e. \( \mathbb{I}_A(x) = 1 \) if \( x \in A \) and otherwise zero. If the GPD model holds, \( e_s(u) \) is linear in the threshold \( u \), with the slope being determined by the shape parameter \( \xi \), see e.g. Embrechts et al. (1997), and \( \hat{e} \) becomes approximately linear. However, as described in greater detail in Ghosh and Resnick (2010), there are some problems associated with the use of the ME plot. These are in particular that the ME function is only well defined for \( \xi < 1 \), and – as for the TS plot – the empirical ME function becomes unstable for high values of the threshold. In many applications only values of \( \xi < 1 \) are found, in which case the ME plot is well defined.
4.3.2 Approaches based on GoF tests

A more quantitative way to select the threshold is to use a Goodness of Fit (GoF) test. Classical tests are the KS test and the AD test, see e.g. Pickands (1975), Davison and Smith (1990), Choulakian and Stephens (2001), and El-Aroui and Diebolt (2002). Both tests can be used to test the composite hypothesis, that the \( n \) excesses over a fixed threshold \( u \) follow a GPD. The KS statistic is defined as

\[
D_n = \sup_y |F_n(y) - \hat{G}(y)|, \tag{4.2}
\]

where \( \hat{G} \) is the estimated GPD from the \( n \) excesses, and \( F_n \) the empirical distribution of the \( n \) excesses. The AD statistic is defined by

\[
A_n = n \int_{-\infty}^{\infty} \frac{(F_n(y) - \hat{G}(y))^2}{\hat{G}(y)(1 - \hat{G}(y))} dG(y). \tag{4.3}
\]

Although these statistics are quite powerful in general, the threshold selection is a real challenge. There is typically only a limited amount of data and with increasing threshold the sample size \( n \) decreases even further. For the univariate case there exist tables with the critical values, which depend on the (estimated) shape parameter, see e.g. Choulakian and Stephens (2001). Notice that their critical values for finite samples were obtained by Monte Carlo experiments. We will also obtain critical values via simulation using two slightly different simulation approaches:

1. Fix \( \tau_0 \) and the corresponding threshold \( u \).
2. Estimate the parameters based on the \( n \) excesses above \( u \).

(3a) Simulate \( n \) independent random variables from the GPD with the parameters estimated in step 2.

(3b) Sample \( \hat{n} \) from a binomial distribution of size \( T \) and success probability \( (1 - \tau_0) \). Simulate \( \hat{n} \) GPD random variables with the estimated parameters.

(4) Calculate for the simulated data the KS and AD statistic as before.

(5) Repeat steps 3 and 4 a thousand times and take the 0.95-quantile of the bootstrapped statistic as critical value for \( D_n \) (or \( A_n \)) at \( \tau_0 \).

(6) The procedure can be repeated for every \( \tau \) in a reasonable range.

In step 3a, we simulate conditional on the observed number of excesses. However, if we would have different samples from the same population the proportion above the threshold would probably be different. This is accounted for in step 3b, which
is equivalent to generating \( T \) standard uniform random variables and transform the values exceeding \( \tau_0 \) to GPD data. In the regional setting this approach is generalized for multiple dimensions. Fig. 4.1 shows a typical plot of the KS statistic. It illustrates the difficulties in the selection of the threshold, owing to the low statistical power of the test given the broad alternative (Davison and Smith 1990). For instance, the 0.9 quantile is not rejected, but slightly higher quantiles are. The 0.925 quantile might be considered, because no higher quantile is rejected. However, this seems somewhat low compared to existing literature, e.g. Kyselý et al. (2010).

### 4.4 Threshold selection in the regional POT model

The general idea behind the proposed regional approach is to use spatial averaging in the selection process. Although we demonstrate the regionalization only with the univariate threshold selection methods, described in section 4.3, it can be applied to a variety of other univariate threshold selection approaches, e.g. the approach proposed in Thompson et al. (2009).
CHAPTER 4. REGIONAL THRESHOLD SELECTION

4.4.1 Approaches based on visual inspection

The following version of the TS plot only assumes (H1). For $\tau \in [0, 1)$ we estimate the corresponding site-specific shape parameters $\hat{\xi}_s(\tau), s \in \{1, \ldots, S\}$ based on the excesses over the threshold $u_s(\tau) := \hat{F}_s^{-1}(\tau)$. Then, we can consider the spatially averaged TS plot

$$\left(\tau, \bar{\xi}(\tau)\right)_{\tau \in [0,1)},$$

(4.4)

where $\bar{\xi}(\tau) = S^{-1} \sum_{s=1}^{S} \hat{\xi}_s(\tau)$. Due to the fact that the individual plots $(\tau, \hat{\xi}_s(\tau))$ should be roughly constant for $\tau \geq \tau^0$, we have that the averaged TS plot in Eq. (4.4) is approximately constant too. Moreover, due to the averaging, small-sample size effects are reduced and the constant behavior becomes more apparent. Assuming also (H2) leads to a variant by considering a common estimate for the shape parameter instead of $\bar{\xi}$, compare Chapter 2.

The empirical ME function is linear in the threshold itself and not in the corresponding probability $\tau$. Therefore, we average not only the excesses but also the thresholds for a common $\tau$. Here the assumption (H2), which implies that the slopes in the single ME plots are the same, is critical, for the following reason. Consider the two linear functions:

$$y^{(1)}_i = au_i + b \quad \text{and} \quad y^{(2)}_i = cv_i + d.$$

Then we can not say much about their average as function of the average argument, because

$$(y^{(1)}_i + y^{(2)}_i)/2 = (au_i + cv_i)/2 + (b + d)/2.$$

However, with $a = c$ we have that the average is again linear with the same slope:

$$(y^{(1)}_i + y^{(2)}_i)/2 = a(u_i + v_i)/2 + (b + d)/2.$$

This gives rise to the spatially averaged ME plot

$$\left(\bar{\tau}(\tau), \bar{e}(\tau)\right)_{\tau \in [0,1)},$$

(4.5)

where $\bar{\tau}(\tau) = S^{-1} \sum_{s=1}^{S} u_s(\tau), \bar{e} = S^{-1} \sum_{s=1}^{S} \hat{e}_s(\tau)$ and $\hat{e}_s(\tau)$ is the ME at site $s$ evaluated at level $u_s(\tau)$. The averaged ME plot should be approximately linear for $\tau \geq \tau^0$, and the slope is determined by the common shape parameter $\bar{\xi}$. The strength of the spatially averaged plots lies in the increased detection probability of non-constant behavior in the TS plot and non-linear behavior in the ME plot, when the threshold
is too low.

4.4.2 Approaches based on GoF tests

We want to test the hypothesis (H1), i.e. that the GPD behavior sets in at $F_s^{-1}(\tau_0)$ for all sites $s \in \{1, \ldots, S\}$. Douglas et al. (2000) evaluate the significance of trends in flood data across the US by averaging the test statistic regionally. We follow this idea and average the KS statistic, respectively the AD statistic, computed as in section 4.3.2, over the available sites. Averaging reduces the variance, which leads to an increased sensitivity for too low thresholds. However, there is no standard way to obtain critical values for the averaged test statistic and one has to rely on simulation.

For the simulation of the spatial dependence we use copulas. Assume the data-generating copula $C$ from Eq. (4.1) is known, then one can use this copula for the computation of the critical values without knowing the full marginal distribution. Say we want to compute critical values for the exceedance probability $\zeta_0 = 1 - \tau_0 = 0.05$. First we generate marginally uniform data from the copula $C$, e.g. for 4 sites and 5 days

$$
\begin{pmatrix}
0.11 & 0.23 & 0.45 & 0.96 \\
0.97 & 0.98 & 0.11 & 0.05 \\
0.02 & 0.45 & 0.23 & 0.61 \\
0.89 & 0.92 & 0.96 & 0.95 \\
0.93 & 0.96 & 0.98 & 0.76 \\
\end{pmatrix}
$$

and keep only those entries exceeding $\tau_0$, i.e. in our example

$$
\begin{pmatrix}
- & - & - & 0.96 \\
0.97 & 0.98 & - & - \\
- & - & - & - \\
- & - & 0.96 & 0.95 \\
- & 0.96 & 0.98 & - \\
\end{pmatrix}
$$

These are then transformed to GPD data using the marginal cdf with the estimated parameters. For each site $s$ this is equivalent to the unconditioned simulation of GPD random variables in section 4.3.2. Note that owing to the construction process, we are only interested in events, where at least one component exceeds the $\tau_0$-quantile of its own distribution, i.e.

$$
1 - C(\tau_0, \ldots, \tau_0) = P(F_1(X_1) > \tau_0 \text{ or } \ldots \text{ or } F_S(X_S) > \tau_0). \tag{4.6}
$$

Finally, we determine for each site the test statistic and compute the average over all sites. Repeating the procedure, e.g. 1000 times, leads to critical values for the
averaged test statistic that take the spatial dependence into account. In general the
copula is not known and has to be estimated. Because copula estimation lies outside
the focus of this chapter, we use fixed copulas in the simulation section and use
a heuristic estimate in the application section, to describe the dependence at high
quantiles.

4.5 Simulation study

The evaluation of a threshold selection method is challenging. If the bulk distribu-
tion and the GPD tail are quite different the threshold can easily be detected, maybe
also with univariate methods (Wang 1995). Therefore we consider a distribution,
which has a smooth transition between the bulk part and the GPD tail.

4.5.1 Marginal model

Any continuous cdf $F$ on $[0, \infty]$ can be written as

$$F(x) = 1 - \exp(-H(x)) = 1 - \exp\left(- \int_0^x h(u) \, du\right),$$

(4.7)

where $(H) h$ is the (cumulative) hazard rate. The density is given by

$$f(x) = h(x) \exp(-H(x)),$$

which is of (differentiability) class $C^k$, whenever $h$ is of class $C^k$.

A distribution with differentiable density and GPD tail above threshold $u$ can be
obtained by defining the following hazard rate

$$h(x) := \eta \left(\frac{x-u}{\varepsilon}\right) h_1(x) + \left(1 - \eta \left(\frac{x-u}{\varepsilon}\right)\right) h_2(x),$$

(4.8)

where $\varepsilon > 0$ is the length of the transition interval, $h_1$ the hazard rate of a general
bulk distribution, and $h_2$ the hazard rate of a GPD with shape parameter $\xi$ and scale
parameter $\sigma$, given by

$$h_2(x) = \frac{1}{\sigma + \xi(x-u)}, \quad x \geq u,$$

(4.9)
and $h_2(x) = 0$ otherwise. The use of the transition function

$$
\eta(x) = \begin{cases} 
  1, & x \leq 0, \\
  2x^3 - 3x^2 + 1, & 0 < x < 1, \\
  0, & x \geq 1,
\end{cases} \quad (4.10)
$$

ensures the differentiability of the hazard rate $h$, see Fig. 4.2. Note that the tail, the part of the distribution to the right of $u$, is contaminated over the interval $(u, u + \varepsilon)$.

For a fixed exceedance probability $\zeta_0 = 1 - \tau_0$, e.g. 5%, the corresponding threshold is defined via the following relation:

$$
\zeta_0 = \exp (-H_1(u)) \iff u = H_1^{-1}(-\ln(\zeta_0)). \quad (4.11)
$$

The transition can be smoothed further by considering the following restriction on Mill’s ratio $\phi(x) = 1/h(x)$ at $x = u$:

$$
\phi'(u) = -\frac{h'_1(u)}{(h_1(u))^2} = \xi. \quad (4.12)
$$

This is motivated by the theory of penultimate approximations (Smith 1990a; Wadsworth and Tawn 2012).

As bulk distribution we consider the Weibull distribution

$$
W_{\kappa, \beta}(x) = 1 - e^{-(x/\beta)^\kappa}, \quad (4.13)
$$

with shape parameter $\kappa$ and scale parameter $\beta$. The hazard rate $h_1$ of the Weibull
distribution is given by

\[ h_1(x) = \frac{\kappa}{\beta} \left( \frac{x}{\beta} \right)^{\kappa-1}. \]

For this distribution \( u = \beta(-\ln(\xi_0))^{1/\kappa} \). Restriction (4.12) implies then the following relationship between the Weibull and GPD shape parameters:

\[ \xi = \frac{1 - \kappa}{\kappa} \left( \frac{u}{\beta} \right)^{-\kappa} = \frac{1 - \kappa}{\kappa} (-\ln(\xi_0))^{-1}. \]

Fig. 4.3 shows the cdf and a zoom into the transition interval of the density of the hybrid Weibull–GPD (WGPD) model, for one set of GPD parameters and varying Weibull parameters. We observe that the transition is much smoother for \( \kappa = 0.7 \), than for \( \kappa = 0.8 \). The reason is, that for \( \kappa = 0.7 \) and \( \xi = 0.15 \) Eq. (4.12) is approximately true, which is not the case for \( \kappa = 0.8 \) and the same \( \xi \).

4.5.2 Spatially independent data

After defining the marginal model we first investigate the influence of the number of considered locations on the selected threshold. Therefore, we simulate spatially and serially independent data for 16 locations and 4600 days, e.g. 50 summer seasons. We took for all sites \( \tau = 1 - \xi = 0.95 \), \( \varepsilon = 0.25 \), \( \gamma = 0.5 \), and \( \xi_0 = 0.15 \). The Weibull
scale parameters $\beta_s$ were obtained as a random sample from the uniform distribution between 2 and 4 and the Weibull shape parameters $\kappa_s$ as a random sample from the beta distribution, with parameters 2 and 5, shifted by 0.5 (mode at 0.7). We denote this parameter set by I.

Fig. 4.4 shows the TS plot for one site and the spatially averaged version for eight and sixteen sites for the simulated data. The individual TS plot shows no stabilization, it increases sharply after an almost linear decrease until $\tau = 0.98$. The averaged at-site estimate stabilizes around $\tau = 0.96$ for eight sites and slightly earlier for sixteen sites. Moreover, we observe that the variation in the averaged shape parameter estimate reduces substantially from eight to sixteen sites. Fig. 4.5 shows the (spatially averaged) ME plot for the same data. Here, the reduced variance of the averaged at-site estimates for eight and in particular for sixteen sites becomes even more apparent. One observes only a small change in the slope of the spatially averaged ME plot for sixteen sites around the averaged 0.95-quantile, i.e. it decreases from 0.25 to 0.19.

For the quantitative approach based on GoF tests, we first have to define a threshold selection criterion. We select the smallest $\tau \in [0.9, 1)$, such that the test statistic is smaller than the 95 percentage point of the bootstrapped statistic. For instance, for the precipitation series at Putten the selected probability $\tau$ would be 0.9, because the values of the KS test statistic are always below the corresponding 95 percentage point, as shown in Fig. 4.1. The adequacy of the proposed procedure is explored.
CHAPTER 4. REGIONAL THRESHOLD SELECTION

Figure 4.5: ME plot for one site (blue), eight sites (green) and sixteen sites (red) for simulated data from the WGPD distribution (parameter set I, no spatial dependence). The dashed vertical lines indicate the corresponding (mean) 0.95-quantile. The dotted lines show the fitted linear regression above the (mean) 0.95-quantile.

by simulating $B = 1000$ samples for the 16 locations of the same size from the same marginal distributions. For each simulated sample the threshold was determined using the criterion above for single site data and for the averaged KS statistic over 2, 4, 8, and 16 sites. The GPD parameters were estimated for each single site but in the case of averaging over 16 sites, joint estimation was also considered, using the approach described in Chapter 2.

Fig. 4.6 shows the influence of the number of considered sites on the selected threshold. The more sites are considered the closer the selected threshold is to the true one. This can be explained by the increasing power of the GoF test, when using more data. Moreover, the figure shows the advantage of using the regional similarity in the parameter estimation. In the case of regional estimation the mean selected $\tau$ is 0.945, i.e. the mean exceedance probability is 5.5%, which is a strong improvement compared to the mean exceedance probability obtained for the univariate samples. Note that strong overestimation of the exceedance probability in the univariate case is also reported for different threshold selection methods, see e.g. Hall and Welsh (1985) and Scarrott and MacDonald (2012).

After looking at the influence of the number of locations, we investigate how the threshold selection method performs, when bulk and tail distribution are alike near the threshold level. Therefore, we simulate 1000 samples using the same marginal parameters as before except for the Weibull shape parameter, which we specify such
that Eq. (4.12) holds, i.e. $\kappa = 0.69$. We denote this parameter set by II. For this parameter set, the 0.9-quantile is selected as threshold in more than 90% of the simulated samples, even when considering all 16 sites. Due to the small differences between bulk and tail distribution the method is not able to select the theoretical threshold.

Focusing on the threshold may not be optimal in terms of bias and variance of estimated parameters or return levels. The bias-variance trade-off was studied for return levels. Denote the vector of the true return levels by $r$ and the estimated return levels from sample $i$, based on probability $\tau$, by $\hat{r}_i(\tau)$. Then we compute for each $\tau$ the averaged Euclidean error over the $B$ simulated samples:

$$AEE(\tau) = \frac{1}{B} \sum_{i=1}^{B} ||\hat{r}_i(\tau) - r||_2,$$

where $||.||_2$ denotes the Euclidean (or $l_2$) norm. Fig. 4.7 shows $AEE(\tau)$ for the 50-year return level. For parameter set I $AEE(\tau)$ reaches a minimum around $\tau = 0.95$. This nicely demonstrates the trade-off, if the threshold is too low a large bias is introduced, and if the threshold is too high the variance dominates. However, for parameter set II the $AEE$ is almost constant up to $\tau = 0.945$ and it increases afterwards. This means there is hardly any bias introduced by selecting lower thresholds,
Figure 4.7: Averaged Euclidean error in the 50-year return level as a function of the probability \( \tau \) for the simulated data from the WGPD distribution (no spatial dependence, blue - parameter set I, red - parameter set II). For both parameter sets the GPD parameters were jointly estimated for all sixteen sites. The dashed horizontal lines indicate the averaged Euclidean error using the selected threshold from the KS test (\( AEE^0 \)).

which highlights the effect of the restriction posed in Eq. (4.12) on the smoothness of the transition between bulk and tail distribution.

We are interested how the averaged Euclidean error \( AEE^0 \) of the estimated return level, based on the automatically chosen probability \( \tau(i) \), compares to the minimal \( AEE \) over all \( \tau \). \( AEE^0 \) can be computed by replacing \( \hat{r}_i(\tau) \) by \( \hat{r}_i(\tau(i)) \) in Eq. (4.14). In Fig. 4.7, we see that \( AEE^0 \) is for both parameter sets close to the minimum value of \( AEE(\tau) \). The situation for the 5-year return level is very similar, see Fig. 4.8. The results in Fig. 4.7 and Fig. 4.8 apply to the KS test. The AD test results in slightly smaller values of \( AEE^0 \) as well as \( AEE(\tau) \) in general, but the observed differences are smaller than 2%.

We compared several significance levels for both tests. The values of \( AEE^0 \) are very similar for significance levels between 5 and 20 percent, but increase outside this interval.

### 4.5.3 Spatially dependent data

The modeling of spatial dependence in general and in particular that in extremes is a very active field of research, see e.g. Davison et al. (2012) and Segers (2012). The
dependence of small to medium values should have no influence on the threshold selection. Therefore, classical dependence measures like correlation and Kendall’s tau, that are usually estimated from all data, are not appropriate. In extremes the notion of tail dependence arises. A bivariate random vector \((X_1, X_2)\) with marginal distributions \(F_1\) and \(F_2\), is said to be (upper) tail dependent if (Coles et al. 1999)

\[
l_u := \lim_{v \uparrow 1} P \left( X_1 > F_1^{-1}(v) \mid X_2 > F_2^{-1}(v) \right) > 0. \quad (4.15)
\]

Similarly, we say that the bivariate vector is tail (or asymptotically) independent if \(l_u = 0\). For many copulas one can derive \(l_u\) directly from the copula parameters. Consider e.g. the Gumbel copula with parameter \(\theta \in [1, \infty)\), given by

\[
C_\theta(u, v) = \exp \left( - \left[ (-\ln(u))^{\theta} + (-\ln(v))^{\theta} \right]^{1/\theta} \right), \quad (4.16)
\]

which can be easily generalized to \(d\) dimensions. Here, the tail dependence can be easily computed as \(l_u = 2 - 2^{1/\theta}\) (Nelsen 2006). However, it is known that a bivariate normal random vector is tail independent for every correlation coefficient \(\rho \in (-1, 1)\). Hence, we conclude that a quantile-dependent measure of dependence,
CHAPTER 4. REGIONAL THRESHOLD SELECTION

Table 4.1: Gumbel copula parameter $\theta$ and normal copula parameter $\rho$ used in the simulation of spatial dependent data (only one data set without spatial dependence was simulated).

<table>
<thead>
<tr>
<th>$l^u(0.9)$</th>
<th>0.1</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>1</td>
<td>1.1514</td>
<td>1.5994</td>
<td>2.9254</td>
<td>6.8769</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0</td>
<td>0.3686</td>
<td>0.7366</td>
<td>0.9358</td>
<td>0.9898</td>
</tr>
</tbody>
</table>

\[ l^u(\tau) := P\left(X_1 > F_1^{-1}(\tau) | X_2 > F_2^{-1}(\tau)\right) = \frac{P\left(X_1 > F_1^{-1}(\tau), X_2 > F_2^{-1}(\tau)\right)}{P\left(X_1 > F_1^{-1}(\tau)\right)} , \quad (4.17) \]

might be more appropriate for threshold selection. $l^u(\tau)$ can be simply estimated from the data by counting the excesses over the sample $\tau$-quantile.

We simulate spatial dependent data from the WGPD distribution for 16 sites using parameter set I. To determine the influence of the dependence structure and in particular of tail dependence on the selection of the threshold, we choose to simulate from a 16-dimensional Gumbel copula with parameter $\theta$ and a 16-dimensional normal copula with the same correlation $\rho$ for all pairs. To obtain a fair comparison between the two copulas we prescribe fixed values of $l^u(\tau)$ for $\tau = 0.9$, the lowest possible probability in our setting, and select the copula parameters correspondingly. We fix the following values for $l^u(0.9)$: 0.1 (which corresponds to independent data), 0.25, 0.5, 0.75, and 0.9. The corresponding parameters for the Gumbel and normal copula are shown in Table 4.1.

Fig. 4.9 shows violin plots of the selected probability for different levels of dependence. We see that the method is quite able to select a threshold close to the true 0.95-quantile. Moreover, we see that there is no clear dependence of the selected threshold on the amount of spatial dependence in the data. In contrast to the selected threshold, $AET^0$ increases strongly with $l^u(0.9)$ and is generally larger for the Gumbel copula than for the normal copula, see Fig. 4.10.

4.6 Application to rainfall data

Fig. 4.11 shows the area of the *Vallei en Veluwe* water board in the Netherlands, together with the locations of the 21 rainfall stations in this area. For each station we have summer (JJA) precipitation records available for the period 1951-2009, that were corrected for inhomogeneities owing to changes in local measurement con-
4.6. APPLICATION TO RAINFALL DATA

ditions (Buishand et al. 2013). Thus, we obtain a rainfall matrix \((X_s(t))_{s,t}\) with \(s \in \{1, \ldots, 21\}\) and \(t \in \{1, \ldots, 5428\}\).

In order to exclude temporal dependence, we apply first the declustering algorithm, described in section 3.2.2. Owing to the weak temporal dependence in summer, a separation time of 1 day is taken as sufficient (Kyselý and Beranová 2009). Fig. 4.12 shows the spatially averaged TS plot for the declustered data. The shape parameter estimate is either determined as the average of the individual Maximum Likelihood (ML) estimates or as common estimate, using the estimation approach outlined in Chapter 2. For the latter approach we rely on assumptions (H2,H3), introduced in section 4.2. Both approaches show a strong increase of the estimated shape parameter up to the 0.96-quantile, which indicates that one should consider thresholds above the 0.96-quantile. The common estimate is more stable at large \(\tau\) than the average at-site estimate, thus allows for a better selection of \(\tau_0\). The difference is probably due to the bias in the at-site estimates of \(\xi\). At the 0.96-quantile the individual shape parameter estimate is based on 217 excesses. For a sample size of 200 and a shape parameter of 0.25, Zhang and Stephens (2009) report a bias of -0.013 for the ML estimate.

To apply the quantitative approach, we have to model the spatial dependence. Buishand (1984a) concludes that annual maximum rainfalls over the Netherlands are asymptotically independent if the measurement locations are separated by more than 30km. Ancona-Navarrete and Tawn (2002) confirm this for daily rainfall records.

Figure 4.9: Selected probability \(\tau\) for different dependence levels \(l^u(0.9)\), based on the 95% critical value of the KS test using joint parameter estimation, for simulated data from the WGPD distribution (red - Gumbel copula, blue - normal copula).
Figure 4.10: $AEE^0$ of the 5- (dots) and 50-year (triangles) return level for simulated data from the WGPD distribution (red - Gumbel copula, blue - normal copula).

Figure 4.11: Left: Map of the Netherlands with the area of the Vallei en Veluwe water board surrounded by a red line. Right: A zoom into the map, the pluses indicate the individual measurement locations. The blue star is the measurement site Putten.

in south-west England, adding that there exists a quite strong extremal association at sub-asymptotic levels, which decays only slightly with distance. Thibaud et al. (2013) find that asymptotic independent models perform better even at very small spatial scales for summer rainfall in Switzerland.
4.6. APPLICATION TO RAINFALL DATA

Figure 4.12: Spatially averaged TS plot for the rainfall data. The estimate of the shape parameter is obtained by averaging ML estimates for the individual sites (blue) and by using joint ML estimation (red).

Figure 4.13: Estimated measure of dependence $l_u(0.9)$ versus distance for the rainfall data. The smooth line is obtained by a loess smoother.

Fig. 4.13 shows the estimated dependence measure $l_u(0.9)$ for each pair of sites versus their distance. The mean of this measure is 0.6. For higher levels this decreases but stays above zero, i.e. 0.5 (0.35) at the 0.95(0.99)-quantile. This decay, however, indicates that the normal copula is more appropriate than a tail dependent...
model. To study potential influences of the tail dependence, we model the spatial dependence in three different ways. A multivariate Gumbel copula with parameter $\theta = 1.933$ representing a simple tail dependent model on the one hand and a normal copula with parameter $\rho = 0.8336$ for all pairs representing a simple tail independent model on the other hand. Both choices of the parameter satisfying $I^u(0.9) = 0.6$. Since, these models do not reproduce the decay of the tail dependence with increasing distance, we also use a multivariate normal copula, with individual correlation parameters $\rho_{s_1,s_2}$ for each pair of sites $(s_1,s_2)$. The correlation parameter $\rho_{s_1,s_2}$ is derived from the loess-fit of $I^u(0.9)$, shown in Fig. 4.13. Fig. 4.14 presents the KS statistic for the rainfall data based on the joint estimation of the GPD parameters, together with the 95 percentage points of the statistic based on three different dependence models: The differences between these dependence models are remarkably small and overall point to a threshold above the 0.96-quantile. In particular the choice to model the decrease of dependence with distance does not seem to influence the selected threshold.

4.7 Conclusion

This paper has presented a method to incorporate similarity of multiple data samples into the selection of a threshold for a subsequent POT analysis. This setting is
often encountered in environmental research but also different stock prices from a similar branch constitute a possible example. We proposed spatially averaged versions of the threshold stability and the ME plot. Moreover, we have introduced spatially averaged KS and AD statistics for assessing the goodness of the GPD fit for the threshold excesses. Our method of threshold selection was validated using a smooth WGPD distribution. Simulations from this model show that the use of common GPD parameters reduces the (negative) bias of the selected threshold substantially. Moreover, the averaged Euclidean distance of the 5- and 50-year return level for the selected threshold is close to the optimal averaged Euclidean distance. The regional method was applied to a set of summer precipitation records in the Netherlands. We found that the 0.96-quantile is a suitable threshold. The choice of the threshold is not sensitive to assumptions regarding the spatial dependence of this data set.

**Acknowledgements**

The data set was kindly made available by T. Brandsma.
Trends in moderate rainfall extremes

Abstract. Rainfall extremes are thought to have increased over recent years. Typically linear trends have been used to model the temporal evolution of high quantiles of the daily rainfall distribution, which are used as indicator for moderate extremes. For long records it is important to allow more flexibility. Quantile regression methods are available to estimate monotone trends for single stations. Having multiple stations in a region, the field significance is often of interest. From this perspective we propose a regression approach, that can be used to estimate a common monotone trend for the site-specific quantiles. Moreover, the method allows the construction of confidence bands and testing the hypothesis of an existing non-decreasing trend against the null hypothesis of no trend. The approach is applied to 102 series of daily rainfall over the Netherlands for the period 1910–2009. The results are compared with those from a (regional) Mann–Kendall (MK) test. Significantly increasing trends are found for the winter season and for the whole year. In the summer season trends are less consistent over the region, and are only significant in the western part of the Netherlands. For the summer season linearity of the trend seems less apparent than for winter and for the whole year. However, the deviation from linearity is not significant.

This chapter is based on: M. Roth, T. A. Buishand, and G. Jongbloed (2014): Trends in moderate rainfall extremes: A regional monotone regression approach (submitted to Journal of Climate).
CHAPTER 5. TRENDS IN MODERATE RAINFALL EXTREMES

5.1 Introduction

Many researchers have studied changes in moderate rainfall extremes, using indices such as the number of days exceeding a specific threshold, e.g., 20 mm, or annual/seasonal quantiles (Klein Tank and Können 2003; Moberg and Jones 2005; Alexander et al. 2006). Quantiles refer to the same part of the distribution everywhere and are therefore more suitable for spatial comparison than counts of exceedances of an absolute threshold (Klein Tank and Können 2003). Linear regression has often been used to test for changes in sample quantiles. A difficulty with this approach is that high quantiles from small samples can be seriously biased. Quantile regression can be used to overcome this (Wasko and Sharma 2014).

Linear trend modeling is typically used because of its broad acceptance (compare Hartmann et al. 2013, Box 2.2). However, when considering long measurement records, with possible phases of stagnation, linear modeling is not flexible enough, see e.g. Villarini et al. (2011). The MK test can be used to test the hypothesis of a monotonic trend in the annual or seasonal quantiles against the null hypothesis of no trend. However, it does not provide a visualization of the trend itself.

Monotone quantile regression provides an alternative, that does not rely on the biased annual quantiles and provides a visualization of the trend. Monotone trends can be estimated by order restricted regression (Robertson et al. 1988). While order restricted regression has been relatively often applied to medical data, e.g. growth, toxicological, or pharmaceutical data, it has rarely been used for environmental data. Examples where this is done are Wu et al. (2001) (temperature anomalies), Leitenstorfer and Tutz (2007) (air pollution), and Tibshirani et al. (2011) (temperature anomalies).

Quantile regression is considered to be robust to outliers (Koenker 2005; Chandler and Scott 2011). Though trend estimates based on quantile regression are almost unbiased, they are still noisy for high quantiles owing to data scarcity. Therefore, it can be advantageous to consider a regional approach where the site-specific quantiles exhibit a common relative trend. This common trend is generally more precisely estimated than the trends of the individual sites.

In this study we propose a new algorithm for (regional) monotone regression of precipitation quantiles. This algorithm is used to explore the changes in the 0.95, 0.975, and 0.99 quantiles of daily precipitation in the Netherlands. The $p$-values are compared with those obtained from the (regional) MK test, proposed in Douglas et al. (2000). In section 5.2 we introduce the considered precipitation data. The methods are explained in section 5.3. In section 5.4 we present the results and the comparison with the MK test.
5.2 Data

The domain of the study is the Netherlands (NL) for which we have a high-quality data set of observed daily precipitation sums for a long period. The data consist of 102 records of daily precipitation sums over the country from 1910 to 2009, that were corrected for inhomogeneities owing to changes in local measurement conditions (Buishand et al. 2013). The selected records exhibited at most 5% missing values. The missing values were supplemented with the data from the nearest station. Zolina et al. (2008) showed that missing value effects are negligible if less than 10% of the data are missing.

We apply the procedures locally and on the (sub-) national level. The individual measurement stations and the considered subregions are shown in Fig. 5.1. The subregions are based on the presence of a possible coastal effect. A similar subdivision was applied in Beersma and Buishand (2007) for evapotranspiration data and Daniels et al. (2013) for precipitation data for the period 1950–2009.
Buishand et al. (2013) studied trends over the periods 1910–2009 and 1951–2009 for the Netherlands. Significant increases were found for annual precipitation and the precipitation amounts in the winter and summer half of the year. The strongest increases were found in the winter season, which is consistent with the increases found in Western and Central Europe (Moberg and Jones 2005; Van den Besselaar et al. 2012). The significant increases in the summer season turned out to be restricted mainly to coastal regions. This increase is probably related to the increase in sea surface temperature (Lenderink et al. 2009). Buishand et al. (2013) also found a clear coastal effect in the number of days per year with a precipitation amount $> 30$ mm. Daniels et al. (2013) studied the trends in seasonal mean precipitation and quantiles of the wet-day precipitation amounts for the period 1951–2009. It was found that zones based on distance to the coast gave a more consistent picture for precipitation changes over time than regions based on soil type, topography, or urbanization.

5.3 Methods

We are interested in the evolution of precipitation quantiles. A commonly used approach is to compute annual or seasonal quantiles and to assess the significance of the trend by the MK test for stationarity. The MK test is distribution free, thus, the null distribution of the test statistic does not depend on the probability distribution from which the data are drawn. Moreover, it has high power against a one or two-sided monotone trend alternative (Yue et al. 2002). In the following we first discuss the MK test before we explain the alternative monotone quantile regression method.

5.3.1 Mann–Kendall test

Consider a sequence of independent random variables $Z_i$, distributed according to $F_i$ $(i = 1, \ldots, n)$. In our application $Z_i$ represents a sample quantile for year $i$. The MK test uses the Kendall score $K$, i.e. the difference between the number of concordant pairs ($Z_j > Z_i$ for $j > i$) and the number of discordant pairs ($Z_j < Z_i$ for $j > i$):

$$K := \sum_{j=2}^{n} \sum_{i=1}^{j-1} \text{sgn}(Z_j - Z_i),$$

to test the hypothesis that all distributions $F_i$ are equal. Under the null hypothesis $K$ is approximately normal with mean 0 and variance $n(n-1)(2n+5)/18$, provided the data do not contain ties (Douglas et al. 2000). Annual and seasonal quantiles exhibit some autocorrelation at lag 1 and 2. In order to account for that we use a Monte Carlo permutation test, resampling blocks of three years (Carlstein 1986;
5.3. METHODS

Leander et al. 2014).

We follow Douglas et al. (2000) in their definition of a regional MK statistic for the
determination of the field significance. For a region with \( S \) time series we calculate
first as above for each time series the statistic \( K \) and denote it by \( K_s \) for time series
\( s, 1 \leq s \leq S \). In the second step we consider the average of these: \( \bar{K}_S = \frac{1}{S} \sum_{s=1}^{S} K_s \).
As in the univariate case we use a random permutation test for the computation of
\( p \)-values. To account for the interdependence between samples, we resample spatio-
temporal blocks of three years, i.e. all \( S \) time series are permuted in the same way.

5.3.2 Monotone Quantile Regression

Linear quantile regression

The \( \tau \) sample quantile of a data set \( y_1, \ldots, y_T \), e.g. daily rainfall data, can be com-
puted by minimizing the following objective function (Koenker and Bassett 1978;
Koenker 2005):
\[
\phi(\gamma) := \sum_{t=1}^{T} \rho_\tau(y_t - \gamma),
\]
where
\[
\rho_\tau(x) = \begin{cases} 
(\tau - 1)x & x \leq 0, \\
\tau x & x > 0.
\end{cases}
\]

The value of \( \gamma \) for which \( \phi \) is minimal is the estimator. Estimators of this type are
known as M-estimators. The M-estimator approach provides an alternative for the
computation of quantiles, which are often computed using a sorting approach. Its
main advantage comes to light when considering quantile regression. Analogously
to least squares regression, we obtain a linear quantile regression by computing
(Koenker 2005):
\[
\arg \min_{\beta_0, \beta_1 \in \mathbb{R}} \sum_{t=1}^{T} \rho_\tau \left( y_t - (\beta_0 + \beta_1 \cdot t) \right).
\]
For the median (\( \tau = 0.5 \)) this is equivalent to the minimization of the absolute differ-
ces.

So far the temporal evolution of a quantile is more often studied by first estimat-
ing annual quantiles and then performing a regression analysis based on these. The
difference between quantile regression and the traditional approach of fitting a lin-
ear trend to seasonal quantiles is illustrated with synthetic precipitation data from
an exponential distribution in Fig. 5.2. The figure shows that the linear regression of
seasonal quantiles considerably underestimates the theoretical quantile. This is not
the case for linear quantile regression.
Monotone Quantile Regression

We consider Monotone Quantile Regression (MQR), i.e. we search a monotone regression function instead of the linear one in Eq. (5.3), see Casady and Cryer (1976). The optimization depends on the regression function only via its values at discrete $t$’s. Therefore, we focus on monotone piecewise constant functions and, hence, obtain the following optimization problem:

$$\arg \min_{r_1 \leq \ldots \leq r_T} \sum_{t=1}^{T} \rho_{\tau}(y_t - r_t), \quad (5.4)$$

where $r_t$ represents the value of the monotone function evaluated at time $t$.

A common problem with monotone regression is the so-called spiking problem, i.e. the estimator is not consistent at the endpoints (Pal 2008; Groeneboom and Jongbloed 2013), because the monotonicity constraint does not prevent the estimate from following the (erratic) behavior of the $y_t$’s near the boundary. Moreover, a smoother trend might be more suitable to visualize the changing behavior than a piecewise constant one. To tackle the spiking problem and in order to obtain a smoother trend estimate we introduce a penalty on the second order difference of the vector...
\( r = (r_1, \ldots, r_T) \) (cf. Eilers and Marx 1996; Koenker 2005) and compute:

\[
\arg\min_{r_1 \leq \ldots \leq r_T} \sum_{t=1}^{T} \rho_{\tau}(y_t - r_t) + \lambda P(r), \tag{5.5}
\]

where

\[
P(r) := \sum_{t=2}^{T-1} (r_{t+1} - 2r_t + r_{t-1})^2, \tag{5.6}
\]

is the sum of the squared second order differences. By introducing this penalty, variations at the endpoints are not free anymore as larger parts of \( r \) will be affected. Moreover, for \( \lambda \to \infty \) the solution of optimization problem (5.5) will become linear (any deviation from linearity would lead to an immense increase in the second term). An alternative approach would be to use a penalty on the total variation (Groeneboom and Jongbloed 2013), which is in the monotonic setting equivalent to the difference between the endpoints, and smooth the obtained estimator using e.g. kernel smoothing (Wand and Jones 1994).

**Computational issues**

The computation of (5.3) has been thoroughly addressed by Koenker (2005), using simplex and interior point algorithms. Koenker and Ng (2005) use an interior point algorithm to fit a smooth MQR using splines and a penalty on the total variation. Bollaerts et al. (2006) approach the same problem with a slightly different P-splines method, allowing for nearly monotone trends, by introducing a penalty on monotonicity violations. The fact that \( \rho_{\tau} \) is not differentiable, in contrast to the sum-of-squares loss function, hinders the use of standard gradient based optimization techniques, such as the Newton-Raphson method.

Recently Muggeo et al. (2012), following others, proposed to consider a slight modification of \( \rho_{\tau} \) namely

\[
\rho_{\tau}^{(c)}(x) = \begin{cases} 
(\tau - 1)x & \text{if } x \leq -c\tau, \\
\frac{(1-\tau)x^2}{2c} + \frac{c\tau(1-\tau)}{2} & \text{if } -c\tau < x \leq 0, \\
\frac{c\tau x^2}{2c(1-\tau)} + \frac{c\tau(1-\tau)}{2} & \text{if } 0 < x < c(1-\tau), \\
x & \text{if } x \geq c(1-\tau),
\end{cases} \tag{5.7}
\]

see Fig. 5.3. This differentiable approximation of \( \rho_{\tau} \) facilitates the use of different optimization techniques, instead of simplex or interior point methods. Therefore,
we consider the following optimization problem:

\[
\arg \min_{r_1 \leq \cdots \leq r_T} \phi_\lambda(r_1, \ldots, r_T),
\]

(5.8)

where

\[
\phi_\lambda(r_1, \ldots, r_T) := \sum_{t=1}^{T} \tilde{\rho}_\tau(c) y_t - r_t + \lambda P(r).
\]

Muggeo et al. (2012) state that the choice of \( c \) has only limited impact on the final estimates, owing to the special form of the approximation. In the following we consider only a fixed value of 0.1 for \( c \), see also Appendix A.3.

The monotonicity constraint in optimization problem (5.8) is tackled with the Iterative Convex Minorant (ICM) algorithm (Jongbloed 1998). This algorithm is based on minimizing successive approximations of the objective function \( \phi_\lambda \). To be more specific, for a given iterate \( r^{(k)} = (r_1^{(k)}, \ldots, r_T^{(k)}) \) we consider the following optimization problem:

\[
\arg \min_{r_1 \leq \cdots \leq r_T} \sum_{t=1}^{T} \left[ r_t - r_t^{(k)} + \frac{\partial \phi_\lambda}{\partial r_t}(r^{(k)}) \right]^2.
\]

(5.9)

The unique solution of this problem can be easily constructed, using the theory of order restricted regression (Robertson et al. 1988). First define the cumulative sum
5.3. METHODS

diagram consisting of the points \( P_0 = (0, 0) \) and

\[
P_t = \left( t \sum_{j=1}^{i} \left[ r_j^{(k)} - \frac{\partial \phi_{\lambda}}{\partial r_j}(r^{(k)}) \right] \right),
\]

for \( 1 \leq t \leq T \). Then construct the (greatest) convex minorant of these points. The left derivative of the convex minorant evaluated at point \( P_t \) is denoted by \( r_{\text{New}}^t \). Compared to \( r^{(k)} \) the new vector \( r_{\text{New}} \) is a step in the right direction, i.e. of decreasing \( \phi_{\lambda} \), however, it might be too big, i.e. \( \phi_{\lambda}(r_{\text{New}}) > \phi_{\lambda}(r^{(k)}) \). Therefore, we obtain the next iterate as a convex combination of \( r^{(k)} \) and \( r_{\text{New}} \):

\[
r^{(k+1)} = r^{(k)} + u \left( r_{\text{New}} - r^{(k)} \right),
\]

for some \( u \in (0, 1) \), such that \( \phi_{\lambda}(r^{(k+1)}) < \phi_{\lambda}(r^{(k)}) \). As a starting value for the algorithm we use the constant sample quantile. For more information on the ICM algorithm, see Jongbloed (1998).

When we are interested in the estimation of the long term trend the seasonal cycle is not relevant. Hence, the dimensionality of the optimization problem can be reduced by specifying that the quantile estimates should be constant over the year or season. Instead of the \( T \)-dimensional vector \( r \) we consider an \( n \)-dimensional vector \( \tilde{r} = (\tilde{r}_1, \ldots, \tilde{r}_n) \), with \( n \) denoting the number of years. Owing to the dimension reduction we need less iterations of the algorithm and hence less computation time.

Smoothing penalty

A difficulty in non-parametric frameworks is often the selection of the smoothing penalty. Koenker (2005) presents a modified version of the Schwarz-Bayesian information criterion based on counting the pieces of a piecewise linear fit. However, our fit is not piecewise linear and, thus, this approach cannot be applied. A possible way to determine the optimal smoothing parameter in this situation, is by cross-validation using the quantile verification score, which is the value of the objective function \( \phi \) in Eq. (5.1) evaluated for the independent data at the prediction \( \hat{\gamma} \) (Gneiting and Raftery 2007; Friederichs and Hense 2007).

Regional quantile regression

Apart from local trends we would like to estimate a common trend over a region, which we expect to be more precise than the local trend estimates. The significance of this common trend is a measure of the field significance. The common trend in the \( \tau \)-quantile is estimated after scaling the daily rainfall amounts by the site specific
τ-quantile \( q_s(\tau) \). Analogously as in the univariate case this is done by minimizing the objective function:

\[
\arg \min_{r_1 \leq \ldots \leq r_T} \sum_{s=1}^{S} \sum_{t=1}^{T} \rho_\tau \left( \frac{y_{s,t}}{q_s(\tau)} - r_t \right) + \lambda P(r),
\]

(5.10)

where \( y_{s,t} \) is the daily rainfall amount at site \( s \in \{1, \ldots, S\} \) and day \( t \in \{1, \ldots, T\} \). Note that \( r \) now represents a common trend in the standardized τ-quantile and that \( r_t = 1 \), for all \( t \), in the case of no trend.

**Testing**

MQR can be used to test the null hypothesis that the quantile of interest is constant over time, i.e.

\[
H_0 : \ \hat{r}_i \equiv \beta
\]

against the alternative that it is non-decreasing, i.e.

\[
H_1 : \ \hat{r}_1 \leq \ldots \leq \hat{r}_n.
\]

(5.12)

As for the MK test we use a Monte Carlo permutation test for the calculation of \( p \)-values, thus, capturing the spatio-temporal correlation structure of the data. We evaluate the resampled data by calculating the \( l_2 \) difference between the constant quantile and the monotone fit. Taking the difference in the objective criterion Eq. (5.3) or the \( l_1 \) difference gives very similar results.

An advantage of the MQR over the MK approach is, that we can e.g. test also the null hypothesis that the quantile is increasing linearly over time, i.e.

\[
H_0^* : \ \hat{r}_i = \alpha_0 + \alpha_1 * i, \ \ \alpha_1 > 0,
\]

(5.13)

versus the more flexible one-sided monotone trend alternative. In order to apply also here a permutation type test we have to generate appropriate data samples satisfying the null hypothesis. Therefore, we compute the residuals \( z_t = y_t - r_t \), where \( r_t = \hat{r}_i^{H_0^*} \) if day \( t \) falls in year \( i \), permute these as before in 3-yearly blocks, and add then \( r_t \).

As test statistic we use the \( l_2 \) difference between the estimated linear trend \( \hat{r}^{H_0^*} \) and the non-decreasing trend \( \hat{r}^{H_1} \).
5.4. Results

5.4.1 Local application

For each of the five regions in Fig. 5.1 we selected one station and applied cross-validation, using the quantile verification score (Friederichs and Hense 2007), to determine the smoothing parameter $\lambda$. In the cross-validation we skipped a block of one year or season of the data and estimated the non-decreasing trend for a given $\lambda$, based on the remaining data. The prediction $\hat{\gamma}$ for the skipped block was then defined as the average of the estimated values for the neighboring blocks. Using this prediction, we computed the quantile verification score for the independent data, i.e. the skipped block. This was done for every block, leading to 100 prediction scores for the given smoothing parameter. The procedure was repeated for several smoothing parameters $\lambda \in [0, \infty]$, where $\lambda = \infty$ refers to linear estimation restricted to non-negative slope estimates.

The prediction performances for the considered smoothing parameters are very similar and the optimal parameter varies over stations and quantiles. For some combinations the optimal parameter is quite small and for others very large. In general we can say that the differences in the mean predictive score for different smoothing parameters are very small compared to the annual variation of the score for the optimal parameter. Because of the inconsistent picture for the data at hand and the increased computational burden for large smoothing parameters, we decided to consider a relatively small, fixed value of $\lambda$ (300) for all sites and quantiles.

Then we tested the null hypothesis of no trend in the 0.95 quantile versus the alternative of a non-decreasing trend for each of the 102 stations. Fig. 5.4 shows

Figure 5.4: Stations with significant positive trend in the 0.95 quantile (blue triangle) and non-significant trend (red dot) at the 5% level, as obtained by MQR.
CHAPTER 5. TRENDS IN MODERATE RAINFALL EXTREMES

the results of the test at the 5% significance level for the annual, winter (December–February), and summer (June – August) 0.95 quantiles. While in summer we only have a few significant trends, in winter an overwhelming majority of the sites shows a significant increase. For the whole year most stations also exhibit a significant increase, which can be expected given the increase in the winter season. In general this is in line with earlier work about trends in extreme precipitation in the Netherlands, as discussed in section 5.2. For comparison the spatial distribution of significant trends obtained by the MK test is shown in Fig. 5.5. The differences between the MQR and the MK approach are small for winter and annual quantiles. In summer the differences are somewhat larger. There are more significant trends detected by the MQR framework than by the MK test, but the spatial distribution of stations with a significant trend is more scattered.

The relative change over the period 1910–2009, with respect to the estimate for 1910, was spatially interpolated over the Netherlands, using simple kriging (Gaetan and Guyon 2010). Figures 5.6, 5.7, and 5.8 show the interpolated, relative increase in the 0.95, 0.975, and 0.99 quantile. In summer we observe a clear coastal effect in the west of the country, which is strongest in the 0.99 quantile. A somewhat weaker coastal effect is also visible in winter, where we do not observe any quantile dependence of the effect. For the annual values the coastal effect is less pronounced and can not be observed on the color scale used. The stronger the deviation from linearity the more sensitive is the magnitude of the trend to the choice of the smoothing parameter. In particular, the results for the 0.99 summer quantile turned out to be affected by this choice.
5.4. RESULTS

Figure 5.6: Relative increase in the 0.95 quantile.

Figure 5.7: Relative increase in the 0.975 quantile.
Figure 5.8: Relative increase in the 0.99 quantile.
5.4. RESULTS

5.4.2 Regional application

The smoothing parameters for the different regions and quantiles were chosen after a similar consideration as in the local setting. Table 5.1 shows the relative change over the century for the whole Netherlands (NL) and the subregions defined in section 5.2. The relative change is considerably larger in winter than in summer. Moreover, while the relative change in winter is about the same for the three quantiles it is decreasing for higher quantiles in the summer season and for the whole year. By aggregating the local relative changes, we come to a similar conclusion. Table 5.2 shows the corresponding \( p \)-values of the regional test for stationarity using the MQR approach. We see that all regions show a significant increase at the 5% level for the annual and winter quantiles. In summer we obtain \( p \)-values mostly below 0.05 for both western regions and mostly below 0.1 for the south east region as well as for the whole Netherlands. The results obtained from the regional MK test are quite similar. In fact only 1 (5) out of the 36 combinations of season, quantile, and region give different results at the 10% (5%) significance level.

Fig. 5.9 shows the trend in the standardized 0.99 quantile with pointwise confidence band for the NW region in winter. Although we clearly see two plateaus in the 1920s and late 1960s no large deviation from linearity can be detected, and indeed the trend in the standardized quantile is almost the same as for a linear fit. Fig. 5.10 shows the standardized 0.99 quantile in summer, where the overall trend is smaller than in winter. Moreover, linearity of the trend seems less clear than in winter and the relative change from MQR is larger than for the linear fit. However, the linearity test described in section 5.3 did not reject linearity at the 5% level and for larger smoothing parameters the trend is becoming more linear again.

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>Season</th>
<th>NL</th>
<th>SW</th>
<th>NW</th>
<th>NE</th>
<th>E</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>Summer</td>
<td>22.1</td>
<td>42.2</td>
<td>28.6</td>
<td>11.7</td>
<td>9.0</td>
<td>14.6</td>
</tr>
<tr>
<td>0.975</td>
<td>Summer</td>
<td>27.2</td>
<td>31.1</td>
<td>31.1</td>
<td>16.2</td>
<td>13.5</td>
<td>30.0</td>
</tr>
<tr>
<td>0.95</td>
<td>Summer</td>
<td>38.5</td>
<td>43.8</td>
<td>37.4</td>
<td>22.7</td>
<td>23.0</td>
<td>40.9</td>
</tr>
<tr>
<td>0.99</td>
<td>Winter</td>
<td>37.3</td>
<td>44.7</td>
<td>39.1</td>
<td>46.2</td>
<td>44.1</td>
<td>32.8</td>
</tr>
<tr>
<td>0.975</td>
<td>Winter</td>
<td>37.0</td>
<td>45.6</td>
<td>41.7</td>
<td>37.8</td>
<td>36.7</td>
<td>32.6</td>
</tr>
<tr>
<td>0.95</td>
<td>Winter</td>
<td>43.5</td>
<td>46.0</td>
<td>43.1</td>
<td>29.6</td>
<td>32.7</td>
<td>39.6</td>
</tr>
<tr>
<td>0.99</td>
<td>Year</td>
<td>28.8</td>
<td>27.9</td>
<td>27.7</td>
<td>25.9</td>
<td>23.7</td>
<td>26.3</td>
</tr>
<tr>
<td>0.975</td>
<td>Year</td>
<td>33.2</td>
<td>32.0</td>
<td>27.7</td>
<td>26.7</td>
<td>34.5</td>
<td>32.9</td>
</tr>
<tr>
<td>0.95</td>
<td>Year</td>
<td>37.2</td>
<td>36.4</td>
<td>38.5</td>
<td>32.4</td>
<td>41.2</td>
<td>34.4</td>
</tr>
</tbody>
</table>
Table 5.2: Resulting $p$-values in percent for the MQR test of the stationary hypothesis versus the one-sided monotone alternative.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Season</th>
<th>NL</th>
<th>SW</th>
<th>NW</th>
<th>NE</th>
<th>E</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>Summer</td>
<td>8.6</td>
<td>2.2</td>
<td>1.1</td>
<td>32.0</td>
<td>33.5</td>
<td>22.7</td>
</tr>
<tr>
<td>0.975</td>
<td>Summer</td>
<td>9.2</td>
<td>5.8</td>
<td>2.2</td>
<td>47.6</td>
<td>22.6</td>
<td>7.2</td>
</tr>
<tr>
<td>0.95</td>
<td>Summer</td>
<td>4.3</td>
<td>4.9</td>
<td>2.5</td>
<td>18.0</td>
<td>24.0</td>
<td>5.2</td>
</tr>
<tr>
<td>0.99</td>
<td>Winter</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.1</td>
<td>0.2</td>
</tr>
<tr>
<td>0.975</td>
<td>Winter</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>0.95</td>
<td>Winter</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>0.99</td>
<td>Year</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>0.975</td>
<td>Year</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.95</td>
<td>Year</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Figure 5.9: Monotone (solid) and linear (dashed) trend in the standardized 0.99 quantile in winter (NW region). The gray area indicates a pointwise 95% confidence band for the monotone trend.
5.5 Conclusion

With the presented MQR approach we provide an additional tool to test the assumption of increasing rainfall extremes. MQR represents a unified method for visualizing non-decreasing trend behavior and testing the stationary hypothesis against the one-sided alternative of a monotone trend. The method can be used for local quantile regression as well as for the estimation of a common regional trend, given site-specific quantiles. Similar results regarding the significance of trends were obtained with MQR and the MK test, both in the local and regional application. Unlike the formal MK test MQR also visualizes the trend. Furthermore, MQR does not rely on yearly quantiles which might be substantially biased and allows for additional testing of the trend structure, e.g. nonlinear versus linear behavior. However, while the algorithm for local and regional MQR works fast, it is computationally more demanding than the calculation of the Kendall score. Therefore, Monte Carlo experiments take considerably more time.

Our analysis strengthens the assumption that the annual and winter rainfalls have increased throughout the Netherlands. In summer the behavior is subject to a coastal effect, as previously discovered by Lenderink et al. (2009) and supported here, see also Daniels et al. (2013). The increase in summer seems to accelerate in recent years, however, a significant deviation from linearity could not be detected. This finding supports the widespread use of linear trends in the analysis of high precipi-
tation quantiles. Buishand et al. (2013) detected a significant deviation from linearity in the trend of the spatially averaged annual number of days exceeding 30 mm rain-fall over the Netherlands. This is not in contradiction to the findings reported here, because 30 mm represents a higher quantile than considered in this chapter (in the number of days exceeding 20 mm no significant non-linearity was found). However, it emphasizes, together with the reported summer situation, the need for continued attention regarding non-linearity.

Previously linear quantile regression has been used to determine thresholds for subsequent peaks-over-threshold analysis in local and regional settings, compare Chapter 2. This can be generalized directly using the proposed MQR approach. Moreover, MQR could be a valuable tool for the assessment of the scaling of extreme precipitation with temperature (compare Lenderink and Van Meijgaard 2008; Wasko and Sharma 2014).

Acknowledgments

The data set was kindly made available by T. Brandsma.
Monotone trends in the GPD scale parameter

Abstract. For a sample of independent and identically distributed (i.i.d.) observations, the parameters of the Generalized Pareto Distribution (GPD) can be estimated by the method of Maximum Likelihood. In applications both the assumption of independence and that of identical distributions can be too strong. We keep up the independence assumption in this chapter. Instead of assuming the random variables to be identically distributed, we assume a common shape parameter and an increasing trend in the scale parameter of the GPD. Estimating ordered parameters of distributions is common in the area of isotonic regression. We use ideas and tools from that area to compute Maximum Likelihood estimates of the GPD parameters. We also study the estimator in a simulation study.
6.1 Introduction

Modeling of extremes is crucial in many branches of modern society. Examples include finance, insurance, and the planning of critical infrastructure such as dikes or sewer systems. Often the Generalized Pareto Distribution (GPD) is used to model the tail of the distribution, which is justified by the Pickands–Balkema–De Haan theorem. It states that, under certain regularity conditions, the distribution of i.i.d. excesses over a threshold $u$ can be approximated by a GPD, if $u$ is sufficiently high (Reiss and Thomas 2007). We consider the two-parameter GPD with $\xi \in \mathbb{R}$ and $\sigma > 0$ denoting the shape and scale parameter, respectively. Its cumulative distribution function (cdf) is given by

$$G_{\xi,\sigma}(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi},$$

with support $y \geq 0$ for $\xi \geq 0$ and $0 \leq y \leq -\frac{\sigma}{\xi}$ for $\xi < 0$. For $\xi = 0$ the GPD reduces to the exponential distribution. The density of the GPD in the case $\xi \neq 0$ is given by

$$g_{\xi,\sigma}(y) = \frac{1}{\sigma} \left(1 + \frac{\xi y}{\sigma}\right)^{-\frac{1}{\xi}-1},$$

on its support.

Parameter estimates can be obtained, for instance, using the Maximum Likelihood (ML) framework. This works fine for $\xi > -0.5$ (Embrechts et al. 1997) and does not pose a severe restriction for our subject as applications in hydrology usually exhibit shape parameters in the interval $(-0.5, 0.5)$ (Hosking and Wallis 1987). Therefore, we restrict ourselves to the case $\xi > -0.5$. The likelihood equations can only be solved numerically, usually by the Newton-Raphson approach or variants including gradient descent steps (Hosking 1985; Embrechts et al. 1997). Hosking and Wallis (1987) show that for sample sizes up to 500 the probability weighted moment estimators and moment estimators have smaller root mean squared error than the Maximum Likelihood Estimators (MLEs) for $\xi \in [0, 0.4]$ and $\xi \in [-0.4, 0)$, respectively. However, these approaches lack the flexibility of the ML method, which is necessary for the inclusion of trends.

In many applications, there are reasons to expect a monotone trend in the behavior of extremes. For insurance and infrastructure planning, climate change may lead to such a monotone trend. The change can be described using a monotone function of time, other covariates might be considered as well. In this chapter we consider the problem of estimating a monotone trend in the scale parameters of independent random variables from a GPD with non-decreasing trend in the scale parameter.
6.2 Maximum Likelihood estimation

Suppose that \( Y_1, \ldots, Y_n \) are independent random variables, such that \( Y_i \sim G_{\xi, \sigma_i} \) for some \( \xi > -0.5 \) and \( 0 < \sigma_1 \leq \cdots \leq \sigma_n \). We want to estimate the shape parameter \( \xi \) and the vector of scale parameters \( \bar{\sigma} \in C \), where

\[
C = \{ \bar{\sigma} = (\sigma_1, \ldots, \sigma_n) \in (0, \infty)^n : \sigma_1 \leq \cdots \leq \sigma_n \}.
\]  

(6.3)

For this we consider the ML approach. Although, there exists no explicit expression for the MLE, it can be found numerically. Based on observed values \( \bar{y} = (y_1, y_2, \ldots, y_n) \), the log likelihood for \( \xi \) and \( \bar{\sigma} \) is given by

\[
\ell(\xi, \bar{\sigma}) = \sum_{i=1}^{n} \ln g_{\xi, \sigma_i}(y_i),
\]

(6.4)

where \( g_{\xi, \sigma} \) is the density of GPD as given in (6.2). Note that

\[
\ln g_{\xi, \sigma}(y) = \ln \left[ \frac{1}{\sigma} \left( 1 + \frac{\xi y}{\sigma} \right)^{-\frac{1}{\xi} - 1} \right] = \frac{1}{\xi} \left[ \ln(\sigma) - \ln(\sigma + \xi y) \right] - \ln(\sigma + \xi y),
\]

yielding (for \( \xi \neq 0 \))

\[
\ell(\xi, \bar{\sigma}) = \sum_{i=1}^{n} \left( \frac{1}{\xi} \left[ \ln(\sigma_i) - \ln(\sigma_i + \xi y_i) \right] - \ln(\sigma_i + \xi y_i) \right).
\]

The maximizing argument \( (\hat{\xi}, \hat{\bar{\sigma}}) \) of the likelihood function in Eq. (6.4) is the MLE of \( \xi \) and \( \bar{\sigma} \).

One way to approach maximizing \( \ell \) over \( C \times (-0.5, \infty) \) is via a two-step procedure using the profile (log) likelihood. In this approach, for a fine grid of possible \( \xi \)-values, the profile likelihood is constructed, i.e.

\[
\ell_p(\xi) = \max_{\bar{\sigma} \in C} \ell(\xi, \bar{\sigma}).
\]

(6.6)

For each \( \xi \), the log likelihood \( \ell \) is maximized over \( \bar{\sigma} \). As \( \xi \) is one-dimensional, this profile likelihood can be visualized. Of course, in order for this to be useful, a method should be designed to actually compute the profile likelihood, i.e., to maximize \( \ell \) over \( C \) for fixed \( \xi \).
Lemma 1 For each $\xi \geq -0.5$, there exists a $\vec{\sigma}_\xi \in C$ such that

$$\ell(\vec{\sigma}_\xi, \xi) \geq \ell(\vec{\sigma}, \xi)$$

for all $\vec{\sigma} \in C$.

Consequently, $\ell_p$ given in (6.6) is well defined.

Proof Fix $\xi > 0$ and note that $\vec{\sigma} \mapsto \ell(\xi, \vec{\sigma})$ is continuous. Moreover, use (6.5) to note that for $y > 0$ fixed and $\sigma \downarrow 0$,

$$\ln g_{\xi, \sigma}(y) \sim \frac{1}{\xi} \ln \sigma \to -\infty$$

and for $\sigma \to \infty$,

$$\ln g_{\xi, \sigma}(y) \sim -\ln \sigma \to -\infty.$$ 

Therefore, in maximizing $\vec{\sigma} \mapsto \ell(\xi, \vec{\sigma})$ over $C$, attention can be restricted to a compact subset of $C$, namely $\vec{\sigma} \in C$ for which $\delta \leq \sigma_1 \leq \sigma_n \leq 1/\delta$ for some small $\delta > 0$. This ensures the existence of $\vec{\sigma}_\xi$.

For $\xi = 0$, $\ln g_{0, \sigma}(y) = -\ln \sigma - y/\sigma$, leading to the same conclusion. In case $\xi \in (-0.5, 0)$, the restriction $y \leq -\xi y$ implies that $\sigma \geq -\xi y$. For $\sigma \to -\xi y$ we obtain

$$\ln g_{\xi, \sigma}(y) \sim \left(-\frac{1}{\xi} - 1\right) \ln(\sigma + \xi y) \to -\infty,$$

due to the fact that $\left(-\frac{1}{\xi} - 1\right) > 0$ for $\xi \in (-0.5, 0)$. For $\sigma \to \infty$ we obtain as before

$$\ln g_{\xi, \sigma}(y) \sim -\ln(\sigma) \to -\infty.$$ 

Thus, attention can be restricted again to a compact subset of $C$, namely $\vec{\sigma} \in C$ for which

$$\min_{1 \leq i \leq n} -\xi y_i + \delta \leq \sigma_1 \leq \sigma_n \leq 1/\delta$$

for some small $\delta > 0$.

It is interesting to note that for $\xi \neq 0$, the function $\vec{\sigma} \mapsto \ell(\xi, \vec{\sigma})$ is not concave, see the appendix to this chapter. Therefore, optimization algorithms that need this property cannot be used. In the next section, we will address the problem of computing the function $\ell_p$ and maximizing this in $\xi$ to maximize the full likelihood $\ell$.

6.3 Maximizing the log likelihood

In order to compute $\ell_p$ as defined in (6.6), a maximization has to be performed over the positive cone $C$ in $\mathbb{R}^n$, defined in (6.3). The case $\xi = 0$ is special in this respect.
6.3. MAXIMIZING THE LOG LIKELIHOOD

As can be seen in Section 1.5 in Robertson et al. (1988), the optimization problem for \( \xi = 0 \) is a special case of the Gamma extremum problem. The solution of this problem is given by

\[
\tilde{\sigma} = pr(\bar{y}),
\]

where \( pr \) is the projection operator from \( \mathbb{R}^n \) onto \( C \), defined by

\[
pr(\bar{y}) = \arg \min \{ ||\tilde{\sigma} - \bar{y}||_2 : \tilde{\sigma} \in C \},
\]

(6.7)

where \( || \cdot ||_2 \) is the Euclidean norm.

An elegant way to obtain the projection \( pr(\bar{y}) \) explicitly is via the derivative of the greatest convex minorant of a diagram of points. More specifically, defining \( P_0 = (0, 0) \) and

\[
P_j = \left( j, \sum_{i=1}^{j} y_i \right), \quad 1 \leq j \leq n,
\]

(6.8)
one can construct the greatest convex function lying entirely below the diagram of points. Then taking the left derivative of this function at \( j \), gives \( \tilde{\sigma}_j \). By construction, the vector \( (\tilde{\sigma}_1, \ldots, \tilde{\sigma}_n) \) is in \( C \). The projection gives the unweighted least squares isotonic regression of \( (y_1, \ldots, y_n) \) (Robertson et al. 1988, Lemma 1.2.1).

For \( \xi \neq 0 \) such a connection between the MLE of ordered scale parameters in a GPD model and isotonic regression does not exist. In order to compute \( \ell_p(\xi) \) for values \( \xi \neq 0 \), an iterative algorithm needs to be employed. A possible algorithm that can be used in this setting, is the projected gradient method, developed independently by Goldstein (1964) and Levitin and Polyak (1966) for the minimization of a continuously differentiable function on a convex subset of \( \mathbb{R}^n \). The gradient projection algorithm is defined by

\[
x_{k+1} = pr \left[ x_k - a_k \nabla f(x_k) \right],
\]

(6.9)

where \( a_k > 0 \) is the step size and \( f \) the profile log likelihood. Consider the following Goldstein-Armijo type choice for the step size

\[
a_k = \beta^{m_k} s,
\]

(6.10)

with \( m_k \) the smallest integer \( m \), such that

\[
f(x_{k+1}) \leq f(x_k) - \mu \cdot < \nabla f(x_k), x_{k+1} - x_k >,
\]

(6.11)

where \( s > 0, \beta \in (0, 1), \) and \( \mu \in (0, 1) \) are given scalars and \( < \cdot, \cdot > \) denotes the standard scalar product. Bertsekas (1976) and Gafni and Bertsekas (1982) showed that in this setting every limit point of \( \{x_k\} \) is stationary. The projected gradient
method prescribes for each iterate multiple projections of the scaled gradient.

An alternative algorithm that can be used to compute $\ell_p(\xi)$ is the Iterative Convex Minorant (ICM) algorithm introduced by Jongbloed (1998) and already used for Monotone Quantile Regression (MQR) in Chapter 5. For this approach only one line search using the projected gradient is needed. As stated above the considered negative log-likelihood is not convex and so it is not natural to choose the weights in the algorithm as the diagonal entries of the Hessian. However, the choice of a quadratic approximation is quite flexible in the ICM algorithm so we decide to choose the identity matrix for this. One step in the ICM algorithm can then be described by

$$x_{k+1} = x_k + \alpha_k \left( pr \left[ x_k - \nabla f(x_k) \right] - x_k \right). \quad (6.12)$$

The scaling constant $\alpha_k$ can again be chosen as in Eq. (6.11). The ICM uses only one projection for each iterate, in contrast to the projected gradient method. The name Iterative Convex Minorant algorithm stems from the computation of the (iterative) projections via the greatest convex minorant of a point diagram similar to (6.8).

Note the geometric difference between the gradient projection algorithm and the ICM algorithm. In the first algorithm, in principle a whole line segment connecting the current iterate $x_k$ and and $x_k - \nabla f(x_k)$ is projected, leaving a trace on the cone $C$ that is in general not a line segment, but a ‘broken line’. The ICM algorithm just takes the point $x_k - \nabla f(x_k)$ and projects it on $C$. Then a new iterate is chosen from the line segment connecting $x_k$ and this projection, a line that lies completely within $C$ due to convexity of $C$.

Having got an algorithm that can be used to compute the profile (log) likelihood function $\ell_p$ on a grid of $\xi$-values, the next step is to plot the function $\ell_p$ on such a grid and find its maximum.

We carried out a small simulation study using the shape parameter $\xi^* = 0.2$ and the scale parameter vector $\sigma^*$ as shown in Fig. 6.1. For the implementation of the algorithms we use the expressions for the gradient vector as given in the Appendix. As starting values for the scale parameter vector we considered the original scale parameter vector $\sigma^*$, the mean $\bar{y}$ and $pr(\bar{y})$. The different starting values and both algorithms yielded the same scale parameter estimates. While in many cases the projected gradient method is only slightly slower than the ICM approach, there are cases where the projected gradient method takes considerably longer than the ICM approach, see Fig. 6.2. Surprisingly, there was hardly any difference in the computation time, when starting from $\sigma^*$ and $\bar{y}$. However, taking $pr(\bar{y})$ as starting value, reduces the computation time by a factor of 2 on average.
6.3. MAXIMIZING THE LOG LIKELIHOOD

Figure 6.1: The scale parameter vector used in the simulation.

Figure 6.2: Density of the ratio between the computation time of the projected gradient and the ICM approach. The dashed vertical ratio indicates the mean ratio.
Figure 6.3 shows one point estimate of the scale parameter and a 95% confidence band obtained by a parametric bootstrap, where the shape parameter was assumed to be known. The spiking problem mentioned in Chapter 5 appears here again.

The profile likelihood approach immediately returns profile likelihood asymptotic confidence intervals for the shape parameter, which are often assumed to be more accurate than standard ones (e.g. Coles 2001). Murphy and Van der Vaart (2000) justify the use of the profile likelihood confidence interval for semiparametric models. The profile likelihood confidence interval is based on the fact, that the profile deviance

\[ D_p(\xi) = 2\left(\ell(\hat{\xi}, \hat{\theta}) - \ell_p(\xi)\right) \]

converges to a \( \chi^2_1 \) distribution. Hence, by this it can be deduced that

\[ C_\alpha = \{\xi : D_p(\xi) \leq c_\alpha\}, \]

with \( c_\alpha \) being the \( (1 - \alpha) \) quantile of the \( \chi^2_1 \) distribution, constitutes a \( (1 - \alpha) \) asymptotic confidence interval for the shape parameter. Fig. 6.4 shows the 95% profile likelihood asymptotic confidence interval for one realization of the simulation.

Moreover, we applied the method to rainfall data. In the Netherlands often the number of days with more than 20 mm rainfall is analyzed (e.g. Buishand et al. 2013). Therefore, it seems interesting to study also the excesses over 20 mm. We focus on
6.3. MAXIMIZING THE LOG LIKELIHOOD

The observations from De Bilt and Den Bommel for the 100-year period 1910–2009, which are also included in the data used in Chapter 5. In total we have 393 excesses for De Bilt and 391 for Den Bommel. Fig. 6.5 shows the estimated scale parameter vectors. The scale parameter estimates show spikes at the endpoints in Den Bommel but not in De Bilt. Fig. 6.6 shows profile likelihood asymptotic confidence intervals for the shape parameter in Den Bommel and De Bilt. The shape parameter estimate

Figure 6.4: Profile likelihood with 95% confidence interval for the shape parameter for the simulated data. The dotted red line marks the true shape parameter.

Figure 6.5: Scale parameter estimate for De Bilt (red) and Den Bommel (blue).
for Den Bommel is $0.12 \ [0.032, 0.233]$, which is expected. But the shape parameter estimate for De Bilt is $0.033 \ [-0.093, 0.12]$, which seems to be too low. The spiking might be addressed in a similar way as in Chapter 5 and a more accurate shape parameter might be obtained by a regionalized likelihood function using a common shape parameter and a common form of the scale parameter.

6.4 Conclusion and further directions

We have developed a two-stage procedure to find a maximizer of the log likelihood based on a sample of independent random variables from a GPD with common shape parameter and non-decreasing scale parameter. The first step is to compute the profile (log) likelihood for fixed values of $\xi$. For this step we describe and test two algorithms, a gradient projection algorithm and an ICM algorithm. Open problems that still need to be resolved are uniqueness of the MLE, in the absence of concavity of the log likelihood function. Our conjecture, supported by simulations, is that the MLE is uniquely defined. Another further direction would be to use the MLE to construct significance tests for the null hypothesis that the scale parameters are equal against the alternative that these are increasing. Likelihood ratio tests, but also permutation-based tests can be studied using the algorithms described in this chapter. Moreover, it seems interesting to consider a regional setting, with a common shape parameter and a common form of the scale parameter.
Partial derivatives Consider the first partial derivative
\[
\frac{\partial \ln g_{\xi,\sigma}(y)}{\partial \sigma} = \frac{y - \sigma}{\sigma(\sigma + \xi y)}.
\]
This shows that \(\sigma \mapsto \ln g_{\xi,\sigma}(y)\) is unimodal with maximum \(\sigma = y\) for fixed \(\xi\). The second derivative is given by
\[
\frac{\partial^2 \ln g_{\xi,\sigma}(y)}{\partial \sigma^2} = \frac{-\sigma(\sigma + \xi y) - (y - \sigma) [(\sigma + \xi y) + \sigma]}{\sigma^2(\sigma + \xi y)^2} \]
\[
= \frac{1 - 2\frac{y}{\sigma} - \xi \left(\frac{y}{\sigma}\right)^2}{(\sigma + \xi y)^2}.
\]
It follows that
\[
\frac{\partial^2 \ln g_{\xi,\sigma}(y)}{\partial \sigma^2} = 0
\]
\[
\Leftrightarrow 1 - 2\frac{y}{\sigma} - \xi \left(\frac{y}{\sigma}\right)^2 = 0
\]
\[
\Leftrightarrow \frac{y}{\sigma} = \frac{2 \pm \sqrt{4 + 4\xi}}{-2\xi}
\]
\[
\Leftrightarrow \frac{y}{\sigma} = -\left(1 \pm \sqrt{1 + \xi}\right)/\xi
\]
For \(\xi < 0\) we have the additional restriction \(\frac{y}{\sigma} \leq -\frac{1}{\xi}\). Therefore, it follows that for \(\xi < 0\) only the solution
\[
\frac{y}{\sigma} = -\left(1 - \sqrt{1 + \xi}\right)/\xi
\]
is admissible. This shows, that the second derivative exhibits in general at least one change of sign. Hence, we obtain that the log likelihood is not concave for \(\xi \neq 0\).
A.1 Bootstrap and composite likelihood confidence intervals

For the comparison of bootstrap and composite likelihood confidence intervals a small simulation study was carried out. Each sample consisted of a 10 by 100 matrix of spatially dependent (column vectors) and serially independent (row vectors) GPD random variables, representing 100 joint excesses at 10 different stations. Excesses were simulated at all stations simultaneously to reduce computation time. The spatial dependence of the series was modeled by a normal copula with $10 \times 10$ correlation matrix $R$ with parameter $\rho$, i.e.

$$
R = \begin{pmatrix}
1 & \rho & \rho^2 \\
\rho & 1 & \rho \\
\rho^2 & \rho & 1 \\
\vdots & \ddots & \rho \\
\rho & \rho & 1
\end{pmatrix}.
$$

The common GPD shape parameter and dispersion coefficient were fixed to 0.1 and 0.5, respectively. The 10 different location parameters were obtained from a normal distribution with mean 10 and standard deviation 0.5. The chosen parameters resemble the estimated values in Chapter 2.

For each simulated sample the confidence interval for the shape parameter was
Figure A.1: Mean confidence limits for the shape parameter (blue - bootstrap, red – composite likelihood).

Figure A.2: Coverage probability of the derived confidence interval (blue – bootstrap, red – composite likelihood). The expected coverage probability is 0.95.

calculated using the regional block bootstrap and the composite likelihood method, compare Varin et al. (2011) and Chapter 2. This was repeated 2500 times and then the mean confidence limits were obtained (shown in Fig. A.1). The uncertainty reduction for decreasing $\rho$ is as expected. However, the composite likelihood appears to be somewhat more conservative. While the mean lower limit of the confidence inter-
vals are almost the same, the mean upper limit is smaller for the bootstrap method, in particular for $\rho = 1$. To estimate the coverage probability of the derived confidence intervals the number of samples where $\xi$ is in the confidence interval were counted. Fig. A.2 shows that while the coverage probability of the composite likelihood confidence interval appears relatively stable, slightly above 0.95, the coverage probability of the bootstrap confidence interval increases with decreasing $\rho$ but remains below the expected 0.95. The composite likelihood confidence interval thus seems favorable to the bootstrap one. Also in the case of little spatial dependence a higher coverage probability of the composite likelihood confidence interval compared to that of the bootstrap is observed.

### A.2 Bias in climate model simulations

In Chapter 3 two climate model simulations were used to project rainfall extremes in a future climate. A bias correction was considered which assumes that the bias remains constant over the projection period. Maraun (2012) found, comparing several climate model simulations with the same transient boundary conditions, that the bias in seasonal precipitation sums is in general relatively stable and that bias corrected values are usually preferable to the original values. This is especially the case, when smoothing over larger areas is applied. It is harder to validate the bias correction of extremes, owing to the large natural variability.

We computed the 5- and 50-year return levels for the observations in the period 1951–1980 and the period 1981–2010 to obtain two bias corrected return levels for the climate model simulations. Fig. A.3 shows the 50-year return level for the R/MIROC simulation in the summer season. In Chapter 3 it was noted that this simulation largely overestimates the observed 50-year return level. This bias stays virtually the same for the two considered sub-periods and a strong improvement of the corrected return level compared to the modeled one can be seen. In the summer season this bias reduction is also observed for the 50-year return level of the R/ECHAM5 simulation and for the 5-year return level of both simulations. Fig. A.4 shows the 50-year return level for the R/ECHAM5 simulation in the winter season. Here, the bias is quite different for both sub-periods. This is owing to the large natural variability in rare extremes. In particular a very exceptional event in December 1960 has a large influence on the shape parameter, compare Chapter 2. The bias correction using the sub-period data only is then very uncertain. However, when we look at less extreme values such as the 5-year return level, see Fig. A.5 we see again little difference in the bias for the two sub-periods and bias correction leads again to an improvement over the modeled return level. This small experiment indicates that bias corrected values are favorable compared to uncorrected model output. Moreover, we see that in the
Figure A.3: 50-year return level for the summer season of R/MIROC as obtained with the non-stationary POT model used in Chapter 3 (solid black); 50-year return level for the period 1950–2011 (1951–1980, 1981–2011) assuming stationarity – solid red (green, blue); dotted colors – R/MIROC corrected with observations for the corresponding (sub-) period.

In Chapter 5 only $c = 0.1$ was used for the quantile estimation. In the following we study shortly the influence of $c$ for the estimation of the 0.95 quantile of daily pre-
Figure A.4: 50-year return level for the winter season of R/ECHAM5 as obtained with the non-stationary POT model used in Chapter 3 (solid black); 50-year return level for the period 1950–2011 (1951–1980, 1981–2011) assuming stationarity – solid red (green, blue); dotted colors – R/MIROC corrected with observations for the corresponding (sub-) period.

Figure A.5: Same as Fig. A.4 but for the 5-year return level.

precipitation for the 100-year rainfall record of De Bilt. Fig. A.6 shows the constant 0.95 quantile, obtained with different values of $c$. It is obvious that there is basically no difference from the quantile obtained by the check function and its smooth approx-
Figure A.6: Quantile obtained with different values of $c$. The dashed line gives the value as obtained with the standard quantile check function $\rho_\tau(x)$.

Figure A.7: Evolution of the quantile, obtained with different values of $c$.

imation for $c \leq 1$. Even for $c = 10$, i.e. where $\rho_\tau$ is approximated on the intervals $[-9.5, 0)$ and $[0, 0.5)$, the difference is smaller than 3%. Fig. A.7 shows a very similar pattern for the monotone quantile estimator. Therefore, it is concluded that the use of $c = 0.1$ in Chapter 5 is justified.


Kyselý, J., J. Picek, and R. Beranová (2010). Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary thresh-


BIBLIOGRAPHY


## Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>Anderson–Darling</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian Information Criterion</td>
</tr>
<tr>
<td>BM</td>
<td>Block Maxima</td>
</tr>
<tr>
<td>EV</td>
<td>Extreme Value</td>
</tr>
<tr>
<td>EVT</td>
<td>Extreme Value Theory</td>
</tr>
<tr>
<td>GCM</td>
<td>General Circulation Model</td>
</tr>
<tr>
<td>GEV</td>
<td>Generalized Extreme Value</td>
</tr>
<tr>
<td>GoF</td>
<td>Goodness of Fit</td>
</tr>
<tr>
<td>GPD</td>
<td>Generalized Pareto Distribution</td>
</tr>
<tr>
<td>ICM</td>
<td>Iterative Convex Minorant</td>
</tr>
<tr>
<td>IF</td>
<td>Index Flood</td>
</tr>
<tr>
<td>KS</td>
<td>Kolmogorov–Smirnov</td>
</tr>
<tr>
<td>ME</td>
<td>Mean Excess</td>
</tr>
<tr>
<td>MILE</td>
<td>Maximum Independence Likelihood Estimator</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum Likelihood</td>
</tr>
<tr>
<td>MLE</td>
<td>Maximum Likelihood Estimator</td>
</tr>
<tr>
<td>MK</td>
<td>Mann–Kendall</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>MQR</td>
<td>Monotone Quantile Regression</td>
</tr>
<tr>
<td>POT</td>
<td>Peaks-Over-Threshold</td>
</tr>
<tr>
<td>RCM</td>
<td>Regional Climate Model</td>
</tr>
<tr>
<td>RFA</td>
<td>Regional Frequency Analysis</td>
</tr>
<tr>
<td>TS</td>
<td>Threshold Stability</td>
</tr>
<tr>
<td>WGPD</td>
<td>hybrid Weibull–GPD</td>
</tr>
<tr>
<td>cdf</td>
<td>cumulative distribution function</td>
</tr>
<tr>
<td>i.i.d.</td>
<td>independent and identically distributed</td>
</tr>
</tbody>
</table>
Many practical problems require knowledge of the behavior of extreme values. In particular, the infrastructure we depend upon for food, water, energy and transportation is sensitive to meteorological extremes, such as high precipitation amounts and resulting floods. Most existing systems for water management and other infrastructure have been designed under the assumption of a stationary climate. This basic concept assumes that weather variations occur, but with properties which are constant over time, and which fluctuate around an unchanging steady state.

Nowadays, it is widely accepted that climate change may alter the mean, variability and extremes of relevant weather variables. Still, assuming stationarity is common practice for establishing design criteria for new infrastructure. Trends in extreme rainfall might have substantial influence on these design criteria, and potentially lead to an increased failure probability.

Including trends in the analysis of extreme rainfall adds to the large uncertainty in the estimation of return levels. The peaks-over-threshold approach often makes more efficient use of the available data than the block maxima approach, which reduces the uncertainty. Moreover, to obtain more precise quantile estimates regional approaches have been used for analyzing extreme rainfall in a stationary climate. This dissertation introduces a regional, non-stationary peaks-over-threshold model.

The model relies on a threshold chosen via linear quantile regression. For the excesses Generalized Pareto Distributions are assumed. The ratio of the scale parameter to the threshold and the shape parameter are defined to be constant over the region but may vary over time. A consequence of this is that the ratio between different return levels is constant over the region.

For The Netherlands, observed precipitation and simulated precipitation from a Regional Climate Model are analyzed with the proposed approach. Although a
substantial reduction of the uncertainty compared to at-site approaches is observed, the remaining uncertainty is still considerable, due to the spatial dependence. Furthermore, a method to correct the climate model output for systematic differences compared to observational data is presented.

A common challenge in the peaks-over-threshold approach is the threshold selection. This is owing to a trade-off situation between a poor approximation of the distribution tail and large estimation uncertainty for low and high thresholds, respectively. It seems natural to view also this challenge from a regional perspective. Therefore, regional methods for threshold selection based on graphical tools and, more quantitatively, on spatially averaged Goodness of Fit statistics are introduced. These regional methods lead to an increased sensitivity for too low thresholds. The methods are studied in a simulation setting and further illustrated using precipitation records from the Dutch water board Vallei en Veluwe.

Quantile regression is used to estimate monotone trends in moderate precipitation extremes. While quantile regression in general is quite robust, the trend estimation for high quantiles becomes less reliable. To increase precision of the trend estimation, a regional model for monotone quantile regression is introduced. This model is used to study trends in the 0.95, 0.975, and 0.99 quantiles of daily precipitation in the Netherlands in 100-year records. An increasing intensity is found for winter and annual precipitation. However, in summer the trend is subject to a coastal effect. To allow focusing on monotone trends in even higher quantiles, two methods for fitting a Generalized Pareto Distribution with monotone trend in the scale parameter are presented.
Curriculum Vitae

Martin Roth was born on September 18, 1984 in Regensburg, Germany. He finished secondary school at Bernhard-Strigel-Gymnasium Memmingen in 2004. In October 2004 he began his studies in mathematics and physical geography at the Albert-Ludwigs-University in Freiburg, which he completed in October 2010. From summer 2007 to summer 2010 he worked part-time at the University as a tutor and research assistant. In spring 2008 he carried out a 2-month internship at AON re services, in Hamburg, focusing on nature catastrophe modeling.

In November 2010, Martin started a PhD project at EURANDOM (TU Eindhoven) and KNMI. His research on the statistical analysis of extreme rainfall was part of the Knowledge for Climate research program. Supervisors were Adri Buishand (KNMI), Geurt Jongbloed (EURANDOM/TU Delft), Harry van Zanten (EURANDOM/University of Amsterdam), and Albert Klein Tank (KNMI). Results of his research were presented at conferences in Liberec, Nemcicky, Amsterdam, Shanghai, Siegen, and Hejnice.