Erasmus University Rotterdam

Erasmus School of Economics

Master Thesis Quantitative Finance

Analyzing trends in extreme precipitation for the Netherlands and Northern-Germany

February 3, 2022

Abstract

The many negative consequences of extreme precipitation events make the subject interesting for climate econometrics. This paper analyzes trends in the extreme precipitation for the Netherlands and Northern Germany over the period 1950-2020. First, the trend in the extreme value index is evaluated to get insight into possible changes in the distribution of extremes. This analysis uses local hill estimation using the SAMSEE method to optimize objectivity. Second, the trend in the frequency of extremes is evaluated for stations individually and regions incorporating spatial dependency. This analysis is based on the heteroskedastic extremes framework introduced by Einmahl et al. (2016). We find no clear statistical evidence for trends in extremes except for an increasing trend in the frequency of extremes for coastal regions in the winter period.

Keywords: Precipitation, Varying Extreme Value Indices, Heteroskedastic Extremes

Author	Y.J.A. Wanders
Supervisor	Prof. dr. D.J.C. van Dijk
Second Assessor	Prof. dr. C. Zhou

Contents

1 Introduction

At present day, climate change is one of the topics high on the political, social, and academic agenda. Within the climate change dispute, climate econometrics concentrates on developing methods to increase our knowledge of the climate, the impact of the changing climate, and the human impact thereon (e.g. Cabral et al., 2020; Diebold and Rudebusch, 2021; Konisky et al., 2016). Changes in temperature, wind speeds, precipitation, drought, or any other meteorological topics that fall within climate econometrics are all much discussed. This paper focuses on the changing behavior of precipitation. Precipitation is directly influenced by rising temperatures and so a great gauge for climate change. Additionally, we observe more natural disasters caused by extreme precipitation, most recently the flooding through extreme precipitation in India/Nepal[1], Germany/Belgium/Netherlands[2], and Italy[3]. These extremes of heavy rainfall bring numerous negative physical effects on ecosystems, food supply, infrastructure, buildings, and even human lives. Subsequent risks for the financial sector, transport sector, hospitality sector, and the food sector increase the total costs of heavy rainfall. Therefore, we choose to analyze precipitation, more specifically, trends in extreme precipitation, as it is of vital importance to fully understand the developments in extreme precipitation such that mitigation strategies ranging from management policies to infrastructure adaptation can be correctly developed.

There are two scenarios connected to trends in extreme events. First, one can look at a trend in the extreme value index. The extreme value index measures the heaviness of the right tail and the heavier the tail, the higher the index. A heavier right tail means that there is a higher probability of observing very large outcomes. Hence, we measure if larger extremes become more or less likely over time. Second, one can look at a situation where the extreme value index is constant, but there are heteroscedastic observations in the tail. In this case, we evaluate the trend of frequency of observations above a certain high threshold. Hence, we measure if the frequency of extremes increases or decreases over time. In our analysis, both scenarios are combined.

For our research, we use daily precipitation levels from 1951 until 2020 of 150 gauging stations spread across the Netherlands and northern Germany. This data is provided by Klein Tank et al. (2002)[4]. Of all the stations in this region, these 150 are filtered on proximity, such that they are

[1] https://edition.cnn.com/2021/07/23/india/floods-landslides-maharashtra-intl-hnk/index.html

[2] https://www.worldweatherattribution.org/heavy-rainfall-which-led-to-severe-flooding-in-western-europe-made-more-likely-by-climate-change/

[3] https://www.bbc.com/news/world-europe-59048809

[4] I acknowledge the data providers in the ECA&D project. Klein Tank, A.M.G. and Coauthors, 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int.

evenly distributed across the region. The daily rainfall data is preprocessed to ensure temporal independence. Using this data, the research focuses on finding an answer to the question: "What trends can be observed in the extreme value index and frequency of extreme precipitation in the Netherlands and northern provinces of Germany over the period 1951-2020?"

To analyze a (possible) trend in extreme precipitation, we look at the tail of the distribution of daily rainfall. To make inference on these extreme events one typically uses Extreme Value Theory (EVT). However, classical EVT considers independent and identically distributed observations, but we are interested in trends and thus non-identically distributed observations. More recent studies focus on cases where observations are drawn from different distributions such that there can be tested for trends. To evaluate the extreme value index of extreme precipitation we follow theory from de Haan and Zhou (2020). First, the extreme value index for every station is locally estimated at different points in time, which results in a plot of the extreme value index over time. This plot gives us visual insight into the trend of the index per station. To add statistical evidence to the research we formally test the hypothesis that the extreme value index is constant over time. If the test shows evidence for a varying extreme value index, the trend can be analyzed to draw conclusions on the developments in extreme precipitation at this location. For estimation of the extreme value index, we use a Hill estimator with the addition of the Smooth Absolute Mean Squared Error Estimator (SAMSEE) developed by Schneider et al. (2021). Using the Hill estimator is common in EVT literature, but the SAMSEE method is novel. This method selects the optimal number of upper order statistics to use for the Hill estimator without manual selection of the parameters. This choice for the number of order statistics is an important feature of EVT and with the SAMSEE method, we remove subjectivity.

A constant extreme value index allows us to evaluate the frequency. Here we use theory proposed by Einmahl et al. (2016) and further developed by Ferreira et al. (2017), Cabral et al. (2020), and Einmahl et al. (2020) on heteroscedastic extremes. The relative frequency of extreme events over time is estimated using a scedasis function. This function estimates the frequency of extremes by using the observations that exceed a certain high threshold. The scedasis function can be tested on trends over time and space. First, we test if the scedasis function is constant over time per station, thus evaluating if there is a trend in extremes for every station individually. Thereafter, we extend this to a space-time analysis in which we incorporate location.

The stations with a constant extreme value index are divided into regions based on location,

J. of Climatol., 22, 1441-1453. Data and metadata available at https://www.ecad.eu

average rainfall, the similarity of extreme value index, and threshold. This regionalization is made such that our preprocessing procedure does not lead to a large loss of extremal information. For these regions, we test if the extreme events are equally distributed over the stations, thereby yielding insight into the spatial dependency of extreme precipitation, and we test if the frequency of extremes for the complete region changes over time. Evidence from the tests allows us to draw conclusions on the trend of frequency per weather station over the whole region.

Using these methods we find no statistical evidence for a trend in the extreme value index. Hence, we find no evidence to suspect that precipitation distributions become more heavy-tailed and thus no significant evidence for increasing magnitude. We do find evidence for a changing frequency of extremes in the winter period. This trend is most notable in the coastal regions. In the summer period, we find no evidence for a trend in frequency for individual stations. The space-time analysis shows no regional trends. However, it becomes clear that extremes are more frequent toward the coast in the winter period and less frequent in the summer. We conclude that precipitation extremes are most vulnerable to increases in frequency in the winter, especially for the regions which are already most affected. Thereby, we must also conclude that there is little to no statistical evidence for the structural effect of climate changes on extreme precipitation.

With the novelty of combining extreme value index and frequency analysis, we contribute to the framework of evaluating trends in non-iid extremes. The combination of both gives a broader overview of the development in extremes and also increases the validity of the frequency analysis by testing the assumption of constant extreme value indices. Besides that, the implementation of automatic threshold selection using the SAMSEE method further develops the framework's objectivity and eases the possibility to evaluate extremes over larger varying regions and bigger data sets. Apart from the progress in evaluating trends in extremes by combining recent techniques, the results of our application add to the existing knowledge on extreme precipitation and can add to the political and social discussion. Furthermore, the design of the research also contributes to climatology by proposing a setup for evaluating extremes that can be applied to the analysis of heavy precipitation in other (larger) regions.

The remainder of this paper is organized as follows. In section 2 the existing literature and its relevance to the research are discussed. Subsequently, the methods for analyzing and testing the trend in extremes are introduced in section 3. An introduction of the data is given in section 4. The results are discussed in section 5. Finally, a conclusion and discussion are provided in 6.

2 Literature Review

The increase in mean temperature influences all meteorological processes, of these, precipitation has possibly the clearest direct link. For each degree Celsius that the atmosphere warms, the air capacity for water vapor increases by 7%. The more moisture in the air, the heavier precipitation events can be. This increase in temperature is observed and Pidcock and McSweeney (2021) show that extreme rainfall and extreme heat are the most prevalent of extreme weather events. Additionally, extreme precipitation is an extreme weather condition with very costly direct and indirect effects. The direct physical risks range from endangering of ecosystems, food supply, and flooding (Rosenzweig et al., 2001) to implications for infrastructure (Golz et al., 2016), buildings and even human lives (Schauwecker et al., 2019). Therefore, it is critical to understand the trends in magnitude and frequency to develop mitigation strategies and manage risks properly (Papalexiou and Montanari, 2019). These physical risks influence the stability of the financial sector, hospitality sector, food sector, and transport sector (Batten et al., 2020; Franzoni and Pelizzari, 2016; Lucas et al., 2017) and it is emphasized that it is critical for businesses to identify such climate risks to reduce their vulnerability and effectively build resilience (Tsalis and Nikolaou, 2017; Weinhofer and Busch, 2013). Due to the clear link with climate change, the prevalence of extreme events, and the abundant negative effects, precipitation is chosen as the extreme weather event of interest in this research.

2.1 Tail trend analysis

The scarcity of extreme observations makes analysis a challenge and asks for a different approach than classical statistical inference. Extreme Value Theory is developed to solve exactly this problem by highlighting the extreme part of the data and modeling the extreme observations separately. Before considering current literature on tail trend detection using EVT, it is necessary to get a basic understanding of classical EVT. Extreme value analysis focuses on extreme observations with the goal to explain and predict the behavior of extreme events. It is these extreme events that carry the biggest losses and risks, or as De Haan and Ferreira (2006) put it elegantly "In cauda venenum"[5]. EVT focuses on the limiting distribution of these sample maxima to draw inferences on the extremes. In statistical terms, consider independent and identically distributed (iid) random variables $X_1, X_2, ..., X_n$ with distribution function $F(x)$. Let $M_n = max(X_1, ..., X_n)$

[5]Poison is in the tail

be the maximum. Then for suitable sequences $a_n > 0$ and $b_n \in \mathbb{R}$ it is observed that as $n \to \infty$,

$$\Pr\left(\frac{M_n - b_n}{a_n}\right) \to G(x).$$

For such distribution functions F and G it is said that F is in the (maximum) domain of attraction of G, $F \in \mathcal{D}(G)$. This limit distribution G belongs to the class of extreme value distributions,

$$G_\gamma(x) = \begin{cases} exp(-(1+\gamma x)^{-1/\gamma}), & \text{if } \gamma \neq 0. \\ exp(-e^{-x}), & \text{if } \gamma = 0. \end{cases}$$

This class of distributions depends on one main parameter γ known as the extreme value index. This index divides the class of distributions into three different distributions. For $\gamma > 0$, the distribution is heavy-tailed and we have a Fréchet distribution. For $\gamma = 0$, the distribution is light-tailed and we have a Gumbel distribution. For $\gamma < 0$, the distribution is short-tailed and we have a Weibull distribution. In our research, we assume to have heavy-tailed distributions and thus a positive extreme value index, which is likely for hydro-meteorological extremes (Katz, 2002).

To estimate γ and to model extreme behavior, the data needs to be downsized to only include extremes. There are two approaches to downsizing the data. The older one, the block maxima approach divides the observation period into equal non-overlapping periods, for instance, years, and focuses research on the maximum of each period. By only using the maximum of each period this method is quite wasteful and relevant high observations can be missed (Ferreira and De Haan, 2015). The modern approach towards downsizing the dataset is the Peak over Threshold (POT) approach. The POT method considers all data points above a certain high threshold or that are part of the k upper order statistics. The advantage is that the POT picks up more relevant extremal information. Where the block maxima is dependent on the proper choice of the number of blocks, the POT is similarly dependent on the choice of a parameter, in this case, k. For estimates of γ, a choice of k too small leads to large variance, whereas large values for k lead to high bias in the estimator (Caeiro and Gomes, 2016).

The methods within classical extreme value theory are based on the assumption that observations are independent and identically distributed (iid). However, this condition does not hold in all applications. For instance, data of financial asset returns is often non-iid (McNeil and Frey, 2000). Research shows that this is also the case for precipitation. Not only is it found that a trend in the frequency of extreme events is present, but also that most events are increased by human-induced

greenhouse gas emission (Alexander et al., 2006; Groisman et al., 2005; IPCC, 2021; O'Gorman, 2015; Zolina et al., 2009). These observed trends are unlikely under the assumption of stationarity. Therefore, the identically distributed assumption needs to be relaxed to investigate trends in extreme precipitation. There are two possible situations in which we can observe trends in extremes. There is a trend in the extreme value index ór the extreme value index is constant, but we have heteroscedastic observations in the tail (de Haan et al., 2015). Both these scenarios will be further described in the following paragraphs.

A varying extreme value index has an impact on the probability of getting larger extremes. Recent research by de Haan and Zhou (2020) investigates the possibility of a trend in extreme value indices. In their paper, they analyze independent observations that (possibly) do not share the same extreme value index. Hereby, allowing the extreme value index to change over time and formally test for trends in this index. They propose a non-parametric local estimator for a positive extreme value index, which enables them to plot it over time. In addition to the local estimator, they propose a global test for changes in the extreme value index. When applying their methods to precipitation at Saint-Martin-de-Londres, they do not find evidence for a trend in the extreme value index. We will use their methods to analyze the extreme value index and so investigate the heaviness of the tail of extreme precipitation.

In the second scenario with a constant extreme value index and heteroscedastic observations, there is the possibility to evaluate the frequency of extremes. de Haan et al. (2015) lay the foundation for this area of EVT by introducing non-parametric estimators for the function c, which measures the probability that a certain high threshold is exceeded. The estimation of this function c allows them to interpret how the probabilities of extreme events change over time. Applied to precipitation data from 18 weather stations in the Netherlands and Germany they find that the trend seems to depend on proximity to the sea. This work is further developed by Einmahl et al. (2016) into a framework for estimating and testing trends in the frequency of extreme events. They call this situation 'heteroscedastic extremes' referring to the variation in extremes caused by the mildly varying underlying distributions. Also, they coined the term 'scedasis' for the function c which under their imposed condition can be interpreted as the frequency of extremes. In their framework, they estimate the scedasis function over time and introduce statistical tests for testing if there is a trend in the frequency of extremes. Our research will follow the method of Einmahl et al. (2016) in evaluating the frequency of extremes over time.

The analysis over time is a start towards understanding the trend in frequency. However,

following the results of de Haan et al. (2015), we argue that location is an additional important feature. Therefore, we extend the research on frequency to include spatial dependency. The space-time analysis is developed in recent years by Ferreira et al. (2017), Cabral et al. (2020), and Einmahl et al. (2020). In their analyses, all observations from all stations in the observed region are pooled for estimation. Hence, the function c is now dependent on time and location. Einmahl et al. (2020) propose two tests in their analysis of the space-time dependent scedasis function. First, they test if the distribution of extremes is constant over the various locations. Second, they test if the scedasis function is constant over time for each of the locations.

In their application of the methods on precipitation data, Cabral et al. (2020) find a local upward trend of frequency in the winter season. They find the highest peaks of precipitation in the summer with some local downward trends, although they are not significant. Also, they observe an increased tendency for variability in extremes in the mountain valley regions in the winter. Einmahl et al. (2020) similarly find that frequencies of extreme precipitation are not constant over the stations. This evidence is not there for the summer region. Additionally, they find more evidence for local temporal trends in winter than in summer. In all applications on precipitation data, the set is split for a summer and winter season to deal with seasonality. The patterns in precipitation tend to differ for the summer and winter season and splitting of the data is currently the best option when applying the heteroscedastic extremes framework (Burauskaite-Harju et al., 2012; Cabral et al., 2020).

2.2 Threshold selection

A central part of the methods described above and a fundamental problem of EVT in general is the estimation of the extreme value index. To estimate a positive extreme value index, the Hill estimator is still the most commonly used method in the literature. The Hill estimator is based on a certain high threshold $U_n = X_{(n-k)}$, where $X_{(n-k)}$ is the k-th upper order statistic and n is the total number of observations. The choice of k - the number of upper order statistics to consider in estimation - is essential for the proper estimation of the extreme value index. As mentioned above, the sample path of the estimator as a function of k can be highly unstable. A value for k too small leads to a large variance in the estimation and a choice for k too large leads to a high bias (Caeiro and Gomes, 2016). The proper threshold only selects the observations that are in the tail of the distribution. Schneider et al. (2021) present in an overview of the available solutions for selecting k.

One basic approach is data visualization using, among others, the Hill plot, QQ plot, or the mean excess plot. Here a plot is used to select a stable plateau for the extreme value index. The major drawback of this approach is the subjectivity due to the visual interpretation of the plot. Another approach, is using heuristics such as using the upper 10% of observations (DuMouchel, 1983) or using $k = \sqrt{n}$, where n is the total number of observations (Ferreira et al., 2003). The third class of approaches consists of theoretically motivated procedures that target the optimal sample fraction. To the last class of approaches, Schneider et al. (2021) contribute by introducing two methods for selecting the threshold without manually choosing any tuning parameters. One of these approaches is a smooth estimator for the Absolute Mean Squared Error (AMSE) of the Hill estimator. This Smooth Absolute Mean Squared Error estimator (SAMSEE) is based on a preliminary estimate of the extreme value index and an estimate of the bias of the Hill estimator. Their method performs consistently well over a wide range of processes and we will apply this in our research.

3 Methodology

The section consists of two subsections regarding the different types of trend we are analyzing in the paper. First, the methods for analyzing the extreme value index are introduced together with the SAMSEE method for automatic threshold selection. Thereafter, the methods for the frequency analysis are introduced. The second part is split in the analysis over time for all stations individually and the analysis over space-time for regional analysis.

3.1 Extreme value index analysis

In the first part of the research, we analyze the extreme value indices of precipitation over time. The extreme value index measures the heaviness of the tail, i.e. how likely it is to observe outliers. The higher the extreme value index, the heavier the tail, and the higher the probability of extremer events.

To analyze the extreme value indices for all weather stations we use the theory developed by de Haan and Zhou (2020). First, we introduce the framework and an estimator for the local estimation of the extreme value indices. This estimator can be used for the visual interpretation of a trend in extreme value indices. Then, we describe the global test statistic to test the hypothesis that the extreme value index is constant over time. If we find the extreme value index to be constant,

we estimate the global index and optimal threshold k. Then, the station perfectly lends itself for the frequency analysis as described in section 3.2. A vital part of the method is the proper choice of intermediate sequence k, which indicates the number of order statistics taken into account in the extreme value analysis. The value for k will be estimated using the SAMSEE method described in section 3.1.1. As mentioned, we start by introducing the local estimation of the extreme value index using the Hill estimator.

Consider at days $i = 1, 2, ..., n$ we have independent observations $X_1, X_2, ..., X_n$ following various continuous distribution functions $F_{1/n}, F_{2/n}, ..., F_{n/n}$, that is, $X_i \sim F_{i/n}(x)$. We have a set of continuous distribution functions $F_s(x)$ for $s \in [0, 1]$. We need the context of extreme value theory, thus we assume that the function F_s is in the domain of attraction of a generalized extreme value distribution with corresponding extreme value index, i.e. $F_s \in \mathcal{D}_{\gamma(s)}$. That means, there is a real number $\gamma(s)$ and a positive scale function a such that, for all $x > 0$,

$$\lim_{t \to 0} \frac{U_s(tx) - U_s(t)}{a(t)} = \frac{x^{\gamma(s)} - 1}{\gamma(s)} \tag{1}$$

where $U_s := \left(\frac{1}{1-F_s}\right)^{\leftarrow}$ and '$\leftarrow$' stands for the left continuous inverse function. Here $\gamma(s)$ is the extreme value index, which we assume to be a positive and continuous function on $[0, 1]$. This assumption indicates that we are dealing with heavy-tailed distributions, which is likely to hold in the case of hydro-meteorological extremes (Katz, 2002).

For calculating the extreme value index we find in section 2.2 that it is essential to choose a proper number of upper order statistics (i.e. observations in the tail) $k := k(n)$, where k is an intermediate sequence such that as $n \to \infty$, $k \to \infty$, and $\frac{k}{n} \to 0$. The choice of k has a direct effect on the value of the threshold and thus on the estimator for the extreme value index. The sample path of the extreme value index estimator can be very unstable as a function of k. Therefore, proper estimation is critical. We estimate a global $\hat{k}$ for each station using the SAMSEE method proposed by Schneider et al. (2021) further described in 3.1.1. This global $\hat{k}$ is used in the local estimation of the extreme value index estimator.

For this local estimation of $\gamma(s)$ we use only the observations X_i that are in the h-neighborhood of s. That is, $i \in \mathcal{I}_n(s) = \{i : |\frac{i}{n} - s| \leq h\}$, where $h := h(n)$ is the bandwidth such that as $n \to \infty$, $h \to 0$, and $nh \to \infty$. There will be $2nh$ observations for every $s \in [h, 1-h]$. We order the $2nh$ observations X_i, $i \in \mathcal{I}_n(s)$, the h-neighborhood of s, as $X_{(1)}^{(s)} \leq X_{(2)}^{(s)} \leq ... \leq X_{(2nh)}^{(s)}$. It is essential to choose a proper bandwidth h. A small bandwidth leads to very high variability

in the estimate of the extreme value index and a large value smooths out all interesting features. Unfortunately, we are not aware of any method that can help us in making this choice. Therefore, we base our choice of bandwidth on the choices of bandwidth made in comparable research and on the number of order statistics that is optimal for estimation such that every local extreme value index is estimated using multiple upper order statistics. Then, to estimate $\gamma(s)$, we follow de Haan and Zhou (2020) in choosing the top $2\hat{k}h$ order statistics from the $2nh$ local observations and we have

$$\hat{\gamma}_H(s) := \frac{1}{2\hat{k}h} \sum_{i \in \mathcal{I}_n(s)} \left(\log X_i - \log X^{(s)}_{(2nh-2\hat{k}h)} \right)^+ \tag{2}$$

as the estimator for the local extreme value index. Assuming the conditions regarding the choice of intermediate sequence $\hat{k}$ and bandwidth h in de Haan and Zhou (2020) hold, we ensure that the number of high order statistics used in each local interval tends to infinity, the estimator has no asymptotic bias for the optimal $\hat{k}$, $\hat{k}$ is compatible with the h-variation in the γ function, and the $(1 - \frac{k}{n})$-quantiles of the distribution are sufficiently smooth in short h intervals.

Visual interpretation of the estimated extreme value index can give insight into its trend. However, we need proper statistical tests to find evidence and be sure to evaluate all stations similarly. To test for a constant extreme value index we follow Schneider et al. (2021) in applying test statistic T4 from Einmahl et al. (2016), which is computationally less intensive than the other test for constant extreme value index. The test statistic considers a limited number of partial Hill estimators. The partial Hill estimators are calculated as in equation 2, but with a subset of observations. We divide the sample into m non-overlapping blocks, with $m > 1$ fixed. The choice for m will be based on previous research and checked for robustness. The cut-off points of the blocks are based on the number of observations above the threshold $X_{(n-k)}$. We use the estimator $\hat{C}(s)$, which will also be used in the frequency analysis. This estimator is given by:

$$\hat{C}(s) := \frac{1}{k} \sum_{i=1}^{ns} 1_{\{X_i > X_{(n-k)}\}}. \tag{3}$$

Then, the cut-off points of the blocks are $l_1 \leq l_2 \leq ... \leq l_{m-1}$ with $l_j := \sup\{s : \hat{C}(s) \leq j/m\}$ and $l_0 = 0$ and $l_m = 1$. This ensures that the partial hill estimators are all calculated using the same percentages of extreme observations, that is, the same number of observations above the threshold $X_{(n-k)}$. Hence, each partial hill estimator uses k/m extreme observations. The partial Hill estimator $\hat{\gamma}_{(l_{j-1}, l_j]}$ is calculated using all observations between the cut-off points l_{j-1} and l_j to

define the test statistic

$$T_4 := \frac{1}{m}\sum_{j=1}^{m}\left(\frac{\hat{\gamma}_{(l_{j-1},l_j]}}{\hat{\gamma}_H} - 1\right)^2, \tag{4}$$

where $\hat{\gamma}_H$ is the Hill estimator of the complete sample, $\hat{\gamma}_{(0,1]}$.

Assuming that the second-order condition usually considered in EVT holds uniformly for all $s \in [0,1]$. That is, there exists a continuous negative function $\rho(s)$ on $[0,1]$ and a set of auxiliary functions $A_s(t)$ that are continuous with respect to s, such that,

$$\lim_{t\to\infty} \frac{\frac{U_s(tx)}{U_s(t)} - x^{\gamma(s)}}{A_s(t)} = x^{\gamma(s)}\frac{x^{\rho(s)}-1}{\rho(s)}, \tag{5}$$

holds for any x and uniformly for all $s \in [0,1]$. Additionally, assuming that our estimated $\hat{k}$ is optimal, we have no asymptotic bias in the estimator $\hat{\gamma}_H$. Then we have that, as $n \to \infty$,

$$kT_4 \xrightarrow{d} \chi^2_{m-1}. \tag{6}$$

Using the test statistic and its limit distribution we can statistically test if the extreme value index is constant for each station. A constant extreme value index is a necessary condition for the frequency analysis. Statistical evidence for a varying extreme value index encourages us to evaluate its trend over time but prevents a frequency analysis.

3.1.1 Threshold selection

In this section, we introduce the method to estimate $\hat{k}$ in a way that we minimize subjectivity and maximize the automation of the process. We will use the SAMSEE method as proposed by Schneider et al. (2021). The SAMSEE is a smooth estimator of the Absolute Mean Squared Error (AMSE) of the Hill estimator. We use the method to estimate the global optimal k for each station for estimation of the extreme value index. This $\hat{k}$ is used in combination with the bandwidth h for local estimation of the extreme value index and in combination with the number of blocks m for calculating the test statistic T_4.

Assume equation 5 holds and let k be an intermediate sequence, i.e. as $n \to \infty$, $k \to \infty$, and $\frac{k}{n} \to 0$, then

$$\sqrt{k}(\gamma_H - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{(1-\rho)}, \gamma^2\right) \tag{7}$$

as $n \to \infty$ and with $\lambda := \lim_{k\to\infty} A(n/k)$, where we have left out time indication s for simplicity as the focus will be on the global estimation. For the optimal choice of k, $\lambda = 0$ and there is no asymptotic bias in the estimator.

It follows from equation 7 that the AMSE of the Hill estimator is

$$\mathbb{AE}[(\hat{\gamma}_H - \gamma)^2] = \frac{\gamma^2}{k} + \frac{A(n/k)^2}{(1-\rho)^2}. \tag{8}$$

The method is focused on finding the optimal k that minimizes the equation above. Therefore, we need to estimate the AMSE. For estimation, we introduce an estimate for γ as well as for the bias term $\frac{A(n/k)^2}{(1-\rho)^2}$.

Estimation of γ is done using the generalized Jackknife estimate γ_k^{GJ} as introduced by Gomes et al. (2000). This estimator is shown to have reduced bias compared to the Hill estimator. Estimation of γ_k^{GJ} is, similar to the Hill estimator, based on the number of observations above a threshold and defined as follows:

$$\hat{\gamma}_k^{\mathrm{GJ}} := 2\hat{\gamma}_k^{\mathrm{V}} - \hat{\gamma}_k, \text{ where} \tag{9}$$

$$\hat{\gamma}_k^V = \frac{\frac{1}{k}\sum_{i=1}^{n}((\log X_i - \log X_{(n-k)})^+)^2}{2\hat{\gamma}_k}. \tag{10}$$

$\hat{\gamma}_k^V$ is the de Vries estimator introduced by De Haan and Peng (1998) and $\hat{\gamma}_k$ is the Hill estimator as in 2. The Generalized Jackknife estimate squared and divided by k forms the first element of the estimate for equation 8.

To estimate the bias, Schneider et al. (2021) use two averages of Hill estimators:

$$\bar{\gamma}_{k,K} := \frac{1}{K-k+1}\sum_{i=k}^{K}\hat{\gamma}_i \quad \text{and} \quad \bar{\gamma}_K := \bar{\gamma}_{1,K} = \frac{1}{K}\sum_{i=1}^{K}\hat{\gamma}_i. \tag{11}$$

The first is the average Hill estimator of all Hill estimates for the values from k to K. The second is the average of all Hill estimates for the values from 1 to K. The difference of these averages forms the estimate for our bias term:

$$\hat{b}_{k,K} := \bar{\gamma}_{k,K} - \bar{\gamma}_K \tag{12}$$

In the paper of Schneider et al. (2021) they show that under the assumption that $\rho = -1$, which is motivated by their simulations and which is used to simplify estimation, $\hat{b}_{k,K}$ has the following

property:

$$\mathrm{AE}\left[\hat{b}_{k,K}\right] = \frac{1}{2}\frac{A(n/k)}{(1-\rho)}. \tag{13}$$

The estimator $\hat{b}_{k,K}$ squared times four gives us the second element of the estimate for equation 8. Combining the estimator from equation 9 and equation 13 we yield the smooth AMSE estimator as a function of k. The SAMSEE and optimal k are then given by:

$$\begin{aligned} \mathrm{SAMSEE}(k) &:= \frac{\left(\hat{\gamma}_K^{\mathrm{GJ}}\right)^2}{k} + 4\hat{b}_{k,K}^2 \\ \hat{k}_{\mathrm{SAMSEE}} &:= \underset{1<k<K}{\operatorname{argmin}} \mathrm{SAMSEE}(k) \end{aligned} \tag{14}$$

It only remains to choose an appropriate K that is large enough to allow for minimization and small enough to be an intermediate sequence itself. Under the assumption that $\rho = -1$, we have a relation between the estimators that provide us with a relatively stable function in k that has the same asymptotic expectation as the Hill estimator. This relation is given by

$$\mathbb{AE}\left[\hat{\gamma}_k\right] = \mathbb{AE}\left[\hat{\gamma}_k^{\mathrm{V}} + \hat{b}_{k,K}\right]. \tag{15}$$

We want to find a K for which the relation holds, so to measure the deviation uniformly over all $k \leq K$, we define

$$\mathrm{AD}(K) := \frac{1}{K}\sum_{k=1}^{K}\left(\hat{\gamma}_k^{\mathrm{V}} + \hat{b}_{k,K} - \hat{\gamma}_k\right)^2. \tag{16}$$

Then we look for K such that we minimize the stabilized approximation of AD. It follows that

$$K^* := \arg\min_K \left\{ \sum_{i=-2, i\neq 0}^{2} \left| \frac{\mathrm{AD}(K) - \mathrm{AD}(K+i)}{i} \right| \right\}. \tag{17}$$

3.2 Frequency analysis

In the second part of the research we focus on extreme precipitation frequency by considering distributions with constant extreme value indices, but with non-identically distributed observations in the tail. Central in this analysis is the scedasis function, which measures the frequency of extreme events over time. We deploy a similar framework as in the intensity analysis but now based on the work of de Haan et al. (2015), Einmahl et al. (2016), Cabral et al. (2020), and Einmahl et al. (2020). First, we introduce the estimator for the frequency of extremes over time to analyze

each station independently. Then, the test statistic to test for constant frequency is described. Thereafter, we extend the framework to space-time. Again, the estimator for the frequency of extremes is estimated, but now it is dependent on the location of the stations. Then, we test for equal distribution of extremes over space and test for constant frequency of extremes given spatial dependence. This gives insight into the regional behavior of extreme precipitation events.

3.2.1 Frequency analysis over time

We consider at days $i = 1, 2, ..., n$ we have independent observations $X_1, X_2, ..., X_n$ following various continuous distribution functions $F_1, F_2, ..., F_n$ and that there is a continuous baseline function F_0. These functions all share a common right end point defined as $x^* = \sup\{x : F_i(x) < 1\} \in (-\infty, \infty]$. Then, we have a continuous positive function scedasis function c defined on $[0, 1]$ such that,

$$\lim_{x \to x^*} \frac{1 - F_i(x)}{1 - F_0(x)} = c\left(\frac{i}{n}\right), \tag{18}$$

for all $i = 1, 2, ..., n$. To ensure that c can be interpreted as the frequency of extreme events we impose the condition

$$\int_0^1 c(s)ds = 1, \tag{19}$$

such that c is uniquely defined and similar to a density function. When $c \equiv 1$ we have a homogeneous density and the frequency of extreme events does not change over time.

We assume that F_0 is the domain of attraction of a generalized extreme value distribution with constant extreme value index, $F_0 \in \mathcal{D}_\gamma$ with $\gamma \in \mathbb{R}$. Thus, the domain of attraction condition in equation 1 holds for $\gamma(s) = \gamma$. By combining this condition with equation 18, we can show that

$$\lim_{t \to \infty} \frac{U_i(tx) - U_i(t)}{a(t)\{c(i/n)\}^\gamma} = \frac{x^\gamma - 1}{\gamma}. \tag{20}$$

Hence, all F_i belong to the domain of attraction of the same generalized extreme value distribution and have the same extreme value index. It follows that the impact of the scedasis function is on the scale of extremes instead of on the extreme value index.

To estimate the scedasis function we take a POT approach. We order all observations X_i, as $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$, choose a suitable intermediate sequence $k = k(n)$, and select the upper order statistic $X_{(n-k)}$ as the extreme value threshold. Then, we can estimate c non-parametrically by using a kernel estimation method. Let G be a biweight kernel $G(x) = \frac{15}{16}(1 - x^2)^2$ for $x \in [-1, 1]$

with $G(x) = 0$ for $|x| > 1$. We choose $h : h(n) > 0$ as a bandwidth such that, as $n \to \infty$, $h \to 0$, and $kh \to \infty$. The scedasis function can then be estimated by

$$\hat{c}(s) = \frac{1}{kh} \sum_{i=1}^{n} 1_{\{X_i > X_{(n-k)}\}} G\left(\frac{s - i/n}{h}\right). \tag{21}$$

Kernel density estimators have difficulties at and near its boundaries. In our case, these difficulties are present for s close to 0 and 1, that is, for $s \in [0, h)$ and $s \in (1 - h, 1]$. We use the boundary kernel from Jones (1993) that follows from the linear combination of $G(x)$ and $xG(x)$. Let $a_l(p, g) := \int_g^p u^l G(u)du$, where $p = \min\left(\frac{s}{h}, 1\right)$ and $g = \max\left(\frac{(s-1)}{h}, -1\right)$. Then the biweight kernel with boundary correction is given by

$$G_l(x) = \frac{(a_2(p, g) - a_1(p, g)x)}{a_0(p, g)a_2(p, g) - a_1^2(p, g)} G(x), \tag{22}$$

where $G(x)$ is the biweight kernel as given above. When $s \in [h, h-1]$, then $a_0(p, g) = 1$, $a_1(p, g) = 0$, and the boundary kernel simplifies to $G(x)$. The estimator for the scedasis function is then as follows:

$$\hat{c}(s) = \frac{1}{kh} \sum_{i=1}^{n} 1_{\{X_i > X_{(n-k)}\}} G_l\left(\frac{s - i/n}{h}\right). \tag{23}$$

The estimated function $\hat{c}(s)$ is ideal for visual interpretation of the trend in the frequency of extreme events. However, before a visual interpretation has any statistical ground we need to test if heteroscedastic extremes are present. The integrated function $C(s) = \int_0^s c(u)du$ for $s \in [0, 1]$ is used in our tests. The integrated scedasis function can be estimated by

$$\hat{C}(s) := \frac{1}{k} \sum_{i=1}^{ns} 1_{\{X_i > X_{(n-k)}\}}, \tag{24}$$

for $s \in [0, 1]$. A main result of Einmahl et al. (2016) is the theorem on the asymptotic normality of the estimator $\hat{C}(s)$. We assume that the conditions quantifying the speed of convergence, for which we refer to their paper, hold in our case.

We want to test if the frequency of extreme events changes over time, i.e. if there are heteroscedastic extremes. To examine this, we test the null hypothesis that extreme events are homogeneously distributed. That is, we test if $c(s) \equiv 1$ against the alternative that $c(s) \neq 1$ for all $s \in [0, 1]$. This is equivalent to testing if C is the identity function on $s \in [0, 1]$, i.e. $H_0 : C(s) = s$

and $H_1 : C(s) \neq s$. A rejection of the null hypothesis concludes that heteroscedastic extremes are present. We consider the Kolmogorov- Smirnov type test statistic T_1 from Einmahl et al. (2016):

$$T_1 := \sup_{0 \leq s \leq 1} |\hat{C}(s) - s|. \tag{25}$$

Under the null hypothesis, as $n \to \infty$,

$$\sqrt{k} T_1 \xrightarrow{d} \sup_{0 \leq s \leq 1} |B(s)|, \tag{26}$$

where B is a standard Brownian bridge.

3.2.2 Frequency analysis over space-time

After the frequency analysis over time, we extend the research to a space-time context. Above we have assumed that observations from stations are spatially independent and we have focused on the frequency of extreme events over time. Now we include spatial dependency in the analysis and research the behavior of extremes in a region. We test for differences in the frequency of extreme events over space, i.e. test if some locations are more affected by extreme events than others. In addition, we test if there is a time trend in the space-time scedasis function, i.e. test if given the temporal dependence there is a trend in time.

We incorporate the location of a station into our framework. Consider at locations $j = 1, 2, .., m$ we have independent random vectors $(X_{i,1}, X_{i,2}, ..., X_{i,m})$ for the days $i = 1, 2, ..., n$. Every observation follows (a slightly varying) continuous distribution function $F_{i,j}$, and we have a continuous baseline function F_0 that is in the domain of attraction a generalized extreme value distribution with extreme value index γ. All the functions share a common right end point $x^* = \sup\{x : F_{i,j}(x) < 1\} \in (-\infty, \infty]$. Then, we have a continuous positive function $c(\cdot, j)$ for each station j, such that,

$$\lim_{x \to x^*} \frac{1 - F_{i,j}(x)}{1 - F_0(x)} = c\left(\frac{i}{n}, j\right) \tag{27}$$

holds for $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$. To ensure that the scedasis function is uniquely defined and can be interpreted as the relative frequency of extremes, we impose the condition

$$\sum_{j=1}^{m} C_j(1) = 1. \tag{28}$$

Where for $0 \leq s \leq 1$ and $j = 1, 2, ..., m$, we have

$$C_j(s) := \frac{1}{m} \int_0^s c(u, j) du. \tag{29}$$

It follows from the assumption $F_0 \in \mathcal{D}(G_\gamma)$ and equation 27, that γ is the common extreme value index, i.e. $F_{i,j} \in \mathcal{D}(G_\gamma)$.

To model the spatial dependence between stations, let F_i be the distribution function of $(X_{i,1}, X_{i,2}, ..., X_{i,m})$. Let $U_{i,j}(s) := F_{i,j}^{\leftarrow}(1 - \frac{1}{s})$, where '$\leftarrow$' denotes the generalized inverse function. Then we assume that

$$\tilde{F}(x_1, x_2, ..., x_m) := F_i(U_{i,1}(x_1), U_{i,2}(x_2), ..., U_{i,m}(x_m)) \tag{30}$$

does not depend on i and is in the domain of attraction of a multivariate extreme value distribution. It follows that the tail dependence structure does not depend on the time i. The tail copula used to describe the dependence of extremes between the stations is given by

$$R_{j_1,j_2}(x, y) = \lim_{s \to 0} P(1 - F_{i,j_1}(X_{i,j_1}) \leq sx, 1 - F_{i,j_2}(X_{i,j_2}) \leq sy), \tag{31}$$

where $(x, y) \in [0, \infty]^2 \setminus \{\infty, \infty\}$.

To estimate $c(\cdot, j)$, the scedasis function at location j, we take the same approach as in the space independent case. Although, now we choose as common high threshold $X_{(N-k)}$ the k-th upper order statistic from all pooled $N = n \times m$ observations. We choose an appropriate intermediate sequence k, a proper bandwidth h, and let G be the biweight kernel as before. Then,

$$\hat{c}(s, j) = \frac{1}{kh} \sum_{i=1}^{n} 1_{\{X_{i,j} > X_{(N-k)}\}} G\left(\frac{s - i/n}{h}\right), \tag{32}$$

for $j = 1, 2, ..., m$ and $s \in [0, 1]$. The scedasis functions of the different locations in a region can be averaged to give insight into the overall change in frequency and regional developments. To formally test the frequency, we need a global estimator for C_j, the absolute frequency of extreme events. The estimator $\hat{C}_j$ for C_j is given by:

$$\hat{C}_j(s) := \frac{1}{k} \sum_{i=1}^{ns} 1_{\{X_{i,j} > X_{N-k}\}} \tag{33}$$

First, we test if the scedasis is constant over the various locations. We test the null hypothesis $H_0 : C_j(1) = \frac{1}{m}$ against the alternative $H_1 : C_j(a) \neq \frac{1}{m}$, for all $j = 1, 2, ..., m$. Let 1_m be the m-unit vector, I_m the identity matrix of dimension m, and $M := I_m - \frac{1}{m} 1_m 1_m'$. In Einmahl et al. (2020) the authors prove that under H_0, $D = \sqrt{k}(\hat{C}_1(1) - \frac{1}{m}, ..., \hat{C}_m(1) - \frac{1}{m})'$ is asymptotically m-multivariate normal with zero mean vector and covariance matrix $M\Sigma_1 M'$, where $\Sigma_1 = \Sigma(1, 1, 1, 1)$. The entries of the covariance matrix are

$$\sigma_{j_1,j_2}(t_1, t_2, s_1, s_2) := \frac{1}{m} \int_0^{s_1 \wedge s_2} R_{j_1,j_2}(t_1 c(u, j_1), t_2 c(u, j_2)) du. \tag{34}$$

We assume that Σ_1 is invertible, thus $\text{rank}(M\Sigma_1 M') = \text{rank}(M) = m - 1$, and we focus on the first $m - 1$ components of D given by D_{m-1}. Ultimately, the test statistic is given by

$$T_n := D'_{m-1}((M\hat{\Sigma}_1 M'))_{m-1})^{-1} D_{m-1} \tag{35}$$

with Σ entries estimated as:

$$\hat{\sigma}_{j_1,j_2} = \frac{1}{k} \sum_{i=1}^{n} 1_{\{X_{i,j_1} > X_{(N-k)}, X_{i,j_2} > X_{(N-k)}\}}. \tag{36}$$

Under the null hypothesis and given the satisfaction of the conditions from Einmahl et al. (2020),

$$T_n \xrightarrow{d} \mathcal{X}^2_{m-1} \tag{37}$$

as $n \to \infty$.

Second, we test for heteroscedasticity of extremes. This test is similar to before, but now using the unified threshold $X_{(N-k)}$. We test the null hypothesis $H_{0,j} : C_j(s) = sC_j(1)$ against the alternative $H_{1,j} : Cj(s) \neq sC_j(1)$ for $s \in [0, 1]$. We consider a Kolmogorov-Smirnov-type test based on the process $\sqrt{k}\left(\hat{C}_j(s) - s\hat{C}_j(1)\right) / \sqrt{\hat{C}_j(1)}, 0 \leq s \leq 1$. This gives us the test statistic

$$T_n := \sup_{s \in [0,1]} |\hat{C}_j(s) - s\hat{C}_j(1)| \frac{1}{\sqrt{\hat{C}_j(1)}}. \tag{38}$$

Under the null hypothesis,

$$\sqrt{k} T_n \xrightarrow{d} \sup_{s \in [0,1]} |B(s)| \tag{39}$$

as $n \to \infty$ and for B a Brownian bridge.

4 Data

In this section, we will first describe our initial data selection yielding a set of 150 stations in the Netherlands and Northern Germany. This set will be split into a summer and winter set to deal with seasonality. Then, some descriptive statistics and graphs give us an idea of the data and particularly the extremes. Finally, the declustering procedures are described and the individual declustering procedure is executed.

4.1 Initial data selection

The total data set provided by Klein Tank et al. (2002) consists of over 13,000 weather stations with daily precipitation levels from 1950 to 2020. The weather stations are distributed over multiple countries, where the density of stations per country varies greatly. For interpretability and due to limited computational resources, we need to make a selection of stations. This selection is made on the completeness of the data and the density of the stations. In the evaluation of data completeness, we use the same basic criterium as de Haan et al. (2015). The data from a station is complete if less than 10 days are missing per year. Stations that do not meet this criterium are left out. Regarding the density, we want stations to be evenly distributed over the region, thus not too far apart and not too close together. This ensures that we get a detailed insight into trends of a specific region and that we do not perform analysis for stations that have an almost identical location. Based on these conditions we choose to analyze stations in the Netherlands and stations in northern Germany (on the same latitude as the Netherlands). To ensure the stations are not too close to each other we filtered the stations at a minimal distance of 30km. The filtering procedure is done in the order that the stations are presented in the provided set. After filtering on proximity we end up with a total of 150 stations (36 in the Netherlands; 114 in Germany), distributed over the region as mapped in Figure 1.

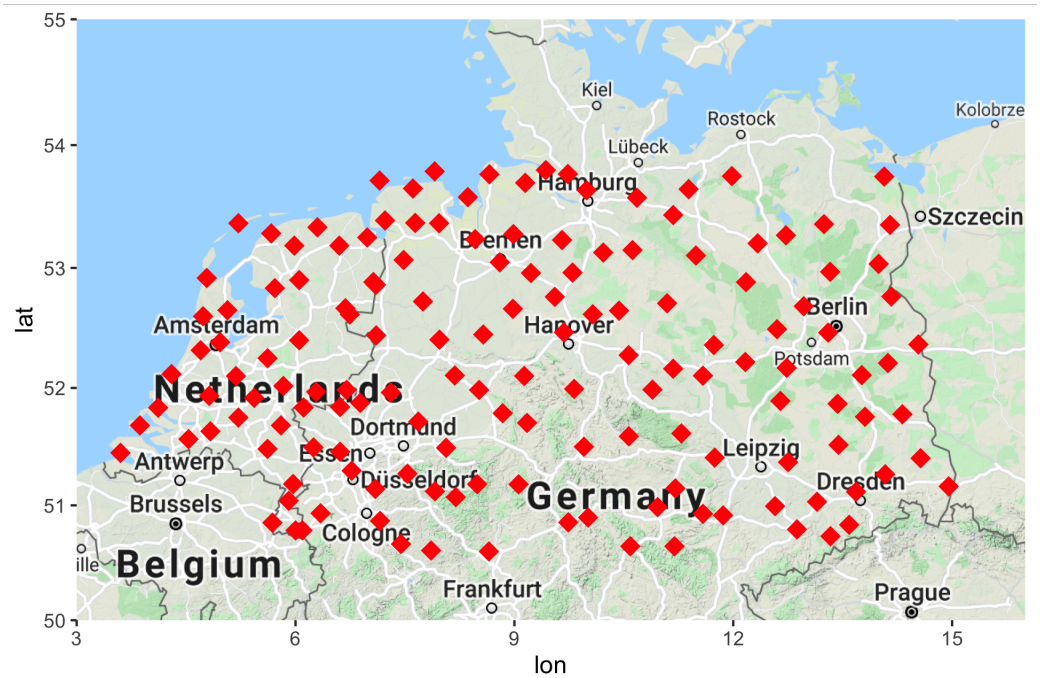


Figure 1: The locations of all 150 stations distributed across the Netherlands and Germany. Latitude and longitude are on the axes.

4.2 Data description

In our research, we are dealing with weather data, which is likely affected by seasonality. This assumption is substantiated by common knowledge on seasonality in weather, previous work regarding extreme precipitation as discussed in section 2, and also found in the box-plots describing the distribution of rain over different months. In figure 2 we plot the distribution of daily rainfall for Station 152 "De Bilt" as an example. Box-plots of other stations show a comparable seasonality trend. We observe a slightly higher median and broader interquartile range in the winter months, while the extremes are higher in the summer months. To deal with seasonality in an EVT setting it is most common to split the data. We follow previous research (Cabral et al., 2020; Einmahl et al., 2020; Ferreira et al., 2017) in setting the months May until September as the summer and November until March as the winter, leaving April and October as transitional periods. Hence, we go forward with two separate data sets, one for the summer period and another for the winter period.

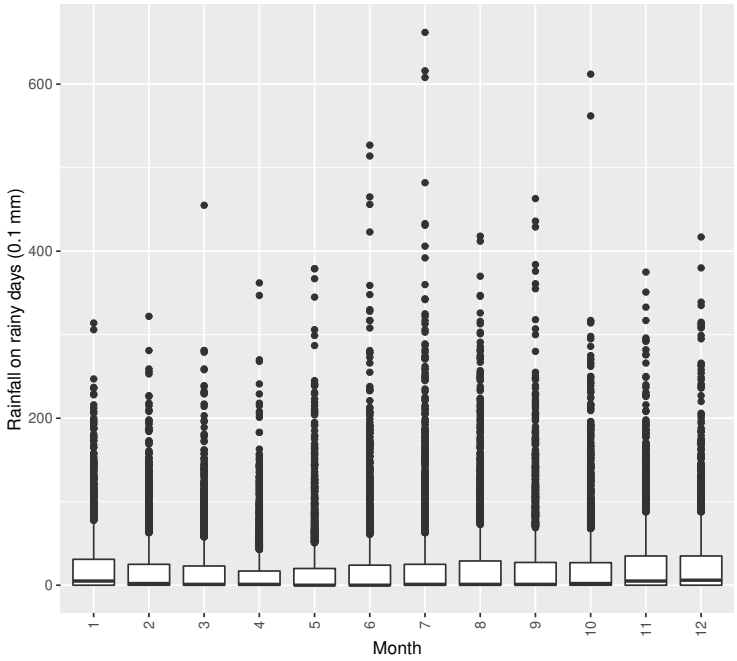


Figure 2: Boxplot of daily rainfall amounts per month of the year for station ST 152 "De Bilt".

The summary statistics in table 1 show this seasonality as well. In table 1 we see that the summer period has fewer days of rain (47% opposed to 57%), but higher quantiles. For the region as a whole, we find that it rains a minimum of 0.1 millimeters on 51% of the days and that the average rainfall on rainy days is 3.9 millimeters. To put this latter number in perspective, the Royal Dutch Meteorological Institute labels a day with more than 50 millimeters as a day of heavy precipitation. Even the 99-th percentile is far from this number, which tells us we are interested in a small sample of extremes, for which EVT is used to make inferences. This finding is supported by the boxplots in figures 3 and 4. We find that there are numerous outliers spread from 20 millimeters per day to over 100 millimeters per day. All stations show the same distribution of observations with median and interquartile region close to zero, and numerous extremes. This informs us that distributions are right-skewed and have a heavy tail. The boxplots of the other stations and seasons are given in Appendix A.1.2. Tables 2 and 3 give us insight into the differences between the two countries. Apart from the number of rainy days we do not observe any significantly large differences between the two countries. Analysis should point out what regions are alike and if there is spatial dependency across the region.

Table 1: Summary statistics for all weather stations together.

	All year	Summer	Winter
Rainy days (1)	13041	5024	5985
% of rainy days (2)	51%	47%	57%
Avg. rain per day *(mm)* (3)	2.0	2.2	1.9
Avg. rain on rainy days *(mm)* (4)	3.9	4.8	3.4
Avg. 95th percentile *(mm)* (5)	10.0	11.5	9.1
Avg. 99th percentile *(mm)* (6)	19.9	23.8	16.7
Maximum *(mm)* (7)	222.0	222.0	91.7

Table 2: Summary statistics for the Dutch weather stations.

	All year	Summer	Winter
(1)	14009	5228	6553
(2)	55%	49%	62%
(3)	2.2	2.3	2.2
(4)	4.1	4.8	3.6
(5)	11.0	12.1	10.3
(6)	20.8	23.7	18.4
(7)	132.5	132.5	76.2

Table 3: Summary statistics for the German weather stations.

	All year	Summer	Winter
(1)	12735	4960	5806
(2)	50%	46%	55%
(3)	2.0	2.2	1.8
(4)	3.9	4.8	3.3
(5)	9.7	11.3	8.7
(6)	19.6	23.8	16.1
(7)	222.0	222.0	91.7

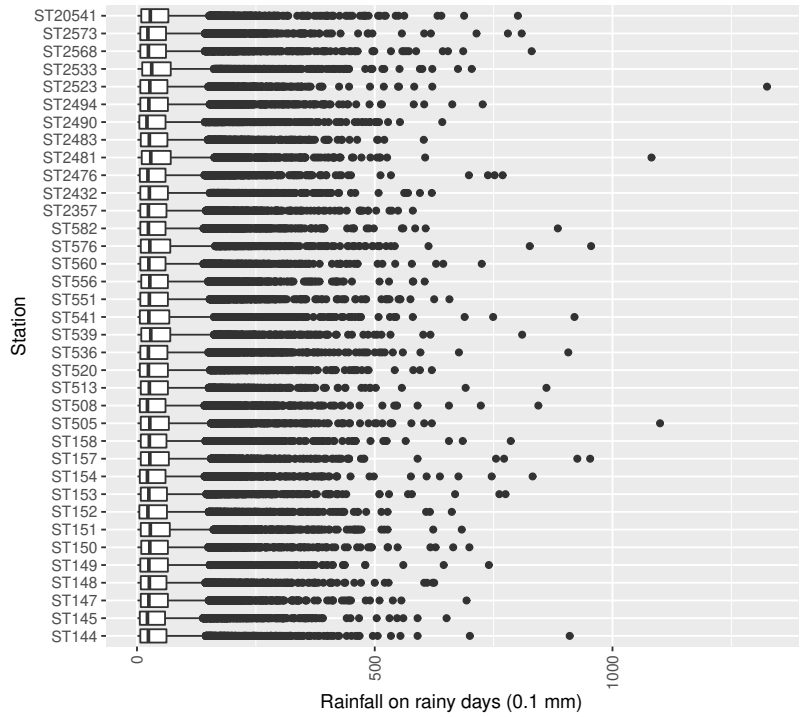


Figure 3: Boxplots of the distribution of daily rainfall amounts on rainy days in the summer for 36 Dutch stations.

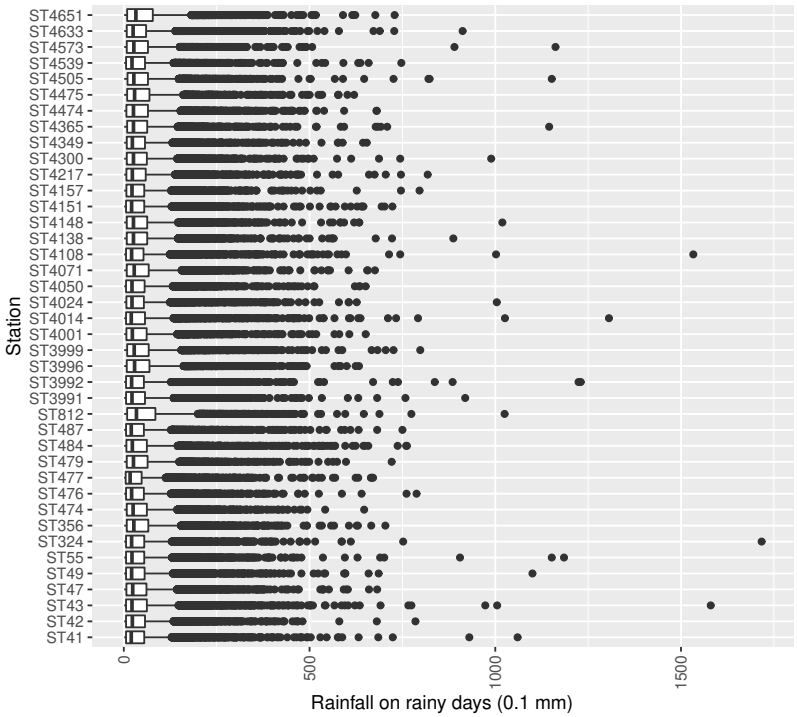
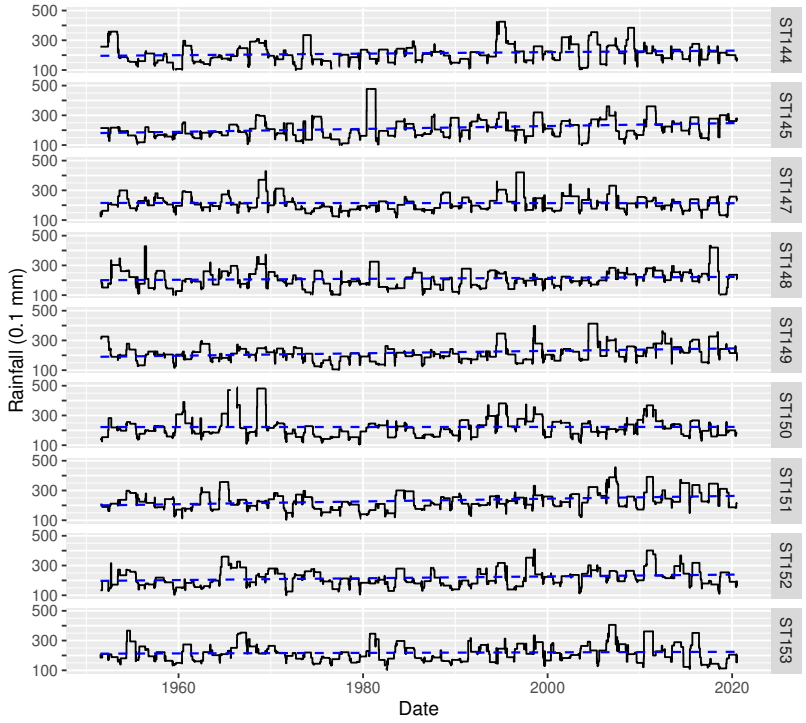
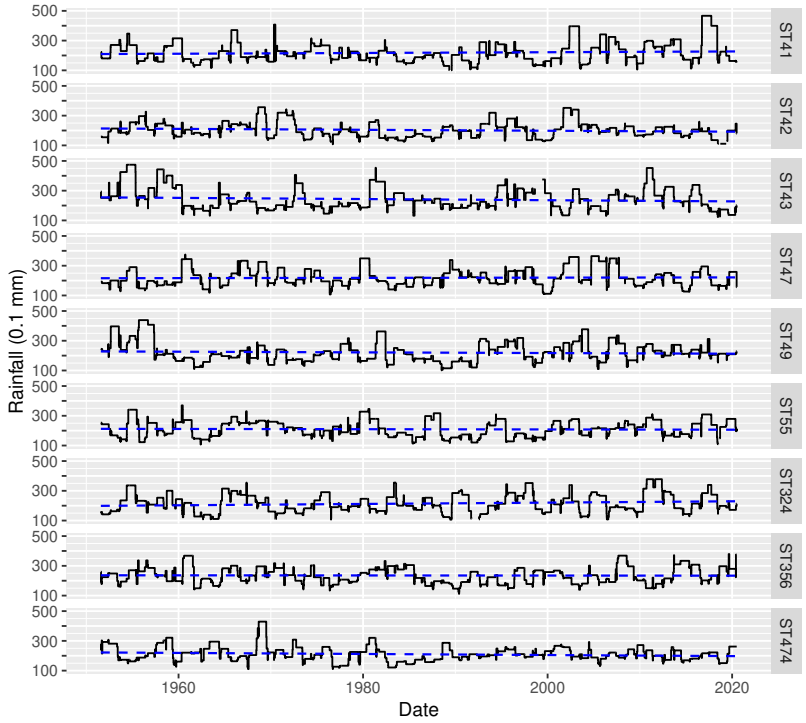


Figure 4: Boxplots of the distribution of daily rainfall amounts on rainy days in the summer for 36 German stations.

We cannot say anything yet about a possible trend. To get a first insight into the change of extreme events over time we plot the 99-th quantile over time using a rolling window of 150 days ($\sim$ five months). In these plots, we add a linear trend such that we can assess if extremes went up or down on average. In figure 5 we observe 9 of the 36 Dutch stations and 9 of the 114 German stations. Plots for other subsets of stations paint a similar picture. The quantiles are very volatile over time, but when focusing on the linear trend we find that the majority of stations have an increasing trend. However, the differences are too small and ambiguous to draw any initial conclusions. For example, ST151 "Hoofddorp" has an increase of (only) 5 millimeters in 70 years. Whereas, ST55 "Schwerin" seems to have a constant 99-th percentile. Hence, for us to comment on the trend in extremes we need to zoom in on the extreme values.



(a) 9 of the 36 Dutch weather stations



(b) 9 of the 114 German weather stations

Figure 5: Plot of the 99-th quantile in the summer using a rolling window of 150 days for 9 of the 114 Dutch (a) and German (b) weather stations. The blue dotted line is the linear trend from 1950 until 2020. The date is on the x-axis, the precipitation in 0.1 millimeters is on the left y-axis, and the right y-axis gives the station ID.

4.3 Declustering procedure

One of the assumptions in our research is that of independent observations. Therefore, we need to remove temporal dependence from the data. To ensure independence, we create gaps between consecutive observations by removing days in the set. A simple way of doing so would be to only include Mondays and Thursdays. However, with such a blunt approach we risk the possibility of losing a lot of extremal information. Since our focus is only on the extremes, we use a different approach to preprocessing the data. Thereby, it is found of daily precipitation that data is approximately independent after two days (de Haan et al., 2015). Given this information, we follow the declustering procedure proposed by Ferreira et al. (2017). In their paper, they apply the declustering procedure to all stations in the set at once, because they are only interested in the space-time analysis and therefore the same days should be removed for all stations. Our focus is first on individual stations. Hence, applying the declustering procedure to the complete set at once would lead to a big loss of extremal information. Therefore, we choose to apply the declustering twice. First, we will apply the process to individual stations such that we have independent observations with optimal extremal information. The yielded set will be used for the extreme value index analysis and the frequency analysis for individual stations. Second, before the space-time analysis, we apply the process to the complete region. The method as described below is for the whole region. The method for individual stations is similar, but for the number of stations $m = 1$.

For each day $i = 1, 2, ..., n$ we order the observations from the m different stations and select for each day a sample maximum M_i. We have n different maxima which we order from largest to smallest, such that $M_{(1)} \leq M_{(2)} \leq ... \leq M_{(n)}$. We pick the largest maximum $M_{(n)}$ and then select the second largest $M_{(n-1)}$. Only if this second maximum is not within two days of the previous maximum it is kept in the set, otherwise, it is discarded. If it is discarded this means that the day is removed from all m different weather stations. Thereafter, we select the next maximum $M_{(n-2)}$ and keep it in the set only if it is not within two days lag of any previous maximum. This process is continued until the smallest maximum.

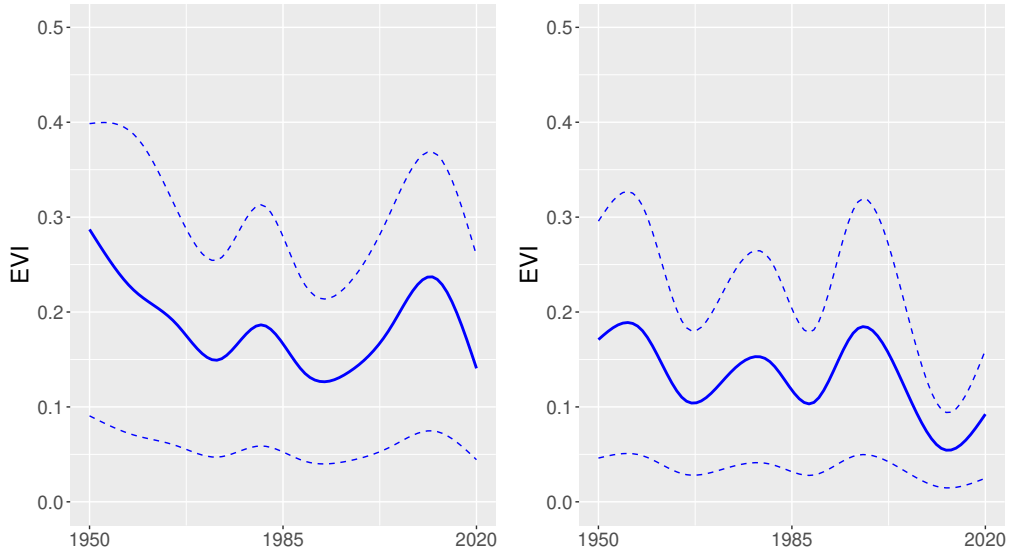
5 Results

We start the results section by looking at the extreme value index and testing if there is a trend present in our observation period. This analysis will be done individually for each station, thus the individually preprocessed dataset is used. This dataset consists of 150 stations with the number of observations n around 3050 for each station for both summer and winter. After the extreme value index analysis, the frequency of extremes individually per station is analyzed. This analysis gives insight into the trend of the frequency of extreme precipitation events. After the time analysis of frequency, we will expand the research to a spatial-time analysis which gives further information on the distribution of extreme events and the regional trends in frequency. The latter analysis is done on two complete regions, which are regionalized based on average precipitation, extreme value index, and threshold. Hence, this analysis is done using the regional preprocessed dataset.

5.1 Extreme Value Index Analysis

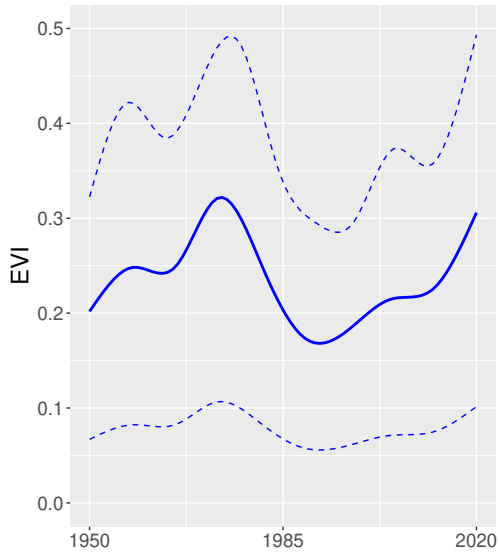
For estimation of the local extreme value indices we need a value for the bandwidth h, an optimal number of upper order statistics k, and a value K^* over which to minimize to find k. The choice for K^* is made based on equation 17 from the SAMSEE method. The estimation of K^* for each station individually is computationally too intensive, but applying the method to a subset of the stations shows that the choice for K^* varies around 400 and that the results of the SAMSEE method are robust to small varying values of K^*. Thus, we choose $K^* = 400$. Then, the SAMSEE method yields an optimal $\hat{k}$ ranging from 33 to 105 for all stations with an average of 49 in the summer and 30 to 63 with an average of 41 in the winter. In percentages, this boils down to 1.6% of all observations in summer are in the tail and 1.3% in the winter. This difference between the winter and summer periods underlines the finding that extremes are more frequent in the summer. Given the small values for $\hat{k}$, we choose a bandwidth that is slightly higher than in comparable research to counter very high variability in the estimates of the extreme value indices. For the bandwidth, we use $h = 0.1$, thus every local extreme value index will be calculated using 20% of the data.

Given the parameters above, the individually declustered set, and the estimated values of $\hat{k}$ for each station, the local extreme value indices are estimated. In figure 6 the local estimated extreme value indices in summer and winter for ST52 "De Bilt" and ST41 "Berlin" are plotted. These two stations are chosen as examples to give insight into the estimation of the local EVI. Other stations can show very different variations or values. However, the EVI values are in the range of 0 to 0.5 for all stations.

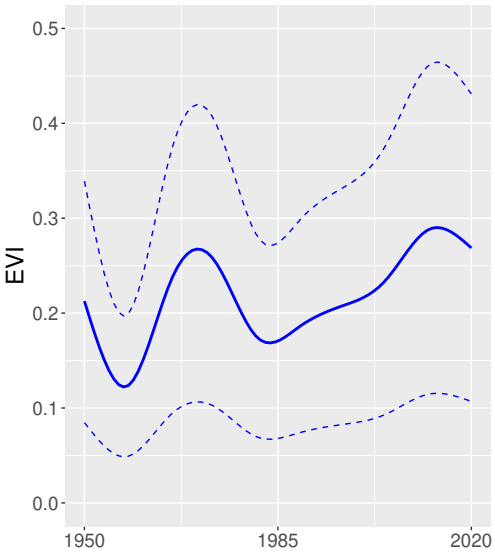


(a) ST152 "De Bilt" - Summer

(b) ST152 "De Bilt" - Winter



(c) ST41 "Berlin" - Summer



(d) ST41 "Berlin" - Winter

Figure 6: Plots of the locally estimated extreme value indices for stations ST52 "De Bilt" and ST41 "Berlin" for summer and winter periods. The dotted lines present the 95% confidence interval. The value of the extreme value index is on the y-axis, the time on the x-axis.

We observe a clear variation in the plotted extreme value indices, but the question remains if there is a trend. For "De Bilt" we see a slight decrease of the extreme value index, which is more notable in winter than in summer. For "Berlin" we observe a small increase, also more notable for winter than for summer. However, the variation is small and it is questionable if it is statistically significant. Our proposed test could find evidence for a changing EVI, but observing the variation in the given plots does not (automatically) imply that there is a trend. The variation could also lead to a rejection of the test. Thus, to ensure that

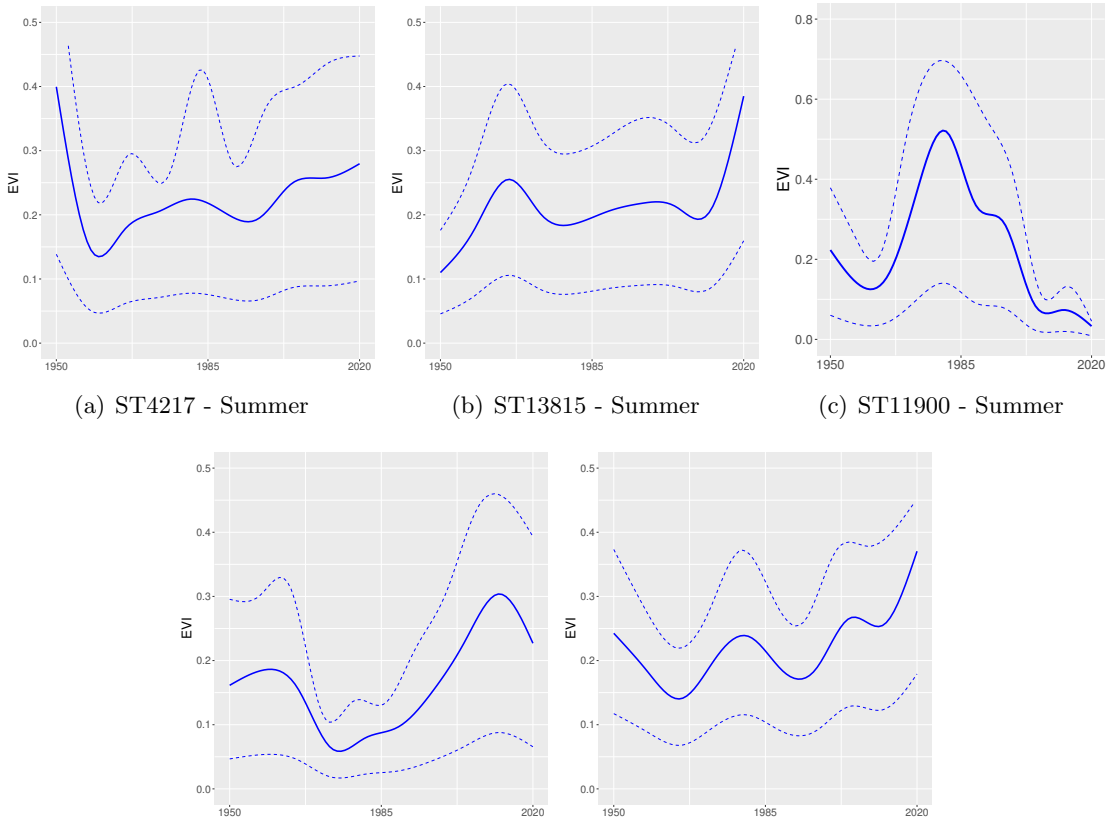
we thoroughly test for constant extreme value index we will use four different values of m, which are the number of blocks in equation 4. The outcome of these tests are given in table 4.

Table 4: Station IDs of the 15 stations with the lowest p-value (in parenthesis) regarding the test for a constant extreme value index for different values of m. The bold IDs are from the stations for which the null hypothesis is rejected with 95% confidence at 3 or more values of m. * attached to IDs with p-value < 0.05.

	Summer							
m =	5		6		7		8	
Station IDs:	**4217***	(1e-04)	**4217***	(7e-04)	**13815***	(0.0061)	**11900***	(0.0096)
	2476*	(0.0064)	**13815***	(0.0112)	12371*	(0.016)	2490*	(0.0144)
	11900*	(0.0219)	**11900***	(0.0316)	4151*	(0.0196)	**13815***	(0.0199)
	13815*	(0.0363)	12167*	(0.046)	12205*	(0.0211)	576*	(0.0202)
	145*	(0.0375)	150*	(0.0478)	12719*	(0.0231)	**4217***	(0.0223)
	4505*	(0.0392)	**13158***	(0.0496)	2490*	(0.0301)	812*	(0.0312)
	12205*	(0.0404)	2490	(0.0536)	**4217***	(0.0325)	4148*	(0.0339)
	13158*	(0.0477)	539	(0.0569)	**4505***	(0.0357)	**4505***	(0.0357)
	12719*	(0.0479)	13579	(0.0617)	145*	(0.0431)	**13158***	(0.0371)
	4300	(0.0592)	4151	(0.0617)	560	(0.0635)	11793*	(0.0473)
	150	(0.0661)	3992	(0.0635)	11933	(0.0647)	13579	(0.0504)
	151	(0.0679)	505	(0.0723)	2476	(0.0738)	3992	(0.058)
	4151	(0.0887)	12205	(0.0758)	324	(0.0826)	539	(0.0697)
	4024	(0.1012)	4677	(0.0758)	11849	(0.0933)	505	(0.0749)
	505	(0.1039)	16015	(0.077)	4024	(0.096)	12719	(0.0778)

	Winter							
m =	5		6		7		8	
Station IDs:	4633*	(0.0452)	11767*	(0.0072)	551*	(5e-04)	4151*	(5e-04)
	14180	(0.0746)	3991*	(0.0139)	14180*	(0.0111)	13083*	(0.001)
	4071	(0.0813)	576*	(0.0163)	13158*	(0.0352)	43*	(0.0174)
	13703	(0.0895)	11898*	(0.0195)	4475*	(0.0367)	12116*	(0.0326)
	12719	(0.0934)	2523	(0.0524)	2523	(0.0586)	4148*	(0.0344)
	576	(0.1022)	13083	(0.0775)	13659	(0.0627)	576	(0.06)
	16015	(0.1121)	14180	(0.0796)	13160	(0.0634)	3992	(0.0614)
	12489	(0.1151)	12337	(0.1056)	324	(0.0667)	4505	(0.0672)
	505	(0.129)	4651	(0.1153)	560	(0.0758)	13659	(0.071)
	4651	(0.1383)	541	(0.1263)	43	(0.0776)	16091	(0.0753)
	12231	(0.1459)	12116	(0.1378)	15198	(0.0831)	14180	(0.0798)
	4151	(0.1492)	4024	(0.1387)	12036	(0.0957)	474	(0.0859)
	4050	(0.1509)	13158	(0.1419)	12489	(0.1043)	158	(0.0981)
	43	(0.1522)	2432	(0.1563)	13703	(0.1223)	4138	(0.1074)
	13659	(0.164)	536	(0.1618)	4014	(0.1227)	15198	(0.1094)

In the winter period, we find that for none of the stations the null hypothesis of a constant extreme value index is consequently rejected. There is no significant evidence against a constant EVI. Thus, we conclude that there is no trend in the winter and the extreme value indices are constant. Subsequently, all stations in winter are eligible for the frequency analysis. In the summer period, there are only two stations for which all 4 tests reject the null hypothesis and 3 for which most of the tests reject the null. We conclude that the EVI of these 5 stations varies over time. Hence, these are not eligible for the frequency analysis. Plots should prove if there is a trend in the EVI or if it only fluctuates heavily around a steady index.



(a) ST4217 - Summer (b) ST13815 - Summer (c) ST11900 - Summer

(d) ST13158 - Summer (e) ST4505 - Summer

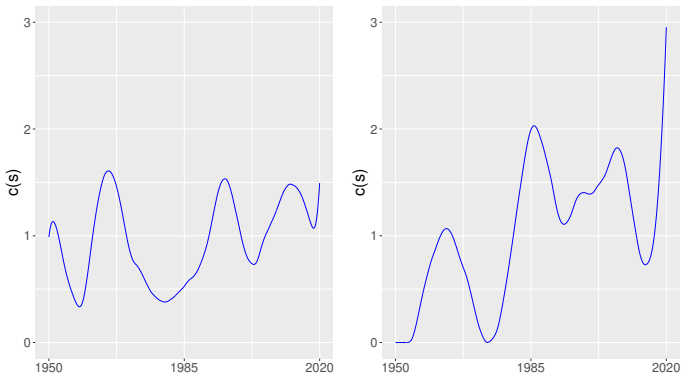
Figure 7: Plots of the locally estimated extreme value indices for stations with a statistically significant varying extreme value index. The dotted lines present the 95% confidence interval. The value of the extreme value index is on the y-axis, the time on the x-axis.

Figure 7 shows that we have an upward trend for four of the stations and a downward trend for the fifth. This does not give a clear view of a trend in the extreme value index. Thereby, the location of the stations are on the most eastern, southern, and western board of our German region. These opposite trends and the fact that the stations are far apart does not provide us with motivation to suspect a general trend in the extreme value index over the period 1951-2020. Therefore, we must conclude that in general there is no trend in the extreme value index of extreme precipitation.

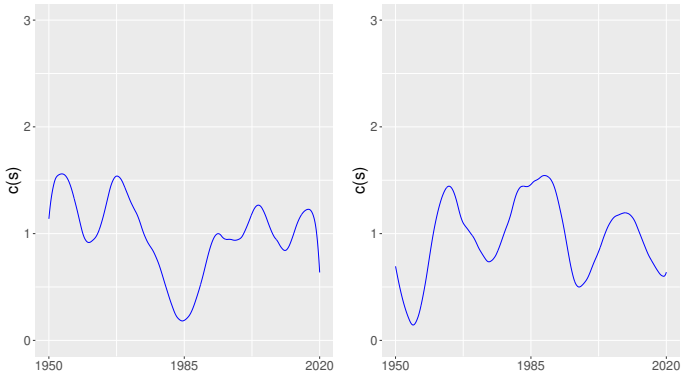
5.2 Frequency analysis for individual stations

Analysis of the frequency of extreme precipitation events is done using the scedasis function. Estimation of the scedasis function is done following the steps in section 3.2.1. For this estimation, we need to make a proper choice for the bandwidth regarding kernel estimation. A bandwidth choice that is too small leads to high variation, while a large bandwidth smooths out all interesting features. Based on the bandwidth choice made in comparable research (Cabral et al., 2020; Einmahl et al., 2016), we choose $h = 0.1$ as the bandwidth. Additionally, we need the number of upper order statistics to consider per station. This value k is identical to those used in the extreme value index analysis, thus based on the SAMSEE method. Using these parameters and the individually prepreprocessed dataset we can estimate the scedasis function for every station. The scedasis function measures the frequency of extreme events, in our case heavy rainfall.

Figure 8 shows the estimated scedasis functions for the locations "De Bilt" and "Berlin". For both locations in both seasons, it is clear that the frequency varies between years, but we are interested in a general trend. For the winter season in "De Bilt" we observe a slight increasing trend in the scedasis function, which would be an indicator of an increasing number of heavy rainfall at this location. An important step remains to see if there are statistically significant trends.



(a) ST152 "De Bilt" - Summer (b) ST152 "De Bilt" - Winter



(c) ST41 "Berlin" - Summer (d) ST41 "Berlin" - Winter

Figure 8: Plots of the scedasis functions over time for stations ST52 "De Bilt" and ST41 "Berlin" for summer and winter periods. The value of the scedasis is on the y-axis, the time on the x-axis.

Testing for a constant scedasis function is done using the test statistic T_1 from equation 25. To calculate the p-values corresponding to the test statistics we run a Brownian Bridge Monte Carlo simulation. We use 100000 simulations where every simulation consists of 1000 steps from 0 to 1. This yields a simulated Brownian Bridge distribution and enables us to calculate the p-values. In table 5 all the stations for which the null hypothesis of a constant scedasis function is rejected at different confidence intervals are listed. Hence, all stations for which we have statistical evidence of variations in the frequency of extremes.

In the summer we observe that the test is rejected with 90% confidence for a total of 22 stations. For a small part of these stations, the test is rejected at the 95% level and for only 2 stations at the 99% level. This result could be an indication of a relatively stable distribution of extreme events over the observations period. Hence, we find no clear trend in the frequency of extreme precipitation in the summer period. In the winter period, we find that 29 stations, almost one-fifth of all stations, are rejected at a 95% confidence level. Fourteen of these stations are even rejected at the 99% level. We conclude that the winter period is much more affected by changing frequencies of extremes. This corresponds to previous work in which more evidence is found for varying extremes in winter as opposed to summer (Cabral et al., 2020; Einmahl et al.,

2016). The question remains if this result is caused by a clear trend or large variations from a constant mean. For this analysis, we will evaluate the scedasis plots from the stations of which the null hypothesis is rejected at the 99% confidence level.

Table 5: Station IDs of the stations for which the test rejects the hypothesis that the scedasis function is constant over time. The results are given for different confidence levels.

	Summer			Winter		
Confidence level:	99%	95%	90%	99%	95%	90%
Station IDs:	487	49	4300	47	42	324
	149	4001	4705	4148	3992	4300
	-	4633	12123	11767	4071	4474
	-	4651	12493	12036	4108	13659
	-	12299	12575	12237	4157	15830
	-	12371	13083	12337	12021	16091
	-	151	13659	149	12116	157
	-	2533	14940	151	12368	513
	-	-	16091	152	13158	551
	-	-	520	154	13311	2476
	-	-	576	505	13688	2481
	-	-	2568	520	147	-
	-	-	-	541	148	-
	-	-	-	2432	2573	-
	-	-	-	-	576	-

Before evaluating the trends, we would like to take a preview of the spatial analysis by displaying the stations from table 5 on a map. It is hard to draw any conclusions from the map in the summer period. The stations are scattered over the map and this could be an indication of the absence of a region that is more prone to trends in the frequency in the summer period. For the winter period, however, we observe a cluster of stations close to the coast from west to north-west. Hence, we find evidence to conclude that the coastal regions are more impacted by changes in extreme precipitation.

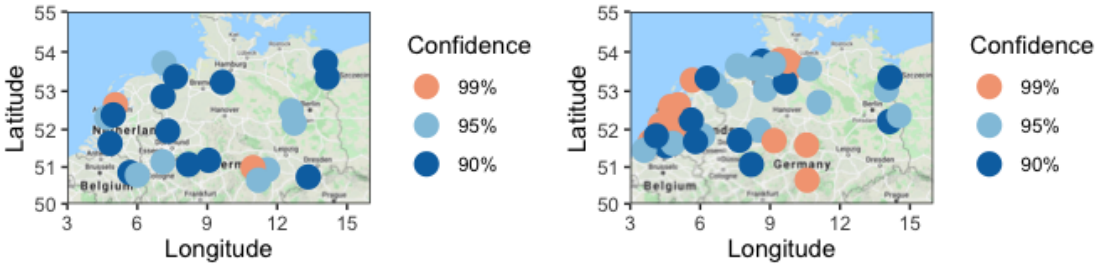
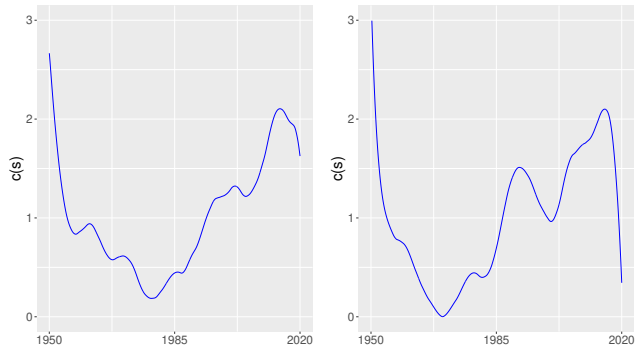


Figure 9: Map of all stations for which the null hypothesis for a constant scedasis function over time is rejected. The map for the summer period is on the left and the map for the winter period is on the right. The maps are color-coded on the confidence level for which the test is rejected.

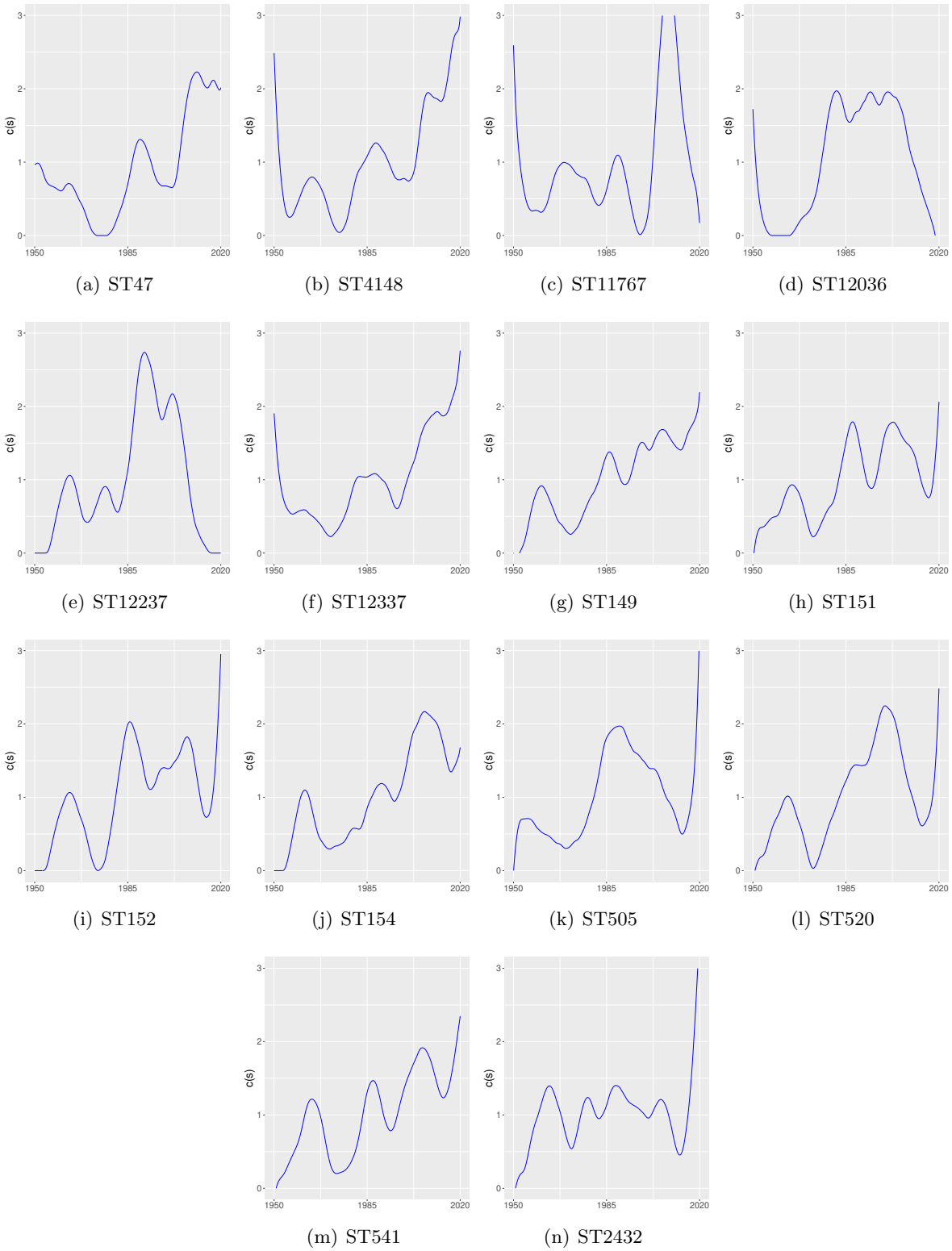
For the stations in figures 10 and 11, we find evidence that the frequency of extremes is not constant. For the summer period, we observe the following characteristics: a peak around 1950, a quick decrease of the frequency thereafter, and a steady average increase since 1985. We find no clear cause for the peak in 1950, however, the rise since 1985 corresponds largely to the rising temperatures since then. This finding indicates an increasing trend, but the low quantity of stations that show a similar statistically significant trend refrains us from concluding that there is a trend present in the summer period. For the winter period it is clear that, given the number of stations with a rejected null hypothesis, there is more fluctuation in frequency. This is probably caused by a higher impact from climate change. We observe from figure 11 that the majority of stations experience an increasing trend over the complete observational period. The stations 11767 (c), 12036 (d), and 12237 (e) do not show such a trend. Interestingly, these stations are the only ones that are not in a coastal area. For the other stations, we can conclude from the individual station analysis that there is a clear increasing trend in the frequency of extremes over the observation period. Hence, the stations close to the sea are subject to climate change in the form of increased frequency of extreme precipitation events. This finding corresponds to the conclusion from de Haan et al. (2015) that trends depend on proximity to the sea. An explanation for this increase in frequency at stations close to the sea as opposed to the more inland stations is the availability of moisture in the air. Higher increases in moisture in the air are expected at locations that are close to the sea. Therefore, these locations are also more vulnerable to increases in extremes (Allen and Ingram, 2002; Lenderink et al., 2009).

.



(a) ST149 - Summer (b) ST487 - Summer

Figure 10: Scedasis plots of the stations for which the null hypothesis of a constant frequency of extremes over the observation period is rejected with 99% confidence in the summer period.



(a) ST47 (b) ST4148 (c) ST11767 (d) ST12036

(e) ST12237 (f) ST12337 (g) ST149 (h) ST151

(i) ST152 (j) ST154 (k) ST505 (l) ST520

(m) ST541 (n) ST2432

Figure 11: Scedasis plots of the stations for which the null hypothesis of a constant frequency of extremes over the observation period is rejected with 99% confidence in the winter period.

5.3 Regionalization

Additional to the individual station frequency analysis we would like to know what the impact of spatial dependency is. To perform a space-time analysis we choose to split the data into regions that are comparable on average precipitation, extreme value index, and threshold. This split is necessary to avoid losing too much extremal information for some stations. In figure 12 the stations are color-coded on a map based on their values for the three characteristics mentioned above. From these maps, we decide to split the data into two regions: a western region with all stations left of the 9.5 longitude line and an eastern region with all stations right of the 9.5 longitude line.

The western region has higher average precipitation, a lower extreme value index, which implies that the extreme values lie closer to their threshold, and a higher threshold. This is in line with a more oceanic climate. Given this information, we would assume, similar to the results found in de Haan et al. (2015), that proximity to the sea influences the distribution and trend in extreme events.

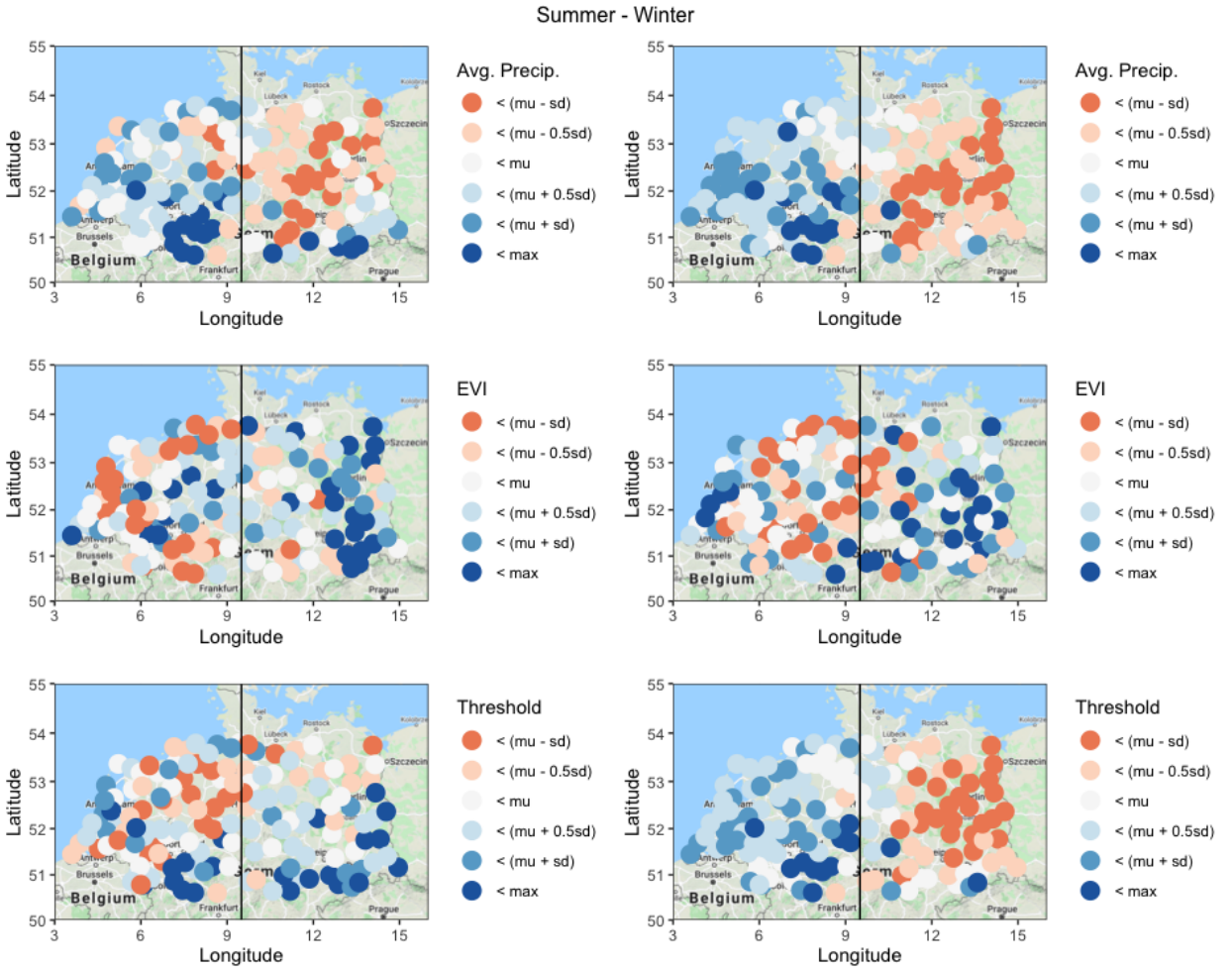


Figure 12: Maps for average precipitation, extreme value index, and threshold for every station per season. The values of these characteristics are color-coded based on mean and deviation. The region split is made on the 9.5 longitude line.

5.4 Frequency analysis with spatial dependency

In the final step of our research, we analyze the trend in the frequency of extremes when spatial dependency is incorporated. We analyze the extremes above a certain region-wide threshold $X_{(N-k)}$, where $N = n \times m$, n is the number of days in the set, and m is the number of stations in the region. Our goal is to find a region-wide trend and analyze the distribution of the extremes over the region, that is which area's experience more extremes. For these analyses, we split the data into a West and East region as discussed in section 5.3 and we have preprocessed the region datasets using the declustering method as described in section 4.3. We ultimately have four independent data sets: West summer, West winter, East summer, and East winter, consisting of 3024 days for 84 stations, 2985 days for 84 stations, 2997 days for 66 stations, and 2969 days for 66 stations, respectively.

For our analysis, we need a choice for the number of upper order statistics k and a proper bandwidth h. The bandwidth used for calculating the scedasis functions is equal for all four sets. We again use $h = 0.1$, similar to Cabral et al. (2020) and Einmahl et al. (2020). The other important parameter k leads to the unified threshold $U = X_{(N-k)}$. Given the extremely large datasets ($N = 254016$ for West summer), limited computational power, and the fact that we only have to find four distinct values of k, we use hill plots. For West summer, we find $k = 1900$ and threshold $U = 35.5mm$, for West winter $k = 2500$ and $U = 25.0mm$, for East summer $k = 2000$ and $U = 33.2mm$, and for East winter $k = 1500$ and $U = 20.6mm$. The hill plots are given in Appendix A.2.1.

We start by testing for equal distribution of extremes over the various locations. That is, we test the null hypothesis $H_0 : C_j(1) = \frac{1}{m}$ for all $j = 1, 2, ..., m$ for each region and season. The test rejects the null hypothesis with 99% confidence for West winter, East summer, and East winter with p-values of $9.27e^{-37}$, $1.13e^{-49}$, and 0.00025, respectively. For West summer the test is less confident with a p-value of 0.0918. Given the test outcomes, it can be concluded that extremes are not evenly distributed over the regions. However, it remains to see where extremes are more frequent. The maps in figure 13, where the values of $C_j(1)$ are color-coded and mapped, give insight into this question. We find for both regions that the majority of extreme events in the summer happen in the southeast of the region, which is the most inland region. In the winter period, we find that most extremes in the West region are at the coast and for the East region we see a similar distribution with fewer extremes further east or inland. We conclude that in the summer extremes seem more frequent land inwards, while in the winter this changes to the coastal regions.

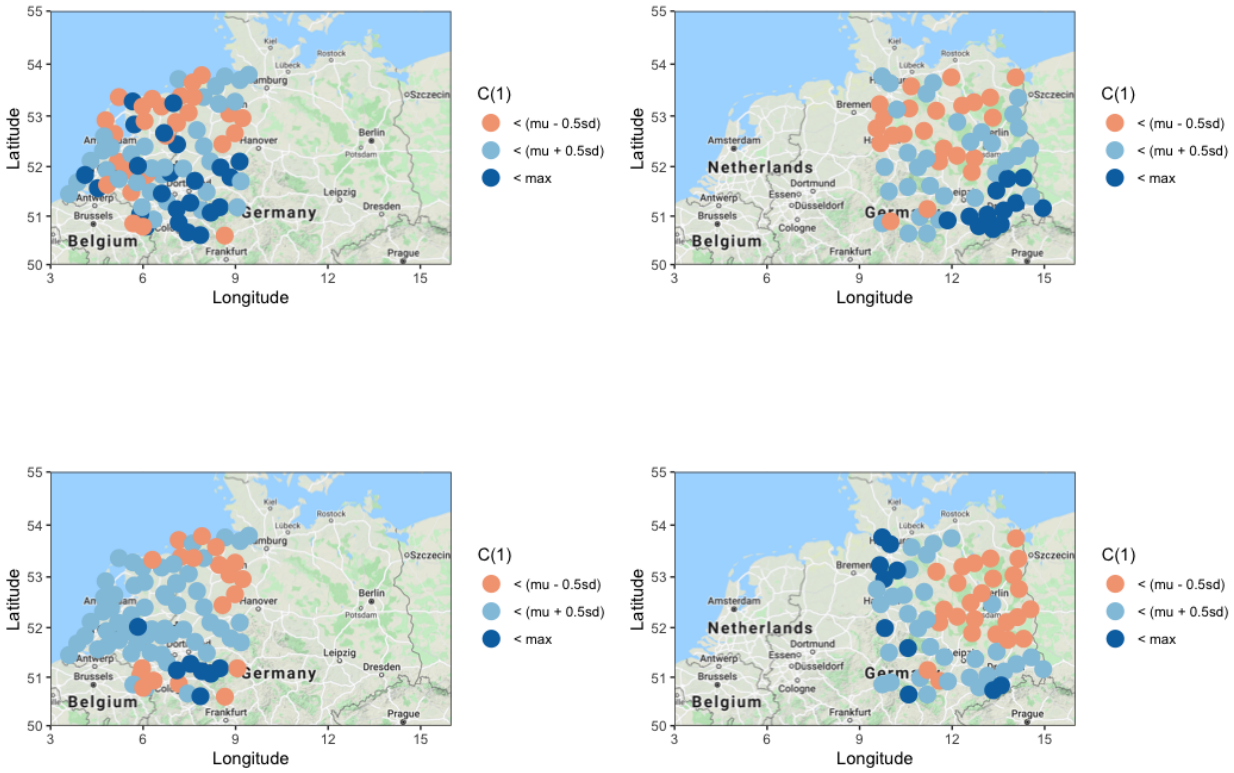
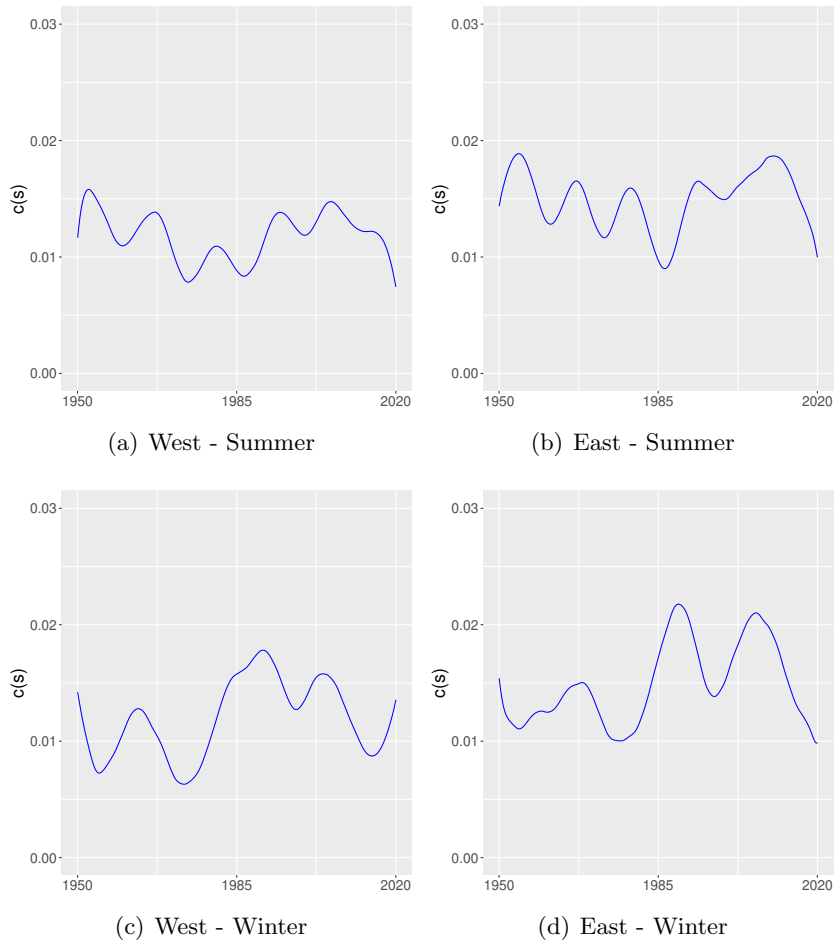


Figure 13: Maps of the number of extremes for every station per season and region. The summer seasons are above and winter seasons are below. The values are color-coded based on mean and deviation.

Finally, we test for constant scedasis for all stations in the region given the unified threshold to ultimately say something about the region-wide trend. For the West region, we find no evidence for any station of changing frequency in the summer. In the winter period, the null hypothesis is rejected with 95% confidence for only six stations. For the East region, we find similar results with two stations in the summer period and five in the winter period. All p-values can be found in Appendix A.2.2. We can conclude that there is no statistical evidence for variation in the frequency for these regions. Visual interpretation supports this conclusion. The average scedasis functions, plotted in figure 14, show no clear trend for the regions.



(a) West - Summer

(b) East - Summer

(c) West - Winter

(d) East - Winter

Figure 14: Average scedasis functions for the West and East region in both seasons. The value of the scedasis is on the y-axis, the time on the x-axis.

6 Conclusion & Discussion

This paper studies trends in extreme precipitation. The focus is on the extreme value index and the frequency of extremes. The first measures the heaviness of the tail of the distribution and the latter measures the relative frequency of observations above a certain high threshold. Recent heavy rainfall disasters show us the impact of extreme precipitation and insight into these trends can help with preparing sufficient countermeasures against extreme precipitation.

We evaluate daily rainfall data of 150 stations across the Netherlands and northern Germany over the period 1950 until 2020. Our analysis shows that there is no statistical evidence for rejecting the hypothesis that the extreme value indices are constant over time. Hence, we conclude that these indices remain equal over the observation period and we see no increase (or decrease) in the heaviness of the tail of precipitation distributions. This finding is equal for the summer and winter periods. We do find a trend in frequency for stations in the winter period. Statistical evidence and frequency plots show that the frequency of extremes increases, most notable for stations near the coast. This trend is only present in the winter. In the summer period, we find no statistical evidence for the changing frequency of precipitation extremes. The spatial analysis shows that in the winter extremes are more frequent towards the coast and in the summer more frequent land inwards. Thus we find that in the winter the region where extremes are most frequent experiences the clearest increasing trend of frequency. We find no evidence for a region-wide increase in frequency.

Our results imply that although small changes in extreme precipitation might be present, it is not statistically evident over our observation period apart from an increase in frequency in the winter for stations near the coast. Given the fact that extremes are heavier in the summer, this has probably no direct implications for safety regarding climate disasters. However, it is a clear sign of changes in extreme precipitation that should warn us and make us prepare business and infrastructure for more frequent heavy rainfall events.

The methodology in this paper has some limitations or gaps that require further research. Although our goal was to minimize subjectivity using the SAMSEE method, the choice for K^*, k in the spatial analysis, and the bandwidth choice still had to be made. The choice for K as well as k can be prevented by having more computer power or a smaller dataset. For bandwidth selection in the kernel estimation, local bandwidth estimation as proposed by Brockmann et al. (1993) could help remove subjectivity in further research. Another extension of the research would be to use a more sophisticated regionalization approach. The method proposed by Saunders et al. (2021) uses extremal dependence to regionalize based on rainfall. Since our focus was on trend detection and not specifically on regionalization, we chose not to use this extensive regionalization method. However, this method would remove subjectivity and enables the researcher to analyze regions in far more detail, thus picking up on smaller changes in extreme precipitation.

References

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., et al. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres*, 111(D5).

Allen, M. R. and Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic cycle. *Nature*, 419(6903):228–232.

Batten, S., Sowerbutts, R., and Tanaka, M. (2020). Climate change: Macroeconomic impact and implications for monetary policy. *Ecological, Societal, and Technological Risks and the Financial Sector*, pages 13–38.

Brockmann, M., Gasser, T., and Herrmann, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association*, 88(424):1302–1309.

Burauskaite-Harju, A., Grimvall, A., and Von Brömssen, C. (2012). Analysing trends in precipitation extremes for a network of stations. *Int. J. Climatol*, 32:86–94.

Cabral, R., Ferreira, A., and Friederichs, P. (2020). Space–time trends and dependence of precipitation extremes in north-western germany. *Environmetrics*, 31(3):e2605.

Caeiro, F. and Gomes, M. I. (2016). Threshold selection in extreme value analysis. *Extreme Value Modeling and Risk Analysis: Methods and Applications*, page 69.

De Haan, L., Ferreira, A., and Ferreira, A. (2006). *Extreme value theory: an introduction*, volume 21. Springer.

de Haan, L., Tank, A. K., and Neves, C. (2015). On tail trend detection: modeling relative risk. *Extremes*, 18(2):141–178.

de Haan, L. and Zhou, C. (2020). Trends in extreme value indices. *Journal of the American Statistical Association*, pages 1–15.

De Haan, L. d. and Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, 52(1):60–70.

Diebold, F. X. and Rudebusch, G. D. (2021). Probability assessments of an ice-free arctic: Comparing statistical and climate model projections. *Journal of Econometrics*.

DuMouchel, W. H. (1983). Estimating the stable index α in order to measure tail thickness: A critique. *the Annals of Statistics*, 11(4):1019–1031.

Einmahl, J. H., de Haan, L., and Zhou, C. (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pages 31–51.

Einmahl, J. H., Ferreira, A., de Haan, L., Neves, C., and Zhou, C. (2020). Spatial dependence and space-time trends in extreme events. *CentER Discussion Paper Nr. 2020-009*. Available at SSRN: https://ssrn.com/abstract=3560128".

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

Ferreira, A., de Haan, L., and Peng, L. (2003). On optimising the estimation of high quantiles of a probability distribution. *Statistics*, 37(5):401–434.

Ferreira, A., Friederichs, P., de Haan, L., Neves, C., and Schlather, M. (2017). Estimating space-time trend and dependence of heavy rainfall. *arXiv preprint arXiv:1707.04434*.

Franzoni, S. and Pelizzari, C. (2016). Weather risk management in tourism industry. *Symphonya. Emerging Issues in Management*, (1):45–55.

Golz, S., Naumann, T., Neubert, M., and Günther, B. (2016). Heavy rainfall: An underestimated environmental risk for buildings? In *E3S Web of Conferences*, volume 7. EDP Sciences.

Gomes, M. I., Martins, M. J., and Neves, M. (2000). Alternatives to a semi-parametric estimator of parameters of rare events—the jackknife methodology. *Extremes*, 3(3):207–229.

Groisman, P. Y., Knight, R. W., Easterling, D. R., Karl, T. R., Hegerl, G. C., and Razuvaev, V. N. (2005). Trends in intense precipitation in the climate record. *Journal of climate*, 18(9):1326–1350.

IPCC (2021). Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change.

Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statistics and computing*, 3(3):135–146.

Katz, R. W. (2002). J3. 5 do weather or climate variables and their impacts have heavy-tailed distributions?

Klein Tank, A., Wijngaard, J., Können, G., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., et al. (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the european climate assessment. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 22(12):1441–1453.

Konisky, D. M., Hughes, L., and Kaylor, C. H. (2016). Extreme weather events and climate change concern. *Climatic change*, 134(4):533–547.

Lenderink, G., Van Meijgaard, E., and Selten, F. (2009). Intense coastal rainfall in the netherlands in response to high sea surface temperatures: analysis of the event of august 2006 from the perspective of a changing climate. *Climate dynamics*, 32(1):19–33.

Lucas, E. C., Da Silva, W. M., Araujo, G. S., et al. (2017). Does extreme rainfall lead to heavy economic losses in the food industry? Technical report.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3-4):271–300.

O'Gorman, P. A. (2015). Precipitation extremes under climate change. *Current climate change reports*, 1(2):49–59.

Papalexiou, S. M. and Montanari, A. (2019). Global and regional increase of precipitation extremes under global warming. *Water Resources Research*, 55(6):4901–4914.

Pidcock, R. and McSweeney, R. (2021). Mapped: How climate change affects extreme weather around the world. `https://www.carbonbrief.org/mapped-how-climate-change-affects-extreme-weather-around-the-world` [Online; accessed 30-September-2021].

Rosenzweig, C., Iglesius, A., Yang, X.-B., Epstein, P. R., and Chivian, E. (2001). Climate change and extreme weather events-implications for food production, plant diseases, and pests. *Global Change & Human Health*, 2(2).

Saunders, K., Stephenson, A., and Karoly, D. (2021). A regionalisation approach for rainfall based on extremal dependence. *Extremes*, 24(2):215–240.

Schauwecker, S., Gascón, E., Park, S., Ruiz-Villanueva, V., Schwarb, M., Sempere-Torres, D., Stoffel, M., Vitolo, C., and Rohrer, M. (2019). Anticipating cascading effects of extreme precipitation with pathway schemes-three case studies from europe. *Environment international*, 127:291–304.

Schneider, L. F., Krajina, A., and Krivobokova, T. (2021). Threshold selection in univariate extreme value analysis. *Extremes*, pages 1–33.

Tsalis, T. A. and Nikolaou, I. E. (2017). Assessing the effects of climate change regulations on the business community: a system dynamic approach. *Business Strategy and the Environment*, 26(6):826–843.

Weinhofer, G. and Busch, T. (2013). Corporate strategies for managing climate risks. *Business Strategy and the Environment*, 22(2):121–144.

Zolina, O., Simmer, C., Belyaev, K., Kapala, A., and Gulev, S. (2009). Improving estimates of heavy and extreme precipitation using daily records from european rain gauges. *Journal of Hydrometeorology*, 10(3):701–716.

A Appendix

A.1 Data

A.1.1 Stations

Table 6: All stations considered in the research with Station ID, Station Name, Country Code, Latitude, and Longitude of the station.

STAID	NAME	CC	LAT	LON
ST41	BERLIN-DAHLEM	DE	52.4638889	13.3016667
ST42	BREMEN	DE	53.0463889	8.7991667
ST43	DRESDEN WAHNSDORF	DE	51.1166667	13.6830556
ST47	HAMBURG FUHLSBUETTEL	DE	53.6350000	9.9900000
ST49	JENA STERNWARTE	DE	50.9266667	11.5841667
ST55	SCHWERIN	DE	53.6441667	11.3883333
ST324	LINDENBERG	DE	52.2094444	14.1202778
ST356	AACHEN	DE	50.7838889	6.0950000
ST474	EMDEN-FLUGPLATZ	DE	53.3891667	7.2266667
ST476	HANNOVER	DE	52.4655556	9.6794444
ST477	MAGDEBURG	DE	52.1030556	11.5841667
ST479	DUSSELDORF	DE	51.2969444	6.7700000
ST484	GORLITZ	DE	51.1630556	14.9530556
ST487	ERFURT-BINDERSLEBEN	DE	50.9844444	10.9630556
ST812	KAHLER ASTEN (WST)	DE	51.1816667	8.4900000
ST3991	GIESSEN/WETTENBERG	DE	50.6008333	8.6505556
ST3992	ANGERMUNDE	DE	53.0330556	13.9930556
ST3996	KOLN-BONN	DE	50.8658333	7.1583333
ST3999	BAD LIPPSPRINGE	DE	51.7866667	8.8394444
ST4001	NORDERNEY	DE	53.7138889	7.1525000
ST4014	COTTBUS	DE	51.7775000	14.3183333
ST4024	DREWITZ BEI BURG	DE	52.2183333	12.1650000
ST4050	ELSFLETH	DE	53.2358333	8.4619444
ST4071	ESENS	DE	53.6491667	7.6091667
ST4108	FRANKFURT/ODER	DE	52.3652778	14.5377778
ST4138	GELDERN-WALBECK	DE	51.4955556	6.2466667
ST4148	GLUCKSTADT	DE	53.7983333	9.4305556
ST4151	GOTTINGEN	DE	51.5016667	9.9527778

ST4157	GRAMBEK	DE	53.5744444	10.6805556
ST4217	BAD HERSFELD	DE	50.8530556	9.7383333
ST4300	KONIGSMOOR	DE	53.2291667	9.6533333
ST4349	LENZEN/ELBE	DE	53.1025000	11.4877778
ST4365	LONINGEN	DE	52.7230556	7.7469444
ST4474	NORDHOLZ (FLUGPLATZ)	DE	53.7644444	8.6575000
ST4475	NORDHORN	DE	52.4408333	7.1055556
ST4505	PAPENBURG	DE	53.0655556	7.4819444
ST4539	RAHDEN-VARL	DE	52.4480556	8.5730556
ST4573	BAD ROTHENFELDE	DE	52.1052778	8.1827778
ST4633	SCHWARZBURG	DE	50.6455556	11.1944444
ST4651	SOLINGEN-HOHENSCHEID	DE	51.1408333	7.0941667
ST4652	SOLTAU	DE	52.9616667	9.7966667
ST4677	WORPSWEDE-HUTTENBUSCH	DE	53.2775000	8.9866667
ST4705	UECKERMUNDE	DE	53.7458333	14.0694444
ST4709	UMMENDORF	DE	52.1616667	11.1766667
ST4732	WANGEROOGE	DE	53.7883333	7.9075000
ST4763	BOCHOLT (MARIENSCHULE)	DE	51.8391667	6.6113889
ST4776	WITTENBERG	DE	51.8902778	12.6463889
ST4796	ZEHDENICK	DE	52.9680556	13.3280556
ST4841	BORKEN IN WESTFALEN	DE	51.8741667	6.8866667
ST4875	BRAMSCHE	DE	52.4050000	7.9730556
ST4880	BRAUNSCHWEIG-GLIESMARODE	DE	52.2755556	10.5666667
ST4951	CELLE (STADT)	DE	52.6163889	10.0738889
ST4954	CHEMNITZ	DE	50.7925000	12.8733333
ST11758	KAMENZ	DE	51.2666667	14.0830556
ST11767	BAD SACHSA	DE	51.5905556	10.5688889
ST11793	ABBERODE	DE	51.6152778	11.2891667
ST11821	AHLDEN	DE	52.7613889	9.5541667
ST11849	ALFELD/LEINE	DE	51.9916667	9.8191667
ST11885	ALTENMEDINGEN-EDDELSTORF	DE	53.1466667	10.6191667
ST11898	ALTMOERBITZ TALSPERRE SCHOEMBACH	DE	50.9930556	12.5730556
ST11900	ALTREETZ	DE	52.7641667	14.1663889
ST11906	AMELINGHAUSEN	DE	53.1266667	10.2180556
ST11915	ANGERN	DE	52.3605556	11.7341667
ST11933	ASBACH	DE	50.6655556	7.4455556

ST11941	ATTENDORN	DE	51.1208333	7.9105556
ST11971	BADERSLEBEN	DE	51.9855556	10.8930556
ST12021	BEETZENDORF	DE	52.7077778	11.0930556
ST12036	BENSHAUSEN	DE	50.6455556	10.5894444
ST12116	BIELEFELD-BRACKWEDE	DE	51.9830556	8.5166667
ST12123	BILLERBECK	DE	51.9616667	7.3069444
ST12167	BLESENDORF	DE	53.2019444	12.3341667
ST12180	BOCKHORN-GRABSTEDE	DE	53.3666667	7.9677778
ST12195	STRIEGISTAL-BOEHRIGEN	DE	51.0305556	13.1488889
ST12205	BOITIN	DE	53.7500000	11.9805556
ST12231	BOXBERG	DE	51.4038889	14.5680556
ST12237	BRAKEL	DE	51.7044444	9.1733333
ST12251	BREDDIN	DE	52.8827778	12.1752778
ST12299	BRUECK-GOEMNIGK	DE	52.1691667	12.7330556
ST12337	BULLENKUHLEN	DE	53.7666667	9.7330556
ST12368	BUTJADINGEN-BURHAVE	DE	53.5791667	8.3619444
ST12371	BUTZOW	DE	52.4925000	12.6000000
ST12379	CARPIN	DE	53.3591667	13.2441667
ST12423	DAHME	DE	51.8666667	13.4300000
ST12489	DOBRA	DE	51.5177778	13.4430556
ST12493	DOERNTHAL	DE	50.7330556	13.3330556
ST12508	DORNSTEDT	DE	51.4091667	11.7458333
ST12575	EDERTAL-EDERSEE	DE	51.1805556	9.0580556
ST12719	EXTERTAL-MEIERBERG	DE	52.1005556	9.1341667
ST12832	FROHNDORF (TALSPERRE)	DE	51.1452778	11.2063889
ST12842	FUERSTLICH DREHNA	DE	51.7563889	13.8055556
ST13066	GROSS OESINGEN	DE	52.6463889	10.4341667
ST13083	GROSSEFEHN-STRACKHOLT	DE	53.3691667	7.6402778
ST13113	GRUENEFELD	DE	52.6830556	12.9666667
ST13158	HAGEN-RUMMENOHL	DE	51.2716667	7.5341667
ST13160	HAGENOW	DE	53.4333333	11.1766667
ST13246	HARTMANNSDORF-REICHENAU	DE	50.8316667	13.5888889
ST13311	HEMMOOR	DE	53.6950000	9.1480556
ST13330	HERINGEN/WERRA	DE	50.8950000	10.0105556
ST13579	JUELICH (KLAERANLAGE)	DE	50.9313889	6.3430556
ST13659	KIRCHHUNDEM-RHEIN-WESER-TURM	DE	51.0705556	8.1969444

ST13688	KLEVE-SCHENKENSCHANZ	DE	51.8350000	6.1102778
ST13703	KLOSTERLAUSNITZ BAD	DE	50.9133333	11.8613889
ST13815	LAAR KREIS GRAFSCHAFT BENTHEIM	DE	52.6150000	6.7441667
ST13972	LOCHUM	DE	50.6094444	7.8583333
ST14047	MAERKISCH-BUCHHOLZ	DE	52.1080556	13.7616667
ST14165	KRUEMMEL	DE	53.2680556	12.7216667
ST14180	MOEHNESEE (MOEHNETALSPERRE)	DE	51.4925000	8.0638889
ST14940	HAREN-RUETENBROCK	DE	52.8605556	7.1025000
ST15198	SELFKANT-HAVERT	DE	51.0391667	5.9066667
ST15217	SIEDENBURG-BORSTEL	DE	52.6616667	8.9791667
ST15621	VERDEN-DAUELSEN	DE	52.9580556	9.2266667
ST15830	WERNE-WESSEL	DE	51.7155556	7.6830556
ST16015	WURZEN	DE	51.3705556	12.7513889
ST16091	GRUENBERG	DE	53.3530556	14.1491667
ST144	WEST TERSCHELLING	NL	53.3666667	5.2166667
ST145	DE KOOY-1	NL	52.9166667	4.7833333
ST147	GRONINGEN-1	NL	53.1833333	6.6000000
ST148	TER APEL	NL	52.8833333	7.0666667
ST149	HOORN	NL	52.6500000	5.0666667
ST150	HEERDE	NL	52.4000000	6.0500000
ST151	HOOFDDORP	NL	52.3166667	4.7000000
ST152	DE BILT-1	NL	52.1000000	5.1833333
ST153	WINTERSWIJK	NL	51.9833333	6.7000000
ST154	KERKWERVE	NL	51.6833333	3.8666667
ST157	OUDENBOSCH	NL	51.5666667	4.5333333
ST158	ROERMOND	NL	51.1833333	5.9666667
ST505	ST. ANNA PAROCHIE	NL	53.2833333	5.6666667
ST508	BERGUMERDAM	NL	53.1833333	5.9833333
ST513	ZOUTKAMP	NL	53.3333333	6.3000000
ST520	SCHELLINGWOUDE	NL	52.3833333	4.9666667
ST536	LEMMER (GEMAAL BUMA)	NL	52.8333333	5.7166667
ST539	GROOT AMMERS	NL	51.9333333	4.8166667
ST541	SCHEVENINGEN	NL	52.1166667	4.3000000
ST551	PUTTEN (GLD)	NL	52.2500000	5.6166667
ST556	TIEL	NL	51.9166667	5.4333333
ST560	DOETINCHEM	NL	51.9666667	6.2833333

ST2573	VLISSINGEN-1	NL	51.4500000	3.6000000
ST576	OOSTERHOUT	NL	51.6333333	4.8333333
ST582	HELMOND	NL	51.4833333	5.6166667
ST2568	MAASTRICHT-2	NL	50.8500000	5.6833333
ST2357	AMMERZODEN	NL	51.7500000	5.2166667
ST2432	HEILOO	NL	52.6000000	4.7333333
ST2476	MILL	NL	51.6833333	5.8000000
ST2481	NIEUW HELVOET	NL	51.8333333	4.1166667
ST2483	NIEUWOLDA	NL	53.2500000	6.9833333
ST2490	OLDEHOLTPADE	NL	52.9000000	6.0500000
ST2494	OOSTERBEEK	NL	52.0166667	5.8333333
ST2523	STEENWIJKSMOER	NL	52.6666667	6.6833333
ST2533	VAALS	NL	50.7833333	6.0000000
ST20541	HORST	NL	51.4591667	6.6111111

A.1.2 Box-plots

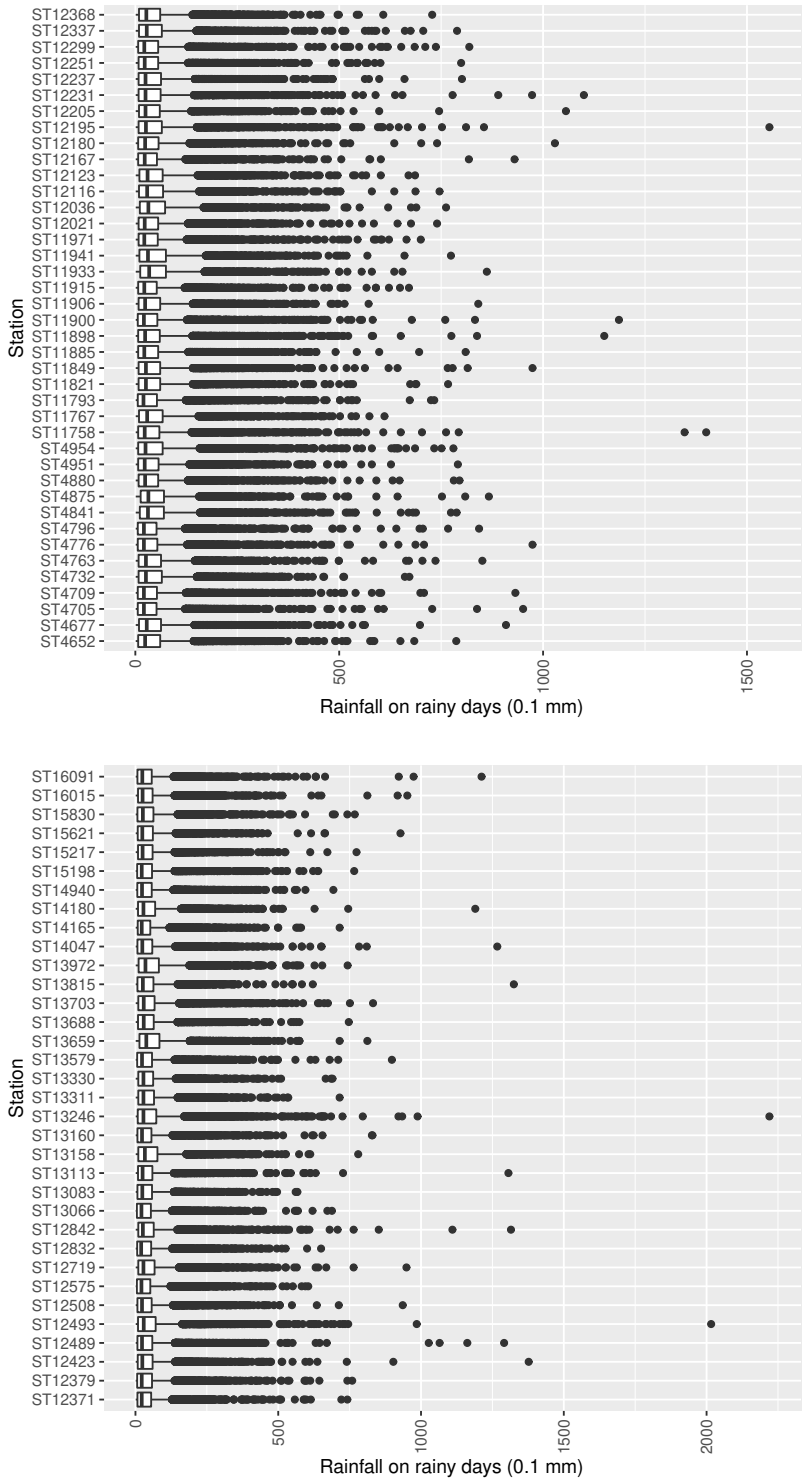


Figure 15: Boxplots of the distribution of daily rainfall amounts on rainy days in summer for the remaining German stations.

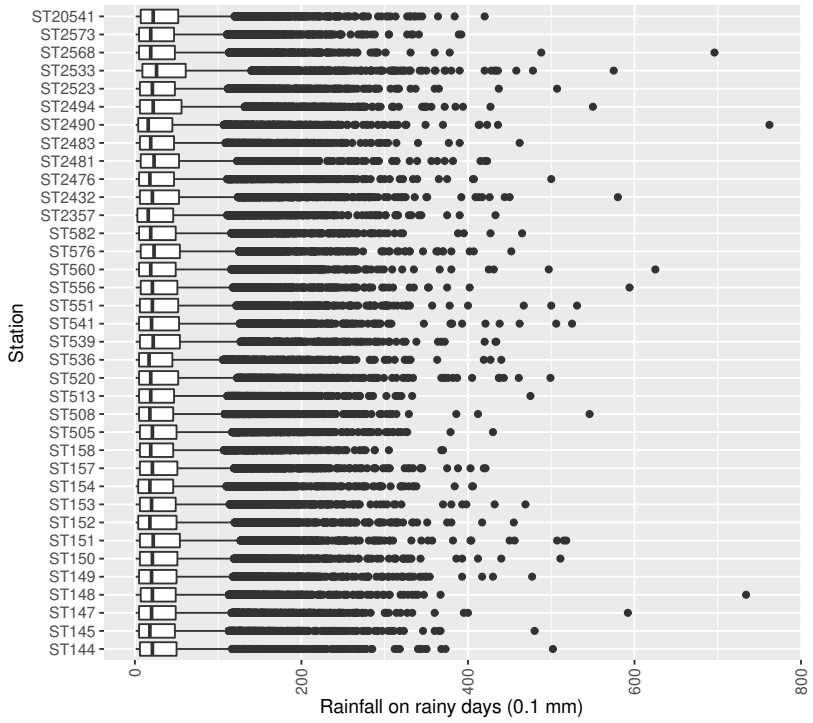


Figure 16: Boxplots of the distribution of daily rainfall amounts on rainy days in winter for the Dutch stations.

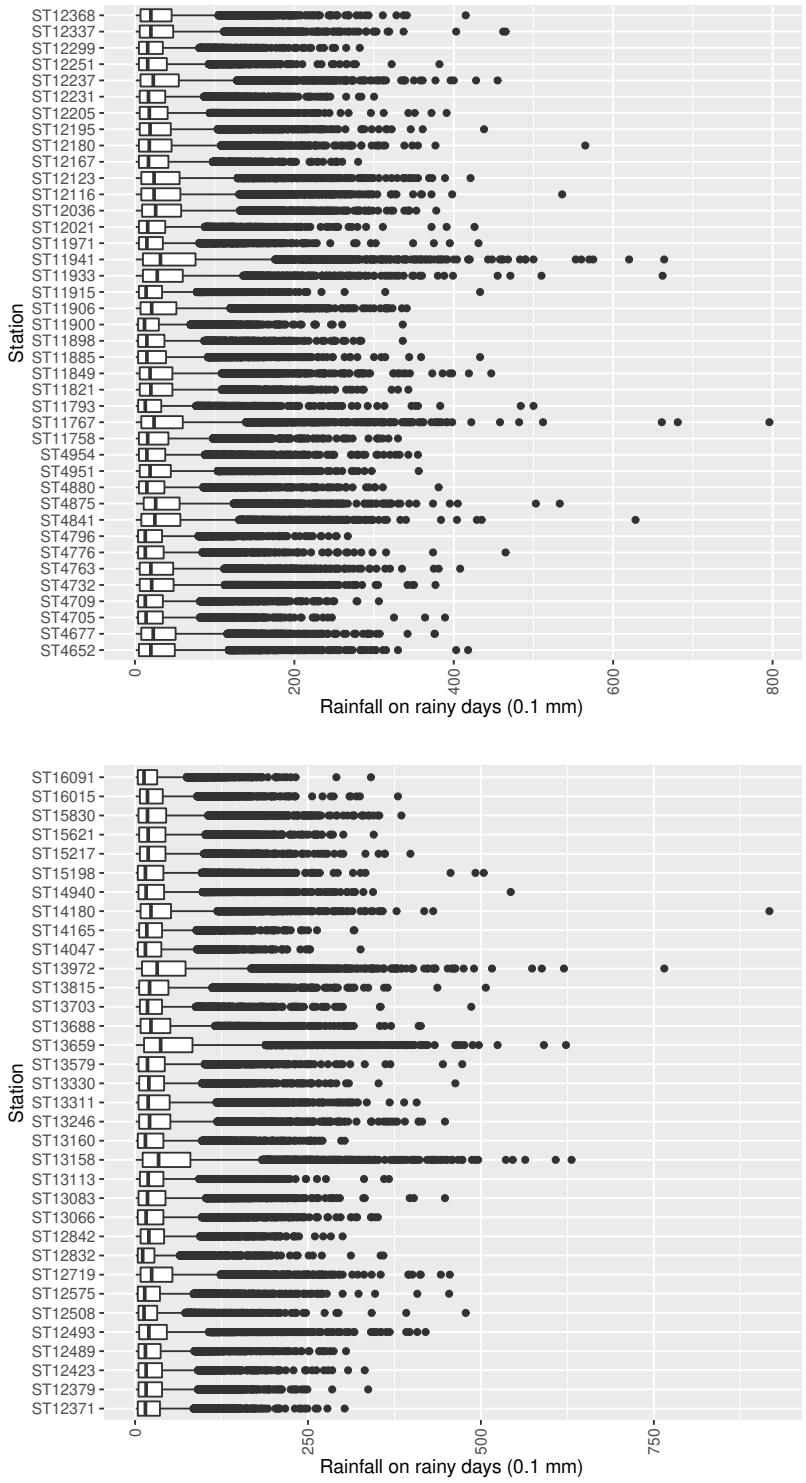
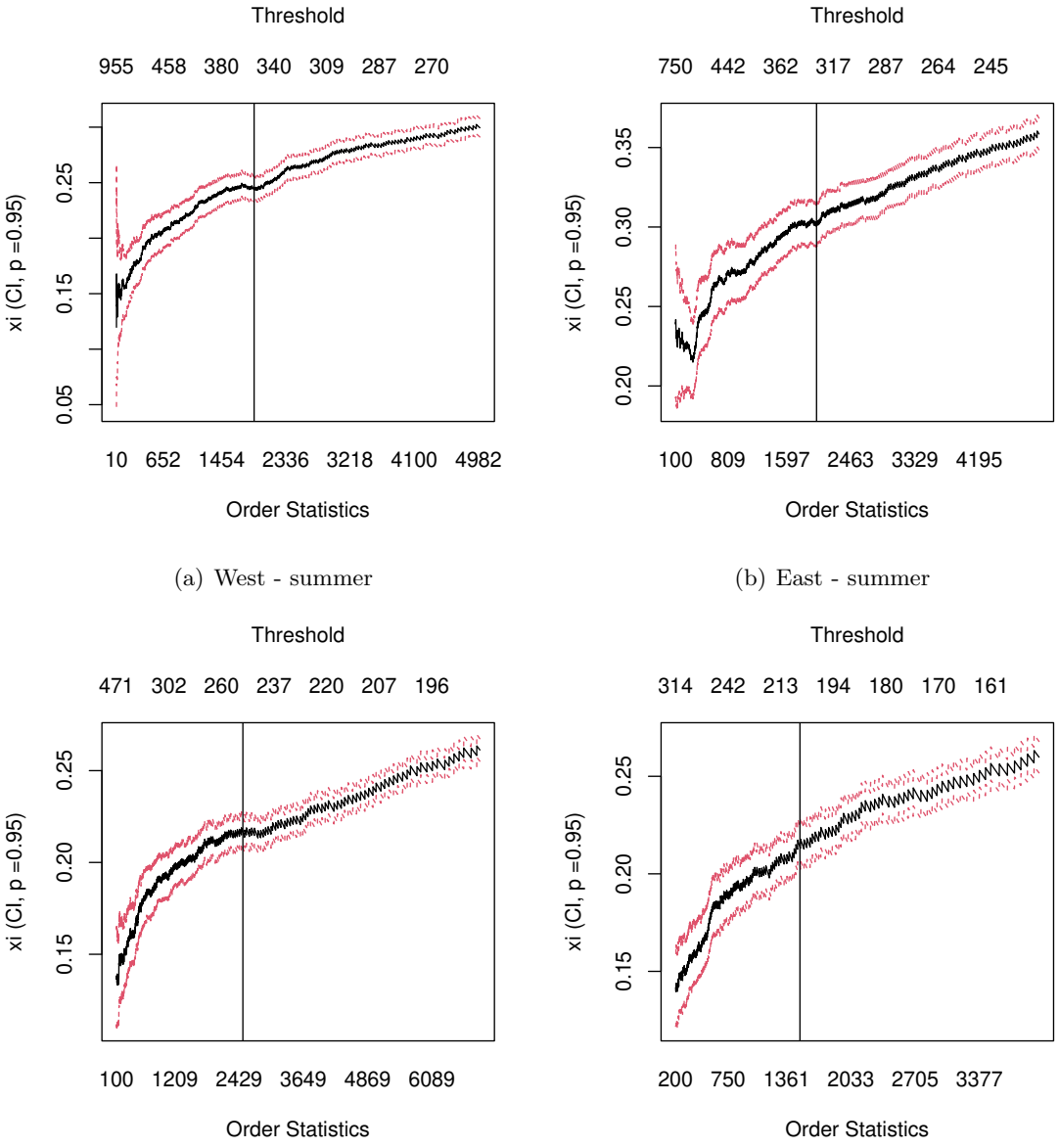


Figure 17: Boxplots of the distribution of daily rainfall amounts on rainy days in winter for the remaining German stations.

A.2 Results

A.2.1 Hill plots



(a) West - summer

(b) East - summer

(c) West - winter

(d) East - winter

Figure 18: Hill plots for the spatial frequency analysis.

A.2.2 p-values

Table 7: p-values for all stations of the corresponding test statistics considered in the research.

STAID	T_4_sum	T_4_win	T_1_sum	T_1_win	T_n_sum	T_n_win
ST41	0.61493	0.36204	0.46298	0.31963	0.80120	0.99225
ST42	0.38025	0.47302	0.11296	0.00306	0.99229	1.00000
ST43	0.47142	0.07762	0.34520	0.25143	0.16740	0.55610
ST47	0.27027	0.73778	0.20701	0.00002	0.65515	0.21936
ST49	0.82652	0.30046	0.00835	0.04546	0.69319	0.99955
ST55	0.48016	0.40443	0.95104	0.08242	0.69319	0.99225
ST324	0.08256	0.06666	0.80034	0.03530	0.76648	0.99955
ST356	0.53185	0.54152	0.11035	0.23780	0.77673	0.71479
ST474	0.46136	0.38180	0.75183	0.84045	0.99229	1.00000
ST476	0.79985	0.43117	0.17759	0.06326	0.95742	0.88183
ST477	0.67237	0.62504	0.36184	0.43030	0.91749	1.00000
ST479	0.30511	0.40603	0.71987	0.46129	0.92704	0.99951
ST484	0.46207	0.28931	0.42155	0.78688	0.18460	0.98477
ST487	0.45142	0.32868	0.00019	0.78035	0.69319	0.99875
ST812	0.14072	0.52420	0.47854	0.02973	0.47412	0.00000
ST3991	0.37891	0.35325	0.44258	0.35558	0.98581	1.00000
ST3992	0.26434	0.15748	0.19744	0.00369	0.76648	1.00000
ST3996	0.54546	0.66535	0.37054	0.13649	0.73911	0.99999
ST3999	0.39709	0.38431	0.58116	0.03309	0.84438	0.64792
ST4001	0.93903	0.66168	0.02492	0.64794	0.96427	1.00000
ST4014	0.26275	0.12268	0.05382	0.74457	0.22320	1.00000
ST4024	0.09603	0.34322	0.06071	0.59392	0.91749	1.00000
ST4050	0.98295	0.20618	0.59086	0.37369	0.97640	1.00000
ST4071	0.72987	0.20349	0.45012	0.00329	0.98581	0.99467
ST4108	0.42526	0.48091	0.15716	0.02680	0.57922	1.00000
ST4138	0.47142	0.30211	0.89682	0.17120	0.97640	0.99987
ST4148	0.39870	0.39559	0.49717	0.00018	0.97640	0.98428
ST4151	0.01960	0.22028	0.28663	0.03344	0.72990	0.59888
ST4157	0.53969	0.70045	0.52324	0.08667	0.98932	0.55610
ST4217	0.03250	0.45834	-	0.21566	-	0.84816
ST4300	0.17364	0.79781	0.15342	0.02549	0.93958	0.27099

ST4349	0.45262	0.41197	0.31572	0.14688	0.89301	0.99999
ST4365	0.82923	0.72735	0.67901	0.06079	0.94789	0.99881
ST4474	0.10203	0.33360	0.16259	0.01093	0.87456	0.99743
ST4475	0.83496	0.03674	0.95190	0.09200	0.73911	0.97584
ST4505	0.03570	0.36963	-	0.53975	-	0.99467
ST4539	0.80242	0.71677	0.94280	0.42532	0.99841	1.00000
ST4573	0.39637	0.47284	0.33136	0.39646	0.97640	0.96538
ST4633	0.24259	0.34510	0.00759	0.41460	0.61765	0.55610
ST4651	0.46636	0.38784	0.01410	0.51901	0.66287	0.02541
ST4652	0.54106	0.58990	0.10973	0.33553	0.89301	0.04897
ST4677	0.49655	0.40182	0.88107	0.42805	0.92704	1.00000
ST4705	0.45458	0.34742	0.10887	0.53198	0.99742	1.00000
ST4709	0.21995	0.54409	0.44503	0.02639	0.65515	0.99225
ST4732	0.65299	0.36355	0.52349	0.10582	0.99606	1.00000
ST4763	0.73371	0.69400	0.50609	0.02786	0.97640	0.99951
ST4776	0.24696	0.57794	0.09768	0.86051	0.89301	0.99996
ST4796	0.38172	0.88755	0.40821	0.13651	0.98932	1.00000
ST4841	0.34518	0.14415	0.30157	0.13589	0.62430	0.71479
ST4875	0.84283	0.65974	0.19210	0.15165	0.97640	0.89187
ST4880	0.57731	0.53758	0.27309	0.59350	0.65515	0.68668
ST4951	0.29767	0.90664	0.08160	0.63047	0.98932	0.51414
ST4954	0.71874	0.66459	0.14211	0.06059	0.18460	0.64316
ST11758	0.85831	0.18417	0.28371	0.08458	0.37493	0.68668
ST11767	0.53437	0.26232	0.21127	0.00002	0.61765	0.00000
ST11793	0.19481	0.25163	0.06165	0.21793	0.76648	0.95843
ST11821	0.84399	0.27799	0.19113	0.09500	0.98932	0.81092
ST11849	0.09334	0.31216	0.74331	0.24685	0.76648	0.12458
ST11885	0.33916	0.40270	0.46964	0.48024	0.97117	0.91176
ST11898	0.11336	0.57989	0.17211	0.06076	0.37493	0.97375
ST11900	0.19081	0.53858	-	0.17697	-	1.00000
ST11906	0.19405	0.34104	0.29359	0.08007	0.83427	0.14043
ST11915	0.42930	0.25316	0.08064	0.61449	0.99742	1.00000
ST11933	0.06467	0.33225	0.09201	0.52439	0.84438	0.51490
ST11941	0.17922	0.53426	0.78727	0.14504	0.90256	0.00000
ST11971	0.25253	0.81806	0.93962	0.84873	0.80120	0.97375
ST12021	0.29102	0.79128	0.13495	0.01399	0.93958	0.93780

ST12036	0.12864	0.09569	0.25432	0.00007	0.80120	0.04272
ST12116	0.60919	0.34711	0.38151	0.00192	0.77673	0.83952
ST12123	0.65386	0.76214	0.03594	0.09303	0.90256	0.83952
ST12167	0.43426	0.21272	0.34847	0.03954	0.95742	0.99996
ST12180	0.52285	0.48413	0.98646	0.45728	0.97640	0.99987
ST12195	0.72781	0.53081	0.60129	0.09775	0.16740	0.51414
ST12205	0.02105	0.81686	0.80733	0.47207	0.97117	0.97375
ST12231	0.79971	0.78849	0.16730	0.61236	0.57922	0.99875
ST12237	0.38272	0.15406	0.48917	0.00013	0.96427	0.61425
ST12251	0.11432	0.29409	0.19703	0.50383	0.83427	1.00000
ST12299	0.87964	0.21191	0.04512	0.57166	0.91749	1.00000
ST12337	0.74822	0.40993	0.71548	0.00115	0.80120	0.43393
ST12368	0.24030	0.39480	0.06436	0.02089	0.97640	1.00000
ST12371	0.01597	0.30452	0.00841	0.62717	0.86429	1.00000
ST12379	0.34636	0.21280	0.75752	0.07708	0.93958	0.99996
ST12423	0.79075	0.82991	0.39871	0.97151	0.76648	0.99955
ST12489	0.20753	0.10430	0.99193	0.08440	0.54244	0.99647
ST12493	0.27089	0.81091	0.15040	0.44961	0.04393	0.01456
ST12508	0.23591	0.51252	0.07498	0.17624	0.80120	0.99647
ST12575	0.22666	0.93065	0.03636	0.16891	0.96427	0.99999
ST12719	0.02313	0.24529	0.73144	0.28083	0.70119	0.61425
ST12832	0.52993	0.53043	0.18998	0.86967	0.89301	1.00000
ST12842	0.25769	0.28659	0.30442	0.11519	0.34564	1.00000
ST13066	0.56678	0.47718	0.06318	0.39041	0.91749	0.51414
ST13083	0.39271	0.57100	0.01302	0.54810	0.99606	0.99998
ST13113	0.19833	0.53915	0.27215	0.77851	0.80120	1.00000
ST13158	0.13981	0.03523	-	0.00520	-	0.00000
ST13160	0.40341	0.06339	0.35176	0.11730	0.65515	0.81092
ST13246	0.90599	0.31917	0.74045	0.36346	0.03400	0.00054
ST13311	0.88064	0.22996	0.12661	0.00086	0.92704	0.95158
ST13330	0.11913	0.96402	0.14885	0.15971	0.98932	0.72915
ST13579	0.09862	0.39440	0.53310	0.62502	0.87456	1.00000
ST13659	0.14895	0.06269	0.08617	0.02383	0.58475	0.00001
ST13688	0.37882	0.38814	0.50710	0.00417	0.99841	0.96538
ST13703	0.26973	0.12226	0.07775	0.08469	0.31760	0.81092
ST13815	0.00608	0.40621	-	0.83343	-	0.97584

ST13972	0.30684	0.19487	0.47255	0.20351	0.34416	0.00120
ST14047	0.81451	0.38113	0.09314	0.36499	0.76648	0.99999
ST14165	0.84803	0.42905	0.15063	0.60086	0.98932	1.00000
ST14180	0.59181	0.01112	0.35965	0.71717	0.94789	0.61425
ST14940	0.35594	0.43200	0.09498	0.37631	0.99229	0.99881
ST15198	0.59919	0.08306	0.81288	0.28964	0.73911	0.99998
ST15217	0.77550	0.62779	0.66799	0.17437	0.99998	1.00000
ST15621	0.73786	0.46564	0.15082	0.17828	0.99229	1.00000
ST15830	0.14419	0.51747	0.16222	0.00594	0.77673	0.86613
ST16015	0.86830	0.57473	0.65948	0.08452	0.69319	0.99225
ST16091	0.56269	0.59741	0.07840	0.03112	0.86429	1.00000
ST144	0.85082	0.20930	0.22155	0.10542	0.99229	0.98428
ST145	0.04312	0.15154	0.05967	0.09404	0.99606	0.99467
ST147	0.24098	0.57232	0.15211	0.01007	0.99841	0.97584
ST148	0.12324	0.51090	0.55593	0.02064	0.99229	0.97584
ST149	0.43609	0.40269	0.00205	0.00131	0.99985	0.91399
ST150	0.10487	0.72926	0.25286	0.33448	0.92704	0.95158
ST151	0.52464	0.63077	0.01008	0.00615	0.90256	0.96538
ST152	0.30245	0.62735	0.14886	0.00096	0.98581	0.83952
ST153	0.24849	0.67025	0.64244	0.34716	0.90256	0.99467
ST154	0.51888	0.38017	0.17840	0.00007	0.90256	0.86613
ST157	0.54728	0.43323	0.45692	0.06365	0.77673	0.93450
ST158	0.15814	0.51808	0.66545	0.45147	0.92704	1.00000
ST505	0.23547	0.39449	0.71681	0.00019	0.84438	0.99743
ST508	0.21310	0.30445	0.16772	0.11366	0.99229	0.99059
ST513	0.53623	0.87752	0.48271	0.01648	0.99841	0.99998
ST520	0.50563	0.73161	0.01533	0.00050	0.94789	0.86613
ST536	0.97785	0.32605	0.46529	0.14541	0.77673	0.99059
ST539	0.09926	0.46017	0.35311	0.05086	0.94789	0.83952
ST541	0.16503	0.40494	0.10903	0.00217	0.92704	0.91399
ST551	0.63017	0.00049	0.22785	0.03592	0.87456	0.71479
ST556	0.54303	0.67773	0.15752	0.61871	0.98581	0.99059
ST560	0.06355	0.07576	0.96869	0.13592	0.96427	0.93450
ST2573	0.45673	0.73610	0.41469	0.02270	0.90256	0.89187
ST576	0.28805	0.58776	0.03730	0.00748	0.99985	0.93450
ST582	0.29165	0.72752	0.16222	0.53587	0.99229	0.99059

ST2568	0.28831	0.48567	0.02326	0.47725	0.99229	0.81038
ST2357	0.71379	0.77644	0.18274	0.05361	0.97640	0.93450
ST2432	0.75248	0.51907	0.08462	0.00655	0.97640	0.97584
ST2476	0.07381	0.31948	0.10025	0.03502	0.97640	0.99987
ST2481	0.22913	0.40065	0.45862	0.11385	0.70119	0.93450
ST2483	0.14219	0.25054	0.57668	0.17927	0.81121	0.83952
ST2490	0.03006	0.47037	0.95412	0.54530	0.99985	0.98428
ST2494	0.15462	0.76717	0.02995	0.85496	0.51009	0.17934
ST2523	0.93425	0.05861	0.08016	0.24498	0.77673	0.99881
ST2533	0.41313	0.84180	0.05115	0.09285	0.99606	0.99998
ST20541	0.48677	0.25016	0.52574	0.55992	0.54697	0.97584