

Extreme value analysis of non-stationary time series

Quantifying climate change using observational data
throughout Germany

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt

der Fakultät Mathematik und Naturwissenschaften
der Technischen Universität Dresden

von

M. Sc. Philipp Müller

geboren am 07.09.1989 in Halle-Saale

Eingereicht am 18.10.2018

Die Dissertation wurde in der Zeit von 10/2014 bis
10/2018 am Max-Planck-Institut für Physik komplexer Systeme angefertigt.

Eingereicht am 18.10.2018

1. Gutachter: Prof. Dr. Holger Kantz
2. Gutachter: Prof. Dr. Jürgen Kurths
3. Gutachter: Prof. Dr. Marc Timme

Abstract

The overall subject of this thesis is the massive parallel application of the extreme value analysis (EVA) on climatological time series. But unfortunately this won't be possible on such a scale with any available software package since they are all affected by conceptual problems in the fit of the generalized extreme value (GEV) or generalized Pareto (GP) distribution using the maximum likelihood method. These issues will be resolved by applying constrained optimization and the improved algorithm is provided as a part of a newly written software package called climex. Using this implementation, as well as the model class of the vector generalized linear models (VGLM), the changes in the extreme events of temperature and precipitation series throughout Germany will be analyzed. To embed the findings of this rather comprehensive study into a global context, a less complete treatment of the temperature series from the ERA-Interim reanalysis data set will be included as well.

Summary

The overall subject of this thesis is the massive parallel application of the extreme value analysis (EVA) on climatological time series. In this branch of statistics one strives to learn about the tails of a distribution and its upper quantiles, like the so-called 50 year return level, an event realized on average only once during its return period of 50 years. Since most studies just focus on average statistics and it's the extreme events that have the biggest impact on our life, such an analysis is key for a proper understanding of the climate change. In there a time series gets separated into blocks, whose maxima can be described using the generalized extreme value (GEV) distribution for sufficiently large block sizes.

But, unfortunately, the estimation of its parameters won't be possible on a massive parallel scale with any available software package since they are all affected by conceptional problems in the maximum likelihood fit. Both the logarithms in the negative log-likelihood of the GEV distribution and the theoretical limitations on one of its parameters give rise to regions in the parameter space inaccessible to the optimization routines, causing them to produce numerical artifacts. I resolved this issue by incorporating all constraints into the optimization using the augmented Lagrangian method. With my implementation in the open source package **climex** it is now possible to analyze large climatological data sets. In this thesis I used temperature and precipitation data from measurement stations provided by the German weather service (DWD) and the ERA-Interim reanalysis data set and analyzed them using both a qualitative method based on time windows and a more quantitative one relying on the class of vector generalized linear models (VGLM).

Due to the climate change a general shift of the temperature towards higher values and thus more hot and less cold extremes would be expect. Indeed, I could find the location parameters of the GEV distributions, which can be thought of as the mean event size at a return period of approximately the block size of one year, to increase for both the daily maximum and minimum temperatures. But the overall changes are far more complex and dependent on the geographical location as well as the considered return period, which is quite unexpected. E.g. for the 100 year return levels of the daily maximum temperatures a decrease was found in the east and the center of Germany for both the raw series and their anomalies, as well as a quite strong reduction for the raw series in the very south of Germany.

The VGLM-based non-stationary EVA resulted in significant trends in the GEV parameters for the daily maximum temperatures of almost all stations and for about half of them in case of the daily minima. So, there is statistically sound evidence for a change in the extreme temperatures and, surprisingly, it is not exclusively towards higher values. The analysis yielded several significant trends featuring a negative slope in the 10 year return levels.

The analysis of the temperature data of the ERA-Interim reanalysis data set yielded quite surprising results too. While in some parts of the globe, especially on land, the 10 year return levels were found to increase, they do in general decrease in most parts of the earth and almost entirely over the sea. But since we found a huge discrepancy between the results of the analysis using the station data within Germany and the results obtained for the corresponding grid points of the reanalysis data set, we can not be sure whether the patterns in the return levels of the ERA-Interim data are trustworthy.

Zusammenfassung

Das Ziel dieser Arbeit ist die massiv parallele Anwendung der Extremwertanalyse (EVA) auf klimatologischen Zeitreihen. Dieser Bereich der Statistik beschäftigt sich mit den Schwänzen von Wahrscheinlichkeitsverteilungen und deren großen Quantilen, wie z.B. dem sogenannten 50-jährigen Return Level. Dies ist ein Ereignis, welches im Mittel nur einmal innerhalb seiner Return Periode von 50 Jahren realisiert wird. Da sich aber die Mehrheit der wissenschaftlichen Studien auf die Analyse gemittelter statistischer Größen stützen, aber es gerade die extremen Ereignisse sind, welche unser Leben maßgeblich beeinflussen, ist eine solche EVA entscheidend für ein umfassendes Verständnis des Klimawandels. In der Extremwertanalyse wird eine Zeitreihe in einzelne Blöcke geteilt, deren Maxima sich bei hinreichend großer Blocklänge mittels der generalisierten Extremwertverteilung (GEV) beschreiben lassen.

Die Schätzung ihrer Parameter ist auf solch massiv parallelen Skalen jedoch mit keinem der verfügbaren Softwarepakete möglich, da sie alle vom selben konzeptionellen Problem der Maximum Likelihood Methode betroffen sind. Sowohl die Logarithmen in der negativen log-Likelihood der GEV Verteilung, als auch die theoretischen Beschränkungen im Wertebereich eines ihrer Parameter machen Teile des Parameterraumes für den Optimierungsalgorithmus unzugänglich und führen zur Erzeugung numerischer Artefakte durch die Routine. Dieses Problem konnte ich lösen, indem ich die Beschränkungen mittels der augmented Lagrangian Methode in die Optimierung integrierte. Mittels dem verbesserten Fit, den ich in dem Open Source Paket **climex** zur Verfügung stellte, ist es nun möglich beliebig viele Zeitreihen in einer parallelen Analyse zu behandeln. In dieser Arbeit verwende ich Temperatur- und Niederschlagszeitreihen des deutschen Wetterdienstes (DWD) und den ERA-Interim Reanalyse Datensatz in Kombination mit sowohl einer qualitativen Analyse basierend auf Zeitfenstern, als auch einer quantitativen, welche auf der Modellklasse der Vektor-generalisierten linearen Modellen (VGLM) beruht.

Aufgrund des Klimawandels ist intuitiv eine Verschiebung der Temperaturverteilung zu höheren Werten und damit mehr heiße und weniger kalte Temperaturextreme zu erwarten. Tatsächlich konnte ich für die täglichen Maximal- und Minimaltemperaturen einen Anstieg des Location Parameters finden, dem man sich als mittlere Ereignisgröße für eine Return Periode gleich der verwendeten Blocklänge von einem Jahr versinnbildlichen kann. Im Großen und Ganzen sind die Änderungen jedoch deutlich komplexer und hängen sowohl vom Ort, als auch von der Return Periode ab. Z.B. verringern sich die 100 jährigen Return Level der täglichen Maximaltemperaturen im Osten und im Zentrum Deutschlands für sowohl die unprozessierten Zeitreihen, als auch für deren Anomalien, und weisen eine besonders starke Reduktion im Süden des Landes für die unprozessierten auf.

Durch die VGLM-basierte, nicht-stationäre EVA konnte ich zeigen, dass nahezu alle Stationen für die täglichen Maximaltemperaturen, sowie rund die Hälfte aller Stationen für die täglichen Minimaltemperaturen, signifikante Trends in den Parameters der GEV Verteilung aufweisen. Somit war es mir möglich statistisch fundierte Beweise für Veränderungen in den extremen Temperaturen finden, die jedoch nicht ausschließlich in einer Verschiebung zu höheren Werten bestanden. Einige Stationen wiesen eine negativen Trend in ihren 10 jährigen Return Leveln auf.

Die Analyse der Temperaturzeitreihen des ERA-Interim Reanalyse Datensatzes ergab

ebenfalls überraschende Resultate. Während in einigen Teilen der Welt, hauptsächlich an Land, die 10 jährigen Return Level steigen, sinkt ihr Wert für den Großteil der Zeitreihen und fast über den gesamten Ozeanen. Da jedoch eine große Diskrepanz zwischen den Ergebnissen der Stationsdaten des DWD und den dazugehörigen Rasterpunkten im ERA-Interim Datensatz besteht, konnte nicht abschließend geklärt werden in wie weit die Resultate der Rasteranalyse der Natur entsprechen.

Contents

1. Introduction	3
2. Extreme value analysis	5
2.1. Theoretic foundations	5
2.1.1. The generalized extreme value (GEV) distribution	7
2.1.2. The generalized Pareto (GP) distribution	8
2.2. Parameter estimation	10
2.2.1. Maximum likelihood-based estimation	10
2.2.2. Maximum likelihood-based estimation in the EVA	13
2.2.3. Error estimation revisited	15
2.2.4. Initialization of the parameter optimization	23
2.2.5. Gumbel vs. GEV hypothesis	25
3. Optimization	30
3.1. Unconstrained optimization	30
3.1.1. Nelder-Mead	31
3.1.2. BFGS	33
3.1.3. Problems in fitting GEV and GP distributions	36
3.2. Concepts of constrained optimization	40
3.2.1. Augmented Lagrangian methods	41
3.3. Constrained optimization in the EVA	42
3.3.1. Application of the augmented Lagrangian method to the GEV and GP distribution	42
3.3.2. Numerical comparison of the optimization results	44
4. Short-range correlations in the EVA	48
4.1. Block maxima in stationary series with short-range correlations	49
4.2. Threshold exceedances in stationary series	52
4.3. Deseasonalization techniques	57
4.3.1. Decomposition	60
4.3.2. Loess	61
4.3.3. STL	61
4.3.4. Fourier-based methods	62
4.3.5. Further notes	64
5. Non-stationary analysis in the EVA	65
5.1. Non-stationarity vs. long-range correlation	65
5.2. Non-stationary modelling in the EVA	66
5.2.1. VGLM	67
5.2.2. VGAM	68

5.3.	Application to EVA	68
5.4.	Model selection	69
5.4.1.	Information criteria	69
5.4.2.	Likelihood ratio test	70
6.	Pitfalls of the EVA	72
6.1.	Inspection of the gauge data of the river Elbe	72
6.2.	Exploration of system dynamics	74
6.3.	Summary of the pitfalls in the EVA	77
7.	Analysis of the temperature and rainfall data provided by the DWD	78
7.1.	Preprocessing of the DWD data	79
7.2.	Structure of the analysis	81
7.3.	Temperature anomalies of the Potsdam station anomalies	83
7.3.1.	Time series and overall distribution	83
7.3.2.	Non-stationary EVA and significance	86
7.4.	Raw temperature series of the Potsdam station	91
7.4.1.	Time series and overall distribution	91
7.4.2.	Non-stationary EVA and significance	94
7.5.	Analysis of the temperature series within Germany	96
7.5.1.	Temperature anomalies	96
7.5.2.	Raw temperature series	102
7.6.	Analysis of the Potsdam precipitation data	107
7.6.1.	Time series and distribution	107
7.6.2.	Non-stationary EVA and significance	111
7.7.	Temporal evolution of the precipitation throughout Germany	114
8.	Application of the non-stationary EVA to the ERA-Interim data set	119
8.1.	The concept of reanalysis	119
8.2.	The ERA-Interim data set	120
8.3.	Comparability between the DWD and ERA-Interim data sets	121
8.4.	Time window analysis of the ERA-Interim data set	127
8.4.1.	VGLM analysis of the ERA-Interim data set	131
9.	Summary	134
A.	Analysis of the temperature and precipitation series provided by the DWD	137
A.1.	Temperature anomalies within Germany	137
A.1.1.	Non-stationary EVA and significance	145
A.2.	Raw temperature series throughout Germany	157
A.2.1.	Non-stationary EVA and significance	165
A.3.	Precipitation throughout Germany	177
A.3.1.	Non-stationary EVA and significance	183

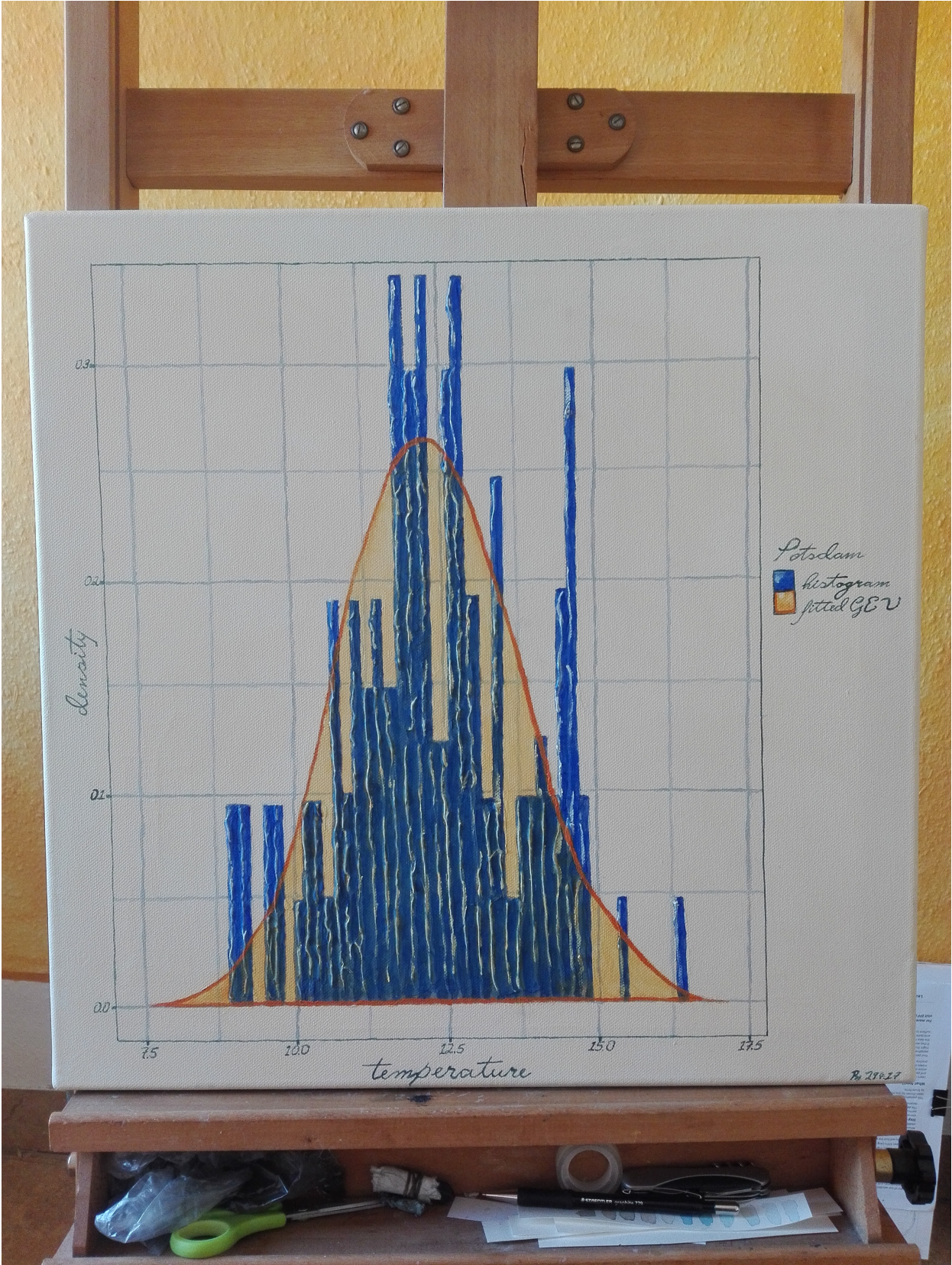
Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die Arbeit wurde am Max-Planck-Institut für Physik komplexer Systeme in der Abteilung *Nichtlineare Dynamik und Zeitreihenanalyse* angefertigt und von Prof. Dr. Holger Kantz betreut. Es hat kein früheres erfolgloses Promotionsverfahren stattgefunden. Ich erkenne die Promotionsordnung der Fakultät Mathematik und Naturwissenschaften der Technischen Universität Dresden vom 23.02.2011 in ihrer letzten Änderung vom 18.07.2018 an.

18.10.2018, Dresden

Philipp Müller



1. Introduction

Our world is constantly changing. Whether we look at climate, economy, or society itself everything is in motion and altering on various time scales. To thoroughly understand those systems and to be prepared for possible declines or catastrophes, one has to apprehend their changes to the largest extent possible. It is therefore not sufficient to restrict our view to bulk properties of distributions, e.g. a shift in the mean temperature, but instead it is very important to access their tails and larger quantiles. Making reliable estimates of such quantiles is thus a task a lot of disciplines, like hydrology, meteorology, or risk assessment for insurance companies, heavily rely on. After all, it is the extreme events causing the greatest harm to society as well as considerable financial losses. But before planing possible prevention measures, one first has to calculate the magnitudes of these extreme events in a robust way.

The main subject of this thesis is to investigate if and how extreme temperature and precipitation events do change throughout Germany. There are several ways to characterize the tails of a distribution and to infer its changes. The most straight forward way is to assume a specific form of the distribution, e.g. Gaussian, and to directly fit it using observed data. This approach has the advantage of being quite robust against outliers due to the sheer abundance of data points available for fitting. But if the true distribution deviates from our assumption, this advantage will turn against us. Since fitting procedures like maximum likelihood (ML), importance sampling, or the method of moments use equal weights for all data points, it is the bulk of the data, which is fitted best. The very tails of the distribution, on the other hand, won't be represented very well.

A second approach is to construct a model representing the physics of the underlying process and to recreate the distribution by running it in large ensembles. But unless one uses a model, which is explicitly tailored to properly describe the tails, its parameters are, again, chosen to fit the bulk of the data rather than its extreme events. For example the state of the art climate models are considered to perform poorly in representing the tails as pointed out by Franzke et al. (2015). Stochastic climate models, which were introduced by Hasselmann (1976), do yield better representations (Sura 2011, @Franzke2015) but lack the profound physical basis making the climate models so appealing in the first place. Therefore, we loose the big advantage of a model-based approach in comparison to the first one, which is the conclusions about the underlying physics of the process drawn from the parameters.

The extreme value theory is the third route to the properties of the tails. It is a purely statistical approach requiring the time series to contain neither long-range correlations nor non-stationarities and will be used throughout this thesis. If the time series is segmented into blocks of equal length, the asymptotic cumulative distribution of the maximal values extracted from the individual blocks can be approximated by the generalized extreme value (GEV) distribution. Using the estimated parameters, one can calculate

the return levels, or larger quantiles, of the time series. Another way to obtain them is to consider only those events of the original series exceeding a sufficiently high threshold as extreme. In this scenario the extracted events are fitted using the Generalized Pareto (GP) distribution. Both extreme value distributions are two sides of the same coin and can be transformed into each other. A detailed introduction into the basic theory and parameter estimation is given in chapter 2.

Since we want to perform a thorough analysis of the extreme temperature and precipitation data throughout Germany, a sole case study of a single station won't satisfy our requirements. Instead, we will take a larger number of stations and apply the analysis on a massive parallel scale. But when doing so one quickly realizes that none of the available software packages is capable of accomplishing this task. They all will throw errors for some but not necessary the same stations. The reason is a previously unknown conceptual problem of the maximum likelihood fit for the GEV and GP distribution. Chapter 3 will, firstly, give an introduction into the basic optimization algorithms used in the available packages. Afterwards, the source of the error is illustrated by a numerical simulation and, finally, a possible solution using the constrained optimization of the augmented Lagrangian method will be provided. This improved algorithm, as well as general tweaks in the estimation of the initial parameters and fitting errors presented in chapter 2, are implemented in my software packages *climex*. It is written in the statistical programming language R and licensed under GPL-3 allowing the reader (and everybody else) to use, modify, and redistribute it free of charge¹.

The background of the actual analysis, which we will perform on climatological time series in the last part of the thesis, will be explained in the following two chapters. Chapter 4 covers the basic stationary analysis in its different flavors including the necessary steps of preprocessing. Chapter 5 explains the non-stationary (time dependent) analysis. It introduces the model class of the vector generalized linear models (VGLM) and their application to the GEV or GP distribution. Before jumping into the extreme value analysis (EVA), a note of caution is sounded in chapter 6. The EVA is a very nice and powerful tool. But it can yield some quite misleading results when applied inappropriately. One of such examples will be illustrated using a case study of the gauge data of the river Elbe in Dresden.

Finally, we will perform the analysis of the temperature and precipitation data in chapter 7. Firstly, we will introduce the station data provided by the German weather service (DWD) and their preprocessing. Afterwards, the overall structure of the analysis will be explained and some selected results will be presented. Since the number of inspected quantities and produced figures is far too large, only a case study of the Potsdam stations and the changes in the 10 and 100 year return levels will be shown. All other findings, like the temporal evolution of the first four moments and GEV parameters, can be found in the appendix A. To embed those regional results into a global context, chapter 8 first introduces the ERA-Interim reanalysis data set and then applies the non-stationary EVA on it. Lastly, the thesis will be concluded with a summary in chapter 9.

¹The source code can be found in <https://gitlab.com/theGreatWhiteShark/climex> and a live example of an included web application can be accessed via climex.pks.mpg.de

2. Extreme value analysis

Most branches in statistics concentrate on the head of probability distribution functions, which is the part of the distribution holding most of its mass. While those methods do certainly well in common tasks like determining the most probable event and its uncertainty, they become increasingly less accurate when dealing with more and more rare events.

To obtain statistically sound insights into the larger quantiles of a distribution and thus into its most rare events, one can use the so called **extreme value analysis (EVA)**, which explicitly focuses on the tails. This quite young branch of statistics was founded by the pioneering papers of Frechét (1927) and Fisher and Tippett (1928). The 1958 book *Statistics of Extremes* by Gumbel (1958) brought the topic to a much broader attention and it quickly developed into a major field with its own journals and conferences.

The big success is especially due to its wide range of applications reaching from the analysis of weather extremes in climate science and earthquakes in geology to the risk assessment in finance and insurance industry. One of its most classical examples comes from hydrology. Imagine you were given the task of building a dyke, which is able to withstand a flood that occurs so seldom that it's realized (on average) only once every 300 years. But there is only as much as 80 years of daily river heights to perform your analysis on. So how can one get insights into the tails of the distribution of the river data and approximate the so called 300 year **return level**?

The EVA is the most promising approach to access the properties of the tails and it comes in two different flavors. In the first one the time series is segmented into blocks of equal size, from which only their maximal values will be extracted. The bigger the size of the blocks gets, the better the distribution of the maxima can be described by the **generalized extreme value (GEV)** distribution. Alternatively, a threshold is applied and all data below it will be discarded. The higher its values, the more appropriate the distribution of the exceedances can be described by the **generalized Pareto (GP)** distribution. After fitting one of these distributions one can use its parameters to obtain an estimate for arbitrary large quantiles of the distribution function of the original data including an approximation of the estimation errors.

The following sections are mostly based on the nomenclature and inspired by the structure of the book *An Introduction to Statistical Modeling of Extreme Events* by Coles (2004). For a more mathematical and rigorous treatment see e.g. *Extreme Value Theory - An introduction* by Haan and Ferreira (2006).

2.1. Theoretic foundations

We start by assuming our data to be a sequence of an independent and identically distributed (iid) random process X with the probability distribution function $F(z) = \Pr\{X \leq z\}$. This series gets split into consecutive blocks of size N and only their

maximal values will be extracted.

$$M_N = \max \{X_1, \dots, X_N\} \quad (2.1)$$

Due to the iid property of the underlying series the maxima M are independently distributed as well and their probability distribution can be calculated directly using F

$$\begin{aligned} \Pr \{M_N \leq z\} &= \Pr \{X_1 \leq z, \dots, X_N \leq z\} \\ &= \Pr \{X_1 \leq z\} \times \dots \times \Pr \{X_N \leq z\} \\ &= \{F(z)\}^N. \end{aligned} \quad (2.2)$$

In other words, if the distribution function of the process X is known, the one for the maxima M can be calculated as well. But in practice it is not possible to write down an explicit equation of F for most systems. So the distribution has to be estimated. As mentioned in the introduction, this will lead to an explosion of the fitting errors in the tails of the distribution since these deviations too are raised to the power of N .

In extreme value analysis a different route for determining the distribution of the M is used. For asymptotic block lengths ($N \rightarrow \infty$), a renormalization of the maxima M to

$$M_N^* = \frac{M_N - b_N}{a_N}, \quad (2.3)$$

and sequences of appropriate normalization constants $\{a_N > 0\}$ and $\{b_N\}$ the distribution $F^N(z)$ can be described by a family of non-degenerated limit distributions, the so called **extreme value distributions**. The renormalization in equation (2.3) is indeed necessary since otherwise the resulting probability distribution function would be a point mass at its upper end point.

Theorem 2.1

Let $M_N = \max \{X_1, \dots, X_N\}$ be a sequence of block maxima of length n taken from an independent and identically distributed series X .

If sequences of constants $\{a_N > 0\}$ and $\{b_N\}$ exists so that

$$\Pr \left\{ \frac{M_N - b_N}{a_N} \leq z \right\} \rightarrow G(z) \text{ as } N \rightarrow \infty,$$

with G being a non-degenerate cumulative distribution function, then G is a member of one of the following families:

$$\begin{aligned} I : G(z) &= \exp \left\{ - \exp \left[- \left(\frac{z-b}{a} \right) \right] \right\}, \quad -\infty < z < \infty \\ II : G(z) &= \begin{cases} 0, & z \leq b \\ \exp \left\{ - \left(\frac{z-b}{a} \right)^{-\alpha} \right\}, & z > b \end{cases} \\ III : G(z) &= \begin{cases} \exp \left\{ - \left[- \left(\frac{z-b}{a} \right)^\alpha \right] \right\}, & z < b \\ 1, & z \geq b, \end{cases} \end{aligned}$$

with $\alpha > 0$.

In a way theorem 2.1 can be thought of as an analogue to the central limit theorem. In the latter one adds independent and identically distributed (iid) random variables of finite variance. The more do take part in the summation process, the better the sum is distributed according to the Gaussian distribution. In the extreme value analysis the summing is replaced by blocking and the extraction of the maximal values. But the procedure is the same: The bigger the block lengths, the better the result can be described by its asymptotics, the family of extreme value distributions.

The independence from the probability distribution function F of the underlying series is the big advantage of this method. Given a sufficiently large block length, the tail of the distribution of the data can be approximated using the families of extreme value distribution in theorem 2.1 or their generalization, the **generalized extreme value (GEV)** distribution.

2.1.1. The generalized extreme value (GEV) distribution

The generalized extreme value distribution unifies the extreme value analysis. Prior to its usage it was common practice to either choose one out of three families of extreme value distributions, the Weibull, Gumbel, or Fréchet distribution, or to use hypothesis testing to determine the one providing the best fit. Now, the GEV distribution combines all three of them and which one to pick is controlled by an additional parameter, the shape parameter ξ . For values other than zero the cumulative distribution function (CDF) of the GEV distribution looks as follows:

$$G(z) = \exp \left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}, \xi \neq 0. \quad (2.4)$$

It is fully described by the location μ , scale σ , and shape ξ parameter. For $\xi > 0$ equation (2.4) follows the **Fréchet distribution** and for $\xi < 0$ the **Weibull** one. If the shape parameter becomes zero $\xi = 0$, the CDF reduces to the so called **Gumbel distribution**

$$G(z) = \exp \left\{ - \exp \left\{ - \left(\frac{z - \mu}{\sigma} \right) \right\} \right\}. \quad (2.5)$$

While all three families are heavy-tailed, they have a quite different support (see figure 2.1.1; note that the probability distribution function (PDF) is displayed and not the CDF). The Gumbel distribution is the only member with an unlimited support. But unlike other distributions, like e.g. the Gaussian one, it is not symmetric and features a positive skewness of roughly 1.14 and a heavy tail only towards positive z values.

The Fréchet and Weibull distribution, on the other hand, are only defined over a limited range of z since the term $\left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]$ in equation (2.4) has to remain positive. This results in the end point

$$z_{end} = \mu - \frac{\sigma}{\xi}. \quad (2.6)$$

For the Weibull distribution in figure 2.1.1 with $(\mu = 0, \sigma = 1, \xi = -0.5)$ the upper end point resides at 2 and the distribution is heavy-tailed towards negative z values. The Fréchet distribution, featuring the parameters $(\mu = 0, \sigma = 1, \xi = 0.5)$, has a heavy tail

towards positive z values and a lower end point at -2, which, in contrast to the Weibull one, is only barely visible.

![[Assuming sufficient large block lengths, the distribution of the individual block maxima can be approximated using the **generalized extreme value (GEV)** distribution. It incorporates three different families of extreme value distributions, which are chosen depending on the shape parameter ξ . For $\xi = 0$ the GEV distribution simplifies into the Gumbel distribution, for $\xi > 0$ into the Fréchet distribution, and for $\xi < 0$ into the Weibull distribution. To visualize the relations between the distributions, as well as their upper or lower end points, three PDFs are plotted with location $\mu = 0$, scale $\sigma = 1$, and shape $\xi = \{-0.5, 0, 0.5\}$]([_main_files/figure-latex/density-gev-1.png](#))

After obtaining the parameters of the GEV distribution using one of the methods described in Section 2.2 the high quantiles of the distribution of the underlying series can be calculated explicitly using

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

The quantile z_p is dubbed **return level** in the context of EVA and represents the magnitude of the event, which is realized on average only once per **return period** $1/p$. Using the block length and the quantization of the underlying series X , the return period can be assigned an unit. If e.g. our X would consist of a daily temperature time series and we would like to extract the annual maxima M , the return period is given in $1/p$ years.

2.1.2. The generalized Pareto (GP) distribution

A different route to obtain estimates for the high quantiles of a series in the context of extreme value analysis is to use the so called **peak over threshold** or just **threshold approach**. In there, every event is considered an extreme event if it exceeds a sufficiently high threshold u . The probability for such an extreme event can thus be calculated in the following way

$$\Pr\{X > u + y | X > u\} = \frac{1 - F(u + y)}{1 - F(u)}, y > 0, \quad (2.7)$$

with the same series X we considered in the beginning of the chapter. Equation (2.7) is expressed in terms of a conditional probability in order to normalize the resulting probability.

Now, if the distribution of the maxima M of the underlying series can be approximated by the GEV distribution for sufficiently large block lengths and the events actually exceed the threshold $X > u$, the distribution of the threshold exceedances $X - u$ can be approximated by the **generalized Pareto (GP)** distribution. Its parameters are directly related to the ones of the GEV distribution. The shape parameter of both distributions is asymptotically identical and independent of the actual block size or threshold height. The scale parameter of the GP distribution, on the other hand, is a rescaled version of its counterpart in the GEV distribution.

Theorem 2.2

Let $M_N = \max\{X_1, \dots, X_N\}$ be a sequence of block maxima taken from an independent and identically distributed series X .

If theorem 2.1 is fulfilled and N is large

$$\Pr\{M_N \leq z\} \approx \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\},$$

for $\sigma > 0$, then the distribution corresponding to equation (2.7) can be approximated, for sufficiently high thresholds u , by

$$H(y) = 1 - \left(1 + \frac{\xi \cdot y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}}, \quad (2.8)$$

with $y = X - u > 0$, $(1 + (\xi y)/\tilde{\sigma}) > 0$, and

$$\tilde{\sigma} = \sigma + \xi(u - \mu). \quad (2.9)$$

As for the GEV distribution, the shape parameter plays an important role in determining the overall characteristics of the GP distribution (see figure 2.1 where, again, the PDF and not the CDF of the distributions is displayed). While for positive values the distribution is of the classical **Pareto-type** and has no upper limit, for $\xi < 0$ it's of the **Beta-type** and the upper end point of the GP distribution is

$$y_{end} = u - \frac{\tilde{\sigma}}{\xi}. \quad (2.10)$$

For $\xi = 0$ the distribution has an unlimited support and simplifies to the **exponential** distribution

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), \quad y > 0. \quad (2.11)$$

The return levels of the GP distribution are calculated via

$$y_m = \begin{cases} u + \frac{\tilde{\sigma}}{\xi} \left[(m\zeta_u)^\xi - 1\right] & , \text{ for } \xi \neq 0 \\ u + \tilde{\sigma} \log(m\zeta_u) & , \text{ for } \xi = 0, \end{cases}$$

with $\zeta_u = \Pr\{X > u\}$ being the probability of an event to exceed the threshold u . The return level y_m can thus be interpreted as the average value exceeded once every m occurrences of a threshold exceedance. Since this is usually not a helpful unit for interpreting the results of the EVA, m often is rewritten to represent a temporal unit. To e.g. obtain the N_{ret} -th year return level using the GP distribution, we would express m in terms of $m = N_{ret}n_y$, with n_y being the average number of exceedances per year.

For the sake of a more concise notation the rescaled scale parameter of the GP distribution $\tilde{\sigma}$ in equation (2.9) will be expressed as σ in the remainder of this thesis.

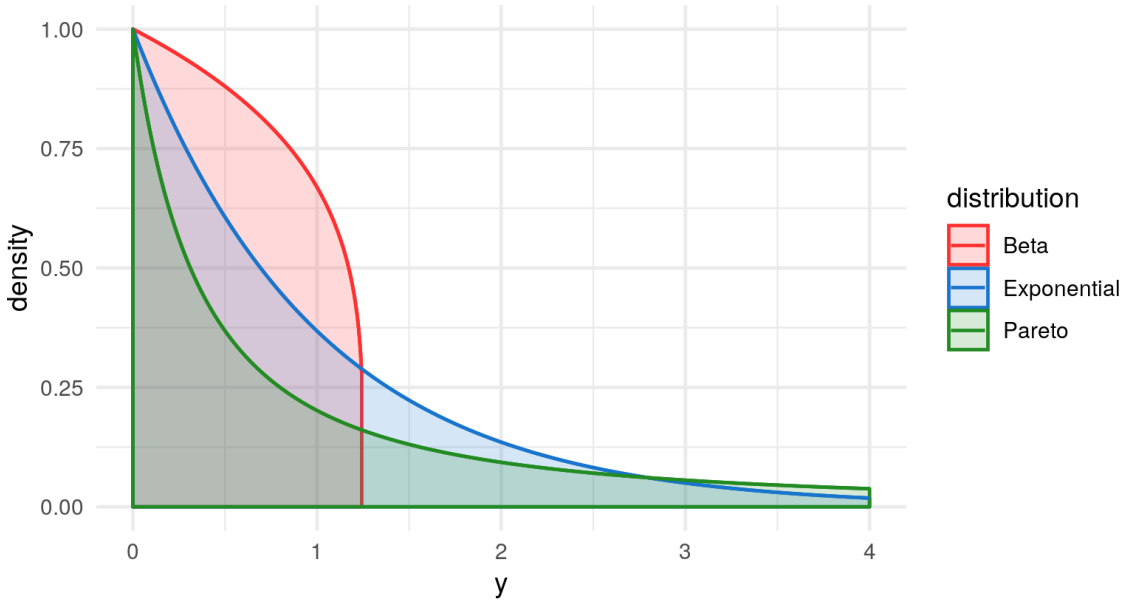


Figure 2.1.: Assuming sufficiently high thresholds, the distribution of exceedances can be approximated using the **generalized Pareto (GP)** distribution. To highlight particularities of this distribution, three different versions with scale $\tilde{\sigma} = 1$ and shape $\xi = \{-0.8, 0, 1.8\}$ are displayed.

2.2. Parameter estimation

The fitting of the parameters of the GEV and GP distribution can be done in a variety of ways, but the most prominent one is the **maximum likelihood (ML)** approach. The major benefits of this well established statistical procedure are its comparably good performance (Hosking, Wallis, and Wood 1985), its asymptotic properties (Smith 1985), and the possibility to incorporate functional dependencies (e.g. trends) into the parameters (Gilleland, Ribatet, and Stephenson 2013). Other approaches, like the method of moments, probability weighted moments (PWM) (Hosking, Wallis, and Wood 1985), L-moments (Hosking 1990), generalized probability weighted moments (Diebolt et al. 2008), or order statistics (Castillo and Hadi 1994), lack at least one of these properties.

There are also frameworks generalizing the usage of the ML approach in the context of EVA, like the method of generalized ML (Martins and Stedinger 2000), penalized ML (Coles and Dixon 1999), and Bayesian approaches (Coles and Tawn 1996). But they require priors or penalty functions assuming the user to have application specific a priori knowledge and thus restricting the use of the algorithms to domain experts only. In addition, Bayesian approaches are quite time consuming since they are relying on Monte Carlo Markov Chain integration (Coles and Tawn 1996).

2.2.1. Maximum likelihood-based estimation

The overall goal of the ML approach is to fit a model, which is most likely to represent the probability density function of the underlying data. The *likelihood*, or whether or not

a model is “likely” to fit the data, is thus determined by the joint probability density to observe the data given the parameters. This probability factorizes since the ML approach requires the underlying data to be a series of independent and identically distributed events x of length n .

$$L(\theta) = f(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n f_i(x_i | \theta) \quad (2.12)$$

Since from a computer science point of view it is usually more convenient to deal with summations instead of multiplications and to perform a minimization instead of an maximization, the estimation is done by minimizing the *negative log-likelihood* l

$$l(\theta) = -\log L(\theta) = -\sum_{i=1}^n \log f_i(x_i | \theta). \quad (2.13)$$

The ML approach itself builds on top of an asymptotic limit law. Only for an infinite length n of our series x the fitted ML estimators $\hat{\theta}_0$ will coincide with the parameters of the true distribution θ_0 . But for larger and larger n , given that the series fulfills a number of regularity conditions, the approximation becomes better and the distance between $\hat{\theta}_0$ and θ_0 negligible.

The trade-off between making the block length or threshold height as big as possible in order for the extracted extreme events to be properly described by an extreme limit distribution and, on the other hand, to get as many extreme events for the fitting procedure as possible, is one of the central problems in EVA. A common way to handle this task is to gradually increase the block length or threshold height, calculate the corresponding shape parameter, and plot the two against each other. As soon as the distribution of the extreme events can be sufficiently described by either the GEV or GP the graph will reach a plateau. Now, the higher we set the threshold or the bigger we set the block size, the lesser is the amount of extreme events we can extract. Thus, the fitting error becomes more and more dominant and the constant behaviour is washed out. Therefore, it's a good rule of thumb to pick a value from the first half of the plateau. But all these advises only hold in the absence of correlations and non-stationarities in the data. If you e.g. deal with temperature data featuring an annual cycle, make sure you use the methods described in chapter 4 to get rid of the short-range correlations.

The error estimates for the parameters $\hat{\theta}$ can be obtained as follows

Theorem 2.3

For $x = x_1, \dots, x_n$ being a series of independent events drawn from a common parametric family of distributions \mathcal{F} and $\hat{\theta}_0$ being the corresponding d -dimensional maximum likelihood estimator of the model parameters, under suitable regularity conditions $\hat{\theta}_0$ is approximately distributed according to a d -dimensional multivariate normal distribution

$$\hat{\theta}_0 \sim MVN_d \left(\theta_0, I_E(\theta_0)^{-1} \right), \quad (2.14)$$

where

$$I_E(x) = \begin{bmatrix} e_{1,1}(\theta) & \dots & \dots & e_{1,d}(\theta) \\ \vdots & \ddots & e_{i,j}(\theta) & \vdots \\ \vdots & e_{j,i}(\theta) & \ddots & \vdots \\ e_{d,1}(\theta) & \dots & \dots & e_{d,d}(\theta) \end{bmatrix}$$

is the **expected information matrix** with

$$e_{i,j}(\theta) = E \left\{ -\frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\theta) \right\}. \quad (2.15)$$

The expected information matrix describes the curvature of the log-likelihood surface and relies on the true values θ_0 , which are of course unknown. Instead, one can replace them with their maximum likelihood estimates (MLE) $\hat{\theta}_0$ to obtain the **observed information matrix** I_O , also called the **Hessian**.

Using theorem 2.3 the standard errors of the MLE can be approximated by the square roots of the diagonal elements $\psi_{i,i}$ of the inverse observed information matrix I_O^{-1}

$$\text{std}(\hat{\theta}_i) = \sqrt{\psi_{i,i}}. \quad (2.16)$$

The approximate confidence intervals of $\hat{\theta}_0$, on the other hand, would be

$$\hat{\theta}_i \pm z_{\frac{\alpha}{2}} \sqrt{\psi_{i,i}}, \quad (2.17)$$

with $z_{\frac{\alpha}{2}}$ being the $(1 - \alpha/2)$ quantile of the standard normal distribution.

In order to obtain the error estimate for the return levels, we need an additional ingredient.

Theorem 2.4

If a scalar function $\phi = g(\theta)$ is applied on the MLE $\hat{\theta}_0$ of θ , then $\hat{\phi}_0 = g(\hat{\theta}_0)$ is the MLE of ϕ (see Coles (2004) for this and all other theorems contained in this chapter).

By combining theorem 2.3 and 2.4, we can derive the so called **delta method** to estimate the error of $\hat{\phi}_0$.

Theorem 2.5

If $\phi = g(\theta)$ is a scalar function and $\hat{\theta}_0$ the MLE of θ with an approximated covariance matrix V_θ , then the MLE $\hat{\phi}_0$ satisfies

$$\hat{\phi}_0 \sim N(\phi_0, V_\phi),$$

where

$$V_\phi = \nabla \phi^T V_\theta \nabla \phi,$$

with

$$\nabla \phi = \left[\frac{\partial \phi}{\partial \theta_1}, \dots, \frac{\partial \phi}{\partial \theta_d} \right]^T$$

evaluated at $\hat{\theta}_0$.

Since both equation (2.1.1) and (2.1.2) are scalar functions of the estimated parameters, we can use theorem 2.5 to obtain an approximation of the fitting error of the return level.

2.2.2. Maximum likelihood-based estimation in the EVA

The negative log-likelihood (NLLH) (2.13) of the **GEV** distribution can be expressed as

$$l(z; \mu, \sigma, \xi) = n \cdot \log \sigma + \sum_{i=1}^n \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} + \left(1 + \frac{1}{\xi} \right) \sum_{i=1}^n \log \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right], \quad (2.18)$$

for $\xi \neq 0$ and with $\sigma > 0$, and as

$$l(z; \mu, \sigma) = n \cdot \log \sigma + \sum_{i=1}^n \left(\frac{z_i - \mu}{\sigma} \right) + \sum_{i=1}^n \exp \left\{ - \left(\frac{z_i - \mu}{\sigma} \right) \right\} \quad (2.19)$$

for $\xi = 0$ and, again, with $\sigma > 0$. Note that (2.18) is only defined for z values with $1 + \xi \left(\frac{z - \mu}{\sigma} \right) > 0$.

To obtain an estimate of the parameters of the GEV distribution, we will minimize equation (2.18) or (2.19) using a numerical optimization scheme. This consists of two main parts covered in later sections of the thesis. Firstly, all numerical optimizations have to be initialized with a parameter combination $(\mu_{init}, \sigma_{init}, \xi_{init})$. While the previously unknown impact of the choice of the initial parameter combinations on the estimation result is discussed in section 3.3, the different heuristics to obtain them are reviewed in section 2.2.4.1. Secondly, an optimization algorithm will gradually improve the initial parameters. An overview of different methods used in EVA, their influence on the optimization result, and why it is strongly advised to use the augmented Lagrangian method is described in chapter 3.

An important note to make in the context of the maximum likelihood estimation of the GEV distribution is *when* to use which of the two equations (2.18) and (2.19). While most books in the field of EVA do not cover this topic, Stuart Coles suggests in his book Coles (2004) to use the Gumbel likelihood (2.19) whenever the shape parameter gets less than a predefined value. But I found this behaviour to produce quite large numerical artifacts. Since the choice of the initial parameter combination is based on a heuristic, its always possible to initialize the shape parameter with the wrong sign. During the optimization the value of ξ gets gradually improved and thus changes sign at some point. Just by bad luck its absolute value can be less than the predefined threshold and the NLLH change from the GEV to Gumbel one. But when it does, there is no way for the algorithm to change back to the GEV one and the shape parameter is permanently trapped at zero. To avoid this behaviour the switch between (2.18) and (2.19) should only occur if the shape parameter is perfectly zero.

The errors of the MLE of the GEV parameters can be obtained using equation (2.16) by inverting the Hessian evaluated at the optimization results and extracting the square roots of its diagonal elements. The error of the return levels, on the other hand, can be calculated via the delta method in theorem 2.5 using

$$\text{std}(z_p) \approx \sqrt{\nabla z_p^T V \nabla z_p},$$

with V being the inverted Hessian (an estimate of the covariance matrix) and

$$\begin{aligned} \nabla z_p^T &= \left[\frac{\partial z_p}{\partial \mu}, \frac{\partial z_p}{\partial \sigma}, \frac{\partial z_p}{\partial \xi} \right] \\ &= \left[1, -\xi^{-1}(1 - x_p^{-\xi}), \sigma \xi^{-2}(1 - x_p^{-\xi}) - \sigma \xi^{-1} x_p^{-\xi} \log x_p \right], \end{aligned} \quad (2.20)$$

for $\xi \neq 0$ and with $x_p = -\log(1 - p)$, as well as,

$$\nabla z_p^T = \left[\frac{\partial z_p}{\partial \mu}, \frac{\partial z_p}{\partial \sigma} \right] = [1, -\log x_p]$$

for $\xi = 0$.

In case of the **GP** distribution the corresponding negative log-likelihood is

$$l(y; \tilde{\sigma}, \xi) = k \cdot \log \tilde{\sigma} + \left(1 + \frac{1}{\xi} \sum_{i=1}^k \log \left(1 + \frac{\xi \cdot y_i}{\tilde{\sigma}} \right) \right), \quad (2.21)$$

for $\xi \neq 0$ and with $\tilde{\sigma} > 0$, and

$$l(y; \tilde{\sigma}) = k \log \tilde{\sigma} + \frac{1}{\tilde{\sigma}} \sum_{i=1}^k y_i, \quad (2.22)$$

for $\xi = 0$ and with $\tilde{\sigma} > 0$. Similar as above the NLLH of the GP distribution features an additional constraint with $1 + \xi y_i / \tilde{\sigma} > 0$.

All the following discussions about the initialization and optimization of the ML equations will apply for both the GEV and the GP distribution. The same holds for the advises given above. When combining the equations (2.21) and (2.22), the second one should only be used if the shape parameter is perfectly equal to zero. Else, numerical artifacts will be generated.

The errors for the GP distribution can be calculated using the inverted Hessian and the delta method too:

$$\text{std}(y_m) \approx \sqrt{\nabla y_m^T V \nabla y_m},$$

with

$$V = \begin{bmatrix} \zeta_u(1 - \zeta_u)/n & 0 & 0 \\ 0 & v_{1,1} & v_{1,2} \\ 0 & v_{2,1} & v_{2,2} \end{bmatrix},$$

and

$$\begin{aligned} \nabla y_m^T &= \left[\frac{\partial y_m}{\partial \zeta_u}, \frac{\partial y_m}{\partial \tilde{\sigma}}, \frac{\partial y_m}{\partial \xi} \right] \\ &= \left[\tilde{\sigma} m^\xi \zeta_u^{\xi-1}, \xi^{-1} \left\{ (m \zeta_u)^\xi - 1 \right\}, -\tilde{\sigma} \xi^{-2} \left\{ (m \zeta_u)^\xi - 1 \right\} + \tilde{\sigma} \xi^{-1} (m \zeta_u)^\xi \log(m \zeta_u) \right] \end{aligned} \quad (2.23)$$

for $\xi \neq 0$, as well as,

$$V = \begin{bmatrix} \zeta_u(1 - \zeta_u)/n & 0 \\ 0 & v_{1,1} \end{bmatrix},$$

and

$$\nabla y_m^T = \left[\frac{\partial y_m}{\partial \zeta_u}, \frac{\partial y_m}{\partial \tilde{\sigma}} \right] = \left[\frac{\tilde{\sigma}}{\zeta_u}, \log(m\zeta_u) \right]$$

for $\xi = 0$.

Despite of the overall good performance and the widespread use of the ML approach, it has some drawbacks. The most restricting one is certainly the limited range of shape parameters the ML approach is applicable to. Its estimates only exist for shapes bigger than -1 and only for values bigger than -0.5 its regularity conditions are fulfilled (Smith 1985). On the other hand, for shape parameters bigger than 1 the first and for values bigger than 0.5 the second moment of the GEV and GP distribution, which are used to determine the initial parameters for the optimization routine, do not exist anymore (Hosking, Wallis, and Wood 1985).

2.2.3. Error estimation revisited

2.2.3.1. Motivation

When inspecting the features and the implementations of available software packages specialized in EVA, one finds that all of them provide the estimates of the fitting errors for both the GEV and GP parameters. But not a single one features the standard error of the calculated return levels using the delta method described in the previous section.

To understand this discrepancy, we will perform a short numerical experiment. Let us consider a fixed scale parameter $\sigma = 1$, a location parameter $\mu = 0$ for the GEV distribution, and a variety of different shape parameters ξ . For each parameter combination we draw 10 series consisting of 100 data points from the corresponding GEV or GP distribution. A length of 100 points serves as a realistic representation of the climatological series of annual maxima covered in chapter 7. In most applications of EVA one has to deal with even less data. The parameters of all those series are estimated using the ML approach and the return levels, as well as their error estimates, are calculated using the functions introduced in the last section. When viewing the results in figure 2.2, the reason for the missing implementation becomes obvious. For shape parameters bigger than 0.3 the fitting errors of the return levels (green) are massively overestimated.

According to the delta method a GEV-distributed series consisting of e.g. 100 points with the estimated parameters of $(\tilde{\mu} = -0.03, \tilde{\sigma} = 0.964, \tilde{\xi} = 1.05)$ has a 100 year return level of 121 with a fitting error of 5879. For a GP-distributed series of the same length and with $(\tilde{\sigma} = 0.964, \tilde{\xi} = 1.05)$ the 100 year return level is 80.6 with a fitting error of 3654. So, why does the delta method fails to yield useful error estimates at larger shape parameters? My guess is that this is due to one of its underlying requirements. To apply the method the log-likelihood has to be symmetric with respect to the resulting MLE. Now, the larger the shape parameters get, the heavier the tails of the distribution will be, and the more asymmetric log-likelihood is becoming.

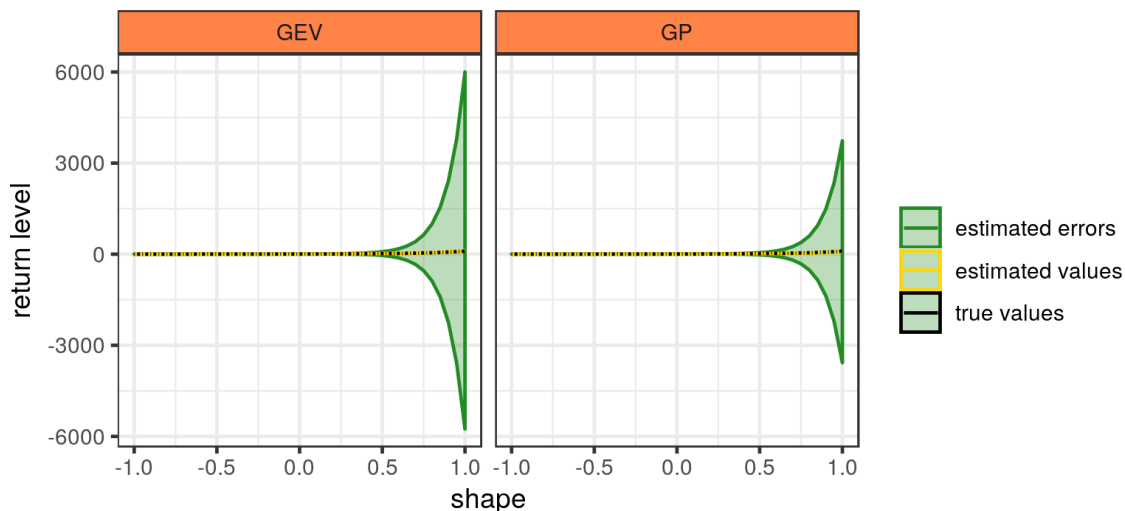


Figure 2.2.: Numerical experiment showing the impact of the shape parameter on the estimated fitting errors of the 100 year return level. For various shape parameters, a scale = 1, and a location = 0 (GEV only) ten time series per parameter combination of the length of 100 points were drawn from either a GEV or GP distribution. Each of them was fitted individually and the resulting estimates of the return levels and their errors were averaged. The black line represents the true return level, which are calculated using the parameter combination used to draw the underlying data. The yellow dots and line are the averaged, estimated return levels. For a shape parameter equal to 1 the standard error of the 100 year return level obtained by the ML method is 5879 for the GEV distribution and 3654 for the GP one.

2.2.3.2. Alternative approaches

To still access the errors of the return levels, I introduced two common statistical approaches to the field of EVA: bootstrapping and a Monte Carlo-based approach.

For **bootstrapping** the series of extreme events, either block maxima or threshold exceedances, is repetitively sampled with replacement. Afterwards, all the different realizations are fitted using the classical ML-based GEV or GP approach and the fitting errors are obtained by calculating the standard deviations of all the individual GEV or GP parameters or return levels.

In the **Monte Carlo**-based approach the plain original series is fitted using ML. Then one takes the fitted parameters and samples a bunch of GEV- or GP-distributed series with the same length as the original one. All those series are fitted using ML and the overall standard errors are obtained by averaging over the individual parameters and return levels of the sampled series.

But before we discuss the differences between both approaches, we will have a look at their performance compared to the delta method in figure 2.3 and 2.4 while applying the same analysis as for 2.2.

For the three GEV and two GP parameters the errors look pretty similar for all three estimation methods. While following the estimates of the delta method almost exactly for shape parameters bigger than -0.5 both the bootstrap and the Monte Carlo approach yield a bigger (GEV) or smaller (GP) error for lower shape values. But since the MLE only fulfill the regularity conditions for $\xi > -0.5$ and the error bars of the delta method for the scale parameters do not enclose the true value for very small shapes, the estimates of both the bootstrap and the Monte Carlo method can be considered more trustworthy.

Another fact figure 2.3 is illustrating quite nicely is the shape-dependent bias of the MLEs of the GEV/GP parameters. The sample mean of the parameter estimates are plotted in yellow while the “true” parameters, which were used for the distribution the ten series per parameter combination were drawn from, are pictured as black lines. Of course, the maximum likelihood estimators of the GEV/GP parameters are consistent. But only in the asymptotic limit of infinitely long series. Due to finite size effect each of those ten realizations has their own distinct global minimum and “true” parameters. What is depicted in figure 2.3 is only the average bias of both the MLEs and the corresponding error estimates.

When closely inspecting the fitted shape parameters and their standard errors near $\xi = -1$, one can see a systematic overestimation. This is an artifact introduced by the more stable augmented Lagrangian optimization procedure, which will be covered in section 3.2.1. But for natural data the shape parameter is in most cases in the range of $-0.5 < \xi < 0.5$ (Hosking, Wallis, and Wood 1985) and the minor deviation at very low shape parameters is a small price to pay. After all the maximum likelihood estimates themselves are just defined up to $\xi = -1$.

Figure 2.4 shows that both the Monte Carlo-based and the bootstrap approach produce much more reasonable error estimates than the delta method. In addition, the errors of the Monte Carlo method are slightly more tight for the GP distribution. So, which one of these two newly introduced methods should be used?

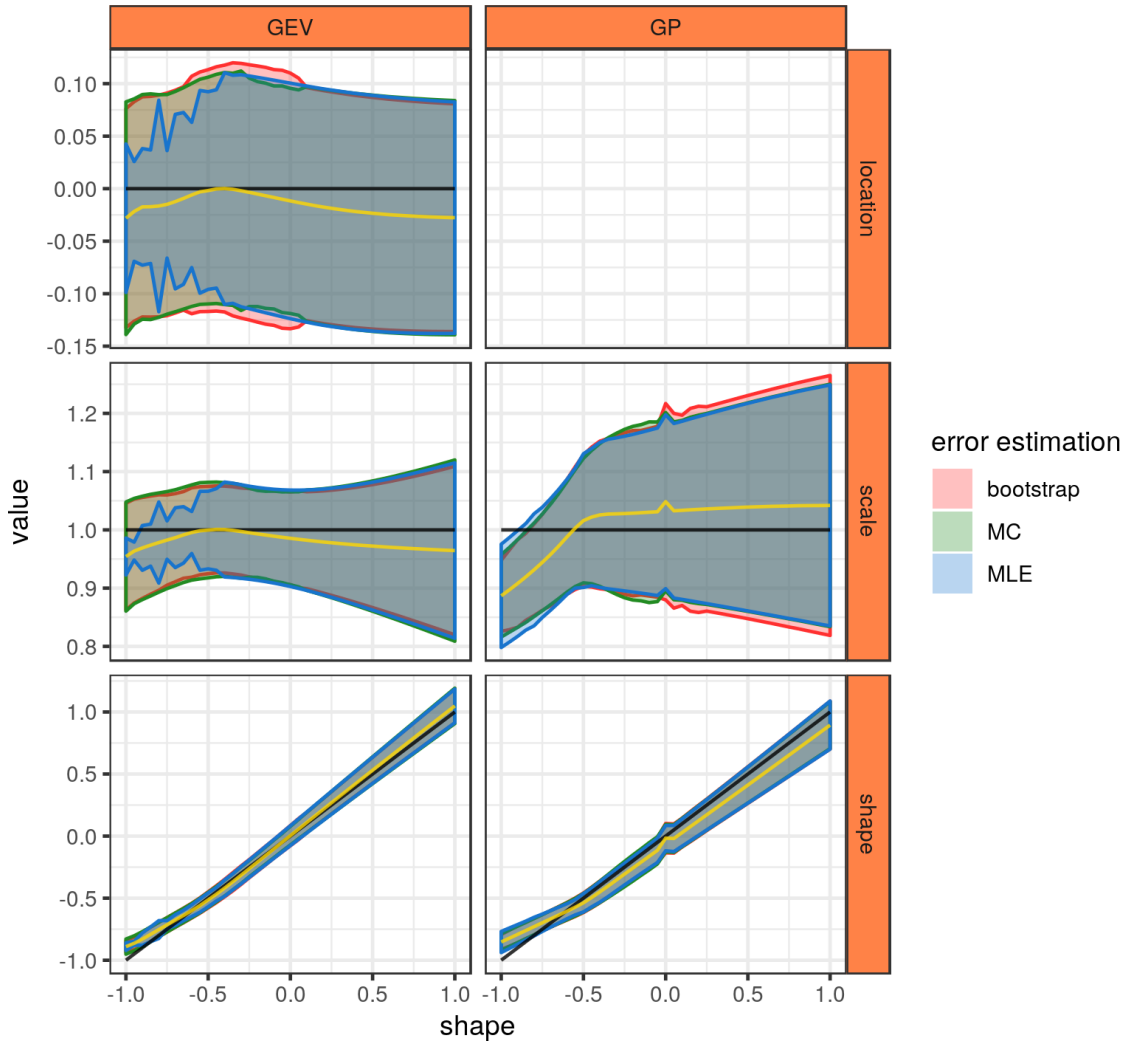


Figure 2.3.: For the same numerical experiment as in figure 2.2 an estimation of the fitting error for the estimated GEV and GP parameters (yellow) are obtained using three different methods: (**MLE**) This approach uses the square roots of the diagonal elements of the estimated Hessian matrix evaluated at the maximum likelihood estimates of the parameters. This is the standard method already shipped with various EVA packages. (**MC**) The new Monte Carlo-based samples 100 series of the same length as the original one and distributed according to the obtained MLE of the GEV/GP parameters. All of those series are fitted and the standard errors of the resulting parameters are reported. (**bootstrap**) The new bootstrap approach samples the original series with replacement 100 times, fits the individual series, and, again, reports the standard errors of the results.

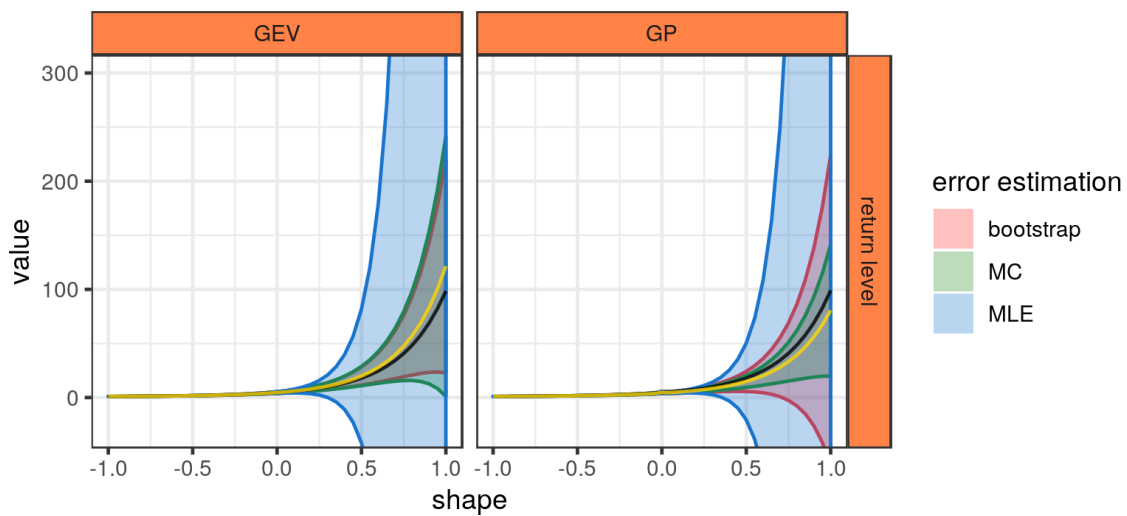


Figure 2.4.: Based on the same numerical experiment as depicted in figure 2.2 the mean estimates of the 100 year return levels (yellow) and the corresponding error bars (mean \pm standard error) are presented. In the (**MLE**) case the errors are obtained by using the delta method using equations (2.20) and (2.23) in combination with the maximum likelihood estimates of the GEV/GP parameters. For the (**MC**) and (**bootstrap**) method the estimates are the standard error of the 100 year return levels calculated for all the 100 times 10 individual series for each parameter combination.

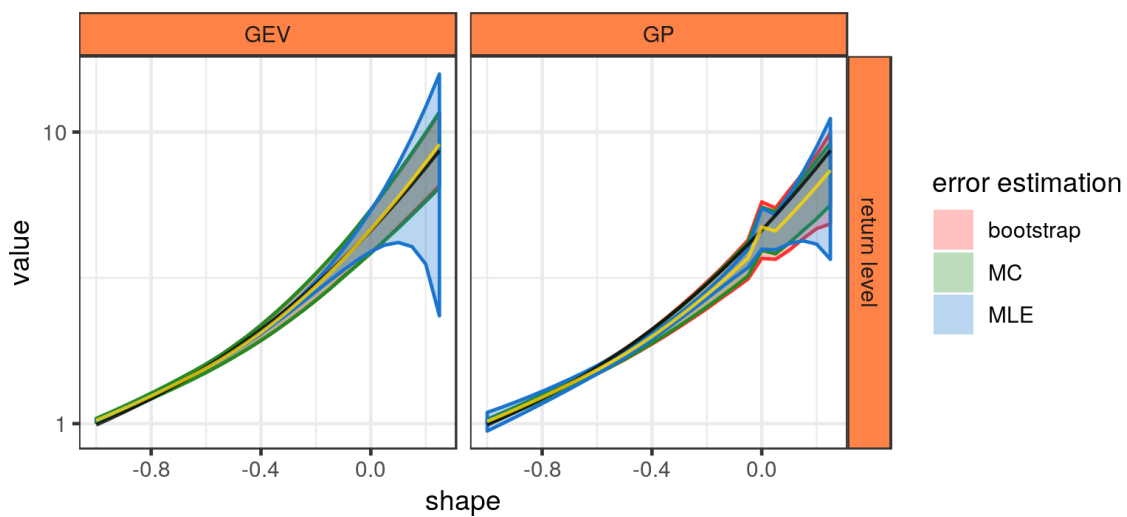


Figure 2.5.: A zoomed version of figure 2.4 for shape parameters up to 0.3 on a logarithmical scale.

2.2.3.3. Justification

When the **Monte Carlo**-based approach is used, the obtained error estimates are not the actual fitting errors. Instead, they are the average fitting errors of an uncorrelated and stationary series distributed according to the fitted GEV or GP parameters and with the same length as the original series. The uncertainty of the fitting of the GEV/GP parameters and the corresponding return levels is not directly represented using this quantity. But in our opinion this approach still has some advantages over bootstrapping, since the underlying series is a sample of a heavy tailed distribution and thus small deviations in the parameters can result in a large change in statistics of a drawn series.

In the **bootstrapping** procedure (Hastie, Tibshirani, and Friedman 2009) the fitted series is randomly sampled with replacement and fitted various times. The standard deviations of the obtained parameters and return levels are then assumed to approximate the fitting errors of the full series, as described beforehand. What yields reasonably good estimates for cases involving short tailed distributions, like the Gaussian one, might lead to bad representations for the GEV or GP distribution.

A naive assumption would be: the bootstrap approach yields a good approximation of the statistical properties of the bulk of the data, whereas the Monte Carlo-based approach is more likely to represent the tails of the distribution. Let's test this assumption with a small numerical experiment.

We will perform a GP analysis on the of daily gauge values of the river Elbe in the city of Dresden¹ in figure 2.6. The threshold for this series will be set to 420 cm, which results in a total number of threshold exceedances of 73 after declustering. Fitted to a GP distribution we obtain a scale of $\sigma = 126.9$ and a shape of $\xi = -0.129$.

The bulk of the data is residing at relatively low values in the range between 0 and 250 cm. From roughly 250cm and 500cm, on the other hand, we only have about four to five events, with (a) and (b) as the biggest ones, constituting the tail of the GP distribution, which is about to get fitted to the data. But since the shape parameter, and thus also the corresponding return level, is strongly depending on the particularities of the tail, omitting one of these events or sampling them multiple times will result in quite large fluctuations of the return levels.

Figure 2.7 shows the impact of the largest events in the estimated 100 year return level. For the original series (upper figure) return level is 890.2 cm. If, instead, the largest event is removed from the series, the value reduces to 823.8 cm and if, on the other hand, the second largest event is occurring two times the 100 year return level is 920.5 cm. Just by the presence of these two values the large quantiles can be altered significantly.

To test whether the bootstrap or the Monte Carlo-based method produces more accurate results, we now will sample the exceedances in figure 2.6 100 times with replacement (bootstrap) and also generate 100 series with the same length constituted by the GP parameters estimated for the gauge series (Monte Carlo). When calculating the mean value of all average gauges of the sampled series and the Monte Carlo samples, we obtain a value of 110.39 cm and 112.1 cm, which is quite close to the average gauge of the original series of 112.56 cm. If we, instead, calculate the mean 100 year return levels of all the sampled series we obtain 854.73 cm for the bootstrap and 844.3 cm for the Monte Carlo method . This, however, is an underestimation of the return level of the original series,

¹This data was obtained from the Wasserstraßen- und Schifffahrtsverwaltung_des Bundes (WSV)

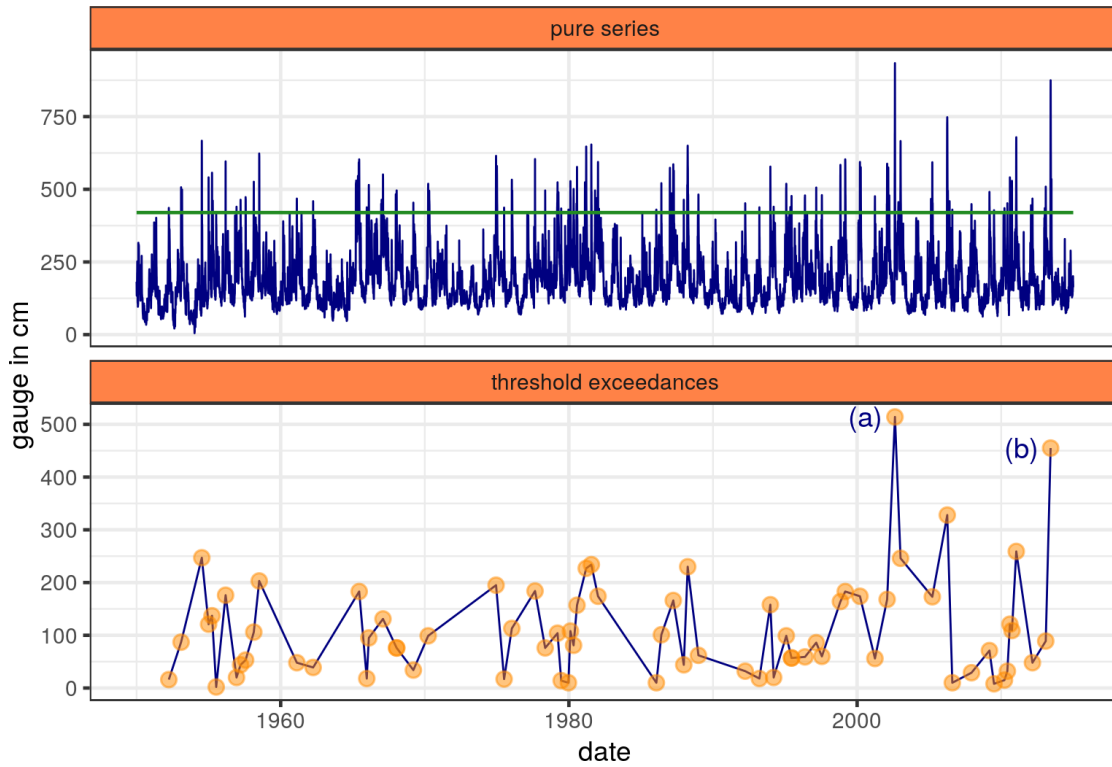


Figure 2.6.: Daily series of the gauge of the river Elbe in the city of Dresden. The lower figure shows only those events remaining after declustering the exceedances over the green threshold in the upper figure.

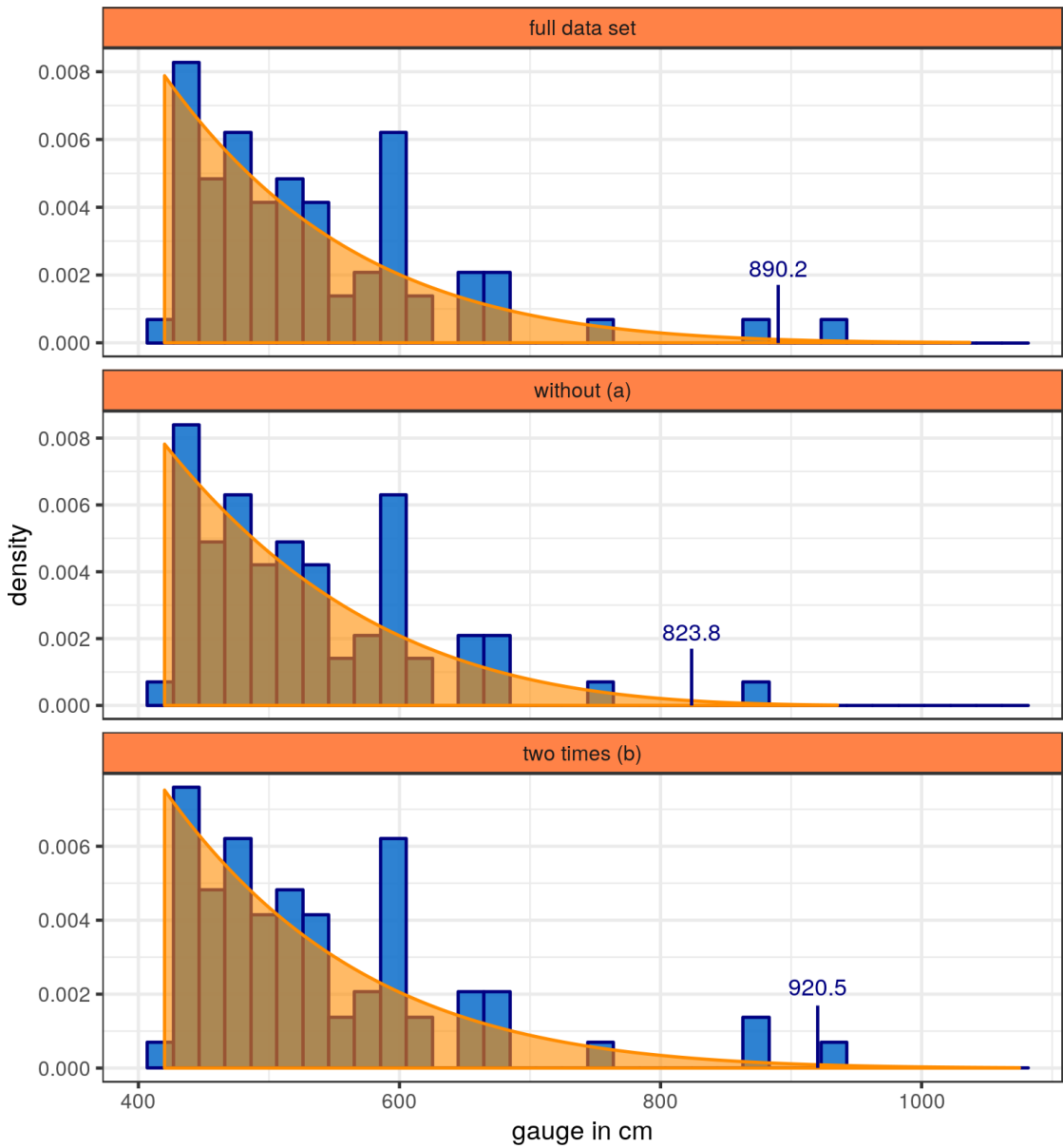


Figure 2.7.: This figure depicts the histogram (blue) of the threshold exceedances shown in figure 2.6 for the full series (upper figure), after the removal of the biggest gauge value (middle), and for the full series with the second largest event appearing two times (bottom). For all three series the GP parameters were estimated and the corresponding probability density plotted in orange. In addition the 100 year return levels were calculated and marked with a black, vertical line and the corresponding value.

Table 2.1.: Maximum likelihood and error estimates of the GP parameters and 100 year return level for the threshold exceedances of the full gauge series. All parameters, except of the dimensionless shape, are given in cm.

	scale	shape	100 year return level
maximum likelihood estimates (MLE)	126.86	-0.13	890.23
delta method error estimate	18.93	0.09	4719.16
Monte Carlo error estimate	26.49	0.14	64.31
bootstrap error estimate	19.61	0.17	83.66

which is 890.2 cm. Keeping the results of this short analysis in mind, let's have a look at the error estimates for the GP parameters and the 100 year return level obtained using the three different methods of error estimation described in the previous sections.

The error estimates for the shape and scale parameters are roughly the same for all three approaches, with a tendency for the bootstrap and Monte Carlo-based method to be more pessimistic in terms of the error of the shape parameter. Since the underlying data is not generated but measured, there is no way of determining which of those values is the correct one. But in the end they all seem plausible. For the error of the 100 year return level things are different. The estimate provided by the delta method clearly exaggerates the true value. The Monte Carlo and bootstrap approach, on the other hand, yield quite plausible results.

But which one to trust? Even when taking a closer look at their differences for shape parameters smaller than 0.3 (see figure 2.5) the answer to this question is hard to determine.

In principle I recommend to use of the Monte Carlo-based approach over bootstrap, since its less time consuming and provides tighter error bars for bigger shape parameters. But since those usually do not occur in measured data, the answer is a matter of taste.

2.2.4. Initialization of the parameter optimization

When optimizing a function, like the ML estimators of the GEV and GP parameters in section 2.2.2, both the optimization algorithm (see chapter 3) and the initial values of the parameters are of utmost importance. Most optimization routines are structured in the following way. At first, the algorithm takes big steps in the direction of decreasing values of the function (I will only consider minimization search routines in here). After a couple of steps the algorithm gradually decreases the step size until it finally converges to a value close to a local or global minimum or until a preset number of maximal iteration steps was taken.

But what exactly is considered big or small? In gradient-based methods this is usually determined using the gradient of the function at the evaluation point. Now, imagine we have to optimize not a quadratic function, but a fairly complex one, with large plateaus of an almost negligible slope and a steep valley containing the global minimum. Even if this function has no local minima, an optimization routine is not guaranteed to end up in the global one. Starting somewhere in the valley, it will of course do. But when the optimization routine will be started far, far away somewhere at the plateau, it is very

likely to not reach the minimum. The likelihood functions for both the GEV and GP unfortunately are rather complex ones featuring a similar topology, which might cause the optimization routine to fail (see chapter 3 for a thorough discussion). Therefore, it is important to pick a parameter combination as close as possible to the global minimum to start the algorithm.

The most simple and widespread heuristic for the initialization of both the GEV and GP parameters is the so called **method of moments**. There, the series x is assumed to be distributed according to the Gumbel or exponential distribution (both have a shape parameter $\xi = 0$) since their first two moments can be used to determine the location and the scale parameter.

$$\begin{aligned}\sigma_{init} &= \frac{\sqrt{6 \operatorname{var}(x)}}{\pi} \\ \mu_{init} &= \operatorname{mean}(x) - 0.57722\sigma_{init},\end{aligned}$$

for the GEV distribution and

$$\sigma_{init} = \sqrt{\operatorname{var}(x)},$$

for the GP one. The shape parameter is set to a constant value independent of x . For most available fitting packages in the context of EVA ξ is set to values between 10^{-8} and 0.1.

Some notable exceptions in the initialization in the context of EVA is the Mathworks Statistics and Machine Learning Toolbox R2016b in Matlab, relying on a variant of the q-q plot, and the `extRemes` by Gilleland and Katz (2016) and `climex` package, written by myself Müller and Kantz (2018), in **R**, which use the method of L-moments (Hosking 1990). The usual procedure is to estimate the initial parameters with both the method of moments and the more sophisticated approach and to pick the parameter combination yielding the smallest possible negative log-likelihood.

These combined approaches are not just yielding parameter combinations more closer to the global minimum, but also result in a more robust fitting routine. Due to logarithms in equation (2.18) and (2.21) the NLLH is not defined throughout the entire parameter plane. It is possible to choose initial parameters in such a way the likelihood is not well defined and thus causing the optimization routine to fail. Although it does not happen very frequently in measured data, such bad initializations do take place while performing parameter scalings or numerical experiments. In case neither the method of moments nor the more sophisticated L-moment one is returning a valid parameter combination, `climex` package performs a seeded Markov walk on the combination estimated by the L-moments algorithm to nevertheless find an acceptable starting point.

2.2.4.1. Estimating the shape parameter using the skewness

In addition I found another, previously unknown, way to improve the method of moments in order to determine a more appropriate initial parameter combination. In this scheme the sign and the magnitude of the shape parameter is assigned according to the skewness of the series of extreme events x .

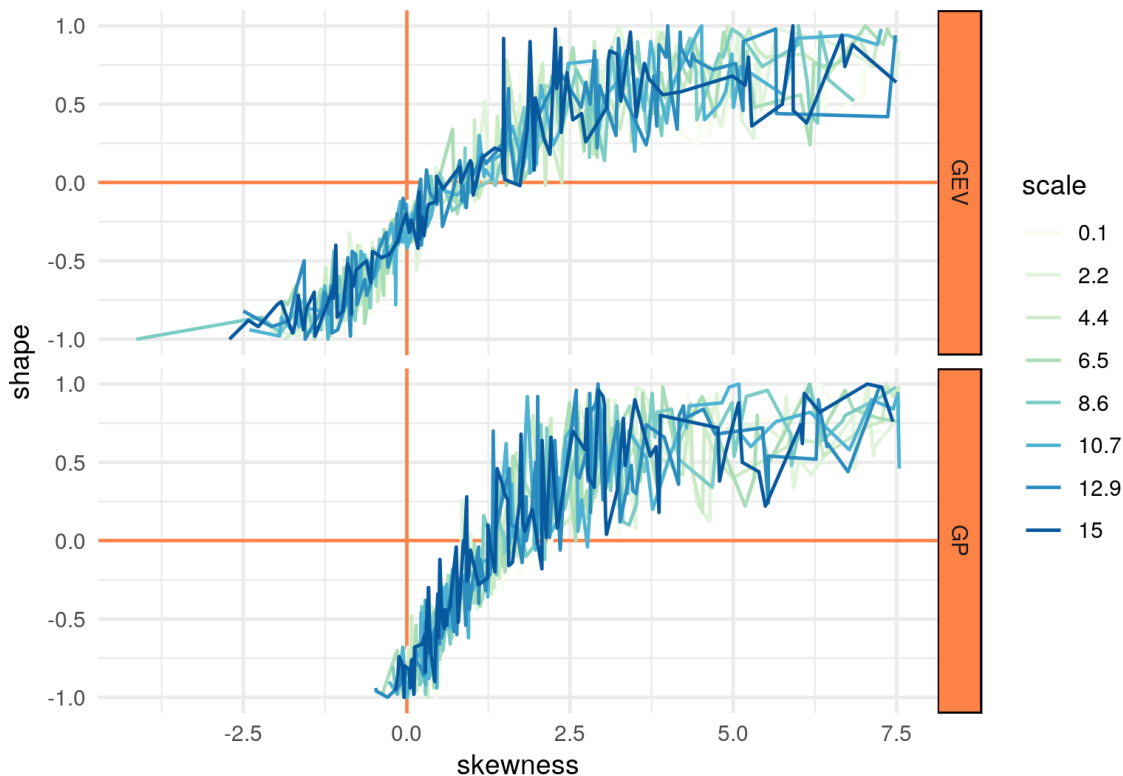


Figure 2.8.: This figure depicts how the skewness of a series of extreme events can be used to estimate its corresponding shape parameter. For eight different scale (GP) or location and scale (GEV) parameter combinations series, consisting of 60 points, were drawn for various shape parameters. For each of those the skewness was estimated. As can be seen in the figure, the skewness can be used to estimate the sign and the overall magnitude of the shape parameter of the corresponding series. There was no smoothing applied to the resulting skewness values on purpose. In practice only a single value is available for choosing the shape parameter.

To verify this statement, let's take 8 different combinations of location and scale (GEV) or scale (GP) parameters, draw 60 points from the corresponding distribution for a range of shape parameters, and calculate the skewness of all the resulting series (see figure 2.8).

I found the magnitude and the sign of the skewness to correlate with the shape parameter. Since the variance of the skewness is quite high for shorter series, it wouldn't be wise to implement a functional dependence to determine the shape parameter based on the estimated skewness. I would suggest to use multiple switch/case statements to assign shape parameters depending on the interval of the skewness estimator, instead of a default value for all series (see tables 2.2 and 2.3).

2.2.5. Gumbel vs. GEV hypothesis

There are many difficulties one has to face when fitting extreme value distributions, like the trade-off between the block length/threshold height and the amount of available points or possible correlations and non-stationarities in the time series. But at the core

Table 2.2.: Heuristic to estimate the initial value of the shape parameter of the GEV distribution according to the skewness

skewness	shape
> 4	0.7500
> 2.5	0.6000
> 1.6	0.3500
> 0.7	0.0010
> 0.1	-0.1000
> -0.2	-0.2775
> -1	-0.4000
≤ -1	-0.7500

Table 2.3.: Heuristic to estimate the initial value of the shape parameter of the GP distribution according to the skewness

skewness	shape
> 4.5	0.75
> 2.5	0.50
> 1.8	0.25
> 1.5	0.05
> 1.2	-0.05
> 0.8	-0.25
> 0.2	-0.50
≤ 0.2	-0.75

Table 2.4.: Rejected hypotheses at the significance level of 0.05%

rainfall	temperature
102/470	289/470

of each fit there is always a hypothesis determining the distribution to fit to the data. If one chooses the wrong one, the results will deviate from the true values and will be at least misleading.

Regardless of the availability of a large number of software packages in various computer languages providing an implementation of both the fitting of the GEV and GP distribution using ML, there are still quite a number of people relying on their special cases (the Gumbel or exponential hypothesis) when performing extreme value analysis (Hirabayashi et al. 2013; Sarkar, Singh, and Mitra 2011; Steenbergen, Koster, and Geurts 2012; Valor et al. 2010). Although one loses generality when applying those specialized versions, there are some arguments, which at first sight would favor their usage. On the one hand, it is sometimes claimed (at least in oral communications) that the shape parameter in climatological time series is approximately zero. On the other hand, a fit of two instead of three parameters is more robust in a statistical sense when dealing with very small data sets (Hirabayashi et al. 2013; Abu-Mostafa, Magdon-Ismail, and Lin 2012). While the latter argument is certainly true, the first one can be proven wrong by testing the hypothesis of using the Gumbel distribution instead of the GEV one to describe the data (see figure 2.9 (a)).

As an example the publicly available daily maximum temperature and precipitation time series provided by the German weather service (DWD) (Deutscher Wetterdienst 2018) were chosen. To restrict the size of the upcoming analysis, I choose to only show the test of the usage of the Gumbel distribution against the more general GEV one. The analyses for the exponential distribution compared to the GP one yields similar results.

Firstly, the anomalies of the temperature series (see section 4.3.1) are calculated and all incomplete years are removed. In order to perform a sound numerical estimation of the GEV and GP parameters, all stations with less than 30 years of data are discarded. I choose a minimum length of 30 years since it is the default period considered by most studies in the context of climate analysis (Hansen, Sato, and Ruedy 2012) and it is common practice to have a data set of at least ten times the number of fitting parameters to avoid overfitting (Abu-Mostafa, Magdon-Ismail, and Lin 2012). For each of the remaining 470 time series the annual maxima are extracted and the corresponding GEV and Gumbel parameters are fitted using the **climex** package in the statistical programming language **R**. To see, whether the usage of the Gumbel hypothesis instead of the GEV one would have an impact on the results, the 100 year return level of each time series is calculated using the parameters fitted by both approaches. The corresponding fitting errors are calculated using the Monte Carlo-based approach, introduced in section 2.2.3.2. The results are shown in figure 2.9 with the estimated shape parameters in (a) and the estimated 100 year return levels in (b).

In order to determine whether the different series are best described by either the Gumbel or the GEV distribution, I will adopt the usage of the Gumbel distribution to describe the data as the null hypothesis. In the analysis I will then fit both the

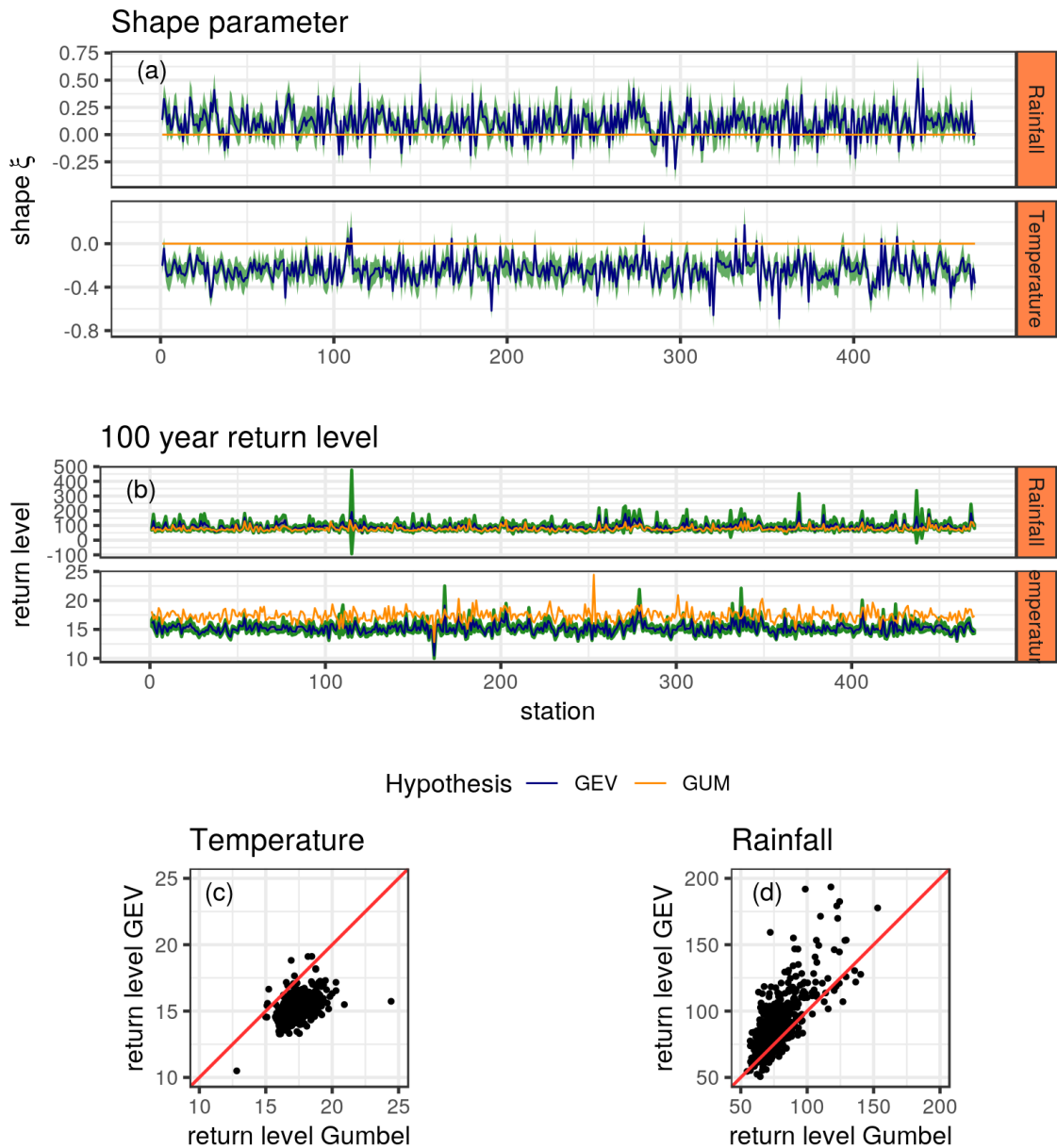


Figure 2.9.: Estimates of the shape parameters and 100 year return levels using either the GEV (blue) or Gumbel (orange) hypothesis for 470 weather stations provided by the German weather service DWD (Deutscher Wetterdienst 2018). Each time series consists of at least 30 complete years of temperature anomalies [$^{\circ}\text{C}$] or rainfall data [mm]. All incomplete years have been removed beforehand. (a) Despite of connecting the resulting shape parameters of the individual stations by a line, their ordering is just alphabetic and there are no distinct connection between neighboring stations in the graphs. The standard errors were calculated using the Monte Carlo-based (MC) approach, which was discussed in section 2.2.3.2. For many stations the null hypothesis of having a shape parameter equal to zero can be rejected for both their temperature and rainfall data. Figure (b) provides the 100 year return levels estimated with both the GEV and the Gumbel distribution. (c) In addition, the usage of the Gumbel hypothesis for the temperature series generally overestimates the return level, while (d) it underestimates the values for the rainfall ones.

Gumbel distribution and the GEV distribution to the data using maximum likelihood and construct a hypothesis test from the fitted values using the **likelihood ratio test**. Via this approach we can test whether we reject the hypothesis of using the Gumbel distribution to describe the data in favor of using the more complex model of the GEV distribution.

In many of the examined time series, 289 of 470 in case of the temperature anomalies and 102 of 470 for the rainfall data, the hypothesis of using the Gumbel distribution over the GEV one has to be rejected. Since the number of points entering the fits are quite limited for most of the stations and the hypothesis test is a rather conservative approach, the number of rejections are quite large and point towards the general usage of the GEV distribution over the Gumbel one. In addition, it is claimed by several groups that it is very difficult to verify $\xi \approx 0$ due to an insufficient amount of data (Yee 2015).

When looking at the estimates of the return levels in figure 2.9 one can find a quite large systematic difference between the results obtained by the GEV and the Gumbel fit. The GEV estimates are bigger than the Gumbel ones for positive shape parameters (obtained by the GEV distribution) and smaller for negative ones.

The impact of the choice of the hypothesis can be best expressed by the mean difference between the 100 year return levels obtained using the GEV and Gumbel distribution. While the mean temperature anomaly, exceeded on average only once every 100 years, is 15.2 °C using the GEV hypothesis, it is 17,3 °C using the Gumbel one. For the precipitation data with 89.8 mm (GEV) and 77.5 mm (Gumbel) the difference is of the same order of around 13 percent.

In summary, for most of the considered stations using the Gumbel hypothesis to access their return levels would yield wrong results and has therefore to be avoided.

3. Optimization

Since the onset of the digital age optimization routines became a more and more important part of our world. Nowadays, their application is not restricted to various fields of research anymore, but performed countless times a day by our computers, smart phones, smart watches, and even some of our refrigerators.

In the field of the extreme value analysis (EVA) optimization routines were applied since the earliest days to obtain the parameters of the generalized extreme value (GEV) or generalized Pareto (GP) distribution describing the data best. With several generations of software packages specifically designed for EVA at hand, one would expect the fitting routines to be well matured and sound and to yield the best possible results. Surprisingly I found this is not the case.

In this chapter I will first provide a short introduction into the basics of optimization and a thorough explanation of the algorithms commonly used with EVA in section 3.1. Afterwards, I will highlight their drawbacks and describe the numerical artifacts one encounters when using those methods (3.1.3). Fortunately all these errors can be avoided by using a constrained non-linear optimization routine. After an introduction to the basic of constrained optimization in section 3.2 a description of the method of augmented Lagrangian multipliers will be given in 3.2.1. This method will be the one used throughout the entire analysis part of this thesis. In the end, the results obtained by non-linear constraint optimization will be compared to the ones of the previously used algorithms in 3.3.2.

3.1. Unconstrained optimization

The overall goal of optimization is to find the global minimum of an objective function $f(x)$ in the most efficient and precise way.

But this task is usually very hard to accomplish. Not that $f(x)$ is just a general and usually quite complicated function, we also only have *local* information available. This means we initialize the algorithm at a given point x_0 and have to guess using the available information in which direction the minimum might be located. Since the algorithm has a very limited knowledge about the function and the optimization should be done as fast as possible, most routines are only able to detect a local minimum.

Definition 3.1

*If there is a neighborhood \mathcal{N} of \hat{x} with $f(\hat{x}) \leq f(x)$ for all $x \in \mathcal{N}$, then the point \hat{x} is a **local minimum** of the function $f(x)$.*

There are two general types of optimization algorithms: those based on **line search** and those based on **trust regions** (Nocedal and Wright 2012, 19). In a line search-based routine the algorithm chooses a direction in the parameter space based on the local

information and searches along this direction for a point yielding a lesser function value than the current position. In an optimization based on the trust region method the objective function is approximated by a model function (usually a quadratic one). If this function yields a good approximation, it is expanded. If not, it is contracted. In the end the user obtains a region reasonably described by a known function, the minimum of whose can be calculated analytically and is thus reached within one step. Out of those two iterative concepts the line search is the one most commonly used.

In the context of EVA most packages do rely on the **Nelder-Mead** algorithm introduced by Nelder and Mead (1965). Only the more sophisticated *extRemes* package by Gilleland and Katz (2016) in R is using the **BFGS** (Press et al. 2007) algorithm instead. Those two will be introduced in the following sections. For a thorough treatment of the initialization process of the optimization of GEV or GP distributions please see section 2.2.4.

3.1.1. Nelder-Mead

The Nelder-Mead algorithm is the default one used by countless fitting packages in many different programming languages. This is mainly due to two of its properties:

1. It doesn't require derivatives but only function evaluations.
2. It is quite robust algorithm.

The algorithm can be classified as a so called **simplex method**. It generates a convex hull of $\{x_1, x_2, \dots, x_{n+1}\}$ points in a \mathbb{R}^n parameter space. In every iteration step the point x_i , which is holding the worst function value, is discarded and replaced with one yielding a better (lower) value. This replacement is either obtained by reflection, expansion, or contraction of the simplex with respect to x_i . In order to perform these tasks, the algorithm uses function evaluations a lot and should only be applied to problems requiring a small amount of computational resources (both in CPU time and memory).

In general the performance of the Nelder-Mead routine is reasonable, but stagnation at non-optimal points can occur (Nocedal and Wright 2012). At times, *restarting* the optimization can be used to escape the stagnation (Kelley 1999). Another approach that can be used in such cases is the so-called *half-stepping* (Yee and Stephenson 2007, 15). There the algorithm is only allowed to take very small steps at the beginning of the fitting procedure. This way the algorithm is less likely to take large steps in a direction of the parameter space, in which the likelihood function is not well conditioned anymore.

To provide a more applied picture of the Nelder-Mead routine, I will present a version implemented in the **dfoptim** package¹ of the statistical language **R** applied to the GEV distribution. Different implementations probably feature different parameter values, but the overall concept remains the same. The individual points in the three dimensional parameter space are linked to the GEV parameters in the following way: $x_i = (\mu, \sigma, \xi)$.

3.1.1.1. Initialization

Firstly, the simplex gets initialized by calculating $n + 1$ vertices using the supplied initial parameter combination $\theta_0 = (\mu_0, \sigma_0, \xi_0)$.

The first vertex is simply set equal to the θ_0

¹The **dfoptim** package can be found at cran.r-project.org/package=dfoptim

$$x_{n+1} = \theta_0.$$

The other n initial vertices of the simplex are created by shifting the initial parameters by a constant value α_2 followed by an displacement towards all the individual directions of the parameter space by a constant distance of α_1 .

$$x_i = \theta_0 + \vec{1}\alpha_2 + e_i\alpha_1 \quad (3.1)$$

with $i = 1, \dots, n$, e_i being unit vector along the i -th direction of the parameter space, $\vec{1}$ being a vector of all ones, and

$$\alpha_1 = \frac{\max\left(1, \sqrt{\sum \theta_0^2}\right)}{n\sqrt{2}} \left(\sqrt{n+1} + n - 1\right) \quad (3.2)$$

$$\alpha_2 = \frac{\max\left(1, \sqrt{\sum \theta_0^2}\right)}{n\sqrt{2}} \left(\sqrt{n+1} - 1\right), \quad (3.3)$$

with n being the dimension of the parameter space and θ_0 being the initial parameter vector.

3.1.1.2. Iterative updates

To relate the index of the individual vertices with their corresponding function values, we will order the $n + 1$ vertices in every iteration step in such a way that

$$f(x_1) \leq f(x_2) \leq \dots \leq f(x_{n+1})$$

The update is now performed in four different steps:

1. **Reflection:** For the sake of illustration, we assume our simplex to be located in a smooth part of the parameter space containing hills and bumps and the worst point x_{n+1} being the most uphill one. Since all the other points are located more downhill, the most straight forward way of optimization is to calculate the centroid of the remaining vertices \bar{x} and to reflect x_{n+1} on it.

$$x^r = (1 + \rho)\bar{x} - \rho x_{n+1},$$

with $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and the weight of the reflection step $\rho = 1$.

If the resulting value $f(x_r)$ is less than the worst value $f(x_{n+1})$ but bigger than the best one $f(x_1)$, the new point x_r will be used as the new vertex and the iteration loop continues.

2. **Expansion:** If, on the other hand, $f(x_r)$ returns the lowest value of the function yet $f(x_r) < f(x_1)$, all points might be located at an extended linear hyperplane. Since we have not reached a saddle point of such a plane yet, we might obtain an even better update x_e by expanding the reflection further down the current direction of line search.

$$x_e = (1 + \rho\chi)\bar{x} - \rho\chi x_{n+1},$$

with $\chi = 2$. In case $f(x_e) > f(x_r)$, x_r is used for the update instead and the iteration loop continues.

3. **Outer contraction:** Sometimes the n points yielding the smallest function values might be already located in a bottom valley or near a saddle point. Then the reflection step would overshoot the minimum and result in value smaller than the original one, but larger than all the others $f(x_n) < f(x_r) < f(x_{n+1})$. To circumvent this problem, we reflect the point x^{n+1} not using its distance to the centroid \bar{x} , but with a smaller value.

$$x_{c_o} = (1 + \rho\gamma)\bar{x} - \rho\gamma x_{n+1},$$

with $\gamma = 0.5$.

4. **Inner contraction:** If, instead, the function value at the reflected position x_r is even bigger than the original one $f(x_r) > f(x_{n+1})$, the worst point will be moved towards the centroid \bar{x} without any reflection at all.

$$x_{c_i} = (1 - \gamma)\bar{x} + \gamma x_{n+1}$$

3.1.2. BFGS

The BFGS algorithm (after its creators Broyden, Fletcher, Goldfarb, and Shanno) is the most popular (Nocedal and Wright 2012) variant of the quasi-Newton algorithm discovered by Davidon (1991) in the mid 1950s. Unlike the original quasi-Newton method, it only requires the calculation of the first but not of the second order derivatives of the objective function f .

But let's start by reviewing the **classical Newton algorithm**. There, we approximate the objective function by a quadratic model m_i at the current parameter combination x_i .

$$m_i(p) = f_{x_i} + \nabla f_{x_i}^T p + \frac{1}{2} p^T B_{x_i} p, \quad (3.4)$$

with B_{x_i} being the symmetric and positive definite Hessian of f at x_i and i being an integer number keeping track of the iteration step. This convex, quadratic optimization problem can be solved explicitly by

$$p_i = -B_{x_i}^{-1} \nabla f_{x_i}. \quad (3.5)$$

Using the minimizer p_i , the parameter combination x_i can be updated by

$$x_{i+1} = x_i + \alpha_i p_i, \quad (3.6)$$

with a step length α_i obtained using an one-dimensional line search algorithm along the direction p_i .

The difference between the **quasi-Newton** algorithm and the original one is that the calculation of the Hessian at each iteration step is not required anymore. Instead, the matrix B_{x_i} is updated using the values and the gradient of the objective function at

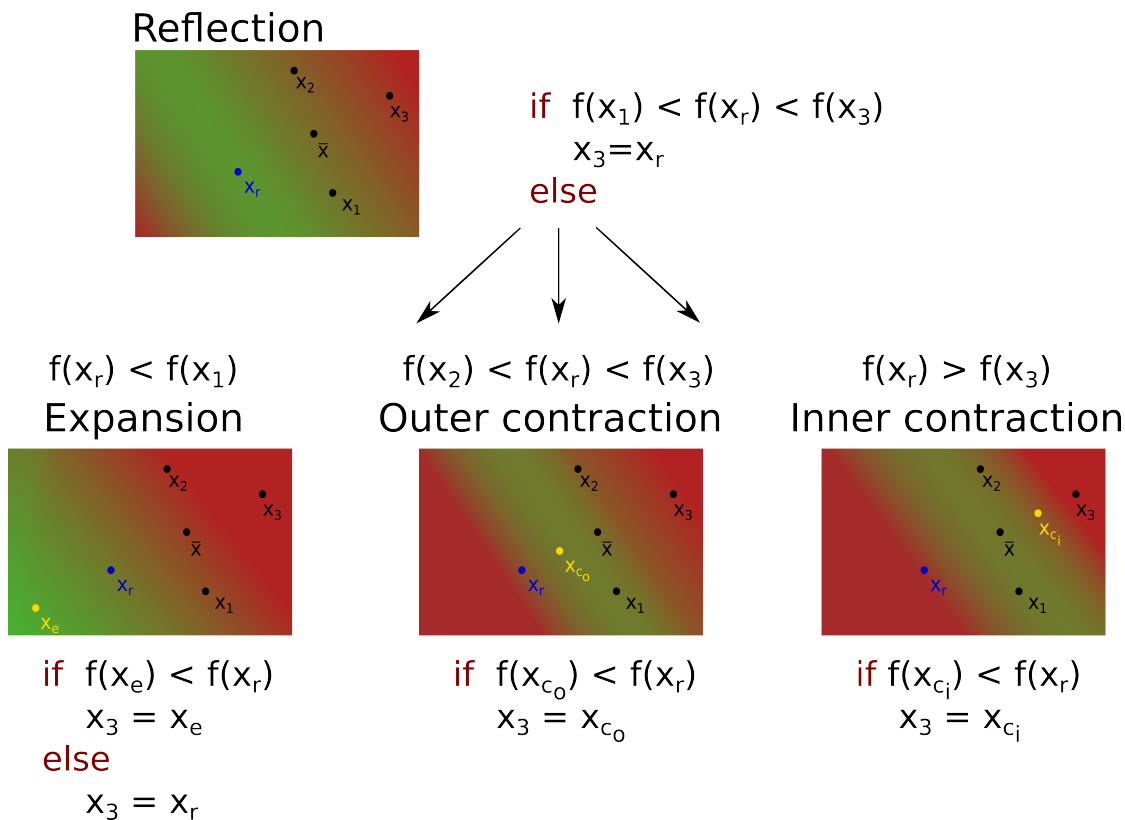


Figure 3.1.: A schematic illustration of the Nelder-Mead optimization algorithm in two dimensions. The background color corresponds to the value of the object function (red reads as high function values and green as low) for the particular parameter combination (vertex of the simplex) and the displayed patterns fit the different scenarios covered by the algorithm. It is performed in the following way. The points x_1, x_2 , and x_3 are the input vertices with $f(x_1) < f(x_2) < f(x_3)$. The vertex x_3 , which holds the worst value, is first reflected at the centroid \bar{x} of x_1 and x_2 . Depending on the value of the reflected vertex x_r , the expansion, outer or inner contraction subroutine is invoked, or x_3 is updated right away.

the current and previous position and thus using information about the curvature. It is this updating procedure that separates the **BFGS** algorithm from other quasi-Newton routines. But more on this later.

While the classical method converges more rapidly, quasi-Newton algorithms typically are still faster, because they avoid the calculation of the second derivatives and the solution of a linear system (Nocedal and Wright 2012).

For a more thorough treatment of both the theory and the implementation details please see Nocedal and Wright (2012, chap. 6).

3.1.2.1. Initialization

With the initial parameter combination $x_0 = \theta$ supplied, the only thing missing is an approximation of the initial Hessian B_{x_0} .

There are several ways to handle this. One could either calculate it using equations provided by the user or finite differences, or just start with a multiple of the identity matrix, where the factors are chosen to handle the different scaling of the parameters (Nocedal and Wright 2012). In the *optim* function in **R** this task is solved by setting B_{x_0} equal to the identity matrix.

3.1.2.2. Iterative updates

Once started, the algorithm continues optimizing the parameter combination x_i until the norm of the gradient of the objective function becomes less than a predefined value (e.g. 10^{-6}).

In each iteration

1. a **new search direction**

$$p_i = -H_{x_i} \nabla f_{x_i} \quad (3.7)$$

is calculated using the inverse of the approximated Hessian $H_{x_i} = B_{x_i}^{-1}$.

2. Afterwards, the **parameter combination** x_i is **updated** using

$$x_{i+1} = x_i + \alpha_i p_i, \quad (3.8)$$

with the best step length α_i along the direction p_i obtained by an one-dimensional line search algorithm.

3. Finally, the inverse Hessian H_{x_i} is updated by minimizing the distance (a weighted Frobenius norm) between the update and H_{x_i} , while assuring the update is symmetric and positive definite, and fulfills the **secant equation** (3.11).

$$\min_H \|H - H_{x_i}\| \quad (3.9)$$

subject to

$$H = H^T \quad (3.10)$$

and

$$Hy_i = s_i, \quad (3.11)$$

with $y_i = \nabla f_{x_{i+1}} - \nabla f_{x_i}$ and $s_i = x_{i+1} - x_i = \alpha_i p_i$.

3.1.3. Problems in fitting GEV and GP distributions

When I started my PhD I wanted to apply the EVA in a massive parallel way by e.g. calculating the 100 year return levels of all stations within Germany. But when performing the analysis on the many hundreds of series a small number of the fits always failed. At first I blamed the *extRemes* package by Gilleland and Katz (2016) I used and switched to the *ismev*² instead. Using this package, the previously failing fits could be obtained. But now some other series could not be fitted anymore and caused the *ismev* package to throw an error. After some trail and error I decided to delve into the fitting routines and to search for the root of the problem.

The problem of the failing fits in the *ismev* package, which uses the Nelder-Mead algorithm 3.1.1, was due to a poor error handling within the package. After each fit it tried to estimate the errors of the GEV or GP parameters using the maximum likelihood method in section 2.2.2. It obtains the Hessian at the maximum likelihood estimate (MLE) of the parameters, inverts it, and calculates the square roots of the diagonal entries. But sometimes the Hessian returned by the fitting routine was badly conditioned and could not be inverted. Instead of implementing a fallback to a numerical approximation of the Hessian at the MLE, the package just throws an error, which can be fixed quite easily.

The nature of the error experienced with the *extRemes* package, which uses the BFGS routine 3.1.2, is way more fundamental and has implications in all kinds of unconstrained optimizations in the EVA. Due to the logarithms in negative log-likelihood functions of the GEV (2.18) and GP distribution (2.21), some parts of the parameter space, e.g. those corresponding to $\sigma < 0$, can not be accessed. Depending on how far away from the global optimum the optimization is initialized, the algorithm can get stuck at the boundary between the allowed and forbidden part of the parameter space, causing the fitting to fail or to return a wrong estimate.

In addition the MLE of the GEV and GP parameters are only defined for shape parameters bigger than -1 (see section 2.2.2). But in the beginning of the optimization procedure the algorithm sometimes takes a big step and moves to lower shape values. There the forbidden regions are way more vast than the ones in figure 3.2 and the route the algorithm takes is quite likely to end up at a boundary.

To illustrate the effect of the boundary, we will perform a short numerical experiment. In figure 3.2 I sampled a series of 30 points distributed according to a GEV or GP distribution with $\mu = 11.727$ (GEV only), $\sigma = 1.429$, and $\xi = -0.261$. This one single series will now be fitted over and over again with a variety of different initial parameter combinations. Each pixel in figure 3.2 represents the parameter combination used for initialization and its color indicates the resulting negative log-likelihood. Since the goal is to minimize the latter one, only those parameter combinations corresponding to the darkest blue can be considered are working starting points in order to reach the global minimum (marked with a yellow cross). The white areas are the forbidden parts of the parameter space. In addition, initial parameter combinations causing the optimization

²ismev version 1.41 <https://cran.r-project.org/web/packages/ismev/index.html>

routine to fail are marked red.

I choose a length of 30 points, because it represents the number of annual maxima in the default period considered by most studies in the field of climate science (Hansen, Sato, and Ruedy 2012). It is also a rule of thumb to have a set of at least ten times the number of fitting parameters to avoid overfitting (Abu-Mostafa, Magdon-Ismail, and Lin 2012). Since the EVA is quite data-consuming and most station data only feature several dozens of years, a larger number of points in the analysis might not represent the general use case well.

The pictures obtained for the Nelder-Mead and the BFGS algorithm are quite different. While the BFGS routine only rarely returns a non-optimal parameter combination but throws an error for more than half of the initial parameter combinations, the optimization never aborts using Nelder-Mead but a lot of the resulting parameters are non-optimal. But both outcomes are highly undesirable and happen, at least in the slice of the shape-location and scale-location plane of the GEV distribution, for similar initial parameter combination. What is causing these optimizations to fails?

Figure 3.3 shows the negative log-likelihood of a failed optimization using the Nelder-Mead algorithm including the returned MLE (black cross) and the components of gradient evaluated at the MLE (black arrows). In the slice of the shape-scale plane (middle) one can clearly see that the fitting routine was not able to return the global optimum. Instead, it got stuck on the boundary to the forbidden region of the parameter space (white). Along this boundary the negative log-likelihood is the smallest with a steep increase in orthogonal direction and only a very small decrease parallel to the boundary pointing towards the optimal value.

Gradient-based methods will try to improve the parameter combination by performing a step along the gradient. But regardless of which step size they choose, the step will always lead into a region where the negative log-likelihood is not defined since we are already very close to the boundary. Therefore, the BFGS methods will throw an error instead of returning a wrong estimate. This behavior is not specific to quasi-Newton routines but also holds for conjugated gradient algorithms (Press et al. 2007) (possibly with a more robust error handling).

The options of the Nelder-Mead algorithm to update the point marked with a black cross in figure 3.3 are shown in figure 3.4.

The direction in the parameter space connecting the worst vortex x_4 and the centroid of the remaining vertices \hat{x} passes the boundary in an almost orthogonal way and x_4 is already so close to the forbidden region only the inner contraction x_{c_i} would yield a valid update. But since the simplex is shrinking more and more along the displayed direction, the optimization routine will return the value of the best vertex x_1 at some point while still being trapped at the boundary.

Since the update of the BFGS algorithm is always performed according to the gradient, restarting the routine (see section 3.1.2.1) would not yield any improvement, in contrast to the Nelder-Mead algorithm. There, the initial simplex is created only at points the negative log-likelihood can be evaluated and the size of the simplex corresponds to the values of the initial parameter combination (see section 3.1.1.1). Restarting a Nelder-Mead-based fit of the GEV or GP distribution can thus be used to escape the boundary. The fresh representation of the simplex is way larger than the previous one and likely to reach parts of the parameter space yielding even better results than initial

3.1. Unconstrained optimization

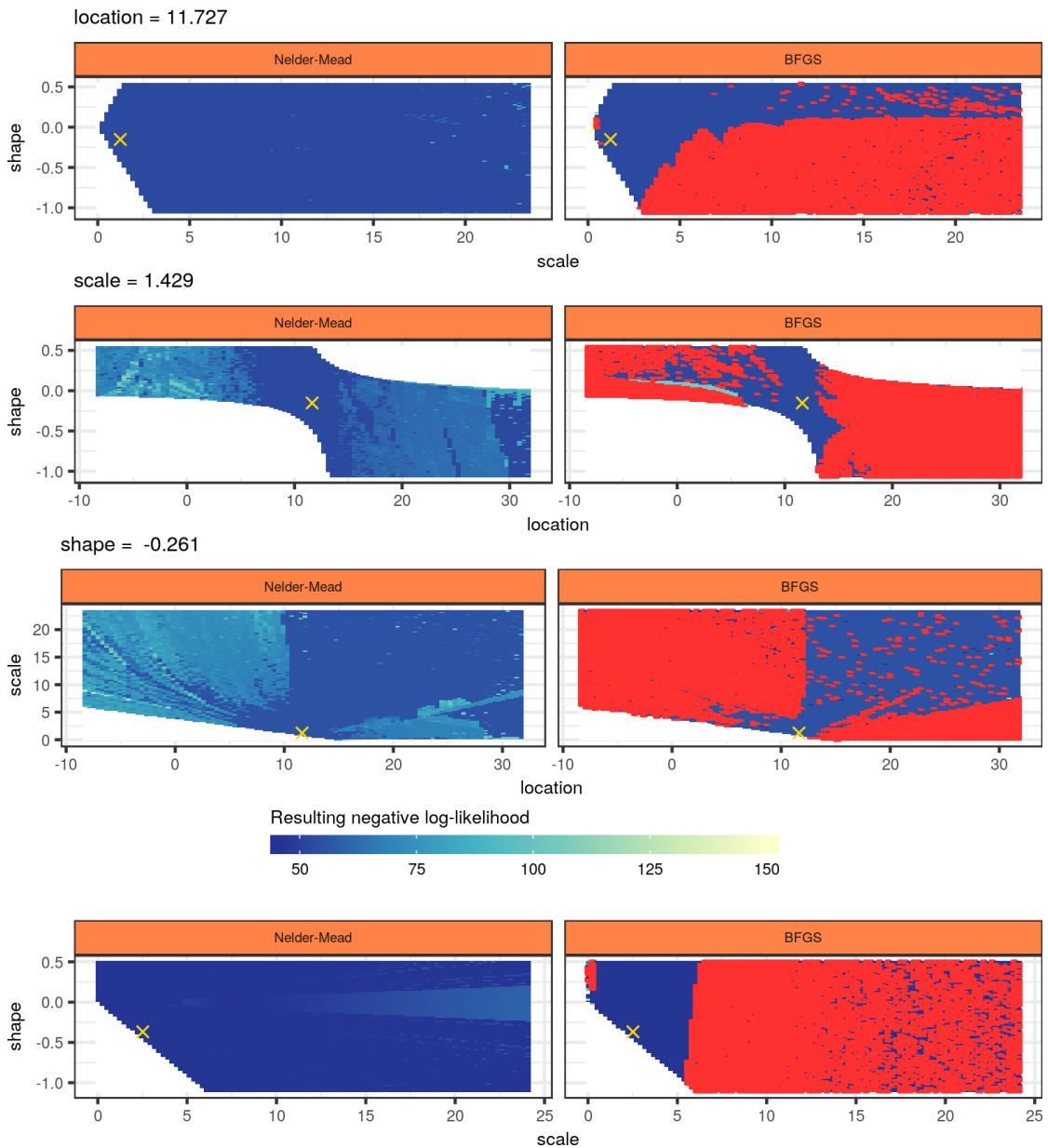


Figure 3.2.: Impact of the initial parameter combinations on the fitting of the GEV (upper six figures) and the GP (lower figures) distribution. A series of 30 points was drawn at random from both a GEV and GP distribution and the minimization of the corresponding negative log-likelihood was initialized at a range of different parameter combinations. The position of each pixel corresponds to the parameter combination used to initialize the optimization and its color indicates the negative log-likelihood obtained in the end of the optimization. In an ideal setting the optimization would be independent of the starting point and all of the explored parameter space would be colored uniformly. But due to the forbidden regions in the parameter space (white area) and existence for the MLE for shape parameters bigger than -1 only, some initial parameter combinations do not reach the global optimum (indicated by a yellow cross) and result in a bigger value (light blue). In case of the BFGS routine (right column), the algorithm is not robust against the choice of the starting points and the red pixels indicate parameter combinations resulting in an error and an abortion of the optimization routine.

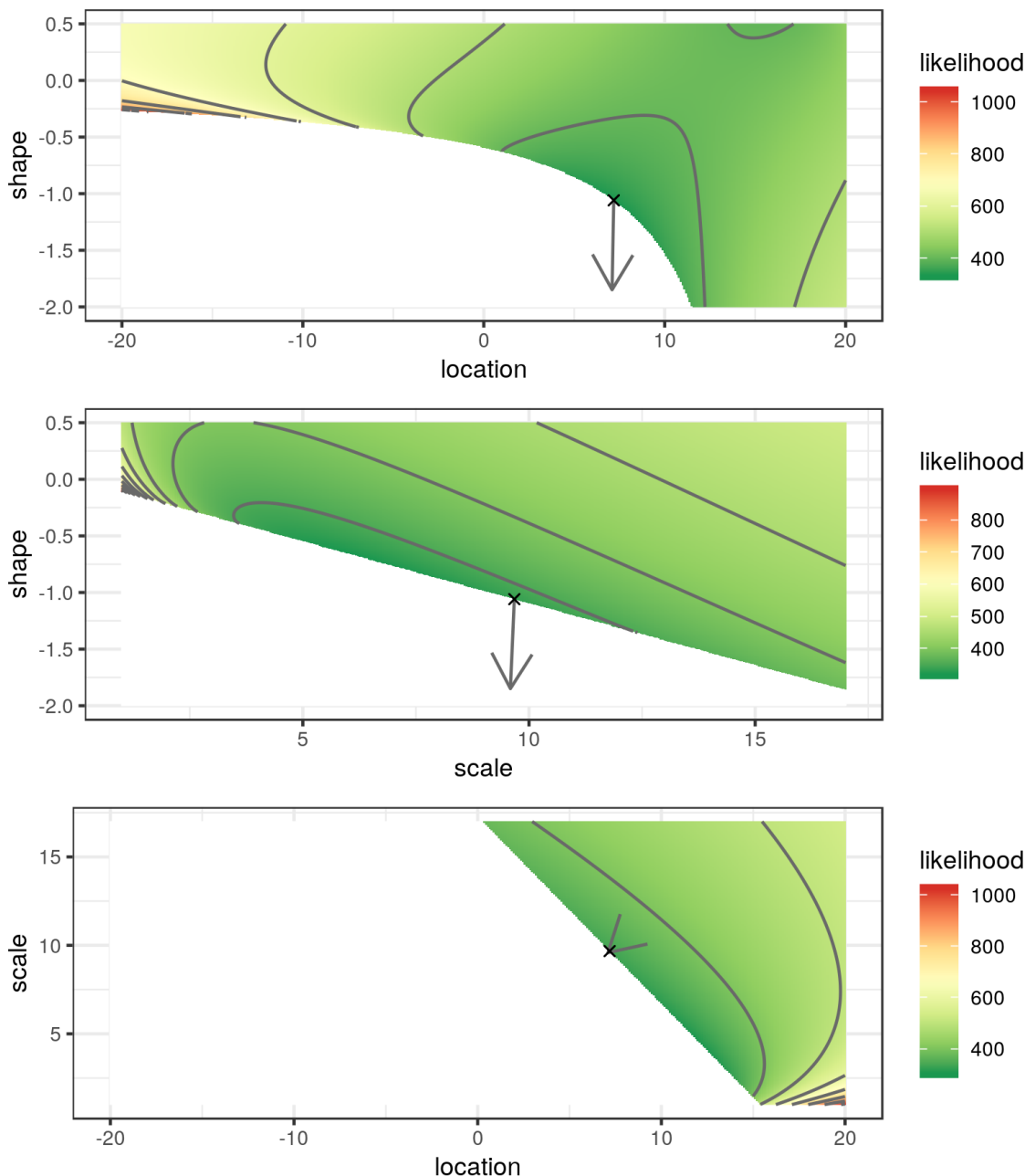


Figure 3.3.: An example of a failing optimization performed using the Nelder-Mead algorithm implemented in the **R** package **dfoptim**. The displayed negative log-likelihood of the GEV distribution was obtained using equation (2.18) on the annual maxima of the daily temperature anomalies of the Potsdam station. The optimization was started at $\{\mu_0 = 0.1, \sigma_0 = 1.429, \xi_0 = -0.01\}$ and did terminate at the parameter combination indicated by the black cross, which is also the intersection of the displayed slices through the three-dimensional parameter space. The global minimum of the negative log-likelihood function, obtained using a different set of initial parameters or a more robust optimization routine, is located at $\{\hat{\mu} = 11.727, \hat{\sigma} = 1.429, \hat{\xi} = -0.261\}$ (not shown in this figure). For each slice the projected component of the gradient at the optimization result normalized to unit length is drawn as an arrow. Note that since the dominant component points into the direction of the shape parameter, the length of its projection onto the scale-location plane is negligible.

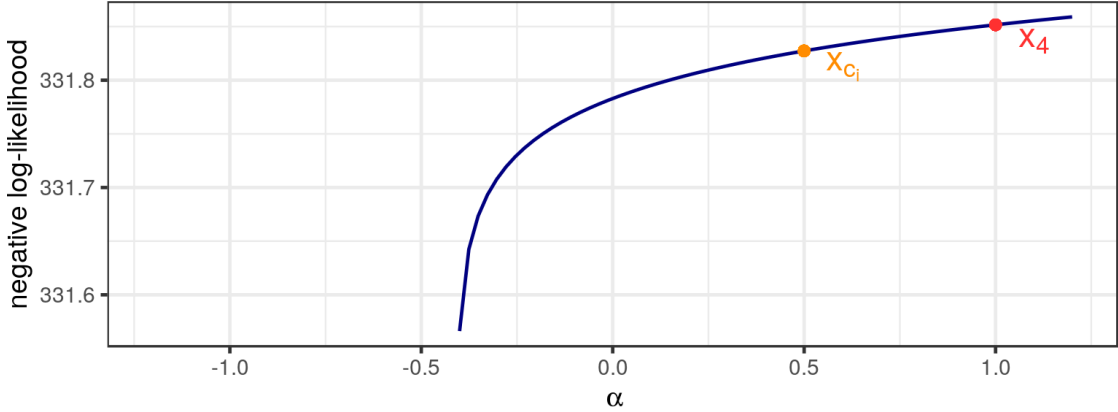


Figure 3.4.: The negative log-likelihood of the objective function along the line crossing both the worst vertex x_4 and the centroid of the remaining points \bar{x} of the final simplex of the Nelder-Mead optimization in figure 3.3. Since the likelihood function of the GEV distribution has three dimensions, a convenient visualization of the resulting four-dimensional simplex is not possible. Instead, the negative log-likelihood along the direction in which the worst point of the simplex would be updated ($\bar{x} + \alpha(x_4 - \bar{x})$) is displayed. Due to the logarithms in the equation of the likelihood (2.18), its values are just defined for α bigger than approximately -0.4. Therefore, only the inner contraction would be possible. But the change in negative log-likelihood is too small to permit further iterations.

parameter combination near the boundary.

A proper and previously neither implemented nor used way of fitting parameters in the EVA would be to use the Nelder-Mead algorithm and restart it until there is no significant change in the resulting parameter combination anymore.

But this procedure, as well as the method of half-stepping, is not guaranteed to find the global minimum. A much better way would be to use a constrained optimization algorithm instead, which hasn't been used previously in EVA either.

3.2. Concepts of constrained optimization

Some optimization problems can not be solved by the straight forward procedures introduced in section 3.1 since there are additional constraints, which have to be fulfilled by the solution. Such problems are encountered quite frequently in real-world applications. E.g., in many physical calculations the mass or energy is assumed to remain constant (**equality constraints** \mathcal{E}) or certain quantities, like rainfall or the population, can not be negative (**inequality constraints** \mathcal{I}).

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} c_k(x) = 0, & k \in \mathcal{E} \\ c_k(x) \geq 0, & k \in \mathcal{I} \end{cases}$$

The inequality constraints c_k , with $k \in \mathcal{I}$, are in general harder to deal with. Therefore, they are often converted into equality constraints instead, such as $c_{-,k} = \min(c_k, 0) = 0$

(Lange 2013).

The overall goal of the constrained optimization is to find the minimum of the objective function in the so called **feasible region**, where all constraints imposed on the functions are fulfilled. Since our task will be to prevent the logarithms in the negative log-likelihood function of the GEV and GP distribution, which include a non-linear statement, from becoming negative, ordinary box constraints, allowing for an upper and lower bound of certain parameters, are not sufficient. Instead, we need a framework, which allows for integrating non-linear constraints into the optimization procedure. This can be achieved by joining the objective function and the constraints into a single formula, which then can be minimized using the algorithms we discussed in section 3.1. The most prominent methods implementing this procedure are penalty methods and barrier methods.

In the **penalty method** a non-negative penalty $p(x)$ is added to the objective function *outside* of the feasible region $f(x) + \lambda_p p(x)$. An outer loop chooses the penalty constant λ_p and an inner one performs an unconstrained optimization of the sum of the objective function and the penalty terms. Whenever the optimization result of the inner loop does not fulfill the convergence criteria of the fit, λ_p is increased in the outer loop and the inner one is rerun. The **barrier method**, on the other hand, does change the perceived function values *inside* of the feasible region $f(x) + \lambda_b b(x)$. Here, a barrier function $b(x)$ is finite at the inside of the feasible region and infinite on its boundary and the weighting constant λ_b is decreased during the optimization. For a thorough introduction to these methods see Lange (2013, chap. 13).

3.2.1. Augmented Lagrangian methods

A widely used and numerically stable penalty method is the so-called **augmented Lagrangian method**.

$$\mathcal{L}_A(x; \lambda, \nu) = f(x) - \sum_{k \in \mathcal{E}} \lambda_k c_k(x) + \frac{\nu}{2} \sum_{k \in \mathcal{E}} c_k^2(x), \quad (3.12)$$

with ν being the penalty parameter and λ_k being estimates of the Lagrangian multipliers.

But in order to gain a better understanding of the different terms involved in equation (3.12), let's review its predecessor, the **quadratic penalty method**, first.

$$\mathcal{Q}(x; \nu) = f(x) + \frac{\nu}{2} \sum_{k \in \mathcal{E}} c_k^2(x), \quad (3.13)$$

In this algorithm all violations of the constraints c_k are squared and added to the objective function $f(x)$, which has to be minimized. Thus only those constraints, which are violated, contribute to the sum in equation (3.13). In the parlance of optimization community they are called *active*. Using this trick we do not have to distinguish between equality and inequality constraints anymore.

The problem with the quadratic penalty method is that the minimization of \mathcal{Q} becomes more and more difficult the larger the penalty parameter ν gets. This is especially true for quasi-Newton and conjugated gradient methods, since the Hessian of \mathcal{Q} becomes ill-conditioned near the minimum (Nocedal and Wright 2012).

The ill-conditioning can be reduced by adding the second term in equation (3.12) featuring estimates of the Lagrange multipliers λ_k . This extension of the quadratic penalty method does make the **augmented Lagrangian method** far more stable and even faster in terms of convergence. See Nocedal and Wright (2012, chap. 17) for a detailed description including proofs of the properties of the method.

3.3. Constrained optimization in the EVA

As discussed in section 3.1.3 both the boundary to the forbidden regions, introduced by the logarithms in (2.18) and (2.21), and the existence of the MLE of the GEV and GP parameters for shapes bigger than -1 only can result in numerical artifacts. Depending on the initial parameter combination, the optimization algorithm can get stuck at the boundary to the forbidden region. A rule of thumb to circumvent this issue is to choose a starting point as close as possible to the unknown global optimum using various heuristics (see section 2.2.4), to use the Nelder-Mead algorithm, and to restart the optimization until the change in the resulting MLE of the GEV or GP parameters is less than a predefined precision constant. But the restarting of the optimization is neither guaranteed to reach the global minimum nor does it feel like a proper solution of the problem. Luckily, one can do better by using a constrained optimization instead.

3.3.1. Application of the augmented Lagrangian method to the GEV and GP distribution

To avoid those problems, one has to introduce both the boundary to the forbidden regions and the lower bound of the shape parameters as constraints. The arguments of the logarithms in the negative log-likelihood of the GEV (2.18) and GP distribution (2.21) have to be strictly positive and the shape parameter must not take values equal to or lower than -1. Since the second logarithm in equation (2.18) holds a non-linear expression, methods only working on linear constraints, like e.g. box constraints, are not sufficient. Instead, we will use the augmented Lagrangian method introduced in section 3.2.1. It seems tempting to just omit the one non-linear constraint and to apply box constraints. But I performed the analysis in figure 3.6 with both all the constraints mentioned above and just the linear ones. The results obtained when omitting the non-linear constraint in equation (2.18) were much more poor. So this simplification should not be used in the EVA.

As discussed in section 3.2.1, in the augmented Lagrangian method we will add the constraints to the objective function in order to transform the constrained optimization into an unconstrained one. Since the concept is a quite general one, I will formulate the resulting optimization problem for the GEV distribution only and omit the GP one as well as their special cases for $\xi = 0$.

$$\begin{aligned}
 l(z; \mu, \sigma, \xi) &= n \cdot \log \sigma + \sum_{i=1}^n \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \\
 &\quad + \left(1 + \frac{1}{\xi} \right) \sum_{i=1}^n \log \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right] \\
 &\quad + \sum_{k \in \mathcal{E}} \lambda_k c_k(z; \mu, \sigma, \xi) + \frac{\nu}{2} \sum_{k \in \mathcal{E}} c_k^2(z; \mu, \sigma, \xi), \quad (3.14)
 \end{aligned}$$

where λ_k are the Lagrangian multipliers, ν is the penalty parameter, c_k are the violations of the constraints, and \mathcal{E} is the set of all active constraints. The $n + 2$ constraints are triggered as follows

$$c_i : \xi \left(\frac{z_i - \mu}{\sigma} \right) > -1 + \epsilon_1 \quad (3.15)$$

$$c_{n+1} : \sigma > \epsilon_2 \quad (3.16)$$

$$c_{n+2} : \xi > -1 + \epsilon_3, \quad (3.17)$$

with where $i \in \{1, \dots, n\}$, $\epsilon_1 = \epsilon_3 = 0.05$, $\epsilon_2 = 0.03$, and n is the number of data points in the series of maxima z . By checking whether or not the constraint is violated and permitting only two different kinds of outputs, either zero or the violation of the constraint, the inequality constraints of the GEV distributions are transformed into equality ones.

A constraint only contributes to (3.14) if it is violated/active. Thus, it is not sufficient to restrict the scale parameters to value bigger than zero, the shape parameter to values bigger than -1, and so forth. We have to slightly adjust the thresholds of constraint violation by introducing the constants $\epsilon > 0$ in order to prevent the algorithm from entering the forbidden regions or the part of the parameter space, where the ML estimator is not defined anymore. Formally, the ϵ could be of any small values but the ones used above showed to be the most appropriate ones throughout the analysis of ours.

As a drawback this further decreases the addressable parameter range of the ML optimization. But since we are now able to find the global minimum of the negative log-likelihood for almost all initial parameter combinations, this is a price we are glad to pay. In addition, the shape parameter is assumed to fall in the range between $-0.5 < \xi < 0.5$ for natural data (Hosking, Wallis, and Wood 1985) and thus the further restrictions should not affect any actual analysis. Within the feasible region of the constrained optimization the original likelihood of the GEV or GP distribution is not altered and thus the obtained results will match the unconstrained ones, which do not get trapped at the boundary of the forbidden region.

In the context of EVA the penalty parameter ν gets updated four to five times while no update of the Lagrangian multipliers takes place. This causes the overall non-linear constrained optimization to last only twice as long as the classic unconstrained one³.

³this benchmark was performed using the *optim* function of the **stats** package in **R** (R Core Team 2017) version 3.4.0 for the baseline *Nelder-Mead optimization* and the **alabama** package version 2015.3-1 for the optimization using the augmented Lagrangian method

Due to the possibility of ill-conditioned Hessians, as well as to the overall robustness regarding the forbidden regions, we strongly encourage to perform the inner routine of the augmented Lagrangian method using the Nelder-Mead instead of the BFGS algorithm.

3.3.2. Numerical comparison of the optimization results

Let's recall our motivation. The overall goal of the non-linear constrained optimization procedure was to make the results independent of the initial parameter combinations and to avoid the artifacts shown in figure 3.2. By introducing constraints via equation (3.14) we prohibit the optimization algorithm from entering the part of the parameter space the log-likelihood of the GEV and GP distribution is not defined anymore and keep the shape parameter bigger than -1 at all points. Now, the algorithm should find the global minimum more easily.

The impact of the constraints is displayed in figure 3.5. In the upper half of the picture you can see the original negative log-likelihood of the GEV distribution of equation (2.18) including some potential lines. The result of the unconstrained optimization is marked with a black cross. Judging from the potential lines it is obvious, especially in the scale-shape plane, that the optimization result does not correspond to the global optimum. Instead, the routine got stuck at the boundary to the forbidden region like in figure 3.3. Since the parameter space of the GEV distribution is three-dimensional, only three orthogonal slices intersecting at the global optimum are plotted.

To avoid these numerical artifacts, the penalized likelihood corresponding to equation (3.14), shown in the lower part of the figure, can be used. It was calculated for the very same underlying series and with all Lagrange multipliers λ_k set to zero and the penalty parameter ν set to 1000. These are values typically encountered during the last outer iteration of the augmented Lagrangian method in the EVA. While the penalty introduced close to the boundary to the forbidden region is not visible, the one for shape parameters smaller than -1 is. It is the latter one ensuring that the optimization routine does not step into a badly conditioned part of the parameter space right in the beginning of the optimization. But the constraints close to the boundary are nevertheless quite important. Without them the algorithm would still get trapped at the boundary in some runs.

On the other hand, figure 3.5 also very nicely illustrates how little the negative log-likelihood is changed. Apart from the discussed regions of shape parameters below -1 and close to the boundary it is not changed at all and the optimization using the augmented Lagrangian method does yield the same results as the unconstrained one started at working initial parameter combination.

In order to compare the performance of the augmented Lagrangian method with the classical, unconstrained routines in the context of the EVA, I repeated the numerical experiment of section 3.1.3 and figure 3.2. In detail, I used the augmented Lagrangian method implemented in the **alabama** package⁴.

As shown in figure 3.6, the optimization routine is now able to reach the global optimum, marked by a yellow cross, for almost all initial parameter combinations. Therefore, we got almost entirely rid of the numerical artifacts introduced by the unconstrained routines and it is thus the method of choice when fitting the GEV or GP distributions.

⁴**alabama** version 2015.3-1 in **R** 3.4.0. was used in all calculations

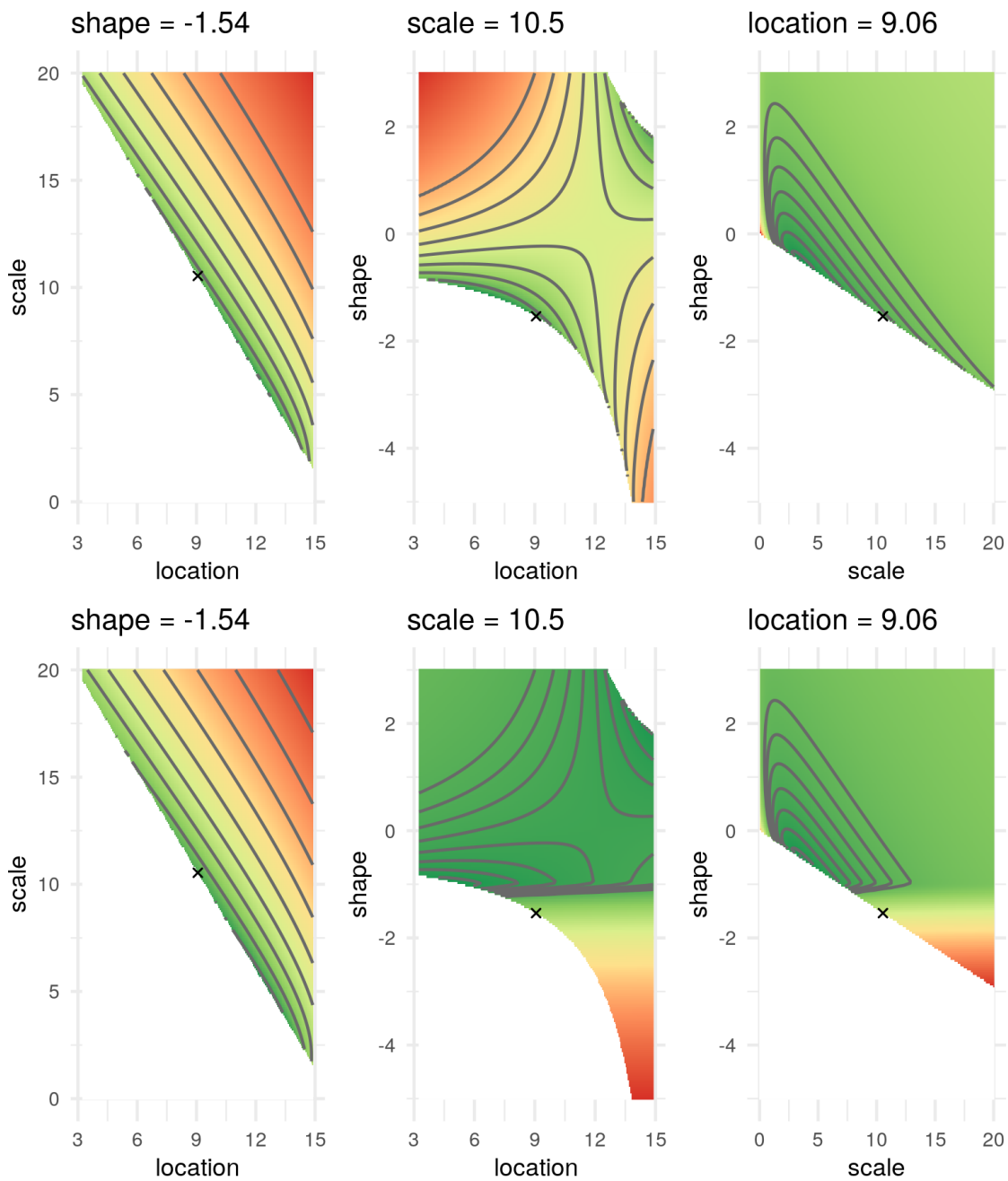


Figure 3.5.: The negative log-likelihood of the GEV distribution obtained using both the raw equation (2.18) (upper half) and the constrained one in (3.14) using the augmented Lagrangian method (lower half). The result of the failing unconstrained optimization is marked with a black cross and can obviously be avoided by applying the constraints. Apart from the regions of the parameter space in which the constraints are active (close to the boundary to the forbidden region and for shape parameters lesser than -1), the likelihood and the contour lines stay the same. Thus, the augmented Lagrangian method only avoids numerical artifacts but does not alter the results of the classical likelihood approach.

3.3. Constrained optimization in the EVA

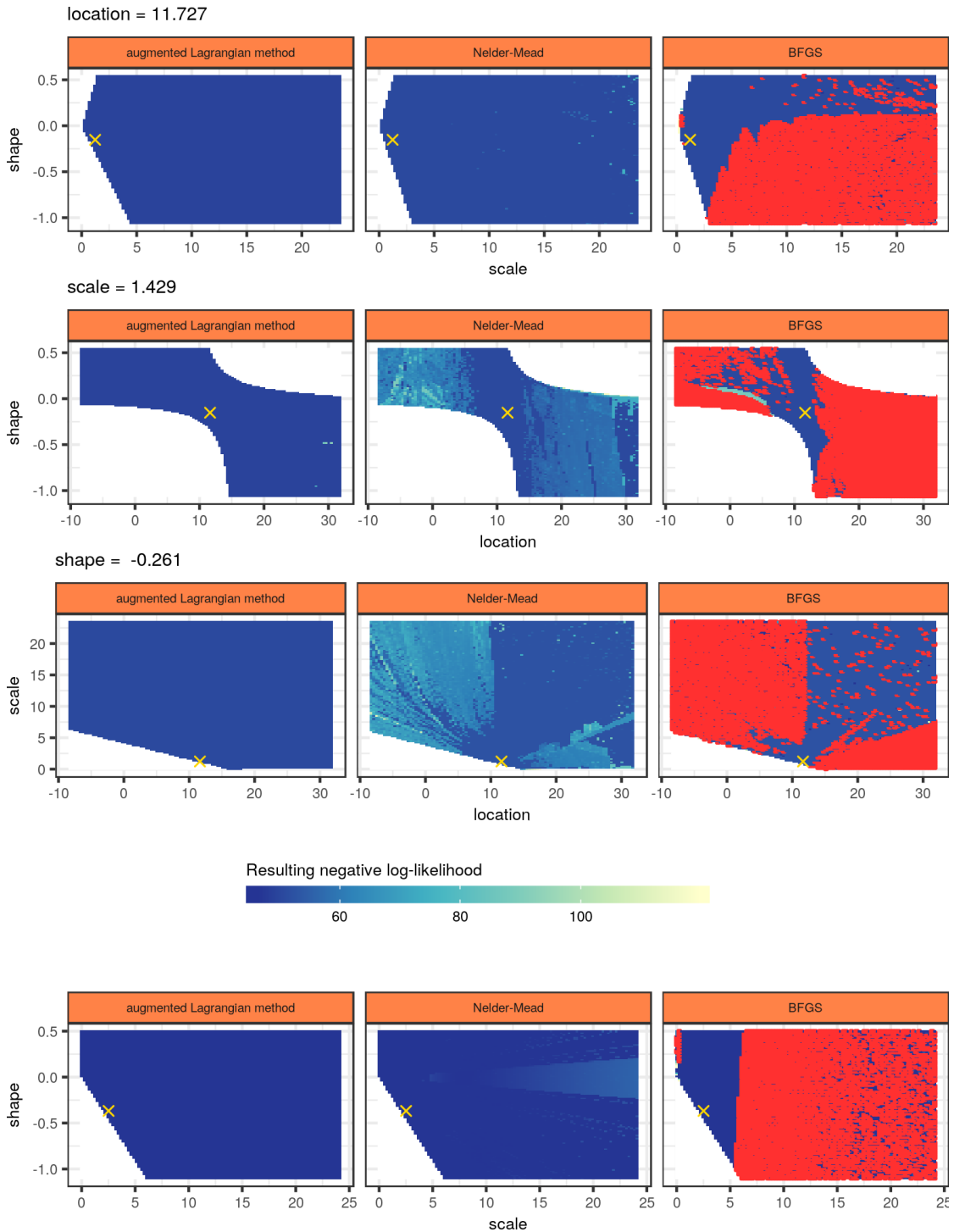


Figure 3.6.: The same analysis as in figure 3.2, but also performed with the augmented Lagrangian method (left-most column), which is presented in section 3.2.1. Using this algorithm, the global minimum of the parameter space could be retrieved from almost all initial parameter combinations.

In terms of performance the augmented Lagrangian method does approximately take twice as long as the constrained optimization and needs around 30 milliseconds to run on a single CPU.

4. Short-range correlations in the EVA

All the derivations and algorithms introduced in chapter 2 required the underlying data to be independent and identically distributed. While these conditions might be met by some artificial series in thought experiments or proofs of concepts, they are almost certainly not fulfilled by real data. In this chapter I will address this issue by extending the theory of the EVA in such a way it also copes with measured data of complex systems.

We will soften the requirements by asking the series to be only identically distributed. Two events, X_N and $X_{N+\Delta N}$, will have a conditional probability $Pr\{X_{N+\Delta N}|X_N\}$ solely depending on the time difference ΔN but not on the time point N itself. This statistical property is also called **stationarity**. Further, we require those dependencies to decay reasonably fast in time, e.g. exponentially, so more distant events can be considered independent. We thus assume our time series X to only contain **short-range correlations**. If, on the other hand, the dependencies decay polynomially in time or even slower, there are **long-range correlations** present in the data.

In the extreme value analysis short-range correlations can be handled rather easily since we are handling the maximum values of non-overlapping blocks. Given that the correlation length is much smaller than the size of the blocks, the maxima can be still considered both independent and identically distributed. If there, however, are long-range correlations present in the data, the maxima are not independent anymore. Thus, the likelihood, which is the joint probability of all observed events conditioned on a particular parameter combination (see equation (2.12)), is not factorizable and we can not use equation (2.18) to estimate the parameters of the GEV distribution anymore, as we will see in section 4.1. Dealing with short-range correlations when fitting the GP distribution is more subtle and will be treated in section 4.2.

Are these new assumptions reasonable when describing the climate system? From climate modelling and the daily endeavors of many companies, which try to foretell both temperature and rainfall for the near future, we know that the horizon of predictability for atmospheric temperatures is around one to two weeks. For precipitation it is even less. Both time spans are very small compared to the usual block size of one year. Therefore, assuming the data to contain only short- but no long-range correlations is reasonable. Please note that there are studies indicating the presence long-range correlations in temperature data by e.g. Massah and Kantz (2016). But since the methods yielding those findings are still debatable and under active development and due to the lack of alternatives, we have to neglect the possibility of their presence and treat our results with care.

A horizon of predictability of several weeks in the atmospheric temperature data is not entirely true. In the northern hemisphere a very high temperature (occurring most probably during summer) will not only indicate that weather will most likely be hot during the next week as well but also that it will be again quite warm one year into the future and quite cold in six months. This is due to the annual cycle of the sun, the most

strongest and dominant of all climatological periodicities. For a proper handling within the EVA we have to remove it from our data. Several techniques to accomplish this task are presented in section 4.3.

Apart from the short-range and maybe long-range correlations there is a very prominent non-stationarity in the data called the climate change. Both the natural variability of the climate and the human-made contributions will cause our data to be not identically distributed anymore. For this chapter we will neglect its presence, as usually done by the extreme value community, and return to its treatment in chapter 5 when discussing the model class of the VGLM.

4.1. Block maxima in stationary series with short-range correlations

Firstly, we focus on how to extend the GEV approach to cope with stationary time series. To quantify the degree of correlation within a series, we introduced the so-called $D(u_N)$ condition. It ensures that exceedances over a sequence of thresholds u_N are almost independent and that their correlation decays sufficiently fast as the block length N is going towards infinity.

Definition 4.1

A stationary series $X = X_1, X_2, \dots$ satisfies the $D(u_N)$ condition if, for all $i_1 < \dots < i_p < j_1 < \dots < j_p$ with $j_1 - i_p > l$,

$$|Pr\{X_{i_1} \leq u_N, \dots, X_{i_p} \leq u_N, X_{j_1} \leq u_N, \dots, X_{j_p} \leq u_N\} - Pr\{X_{i_1} \leq u_N, \dots, X_{i_p} \leq u_N\} Pr\{X_{j_1} \leq u_N, \dots, X_{j_p} \leq u_N\}| \leq \alpha(N, l),$$

where $\alpha(N, l_N) \rightarrow 0$ for some sequences l_N such that $l_N/N \rightarrow 0$ as $N \rightarrow \infty$.
(see Coles (2004) Definition 5.1)

This corresponds to the requirement that our series X only contains short- but no long-range correlations.

Next, we have to incorporate this condition into the extreme value limit theorem we already established in 2.1.

Theorem 4.1

Let $X = X_1, X_2, \dots$ be a stationary process and $M_N = \max\{X_1, \dots, X_N\}$ the corresponding series of block maxima with a block length N .

If $\{a_N > 0\}$ and $\{b_N\}$ are sequences of constants such that

$$Pr\left\{\frac{M_N - b_N}{a_N} \leq z\right\} \rightarrow G(z),$$

with $G(z)$ being a non-degenerated distribution function, and if the $D(u_N)$ condition with $u_N = a_N z + b_N$ is satisfied for all real z , then G is a member of the family of extreme value distributions.

(see Coles (2004) Theorem 5.1)

The GEV distribution $G(z)$ in theorem 4.1 is closely related to the limiting distribution of the independent version X^* of the series X . To illustrate the effect of the short-range dependencies on the estimated GEV parameters, we will have a look at an example taken from book *Introduction to Statistical Modeling of Extreme Events* written by Coles (2004, 94–95). Here, we will sample two series, an independent and a correlated one, featuring the same marginal distributions.

Example 4.1

We start by defining an independent sequence of random variables Y_0, Y_1, \dots with the common distribution function

$$F_Y(y) = \exp \left\{ -\frac{1}{(a+1)y} \right\},$$

with $y > 0$ and $0 \leq a \leq 1$, and a process X_i as

$$\begin{aligned} X_0 &= Y_0, \\ X_i &= \max \{ aY_{i-1}, Y_i \}, \text{ for } i = 1, \dots, n, \end{aligned}$$

with

$$\Pr \{ X_i \leq x \} = \Pr \{ aY_{i-1} \leq x, Y_i \leq x \} = \exp \left(-\frac{1}{x} \right),$$

for $x > 0$. Therefore, X_i is a stationary series with a marginal distribution of standard Fréchet.

Next, we will define an independent series X_1^*, X_2^*, \dots having a standard Fréchet marginal distribution as well. The corresponding block maxima $M_N^* = \max \{ X_1^*, \dots, X_N^* \}$ are distributed as

$$\Pr \{ M_N^* \leq Nz \} = \left[\exp \left(-\frac{1}{Nz} \right) \right]^N = \exp \left(-\frac{1}{z} \right).$$

The block maxima of the stationary series $M_N = \max \{ X_1, \dots, X_N \}$, on the other hand, are distributed as

$$\begin{aligned} \Pr \{ M_N \leq Nz \} &= \Pr \{ X_1 \leq Nz, \dots, X_N \leq Nz \} \\ &= \Pr \{ Y_1 \leq Nz, aY_1 \leq Nz, \dots, aY_{N-1} \leq Nz, Y_N \leq Nz \} \\ &= \Pr \{ Y_1 \leq Nz, \dots, Y_N \leq Nz \} \\ &= \left[\exp \left(-\frac{1}{(a+1)Nz} \right) \right]^N \\ &= \left[\exp \left(-\frac{1}{z} \right) \right]^{\frac{1}{a+1}}. \end{aligned}$$

Therefore, the distributions of the block maxima of both the independent and stationary series can be related via

$$\Pr \{ M_N^* \leq Nz \} = (\Pr \{ M_N \leq Nz \})^{\frac{1}{a+1}}.$$

A more general statement about the impact of short-range correlations on the fitted GEV distribution can be phrased as follows.

Theorem 4.2

Let X_1^*, X_2^*, \dots be an independent, X_1, X_2, \dots a stationary series, and $M_N^* = \max \{X_1^*, \dots, X_N^*\}$ and $M_N = \max \{X_1, \dots, X_N\}$ the corresponding block maxima for a block length of N . Under suitable regularity conditions

$$Pr \left\{ \frac{M_N^* - b_N}{a_N} \leq z \right\} \rightarrow G_1(z),$$

as $n \rightarrow \infty$ for arbitrary normalization sequences $\{a_N > 0\}$ and $\{b_N\}$ and a non-degenerated distribution function G_1 , if and only if

$$Pr \left\{ \frac{M_N - b_N}{a_N} \leq z \right\} \rightarrow G_2(z),$$

with

$$\begin{aligned} G_2(z) &= G_1^\theta(z) = \exp \left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}^\theta \\ &= \exp \left\{ -\theta \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \\ &= \exp \left\{ - \left[1 + \xi \left(\frac{z - \mu^*}{\sigma^*} \right) \right]^{-\frac{1}{\xi}} \right\}, \end{aligned}$$

$$\mu^* = \begin{cases} \mu - \frac{\sigma}{\xi} (1 - \theta^{-\xi}) & , \text{ for } \xi \neq 0 \\ \mu + \sigma \log \theta & , \text{ for } \xi = 0, \end{cases} \quad (4.1)$$

and

$$\sigma^* = \begin{cases} \sigma \theta^\xi & , \text{ for } \xi \neq 0 \\ \sigma & , \text{ for } \xi = 0. \end{cases} \quad (4.2)$$

and a constant θ defined as $0 < \theta \leq 1$.

θ is the so-called **extremal index** and can be thought of as the inverse of the mean cluster size (the number of successive highly correlated events). For values of $\theta = 1$ the mean cluster size of 1 indicates that every single event can be treated separately. Thus, the data itself is independent.

Although extended to cope with short-range correlations in stationary time series as well, theorem 4.2 looks remarkably similar to theorem 2.1 we stated for independent series earlier on. The fitting procedure, as well as the required likelihood function to fit the GEV coefficients, will thus be the same. The actual values will be affected by the degree of correlation in the data. While the shape parameter ξ will be the same, the location μ and scale σ parameter get rescaled due to the dependencies.

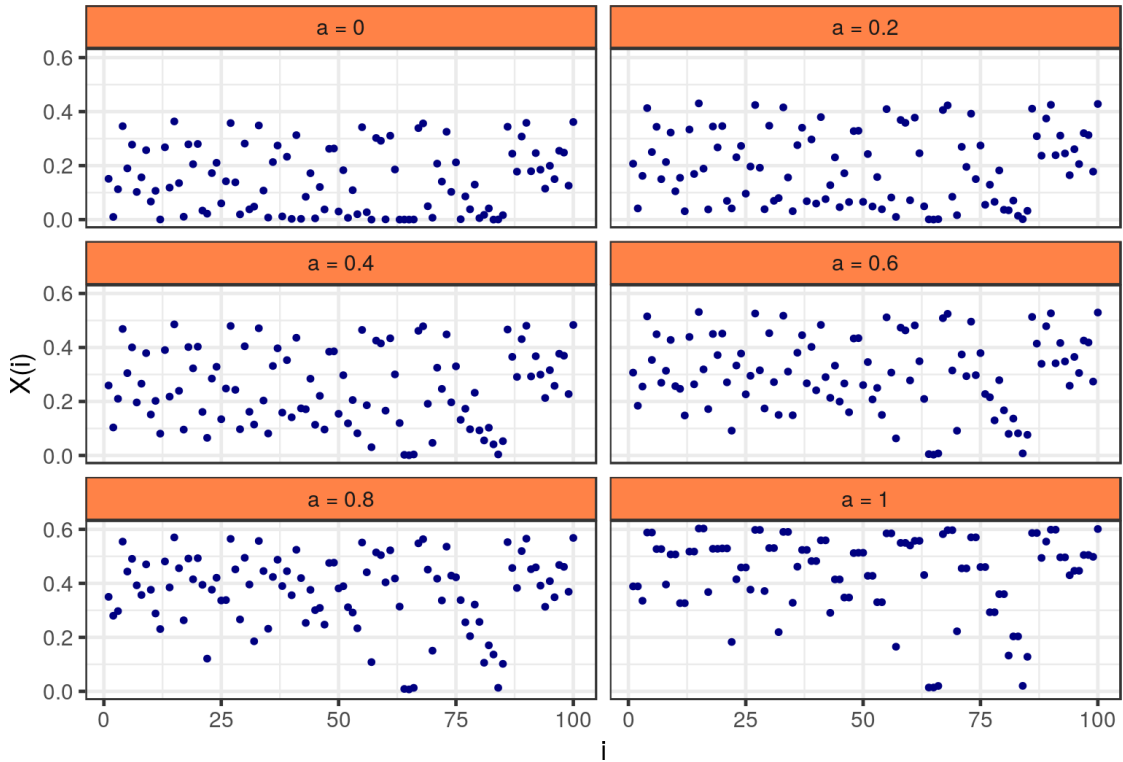


Figure 4.1.: Artificial series showing the impact of the a parameter in example 4.1. For all plots the same seed was used in the sampling of the underlying series y . As the value of the parameter a increases, the correlation between neighbouring points increases as well.

But since we have to estimate them anyhow, we do not have to mind the difference between the independent and the stationary series in the analysis.

As another illustrative example we consider the stationary sequence X of example 4.1. For $a = 1$ the extremal index becomes $\theta = 0.5$. The resulting mean cluster size of 2 can be seen in the bottom right part of figure 4.1, where higher values tend to appear in pairs.

4.2. Threshold exceedances in stationary series

The main idea of how to cope with short-range correlations in block maxima, as discussed in the previous section, can be rephrased as follows: We assume the correlation lengths to be much smaller than the size of the blocks so as a result the extracted maxima are independent. Unfortunately for threshold exceedances in the GP approach things are not that simple.

The most prominent way of dealing with short-range correlations in this context is **declustering** (Reiss and Thomas 2007). There are also attempts to refute the removal of dependencies in the threshold exceedances, e.g. by Fawcett and Walshaw (2007), but these ideas are very uncommon. In the declustering we group exceedances, extract only

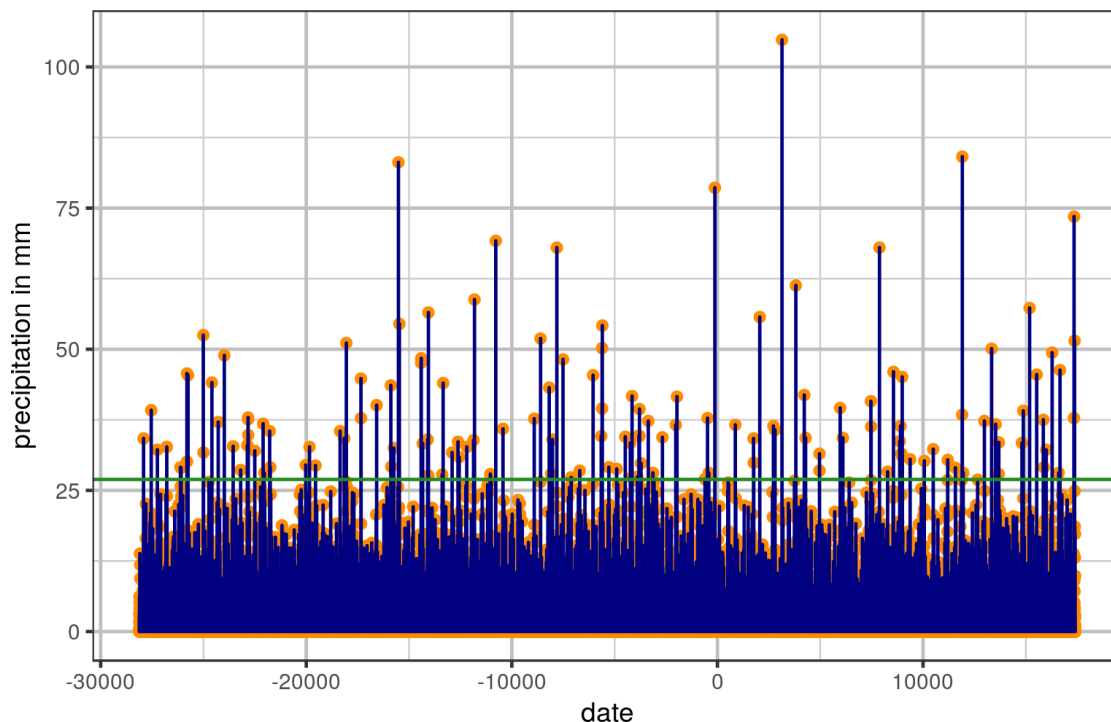


Figure 4.2.: Daily precipitation data of the Potsdam station. The green line indicates the threshold over which all data are considered to be extreme events. The data was provided by the German weather service (DWD)(Deutscher Wetterdienst 2018).

the maximal value of the group, and discard all others. The intricate part of this approach is *how* to define a group. But before diving into the theory, let's start with an example first.

Imagine we want to apply the EVA using the GP distribution to the daily precipitation records of the Potsdam station¹ in Germany in figure 4.2.

Next, we apply a threshold and consider only those days to be *extreme*, on which more than 27 mm of rain occurred in figure 4.3.

We see multiple events, which are potential candidates for the declustering since they appear rather close to each other. But instead of discarding most of them, what would be a wasteful procedure, we have to think about a reasonable way of defining the groups of dependency. For instance, several consecutive days of heavy rain can certainly be grouped together and reduced to a single event. But if, on the other hand, heavy precipitation takes place at the end of one low-pressure system and at the beginning of the next one as well, we want those two events to be recognized as independent and thus as part of different groups.

In the EVA literature there are two main approaches for applying the declustering: the block (Falk, Hüsler, and Reiss 2011) and runs (Leadbetter and Nandagopalan 1987) method. In the **block method** the time series is split in blocks of equal length and all

¹The data is provided by the German weather service (DWD) (Deutscher Wetterdienst 2018) free of charge at <ftp://ftp-cdc.dwd.de/pub/CDC/>

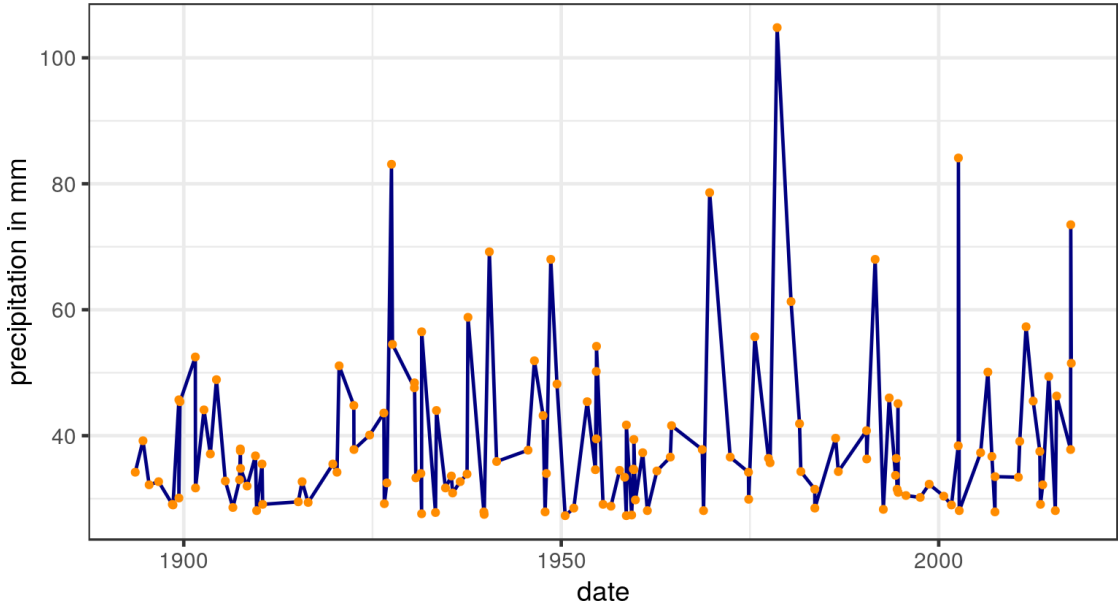


Figure 4.3.: Threshold exceedances of the daily precipitation data of the Potsdam station displayed in figure 4.2. Only those events are shown, which are bigger than 27 mm.

threshold exceedances within a block are assumed to belong to the same cluster. The **runs method** is more dynamic. Here all exceedances are grouped to a cluster unless a predefined number of consecutive events occur *below* the threshold. In general, the choice of just one block length is too rigid to be used with measured data and is therefore only used within theoretical derivations. In contrast, the number of events below the threshold to separate individual clusters, the so-called **run length**, can be tuned to correspond to the intrinsic dynamics of the system (e.g. the mean duration of a heat wave or the mean time a pressure system needs to pass through a region).

To illustrate the performance of both methods and the impact of their free parameters, the block length and the run length, we will apply them on the first year of daily maximum temperature of the Potsdam station also provided by the German weather service (DWD) (Deutscher Wetterdienst 2018). To remove the annual cycle, the anomalies of the time series were calculated (see section 4.3). In addition, we will set the threshold to 5°C . The value itself is too low to be used on the full series in EVA, but since we just want to get accustomed to the different declustering algorithms, it will serve the purpose (see figure 4.4).

Next, we perform the declustering of the threshold exceedances with block lengths of 15 and 30, as well as run lengths of 3 and 10.

While the results for the block method in figure 4.5 are of poor quality (24 clusters for a block length of 15 events and 12 for a block length of 30), the runs method yields more appropriate clusters (26 for a run length of 3 and 12 for a run length of 10).

But what is the most reasonable value of the run length? Should all the exceedances in January 1983 be grouped into one cluster, as in the case of a run length of 10, or is the data described more appropriately using additional clusters during this period? The answer to this question can be given by inspecting the dynamics of the system itself. But

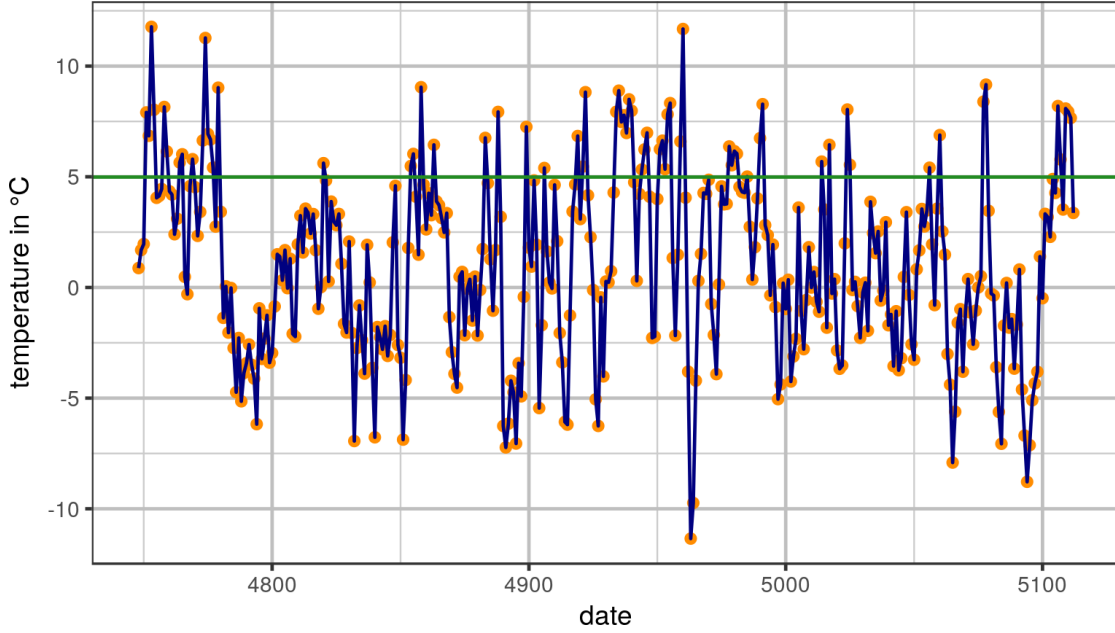


Figure 4.4.: First year of the daily maximum temperature anomalies of the Potsdam station in Germany provided by the German weather service (DWD) (Deutscher Wetterdienst 2018). The green line indicates the threshold above which all data points are considered extreme events.

this is a very tedious thing to do for a high dimensional and complex system, like the climate or the economic market. Fortunately, there is a non-parametric approach based on the runs method by Ferro and Segers (2003), which yields quite reasonable results as can be seen in the last row of figure 4.5.

Their approach is based on describing the exceedance times, the point in time the threshold exceedances did happen, by a point process. The asymptotic distribution (for time series lengths $n \rightarrow \infty$) of the **interexceedance times** T_i , the temporal distance between consecutive threshold exceedances, is a member of an one-dimensional family of distributions with the extremal index as their parameter. Its first moment is $E(T_i) = 1$ and its second is $E(T_i^2) = 2/\theta$ (see Ferro and Segers (2003)). By equating these theoretical moments of the limiting distribution and approximating them using the empirical ones, the extremal index can be estimated without any previous knowledge of cluster numbers or sizes using

$$\tilde{\theta}_n(u) = \begin{cases} \min \{1, \hat{\theta}_n(u)\}, & \text{if } \max \{T_i : 1 \leq i \leq N-1\} \leq 2, \\ \min \{1, \hat{\theta}_n^*(u)\}, & \text{if } \max \{T_i : 1 \leq i \leq N-1\} > 2, \end{cases} \quad (4.3)$$

with u being the applied threshold, n the length of the series, and N the number of observed threshold exceedances. In general, we will use

4.2. Threshold exceedances in stationary series

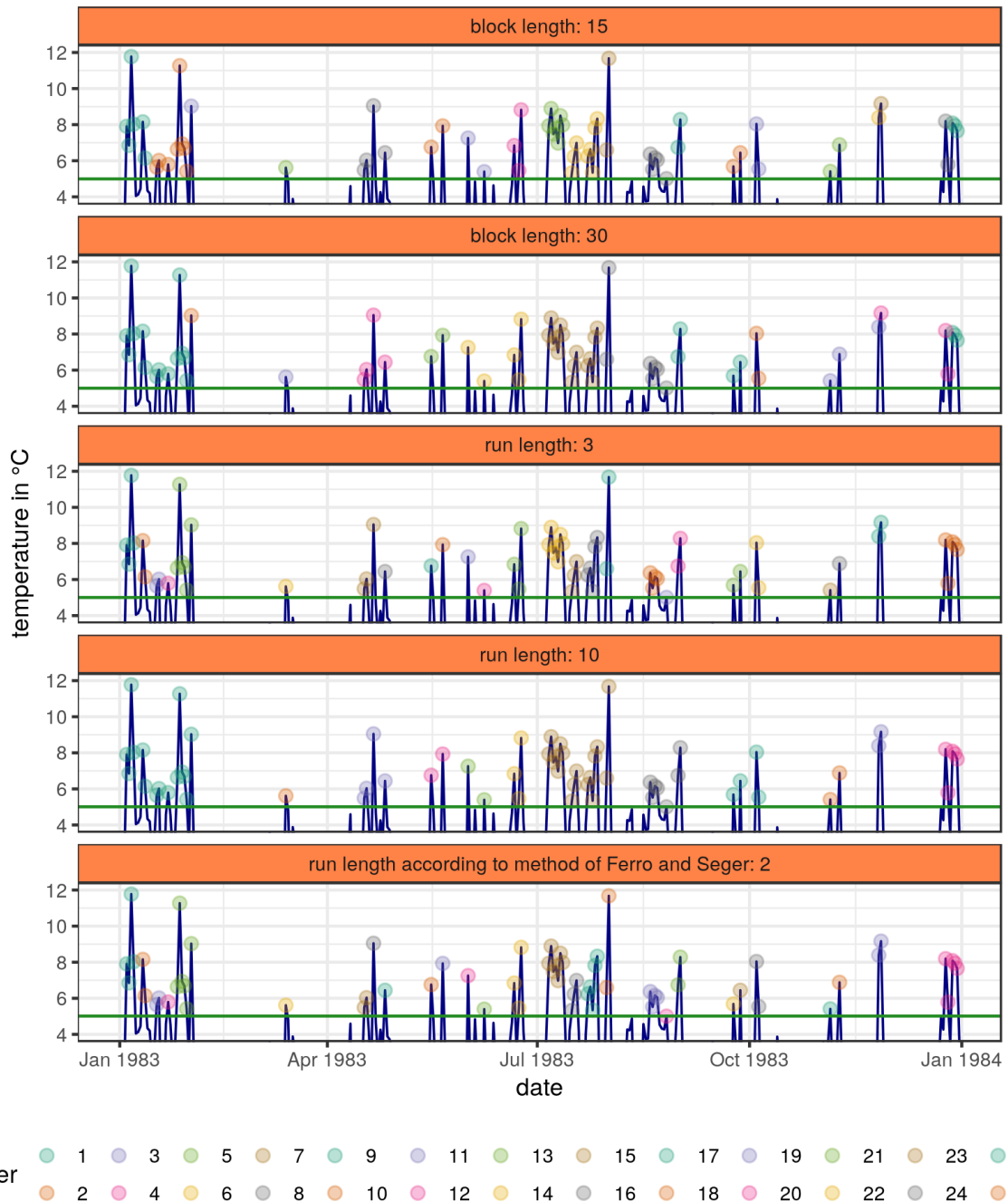


Figure 4.5.: A comparison of different declustering methods performed on the threshold exceedances of the daily maximum temperature anomalies at the Potsdam station of figure 4.4. In the upper two subfigures the **block method** and in the three lower ones the **runs method** was used. The colors represent the cluster individual events are belonging too. From all points within one cluster only the largest one is kept and all others discarded in order to remove short-range correlations from the data. Note that these colors are not unique, but only eight different ones are reused in a constant order. Since the information is not contained in the actual number of the individual cluster but in the fact if an event belongs to the same cluster as the previous or next one, a small number of colors with a high contrast had been chosen.

$$\hat{\theta}_n^*(u) = \frac{2 \left\{ \sum_{i=1}^{N-1} (T_i - 1) \right\}^2}{(N-1) \sum_{i=1}^{N-1} (T_i - 1)(T_i - 2)} \quad (4.4)$$

to estimate the extremal index. Note that this estimator can be larger than one, in which case an extremal index of one will be returned. If, however, the maximal interexceedance time is less than 3, equation (4.4) is undefined and we have to use the less optimal estimator

$$\hat{\theta}_n(u) = \frac{2 \left\{ \sum_{i=1}^{N-1} T_i \right\}^2}{(N-1) \sum_{i=1}^{N-1} T_i^2} \quad (4.5)$$

instead. The benefit of equation (4.4) is that its first-order bias is zero. For a more thorough theoretical treatment of the derivation of the estimators including proofs please see Ferro and Segers (2003).

In this setting, the estimated extremal index can also be thought of as the ratio of interexceedance times occurring *between* clusters, the so-called **intercluster times**, and the total amount of interexceedance times. Therefore, the $C - 1 = \lfloor \theta N \rfloor$ largest interexceedance times are independent intercluster times. If there is a tie and multiple T_i do share the $C - 1$ largest value, the C is decreased until the corresponding value is unique. This results in a total number of C clusters in the series and in the C -th largest T to be equal to the run length we need for declustering.

Since this approach doesn't contain any free parameters and its results are quite robust, it quickly became the default algorithm for declustering in EVA. It also will be used throughout this thesis.

Now, we return to our example of the daily precipitation data at the Potsdam station in figure 4.2 to visualize the effect of the declustering method by Ferro and Segers (2003) on the threshold exceedances.

After calculating the run length (5) using the estimate of the extremal index (0.95) and grouping all threshold exceedances in clusters (132), only the biggest events of each cluster are kept to ensure the independence of the exceedances. Using this procedure only eight events have been dropped in the bottom panel of figure 4.6 compared to the top one, decreasing the number of total threshold exceedances from 140 to 132.

4.3. Deseasonalization techniques

Imagine we want to determine the 100 year return level in a daily temperature series, or to phrase it more stirring, we want to calculate the temperature, which is so high it only occurs once every 100 years on average. Since we will use the annual maxima, it seems we would have $N \times 365.25$ data points to perform the EVA on, with N being the total number of years in our series. But unfortunately this is not the case (at least not for stations in Germany).

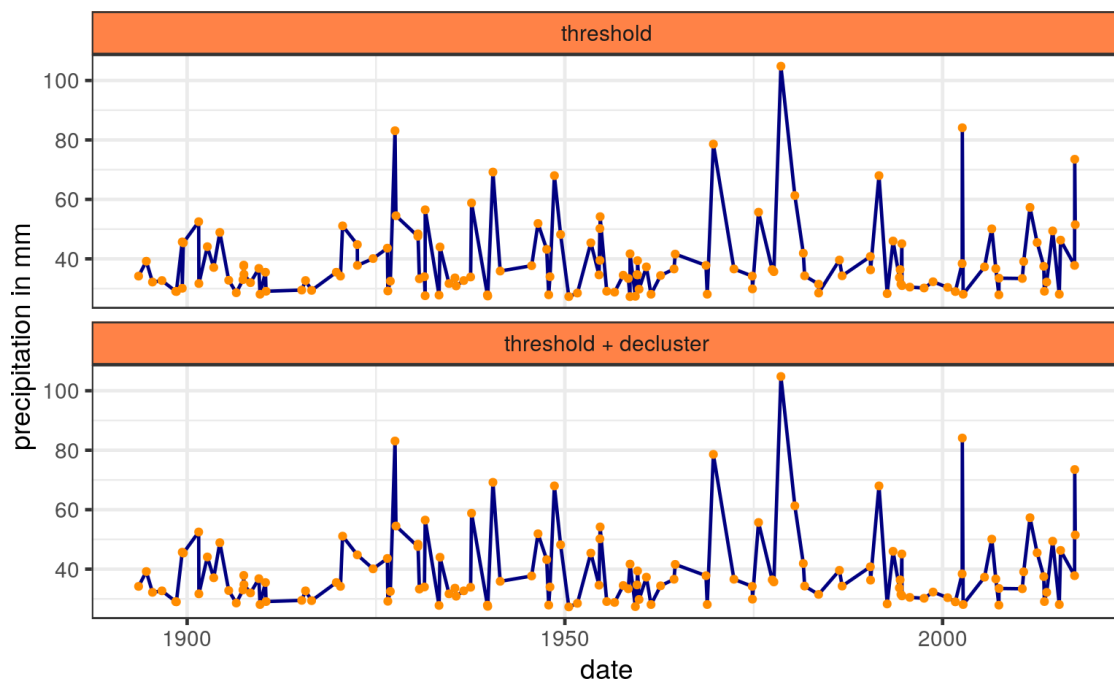


Figure 4.6.: Threshold exceedances of the daily precipitation data of the Potsdam station displayed in figure 4.2. Only those events are shown, which are bigger than 27 mm. While the upper figure displays the raw threshold exceedances, in the lower one the declustering algorithm by Ferro and Segers (2003) was applied to remove the short-range correlations in the data.

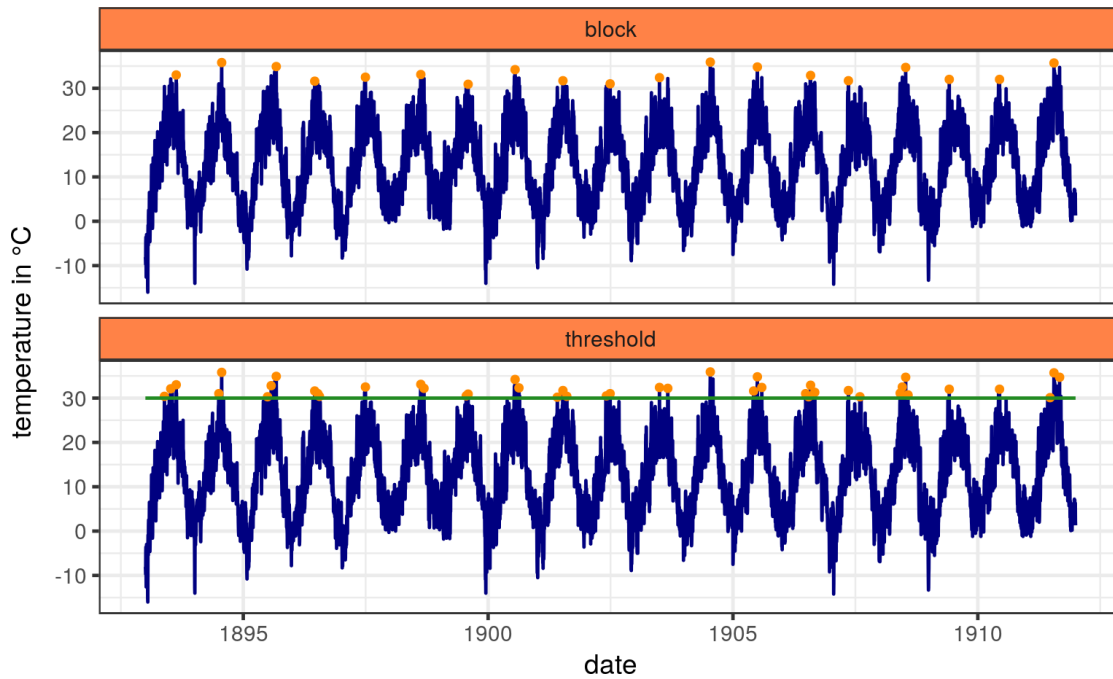


Figure 4.7.: Daily temperature maxima of the Potsdam station from 01.01.1893 till the 31.12.1911. Both the block maxima and the threshold exceedances (including declustering) are marked orange to emphasize the influence of the seasonal cycle on the analysis. The data was provided by the German weather service (DWD) (Deutscher Wetterdienst 2018).

The annual cycle superimposes quite a large modulation on the temperatures, causing both the block and threshold method to only sample data points throughout the summer (see figure 4.7). Therefore, the *effective length* of the series is much shorter (Massah and Kantz 2016).

Since the main goal of this thesis is to draw conclusions about the extreme events in the overall climate, the analysis will be more robust when applied to data with the seasonal cycle already removed. To illustrate this idea in figure 4.8 the time series excerpt of figure 4.7 was deseasonalized using the STL algorithm (see section 4.3.3) and the extreme events were extracted in the same way as beforehand. Now, the extremes of the time series can occur throughout the whole time series, not only summer, and the total number of data points is truly $N \times 365.25$.

In the remainder of this chapter I will provide an overview of the most prominent methods of deseasonalization.

4.3.1. Decomposition

The main goal of most deseasonalization software is to decompose a time series into three (additive and/or multiplicative) components: a **trend**, a **seasonal** part, and a **random** part.

One of the simplest ways to accomplish this decomposition is to use a moving average

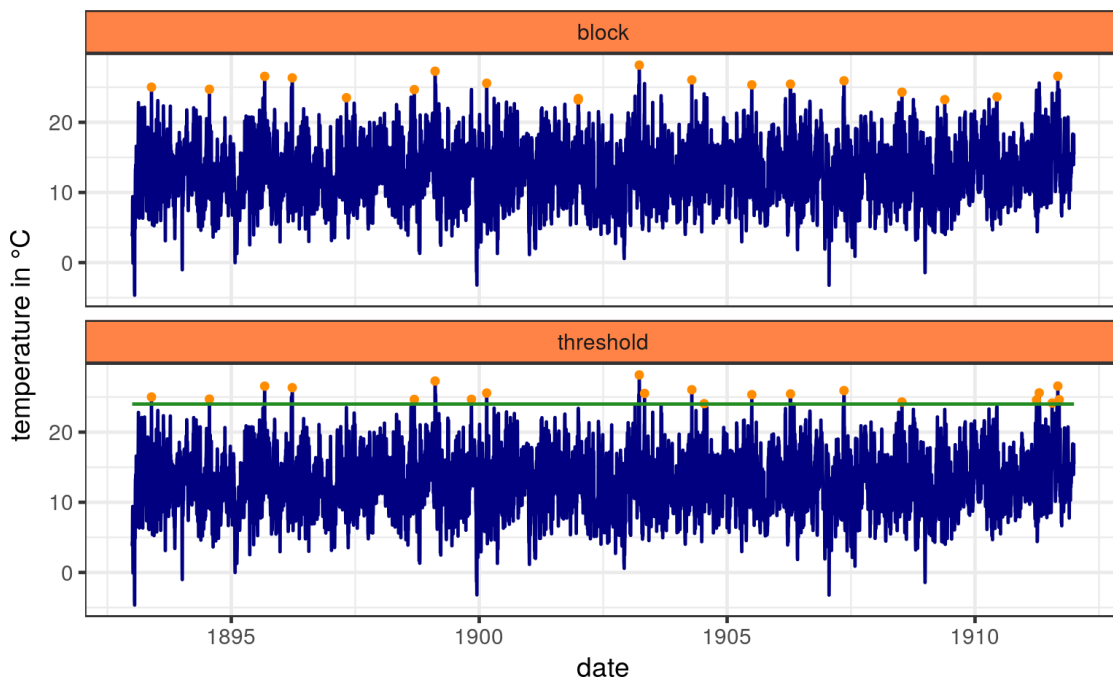


Figure 4.8.: This figure displays the same data as figure 4.7, but with the seasonal cycle removed using the STL algorithm prior to the detection of the extreme events.

to remove the trend and to calculate the anomalies of the remaining series. This approach is implemented in the `decompose` function in **R** (R Core Team 2017).

Firstly, one has to specify the size of one season in *lags*. A lag is the basic time unit of a series. An annual reoccurring season would have a size of 365.25 lags for daily data and a size of 12 for monthly ones. A moving average with a filter width of the size of the season is performed on the time series and the result gets subtracted from the original one.

Afterwards the **anomalies** will be calculated from this *detrended* time series by subtracting the seasonal cycle. The latter is approximated by the mean values of the individual dates throughout all seasons. E.g. to obtain the anomaly at the 1st of January of a particular year, one has to calculate the mean value of all the 1st of Januaries and subtract it from the original one.

A very popular generalization of the decomposition scheme described above is the Seasonal-Trend Decomposition procedure based on Loess (STL) (Cleveland et al. 1990). It is implemented in the `stl` function in **R** and based on the Fortran code of Cleveland et al. But before explaining the algorithm let me give you a summary of the Loess (Local regression) routine.

4.3.2. Loess

Suppose we observed an independent series x_i and a dependent series y_i of length n related via an unknown function $y = g(x)$. The loess regression provides a way to obtain an estimate $\hat{g}(x)$ for all x . Therefore, the evaluation is not bound to the exact locations

of the x_i and the observations are allowed to contain missing values too.

The basic idea of the local regression is, the closer neighbouring points are, the more similar is their functional behaviour. So instead of modelling a very complex unknown function with a sophisticated hypothesis and dozens of free parameters, we will perform local, piecewise regressions. The first free parameter in the loess regression is the number of points q we consider to be local neighbours. Using them we can construct a **neighbouring weight**

$$v_i(x) = W\left(\frac{|x_i - x|}{\lambda_q(x)}\right), \quad (4.6)$$

with the **tricube weight function**

$$W(u) = \begin{cases} (1 - u^3)^3 & \text{for } 0 \leq u < 1 \\ 0 & \text{for } u \geq 1, \end{cases} \quad (4.7)$$

which is classically used by the community, and the euclidean distance $\lambda_q = \|x_i - x\|$ from the evaluation point x to the q -th farthest x_i . In case $q > n$ the distance is

$$\lambda_q(x) = \lambda_n(x) \frac{q}{n},$$

with λ_n being the distance to the farthest point in x_i .

Afterwards a polynomial of degree d , the second free parameter, is fitted to the data using the weights v_i at (x_i, y_i) . The value obtained by evaluating the polynomial at x is the result of the regression routine. So, for tasks like deseasonalizing time series the routine has to be run n times at all the observation sites x_i .

The degree of the regression polynomial d is usually set to 1. Only for very complex and irregular functions a quadratic regression polynomial is used (Cleveland et al. 1990). For $d = 0$, on the other hand, the method reduces to a weighted moving average.

4.3.3. STL

The actual STL routine consists of an inner and an outer loop. In the inner one the series is first getting detrended, then all years will be smoothed using loess regression *separately*, and after all smoothed years are concatenated and a low-pass filtering is applied to prevent a power buildup in the seasonal component.

Since the linear or quadratic regression employed by the inner loop is quite sensitive to outliers or erroneous data points, the outer loop computes **robustness weights** ρ_v to reduce the influence of such points. Firstly, the residuals r_v are calculated by subtracting the trend t_v and seasonal component s_v from the dependent variable y_v

$$r_v = y_v - t_v - s_v.$$

Then, the robustness weights are obtained by

$$\rho_v = B\left(\frac{|r_v|}{h}\right), \quad (4.8)$$

with $h = 6 \times \text{median}(|r_v|)$ and the **bisquare weight function**

$$B(u) = \begin{cases} (1 - u^2)^2 & \text{for } 0 \leq u < 1 \\ 0 & \text{for } u \geq 1. \end{cases} \quad (4.9)$$

The calculated ρ_v will be multiplied with the corresponding neighbouring weights v_i during the next evaluation of the inner loop. The whole algorithm usually only needs about two to three evaluations of the inner loop in total.

4.3.4. Fourier-based methods

Since the annual cycle seems to be represented by a sinusoidal function quite reasonably, one might consider to perform a Fourier decomposition of the detrended data and use the leading frequency terms to remove the seasonality. But unfortunately a fixed wavelength for the sinusoidals is a rather bad assumption. While the approximation of the summer and winter months is quite good, this method tends to produce rather large artifacts in spring and autumn.

The reason for this problem is the onset of the transition between winter and spring or summer and autumn. Especially in the latitudes of Germany and northern Europe, where the *Ferrel cell* and the *Polar cell* of the atmosphere are switching places during summer and winter (Wallace and Hobbs 2006) this process is of rather high variance. If e.g. the onset of spring occurs more early within a year and the summer lasts especially long, Fourier-based methods tend to produce a number of artifacts in both spring and autumn, while smoothing-based methods, as mentioned above, will produce more robust results.

Also when modelling events like the sales of Christmas trees or snowfall, Fourier-based methods would need quite a number of terms to reasonably approximate the seasonal component of the time series. Therefore, most publications applying analysis to real world data prefer methods like anomalies or STL. But for the sake of completeness two of the most popular Fourier-based methods will be covered in the following subsections.

4.3.4.1. Harmonic regression

In the field of time series analysis, especially with autoregressive moving average (**ARMA**) processes (Box, Jenkins, and Reinsel 2008), the detrending and removal of the seasonal component are rather side-effects than actual goals of the analysis. Instead, the users intend to generate a parsimonious model of the series in order to make proper forecasts by taking the correlations within the data into account. A popular implementation can be found in the *ds* function of the *deseasonalize* package by McLeod and Gweon (2013) in **R**.

The basic procedure for fitting ARMA with seasonal dependence (Hipel and McLeod 1994, chap. 13) can be summarized as follows. Firstly, the data is detrended (e.g. using a moving average). Secondly, the detrended series $z_{r,m}$, including $r = 1, 2, \dots, n$ years and $m = 1, 2, \dots, s$ seasons, is deseasonalized. This is done by subtracting the mean value $\tilde{\mu}$ and dividing by the standard deviation $\tilde{\sigma}$ of each individual season

$$w_{r,m} = \frac{z_{r,m} - \tilde{\mu}_m}{\tilde{\sigma}_m}. \quad (4.10)$$

The corresponding values can be obtained by their maximum likelihood estimators

$$\hat{\mu}_m = \frac{1}{n} \sum_{r=1}^n z_{r,m}, \quad m = 1, 2, \dots, s, \quad (4.11)$$

and

$$\hat{\sigma}_m = \sqrt{\frac{1}{n} \sum_{r=1}^n (z_{r,m} - \hat{\mu}_m)^2}, \quad m = 1, 2, \dots, s. \quad (4.12)$$

Note that the term season is, again, used more generally. For an annual cycle and daily data we would have $s = 365.25$ and for a quarterannual cycle and monthly data $s = 3$. For monthly data and an annual cycle this approach thus already requires 24 parameters in order to perform the deseasonalization. For weekly or daily data it soon becomes intractable. Instead, one usually uses a Fourier series to get more parsimonious estimates for the seasonal mean and standard deviation.

The Fourier coefficients of the seasonal cycle can be determined using the estimates of eq. (4.11) and (4.12)

$$\begin{aligned} A_k &= \frac{2}{s} \sum_{m=1}^s \hat{\mu}_m \cos \frac{2\pi km}{s} \\ B_k &= \frac{2}{s} \sum_{m=1}^s \hat{\mu}_m \sin \frac{2\pi km}{s} \\ C_h &= \frac{2}{s} \sum_{m=1}^s \hat{\sigma}_m \cos \frac{2\pi hm}{s} \\ D_h &= \frac{2}{s} \sum_{m=1}^s \hat{\sigma}_m \sin \frac{2\pi hm}{s}, \end{aligned}$$

with $k = 1, \dots, F_\mu$, $h = 1, \dots, F_\sigma$, and the number of corresponding Fourier components $F_\mu \geq 1$ and $F_\sigma \leq s/2$. Now, the new estimates of the mean and the standard error can be obtained by

$$\begin{aligned} \tilde{\mu}_m &= A_0 + \sum_{k=1}^{F_\mu} \left(A_k \cos \frac{2\pi km}{s} + B_k \sin \frac{2\pi km}{s} \right), \quad m = 1, 2, \dots, s \\ \tilde{\sigma}_m &= C_0 + \sum_{h=1}^{F_\sigma} \left(C_h \cos \frac{2\pi hm}{s} + D_h \sin \frac{2\pi hm}{s} \right), \quad m = 1, 2, \dots, s, \end{aligned}$$

where A_0 and C_0 are the average values of the seasonal means and standard deviations.

To get the best trade-off between the number of parameters and the performance of the approximation, an ARMA model is fitted to the remainder $w_{r,m}$ for all possible combinations of F_μ and F_σ . An integrated component should be avoided in the presence of seasonality. In the next step, for each of the results an information criterion (McLeod and Gweon 2013), like AIC or BIC (see 5.4.1), is calculated. Finally, the model featuring the

lowest information criterion and uncorrelated, normally distributed, and homoscedastic residuals is selected. To ensure the latter point, the **Box-Cox transformation** can be applied on the data.

Another way of handling seasonality within the ARMA framework is to use so-called seasonal ARMA (**SARMA**) model (Hipel and McLeod 1994, chap. 12). But this approach incorporates the seasonal component in the model itself and is therefore not applicable for deseasonalization or decomposition schemes.

4.3.4.2. Band spectrum regression

A different way to exploit the pronounced frequencies of the seasonal component is to use **band spectrum regression** (Harvey 1978). The basic idea behind this approach is to transfer both the independent variable x_i and the dependent one y_i into Fourier space and omit all terms corresponding to the frequencies of the seasonal component and its higher harmonics.

An implementation of this procedure can be found in the descomponer package for **R**. But since the approach is very computationally demanding for longer time series, it is not applicable for the long time series handled in the context of EVA.

4.3.5. Further notes

If an even more complex algorithm is sought for, one can use X-13ARIMA-SEATS provided by the U.S. Census Bureau. It's an up-to-date seasonal adjustment software used by many officials and governmental agencies. But its mere complexity goes beyond the scope of this overview chapter.

A remaining question is how to specify the season or its length. For the climatological data at hand the obvious answer is to consider a year due to the annual cycle. But for other time series, e.g. economic ones, things might not be that obvious. In such cases, a careful study of the eigenvalues and frequency responses, which is described by Cleveland et al. (1990), might provide critical hints.

5. Non-stationary analysis in the EVA

An essential requirement, which had to be fulfilled by the data and had been introduced in the previous chapter, is its stationarity and the presence of only short-range correlations. There, we assumed the average statistical properties to remain constant over time. But is this assumption reasonable? Or do we introduce a large error by forcing a stationary model on a clearly non-stationary time series?

After all, the overall climate *is* changing and the presence of the climate change is undeniably real (see IPCC 2014). Using climate models, in which the human-made contribution to the greenhouse gas concentration in the atmosphere can be turned on and off, one can show that the increase of the mean heat content of the ocean and the mean global temperature can indeed be attributed to mankind's actions.

Another question is *how* these changes influence the *local* climate and weather of certain regions of the earth. For temperature series an almost linear trend throughout the last 30 to 40 years is quite apparent (see e.g. figure 7.1). But for precipitation data the picture is less obvious. It can be summarized as follows. The increase of the surface temperature of the ocean causes more evaporation and the amount of moisture, which can be stored in the air, increases with temperature on an approximately exponential scale (Wallace and Hobbs 2006). Therefore the overall amount of moisture in the atmosphere is rising (Trenberth, Fasulo, and Smith 2005; Trenberth 2011). But the consequences are not as simple as an increase of precipitation around the globe in an uniform way. Instead, the higher moisture levels are causing both more extreme precipitations and droughts. Which of those scenarios, if any, might apply to a specific location can only be accessed by performing a thorough analysis of local measurements and/or by inspecting the results of a regional model of the system (e.g. a regional climate model to describe weather phenomena or a hydrological model to describe the gauge and volume flow of a river). For quantities, which are not the primary variables of the climate system, like the height of a river, the wind speed at the bottom of a valley, or the temperature of a lake, reliable statements about the temporal dependence of their statistical properties is even harder to obtain (Storch and Zwiers 1999, 63–68). In chapter 6 I will demonstrate such an analysis as well as the errors one might encounter using the example of the gauge and volume flow of the river Elbe in the city of Dresden.

5.1. Non-stationarity vs. long-range correlation

While it is moderately simple to attribute the share of us humans to the climate change when comparing the results of climate models to observations, it is very difficult to pin it down when solely relying on measured time series. This is mainly due to the massive capabilities of the ocean to store heat and due to the natural variability of the climate system. Apart from the annual cycle and short-range correlations discussed in chapter 4, our climate contains a large amount of oscillation patterns, which are both longer

in time and much more variable in terms of the frequency of their occurrence. Those cycles are e.g. the El Niño and La Niña southern oscillations occurring at a time scale of several years (Wallace and Hobbs 2006), Dansgaard-Oeschger events, found in ice core data by Dansgaard et al. (1993), on a scale of 250000 years, different astronomical cycles on scales of >26000 years (Edwards 2011) and many others (Esper, Cook, and Schweingruber 2002; Jones et al. 1998; Ashkenazy et al. 2013; Budyko 1969; Benzi et al. 1982). Due to the sparsity and uncertainty of measurements reaching back more than 150 years, there might be hundreds of additional cycles we aren't aware of. But since we lack both the data to determine them and the knowledge of the intimate physical and chemical couplings in the climate system to appropriately model them, we have no way to determine them.

To ease the analysis of the climate systems facing the lack of information about many of its internal dynamics, those influences can be termed the **long-range correlations (LRC)** within the climate system or the measurement data. Approaches like the **detrended fluctuation analysis (DFA)** had been used to prove the presence of such dependencies in e.g. temperature series (see Massah and Kantz 2016). By combining all these individual effects in just one long-range correlated error one could in theory distinguish between the human-made trend and the natural variability in the data. But unfortunately a very large amount of data at a lot of different measurement sites is necessary to distinguish between those two effects (Yu et al. 2018). For single time series, like in the analysis of the river Elbe in chapter 6 one thus can only access the non-stationarities in a combined version of both the LRC and the trend introduced by the climate change.

5.2. Non-stationary modelling in the EVA

Due to the tremendous implications of the climate change and the very nature of the climate itself, the non-stationary modelling of extreme value distribution has been a central endeavour of the extreme value community since many years.

The first attempts to introduce time as a covariate of the GEV parameters were done by Smith (1986) based on the work of Weissman (1978). The modelling of the parameters of the GP distribution was covered shortly afterwards by Davison and Smith (1990). From there on the framework of incorporating covariates has gradually been improved to a point it extended the classical generalized linear model (GLM) approach (Coles 2004). With the works of Pauli and Coles (2001) and Chavez-Demoulin and Davison (2005) a smooth dependence of the GEV and GP parameters on covariates were added using spline smoothers. But apart from all these successes a general framework incorporating all these different models was still missing.

Such an unifying model class was introduced by Yee and Hastie (2003) with the **vector generalized linear models (VGLM)** for linear dependencies and by Yee and Wild (1996) with the **vector generalized additive models (VGAM)** for smooth dependencies. Once this framework was applied to the extreme value distributions (Yee and Stephenson 2007), its rigorous mathematical properties and fine-tuned fitting algorithms could be used to model all sorts of dependencies on covariates in the data.

5.2.1. VGLM

The overall goal of this section, as already mentioned, is to extend the extreme value analysis to also cope with non-stationary time series. We will do so by incorporating the explanatory variable time \mathbf{x} as covariates into the modelling of the measured data, which we will call our response \mathbf{y} in this section. This way, we will obtain a time-dependent description of the GEV and GP parameters.

Using the framework of the vector generalized linear models (VGLM) the conditional distribution of a q -dimensional response \mathbf{y} conditioned on a p -dimensional explanatory variable \mathbf{x} can be phrased as

$$f(\mathbf{y}|\mathbf{x}; \mathbf{B}) = h(\mathbf{y}, \eta_1, \dots, \eta_M). \quad (5.1)$$

$\mathbf{B} = (\beta_1 \beta_2 \dots \beta_M)$ is a $p \times M$ matrix containing the unknown regression parameters entering the VGLM, h is a known link function, and η_j are the linear predictors

$$\eta_j = \eta_j(\mathbf{x}) = \beta_j^T \mathbf{x} = \sum_{k=1}^p \beta_{(j)k} x_k, \quad j = 1, \dots, M. \quad (5.2)$$

In contrast to the older class of the generalized linear models (GLM), introduced by Nelder and Wedderburn (1972), the linear predictors can be used to directly model the individual parameters of the distribution function f , which is not restricted to be part of the exponential family of distributions anymore.

The parameters of the VGLM are fitted using their log-likelihood

$$l = \sum_{i=1}^n w_i l_i, \quad (5.3)$$

where w_i are known weights and n is the number of available measurements, in combination with an algorithm called **iteratively reweighted least squares (IRLS)**. As explained in Yee and Stephenson (2007), or in more detail in Yee (2015, 92–96), the fitting is done by constructing a dependent vector $\mathbf{z}_j = \eta_j + \mathbf{W}_j^{-1} \mathbf{d}_j$, where $\mathbf{d}_j = w_j \partial l_j / \partial \eta_j$, and regressing it over the working weight matrix \mathbf{W}_j . The latter is defined as either

$$\mathbf{W}_j = -w_j \frac{\partial^2 l_j}{\partial \eta_j \partial \eta_j^T} \quad (5.4)$$

or

$$\mathbf{W}_j = -w_j E \left[\frac{\partial^2 l_j}{\partial \eta_j \partial \eta_j^T} \right], \quad (5.5)$$

where $E[\dots]$ indicates the expectation value. Equation (5.4) results in the Newton-Raphson and equation (5.5) in the Fisher scoring algorithm, which are two classical routines to solve the Newton algorithm using maximum likelihood (see Jennrich and Sampson 1976).

5.2.2. VGAM

In the framework of the vector generalized additive models (VGAM) the requirements introduced in the VGLMs are relaxed even more. Now, the dependencies on the covariates are not restricted to purely linear but to additive ones instead and the model class is thus an extension of the class of generalized additive models (GAM) introduced by Hastie and Tibshirani (1986). The linear predictors can be modelled as a sum of a smooth functions applied to the covariates

$$\eta_j = \eta_j(\mathbf{x}) = \beta_{(j)1} + \sum_{k=2}^p f_{(j)k}(x_k), \quad j = 1, \dots, M. \quad (5.6)$$

The estimation is performed using vector smoothers.

For a thorough introduction into the model class of either the VGLM or the VGAM please see Yee (2015, chap. 3 or chap. 4) respectively.

5.3. Application to EVA

The application of the framework of both the VGLM and VGAM to model the GEV and GP distribution is straight forward (see Yee and Stephenson 2007). Using the linear predictors in either equation (5.2) or (5.6) the values of the GEV or GP parameters at all time steps i can be calculated via their dependencies on covariates in the explanatory variable \mathbf{x} . This way we obtain a series of n location, scale, and shape parameters corresponding to the individual pair of observed variables (\mathbf{y}, \mathbf{x}) , which represent measurement and time. From these values we can construct the log-likelihood of our model using equation (5.3), which will be fitted with the methods discussed in the previous section. But which of those model classes should be used to describe the data?

Both the model framework of the VGLM and the VGAM do have their merits and shortcomings. The VGAM is primarily used for visual inspection and exploratory data analysis. Its ability to fit the dependencies on covariates using arbitrary smooth functions makes it easy to pin down the actual type of dependency and to decide on a model for the fitting procedure using e.g. VGLM. On the other hand, the VGAM are not a good tool to do predictions on the parameters of the GEV or GP distribution into the future.

That's why, for a thorough analysis, especially when featuring extrapolation, VGLMs are used instead. They do not just allow to fit an arbitrary amount of possible models to describe the dependencies on covariates. They also offer a convenient way to perform statistical hypothesis tests to determine which of the fitted models is the most appropriate one to describe the data. These aspects are essential when performing an automated fitting procedure as in chapter 7. In addition, the link function h in equation (5.1) can be used to incorporate all linear constraints we reviewed in section 3.3 and is supposed to make the log-likelihood function more realistic and thus the quadratic approximation better as well as the estimated error using the delta-method more reasonable (see Yee and Stephenson 2007, 16).

When modelling the parameters of the GEV or GP distribution by their dependence on covariates one has to treat the shape parameter with extra care. It is considered to be quite difficult to estimate it from data (Yee and Stephenson 2007). That's why, it

is should almost always be modelled as a constant value with no dependence on any covariate.

5.4. Model selection

A conceptual problem one has to cope with when using the VGLM framework is how to handle the large pool of models fitted to the data. Which one of these is describing the data best? Is this model also the most plausible one? To answer those questions I will use two different approaches of model selection in this thesis.

But before we delve into the details of model selection, there is an important note to be added. The methods described in the remainder of this chapter only work on models yielding a shape parameter bigger than -0.5 . Just for those the asymptotic properties of the maximum likelihood estimators are fulfilled, which are necessary conditions for the following approaches.

5.4.1. Information criteria

The first one are the so called **information criteria**. The general idea of this approach is that fitting the same data with a model of more and more degrees of freedom will in general improve the resulting maximum likelihood but not necessarily improve “the goodness” of the fit.

Imagine a noisy signal consisting of 20 points sampled along a line. While the fit of an linear model would already yield a plausible result, we can reduce the final negative log-likelihood even further by introducing higher orders polynomials, and hence parameters, in our model. As soon as we will fit a polynomial with a degree of 19 it will perfectly describe the data. But it would be way to wiggly and complex. If we would introduce new measurements, the model would most probably not be able to describe those data well due to the divergence outside of the original range of the time interval. In other words: it would fail to **generalize**. Such a behavior is introduced when choosing an overly complex and sophisticated model to describe a small number of data points. At a certain number of degrees of freedom the model would rather fit the noise than the functional dependence hidden in the data. This phenomenon is widely known as **overfitting**. For linear models one can give a rule of thumb that one needs at least 10 times the amount of data points as the number of parameters of the model to fit them (see Abu-Mostafa, Magdon-Ismail, and Lin 2012).

In order to still fit a large amount of models into the data one can use a so-called **regularization** term. These are positive and additive terms introduced into the negative log-likelihood. The more complex the model becomes, the bigger they get.

The idea of an information criterion is to add regularization terms to the negative log-likelihood of the fitted model and to use the result to judge the quality of the fit in comparison to other models. In our example of the straight line above we would therefore calculate the information criterion for all models we fitted to the data and choose the one yielding the lowest value. An important assumption for the comparison of the different models is that they all were fitted to exactly the same data.

The most popular criteria are the **Akaike information criterion (AIC)** introduced by Akaike (1973)

$$\text{AIC} = 2k - 2 \ln(L), \quad (5.7)$$

with L being the maximum likelihood and k being the total number of model parameters, and the **Bayesian information criterion (BIC)** introduced by Schwarz (1978)

$$\text{BIC} = \ln(n)2k - 2 \ln(L), \quad (5.8)$$

where n is the number of points in the data set. While both criteria perform quite well on actual data, the BIC yields more conservative results than the AIC by preferring models with lesser degrees of freedom. This is due to the additional factor $\ln(n)$, which can be thought of as a normalization of the penalty introduced by the number of model parameters on the likelihood. In order to get some intuition, we will draw a series of 10000 points from a GEV distribution with $\mu = 0$, $\sigma = 1$, and $\xi = 0.1$. The negative log-likelihood $-\ln(L)$ is 16.4 for the first ten points, 157.2 for the first 100, and 1624.7 and 16402.5 for the first 1000 and 10000 points respectively. The penalty term in the AIC, however, will be the same for all subsets. It is of course plausible to describe larger sets of data with more complex models. But to have three parameters yielding the same ratio of penalty for 10 points as 3000 parameters for 10000 points seems not right. That's why a scaling of the contribution of k , which is smaller than the one of the likelihood itself, is included. Apart from these, there are dozens, if not hundreds, of additional criteria published.

5.4.2. Likelihood ratio test

The second approach is to construct a hypothesis test, the so-called **likelihood ratio test**, out of the negative log-likelihood values obtained for different nested models. *Nested* means that the parameter set θ_0 of the more simple model is actually a subset of the parameters of the more complex one θ_1 .

We calculate the **deviance statistic** by

$$D(\theta_1) = 2 (l(\theta_0) - l(\theta_1)), \quad (5.9)$$

where l are the negative log-likelihood values obtained by fitting an arbitrary model to the data. For large n and under suitable regularity conditions, only for shape parameters $\xi > -0.5$, the deviance $D(\theta_1) \sim \chi_d^2$ is distributed according to a χ^2 distribution with a degree of freedom equal to the difference in number of parameters of model 1 and model 0.

Using this fact we can construct the confidence intervals for a specific confidence level α . As soon as the calculated deviance in equation (5.9) is outside of these confidence intervals we have to reject the hypothesis of using the more parsimonious model, described by θ_0 , in favor of the more complex one, described by θ_1 . This way, we can test all models fitted to our data against the most simple one. If there is at least one model, for which the hypothesis mentioned above has to be rejected, it will get adopted as our new null hypothesis and we perform the likelihood ratio test again with all other models, for the which the parameters of the new base model are a subset of their own parameters. This procedure is repeated until none of the constructed hypotheses can be rejected anymore.

In my opinion the approach of the likelihood ratio test is better suitable to perform the model selection in the context of the EVA since it is quite easy to construct nested models. In addition, the hypothesis tests yield a much more solid statistical foundation than the regularization terms of the information criteria do.

6. Pitfalls of the EVA

Every method has its benefits and its shortcomings. This is also true for the methods based on extreme value theory. The benefit is of course that we are able to inspect the very tails of the distribution of a measured process. We can make estimates of the highest quantiles and their fitting errors. But when applying a rather sophisticated analysis method to measured data there is unfortunately no free lunch.

This also holds true for the analysis with the GEV and GP distribution and the price we pay is the reliability of our results. Even when avoiding numerical artifacts with the improved fitting routines introduced in section 3.3 there are a lot of additional sources of errors that can spoil our results. Therefore they always have to be treated with care.

Such additional sources are possible non-stationarities and/or correlations in the data, errors introduced by the measurement device, an insufficient length of the time series (the asymptotic regime of the GEV or GP distribution has not been reached yet), and a measurement period, which is not sufficiently long to properly capture the dynamics of the underlying system. The effect of the last source of uncertainty is the most subtle one. Therefore the remainder of this chapter will be used to illustrate its influence using a case study on the gauge data of the river Elbe in Dresden we already know from figure 2.6.

6.1. Inspection of the gauge data of the river Elbe

Since we are not dealing with simulated data but with a real-world example, it is important to have a look at the data set first.

The series in figure 6.1 exhibits a small seasonal component in the beginning of spring. This is due to the melting of snow in the mountains providing intake to the river. But the contribution of this component is rather small so I decided to analyze the raw series instead.

It was measured at the so-called Augustus bridge in the city of Dresden in Germany (see figure 6.2). The measurement station introduces a constant offset. At a gauge of 1 m the true depth of the river is 2.2 m and the station itself is located at 102.68 m above sea-level.

The assumptions needed in order to apply the EVA require the series to be both uncorrelated and stationary. Since we are only using the annual maximal values, the resulting series can be considered uncorrelated. To decide whether the series can be assumed to be stationary or not is more complicated. While the bulk of the data seems to remain more or less in the same range of values, four out of the five of the largest events occurred during the last 18 years.

To decide about the temporal dependence in the data I performed the non-stationary EVA using the class of vector generalized linear models (VGLM) (see chapter 5), constructed various models to describe the extreme events in the time series, and choose the most plausible model using both the Akaike information criterion (AIC) and the likeli-

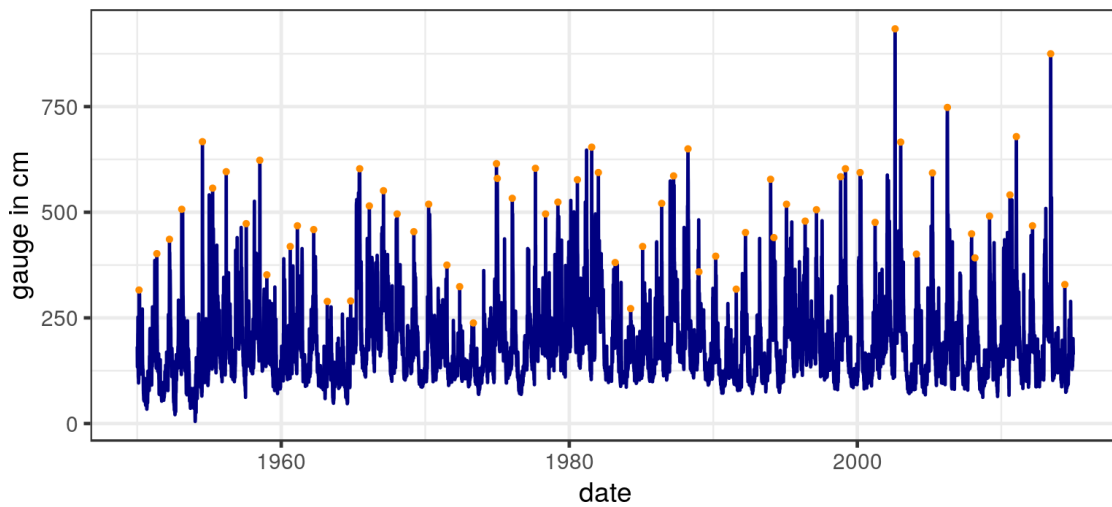


Figure 6.1.: The full series of the gauge measurement of the river Elbe in the city of Dresden is shown as a blue line. The annual maxima are marked using orange dots.



Figure 6.2.: View of the Augustus bridge in Dresden. The measurement station of the gauge series discussed in this chapter is located at the pillar in the center of the river. Photo by Dr. Brend Gross [CC-BY](<https://creativecommons.org/licenses/by/2.0/>)

hood ratio test (see section 5.4). For both measures the model containing no temporal dependence turned out to be the most plausible one leading to the decision to consider the series of the gauge as stationary. But note that what the likelihood ratio test did is not rejecting the hypothesis of the underlying data to be distributed according to e.g. a model with a linear dependence in time instead of a stationary one at a five percent confidence level. This does not necessary mean the series is truly stationary. There maybe be just too few data to proof the statistical significance of a trend in the series.

Is it plausible to consider the series of the river as stationary from a climatological point of view? Due to the climate change the mean temperature of the ocean and the air is increasing. Since the amount of moisture that can be stored by an air parcel is increasing with temperature, the amount of water in the atmosphere and thus also the total amount of precipitation is increasing. This leads, in general, to an increase of both the frequency and the magnitude of floods. But what climate analysis and simulations showed to be true for most parts of the globe does not apply to Europe (Trenberth 2011). There the total amount of precipitation for individual regions seems to remain constant. At the same time the studies show the extreme precipitations to become more extreme while the mean amount of rainfall is declining (Trenberth, Fasulo, and Smith 2005). This could serve as an explanation why four out of the five largest events occurred during the last 18 years. Nevertheless, we will describe the series as stationary according to the findings of the VGLM analysis.

6.2. Exploration of system dynamics

The tricky thing about river data is that the height of the river does not necessarily scale linear with its volume. For small gauges the river basin often can be approximated by e.g. a parabolic shape. But as soon as the height of the river overpasses the boundaries of the river basin and floods the neighbouring regions, the scaling becomes much more complicated. A tiny increase in the gauge can correspond to a tremendous increase of the overall volume of the river.

Apart from the height of the river its flow velocity is usually measured as well. But this quantity may yield an even worse proxy of total volume with respect to flooding events. Imagine the river carries so much water it floods the neighbouring fields. While the flow velocity is still as high as if the basin would be completely filled and no flooding occurred, the water on the fields is barely moving. The water masses in both the basin and the fields could be approximated as separate entities with a small coupling. But since we only measure the flow velocity at a single point in the original river basin, we would introduce a large systematic error every time a flooding event occurs. Instead, we will use the gauge. It yields a far more stronger coupling of the water masses.

The dependence of the total volume of the river on its gauge itself does not cause any trouble as soon as the basin of the river and the neighbouring regions were not altered and can be considered stationary. The gauge represents still a stationary stochastic process and can be described by the extreme value theory. It is the extrapolation of the return levels to values not or barely seen in the data that becomes very complicated. Let's do a second thought experiment. Imagine we are facing 40 years of data of a river, which floods its vicinity on average once every 50 years. In the data we were provided

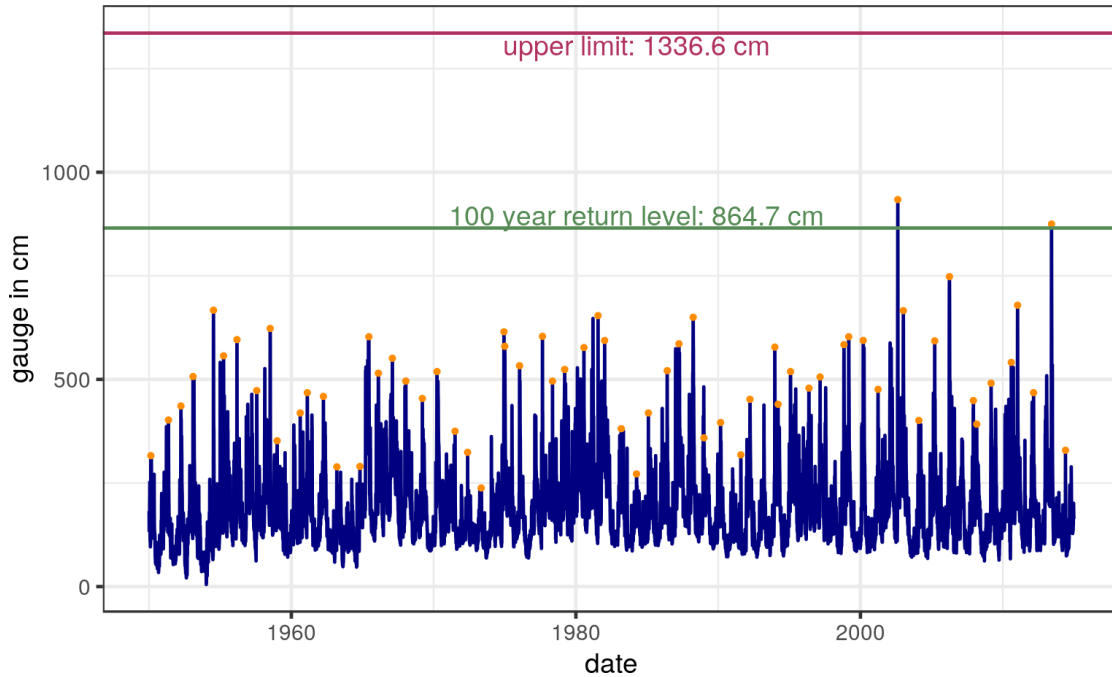


Figure 6.3.: EVA applied to the full series of the gauge of the river Elbe. Using the maximum likelihood estimates (MLE) of the parameters of the stationary GEV distribution both the 100 year return level and the upper limit of the Weibull-type distribution were calculated and marked.

with none of such a flooding events is present and we have the task to estimate the height of a flood occurring on average once every 300 years. Since our approach is data driven, the distribution we fitted did only see events within the approximately parabolic basin. When extrapolating to unseen values by calculating the return levels we will now most probably miss the gauge by a margin since the fitted model has no idea about the upper end of the basin and the change in scaling behavior. The only way to properly capture this scaling is to incorporate flood events into the fitted data. Therefore, the time series containing only 40 years had not explored the dynamics of the underlying system sufficiently in order to enable us to perform predictions using the EVA.

Let's verify these remarks using the gauge series of the river. Firstly, we will analyze the whole series and calculate both the 100 year return level and the upper end of the GEV distribution (since it will be of Weibull-type). The results in figure 6.3 look plausible. The 100 year return level is exceeded twice in this time series containing 65 years of data but the flood in 2002 was the largest one in the history of Dresden and since the return levels are average statistical quantities there is no problem in having a couple of such events even one after another. Reviewing the estimates for the fitting error of both quantities in table 6.1, which were obtained using the Monte Carlo method introduced in section 2.2.3.2, one can clearly rule out any proper usage of the upper limit. Since its error is even larger than its actual value, there are no conclusions to be drawn from it.

Now, let's pretend it is the year 2001 and we have to perform the same analysis in order to estimate the return level, which may be used to construct a dyke, and the

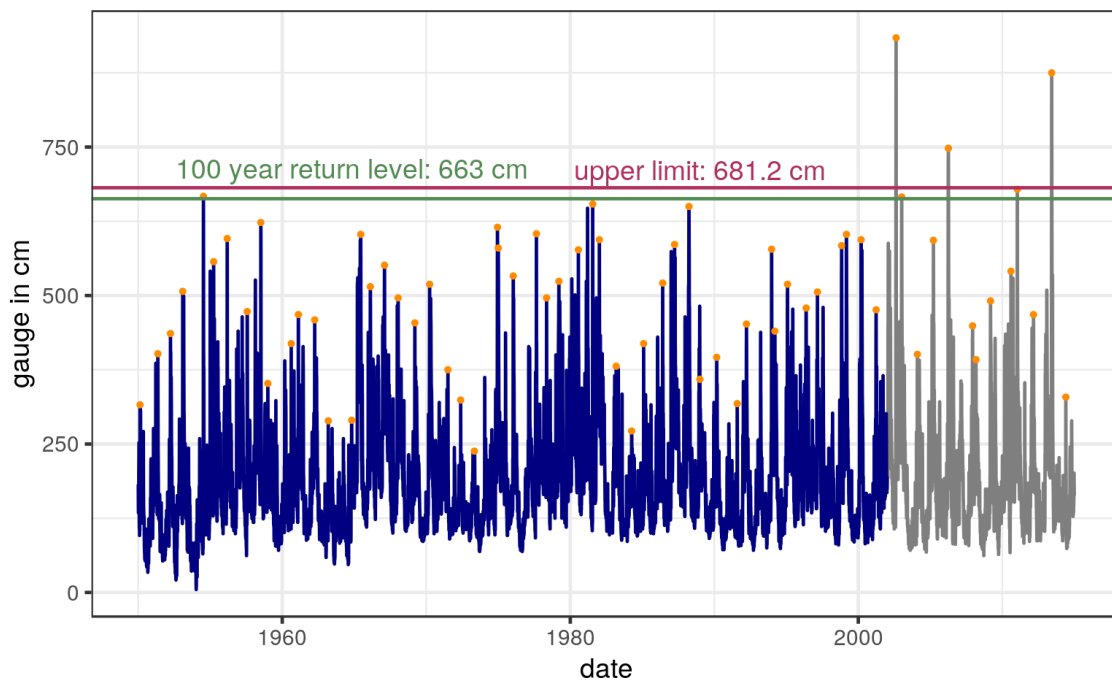


Figure 6.4.: Similar analysis as in figure 6.3. But in here only the data before 2002 was used. All data plotted in grey was excluded from the fitting of the GEV distribution. This results in major differences between the obtained 100 year return level and upper end point of the Weibull-type distribution.

upper limit. As we can see in figure 6.4 both the return level and upper limit again look plausible facing only the data prior to 2002 but totally fail to describe the distribution of the data in the upcoming years. Especially the exceedances of the upper limit, even after considering its fitting error, at three points in time are quite bothering.

Why do the results differ so much? In table 6.1 you can find the GEV parameters, return levels, and upper limits estimated for both the full series and the series before 2002. Including the last 18 years did only have a small effect on both the location and scale parameter. But the shape parameter did change quite drastically. This is due to its sensitivity to events in the tail of the distribution of the annual maximum values. Including or excluding the largest or both the largest and second largest event results in a large difference of the values of the shape parameter and thus also alters the return level and upper limit as well as their error estimates by a margin (see section 3.3).

Another question, which immediately comes to mind, is whether the results of the full series can be trusted. Maybe an even larger event will occur during the next years and our analysis will again be spoiled. I do not think this will be the case. At several locations in Dresden and Saxony in general one can find marks of the peak flood levels of events throughout the last couple of centuries. Although these values might be inaccurate and not suited for incorporation into the analysis itself, they quite nicely show the estimated 100 year return level to be of a plausible value. The actual magnitude will very likely change when incorporating more events sampling the tail of the annual maximum values but the change in GEV parameters will be small compared to the one between the full

Table 6.1.: The MLE of the GEV parameters were fitted to either the full gauge series or only to the data before 2002. In addition, the 100 year return levels and the upper end points of the Weibull-type distribution have been calculated from those parameters and the fitting errors of all quantities have been estimated using the Monte Carlo method. All values, except of the dimensionless shape parameter, are given in cm.

	full series	series till 2002
location	449.99 ± 17.4	460.96 ± 19.15
scale	121.53 ± 12.1	119.40 ± 11.81
shape	-0.14 ± 0.1	-0.54 ± 0.09
100 year return level	864.65 ± 61.4	663.00 ± 11.45
upper limit	1336.56 ± 1896.7	681.19 ± 15.48

series and the series before 2002.

6.3. Summary of the pitfalls in the EVA

By naively applying the EVA to a time series one might obtain results that are totally wrong. Even if the series itself is both uncorrelated and stationary and the size of the blocks is large enough to sufficiently approximate the asymptotic limit of the extreme value theorem. Therefore one has to use all sorts of additional information, like sparse observation data from earlier time, simulation data, and/or physical insights into the system the analysis is applied to. If it shows dynamics on time scales larger than the length of the time series itself, the block maxima might not sufficiently sample the tails of the PDF and thus will lead to erroneous descriptions of the higher quantiles by the EVA.

7. Analysis of the temperature and rainfall data provided by the DWD

The main objective of this thesis is to study the change of extreme weather events in time. The question I want to answer is not *why* there are changes in the climate but rather *if* and *how* it does change. To give statistically sound answers, one needs a lot of data determining the state of the climate at both present and past times. On top of that, the extreme value analysis with its peak over threshold and block maxima method is especially data consuming. That's why we need our time series to be as long as possible. This requirement restrains us from using satellite data, which only became available in reasonable quality after the launch of the TIROS-N satellite in 1978 (Dee et al. 2011).

The data source used within this chapter is the vast archive of measurement stations provided by the German weather service (DWD) (Deutscher Wetterdienst 2018). This comprehensive collection features 1082 different stations, each containing recordings of various climatological quantities, and does approximately cover the area of Germany uniformly. Our analysis will therefore be restricted to observations within Germany and our conclusion may not hold for other countries within Europe, let alone the world.

To nevertheless embed our local findings into the global climate system, chapter 8 will contain an analysis similar to the one in this chapter performed on the ERA-Interim reanalysis data set.

Of all climatological quantities measured at the different observation stations three distinct ones will be analysed and discussed: the daily maximum and minimum temperature and the daily total precipitation. The daily maximum and minimum temperature yield the range of the values realized throughout a single day. It does not have a sufficient temporal resolution to gain insights into the particular dynamics and fluctuations of the temperature in the atmosphere. But it is perfectly suited to draw conclusions about the tails and quantiles of the temperature distributions. The daily total precipitation is the height of rainfall accumulated over a time period of one day. The distribution of this quantity differs a lot from the one of the temperature. While daily, monthly, hourly, or minute-wise temperature measurements are all sampled from one (possibly non-stationary) climatological distribution, precipitation data of different time scales can not be described by a single function. The shape and type of the distribution is depending on the time the rainfall gets accumulated in and is not invariant under transformation of scales. All the conclusions we draw for the right tail of the daily total rainfall will thus only hold for an accumulation window of approximately one day. The hourly or weekly total precipitation might show a completely different behavior.

The comparison of the observation data to the global ones in chapter 8 will only be done for the temperature data. While these are obtained using direct observations entering both the atmospheric model and the adjustment of the forecast step (see section 8.1), rainfall is only a derived quantity. This means that the measurements of the daily

total precipitation do not enter neither the forecast nor the adjustment step. Instead, the atmosphere is getting initialized using, among others, temperature and air pressure measurements and the model simulates its dynamics for over a day. If some clouds did form and the right conditions were met, the forecast step will return a quantity representing the daily accumulated rainfall. But the description of especially cloud formation and dynamics is still an ongoing and very difficult area of research (since it happens on regional scales similar to the resolution limits). The errors of the simulated precipitation are therefore far to large to allow an integration into our analysis.

To obtain insights into the temporal evolution of extreme weather events in the climate change we will perform two different analysis. In the first one, which we will call the *time window* method, the annual maxima of the temperature or precipitation data (or annual minima for the daily minimum temperatures) get extracted and split into time windows of equal size. Within those windows a stationary EVA discussed in chapter 4 is applied and the results of different windows are compared afterwards. The second method is based on the vector generalized linear models introduced in chapter 5 and will be thus called the *VGLM* method. It uses the block maxima/minima of the whole series and performs a non-stationary EVA by fitting a GEV distribution with a linear trend in time to both the location and scale parameter. Using a model with no temporal dependence at all as a baseline the presence of trends in the extreme events of the data can be checked by a hypothesis test called the likelihood ratio test (see section 5.4.2). The benefit of the VGLM approach is that we can make quantitative and statistically sound statements about the existence of possible non-stationarities in the data. But at the same time its first order linear models are only a crude approximation of the true changes of the tails over the last 60 years. In terms of interpretability and explanatory power the more qualitative approach of the time window method is better suited to make statements about the particular nature of the changes in the underlying climatological system. In addition to the EVA the change of the overall distribution of the observed quantity is analyzed using the time window method as well. This will be done to get some general insights into the temporal changes of the climate system as a whole in order both to find reasons for the changes in the extremes and to perform a comparison between the two.

7.1. Preprocessing of the DWD data

The length of the time series used within this analysis, as already has been hinted, is set to 60 years. This seems to be the ideal number as we trade the maximum possible time series length off against the maximum number of involved stations. Only 199 stations out of the 1082 in the data base of the DWD do have a length of at least 60 years. A number that decreases rapidly the more years we require. Considering the length of an individual series 60 years is already sufficiently large. It spans two distinct periods of 30 years, which is the default time period to define climate in atmospheric sciences. An even more important reason to use at least 30 years of data within one time window stems from the maximum likelihood fit of the GEV distribution. In the stationary EVA, applied within the individual time windows, a three parameter GEV distribution is fitted to the data. As a rule of thumb (Abu-Mostafa, Magdon-Ismael, and Lin 2012) at least ten times more data than the number of parameters should be used for the fitting procedure. Bigger

time windows would improve the estimation from a numerical point of view but will be more likely to violate the assumption of stationarity within the window and drastically decrease the number of involved measurement stations.

We already established that the description of the temporal evolution of the GEV parameters by a first order linear model is a rather crude assumption. But an analysis featuring only two time windows wouldn't be any better. Instead, we introduce a third time window having an overlap of 15 years with the other two non-overlapping windows as well. This will be done to both perform analysis of the temporal evolution in more detail and to verify the underlying assumptions of the VGLM approach. If the changes in the parameters do not occur in a consistent way, their descriptions with first order linear models can not be justified. On the other hand, we could also use more complex models, like linear models of second or third order. But the first order approach with a linear trend in time for both the location and scale parameter, while keeping the shape parameter constant, does already contain five different coefficients to be fitted to the data. With only 60 data points at hand we are already on the brink to the realm of overfitting. And indeed, comparing the fit results with more complex models via information criteria (see section 5.4.1) and the likelihood ratio test (see section 5.4.2) yielded no significance to use them over the first order ones.

In order to compare the results of the different stations, an additional constraint has to be introduced: all series have to be complete and cover the same years from 1957 till 2016. The station data of the DWD is considered to be of high quality and extensive temporal coverage. Nevertheless, some of the series contain missing values ranging from single days to a couple of years. This can introduce two different types of errors. Firstly, if only a subset of data is available within a year, the extraction of the minimum or maximum value in the GEV analysis will introduce artifacts and possibly spoil the analysis. Imagine only the first two months of a temperature series being present for a particular year. The maximum value extracted from those winter months will be by no means an approximation of the true annual maximum (at least not at the latitudes of Germany). Thus, including this artifact in the analysis will most likely distort the estimation of the GEV parameters, especially of the shape parameter, which is very sensitive towards outliers. Secondly, if a couple of years are missing in one of the data sets, the obtained results of the analysis will lose their comparability with other stations to some extent. By requiring all series to have exactly the same length and allowing them to cover different years, some might include or miss a year with especially extreme weather events. This point, taken from the very tail of the distribution, will largely affect the value of the shape parameter, which has been shown in chapter 6. But since not all series do share the event, this mismatch in the temporal alignment introduces an additional source of error to the analysis, which can be circumvented at the expense of a smaller number of stations at hand. In the resulting set of complete and temporally aligned stations spanning the years between 1957 and 2016 there are 74 series for the maximum daily temperature, 75 for the minimum daily temperature, and 57 for the daily total precipitation.

In addition to the raw maximum and minimum temperature series the seasonal cycle will be removed by calculating their anomalies (see section 4.3.1). This will always be performed on the whole time series and thus prior to the separation into the distinct time windows. Both quantities, the raw temperature series and the anomalies, provide

two different views on the temperature distribution. The extreme events in the raw series correspond to exceptionally hot days in summer or cold days in winter. While just a minor detail considering the overall climate, those extreme temperatures do have a significant impact on the human life, economy, and health. A large body of literature is devoted to the investigation of their causes and implications (Christidis, Jones, and Stott 2014, @Kamae2014, @Unkasevic2005, @Schar2004, @Schar2004a). The extremes in the temperature anomalies, on the other hand, are not of big concern to the society itself. A day in January being 15 degrees warmer than usual might be welcomed the freezing people or an exceptionally cold day in July might be considered refreshing. But since the anomalies represent the deviations from the mean climatology their extreme events do tell us a lot about the variation of the climate system itself. Especially since they are sampled throughout the whole year and not just in the warmest or coldest months as for the raw data. The precipitation series do have a seasonal cycle as well. But it is way less pronounced and thus not treated separately.

When analyzing the data of the many different stations and e.g. plotting all their mean values into a single distribution we have to ask ourselves about the correlations between series. Since we assume them to be spatially pointwise measurements of a shared overall climate within Germany, it is save to assume that they are correlated. Then again, they show distinct and sometimes even opposite behaviors. In an ongoing project with some colleges we found the series of the daily maximum temperature anomalies of the DWD to be best modelled with a shared trend, a shared long-range correlated noise, and an individual Gaussian white noise with parameters specific to each station (Yu et al. 2018). These correlations prevent us from discussing the changes of the nationwide temperature distribution by lumping together all the different measurements and estimating a single PDF for each climatological quantity. Instead, the first four moments of each individual series will be calculated within the different time windows and the temporal evolution of the distributions of those moments will be used to draw conclusions about the changes in the climate within Germany.

Apart from the preprocessing applied to the data itself, a critical part of the EVA is to decide whether to use the GEV distribution and the block maxima approach or the GP distribution and the threshold exceedances. For our analysis I chose the GEV distribution because both the extraction of the annual maximal/minimal values and the fitting itself can be easily automated. The extraction of the threshold exceedances, on the other hand, requires a manual setting of the threshold height. Both the normalization of the data and the usage of a high quantile (e.g. 0.975) as default values have proven insufficient in previous analysis.

7.2. Structure of the analysis

The first part of the analysis is covering the daily minimum and maximum temperature series using both the anomalies and the raw series. The second, much smaller, one will discuss the annual maxima of the daily total precipitation.

The analysis of all the measurement stations at once is a great opportunity to gather insight into the changes of the overall climate within Germany. But we are also bound to loose some detailed information about the individual stations in the process of spatial

coarse-graining. The analysis of the individual parts, first the temperature anomalies, then the raw temperature series, and finally the precipitation data, will therefore start with a case study. I chose the Potsdam measurement station since its old measurement equipment is still available and thus its calibration is well-checked and considered to be of especially high quality. Since it is located outside of the city of Potsdam it is not affected by the heat island effect.

Afterwards, the whole ensemble of series will be analyzed and compared to the results obtained for the Potsdam station. Although this is a central part of this thesis, the investigated quantities (first four moments and the three GEV parameters plus two return levels for the daily maximum/minimum raw temperatures/anomalies or rainfall) are just too many to present them in this chapter. The proper analysis of the DWD series was therefore moved to the appendix A and only some highlights and important results are mentioned in here. Since, as discussed above, the overall distribution of the climatological quantities can not be accessed by estimating the PDF of all observed data at once due to the spatial correlations between the individual stations, the temporal evolution of the first four moments are analyzed instead. These will serve as proxies for the overall shape and behavior of the climate in Germany and can be interpreted as follows. An increase in just the *mean* corresponds to a shift of the distribution towards higher values while its shape remains unchanged. A larger *variance* reflects a broader distribution while both its symmetry and the ratio of mass between its bulk and tails stays the same. The *skewness* is a measure of symmetry of a distribution and positive values indicated that the right tail is fatter than the left and vice versa for negative values. Thus, an increase in skewness indicates an increase in mass of the right tail with respect to the left. Since only the right tail of the GEV distribution fitted to the daily maximum temperatures and daily total precipitations and only the left tail for the daily minimum temperature are of interest in the EVA, one has to be careful with the interpretation. Larger values of the skewness for the daily minimum temperatures lead to a lighter left tail of the overall distribution. The right tail of the daily minimum temperatures is of no interest in neither the analysis of the extreme events nor the description of the climate itself. The *kurtosis* comes in different flavors. In this analysis Pearson's measure of the kurtosis will be used, which is the fourth moment of the distribution divided by the fourth power of the standard deviation. It can be understood as a comparison between the probability mass of the tails and the bulk of the PDF. For a Gaussian distribution it would have a value equal three. The higher the kurtosis, the sharper the peak of the distribution, the fatter its tails, and thus the slower their decay towards infinity gets. For smaller and smaller values of the kurtosis the distribution does approach the uniform distribution. An increase in kurtosis will therefore correspond to an increase in mass in the tails of the distribution and to larger extreme events at a given return period. While these interpretations are well suited for the analysis of the temperature anomalies and precipitation data they might have less meaning for the raw temperature series, which feature a bimodal distribution. But since the analysis of the overall distribution is mainly serving as an appetizer and source of interpretation for the results of the EVA, no additional measures will be introduced in order to analyze the raw temperature series.

The distributions of the first four moments, colored according to their time window, will be presented in a single figure. To ease the comparison of the Potsdam station with the overall statistics of the German climate, the results of the Potsdam series are

marked with vertical lines. Since the changes in the climate will most probably not occur uniformly in space, both the values of the last time window and the difference between the third and first time window will be plotted on a map. This way possible patterns could be found, which would else remain hidden in the distribution of the moments.

After the analysis of the overall distributions the EVA is performed using the time window approach. Firstly, the temporal evolution of the fitted GEV distribution (Potsdam) or the distributions of the fitted GEV parameters (Germany) are discussed. Afterwards, again, the values of the third time window and the changes will be plotted on a map. Last but not least, the significance of the trend in the extreme events is checked using the VGLM approach and the results are displayed using a map too.

When interpreting the colors of the maps please keep the following in mind. In order to be as consistent as possible but at the same time to allow the colors to be as meaningful as possible, I kept them fixed to a certain degree. The value of the third time window will always be displayed as the inner color of the circle. If it is spanning both negative and positive values, yellow will be the lowest, red the highest one, and orange will correspond to zero. For only negative values yellow will, again, be the lowest one and orange corresponds not to zero but to the largest value present in the data. For only positive values red is the largest and orange the lowest value but not necessary zero. In addition, in some maps, spanning both negative and positive values, bright yellow or deep red might be absent. This is because the color scales between yellow and orange or red and orange are always symmetric. Thus, if the biggest absolute value is 0.7 and corresponds to an observed -0.7 colored in bright yellow, the value corresponding to dark red will also be set to 0.7. Now, if the largest observed positive value is no larger than 0.2, the color range will be already exhausted at a reddish orange. But now the reader can easily look at different pictures without getting fouled by the color scales. The same holds for the difference between the third and first time window, which will be always plotted as the outer color of the circle. But in contrast to the inner one the highest value in here will always correspond to a dark green, the lowest to a dark red, and zero change will always correspond to white.

In addition to multiple colors there will also be two subplots in each figures of the temperature analysis. They will correspond to the analysis of the daily maximum and minimum temperatures. For plots containing time series the daily maxima are always plotted above the minima and for distributions and densities the minima are always left of the maxima. In order to avoid any confusion orange strips were added to all plots labeling what is depicted there.

7.3. Temperature anomalies of the Potsdam station anomalies

As outlined in section 7.2 we will start the analysis of the DWD temperature series with the case study of the Potsdam station.

7.3.1. Time series and overall distribution

We begin with an inspection of the daily temperature anomaly series in figure 7.1. The range of temperature values densely filled by the daily maximum anomalies (upper plot) is wider than the one of the minima. This indicates a larger variance of the daily maxima.

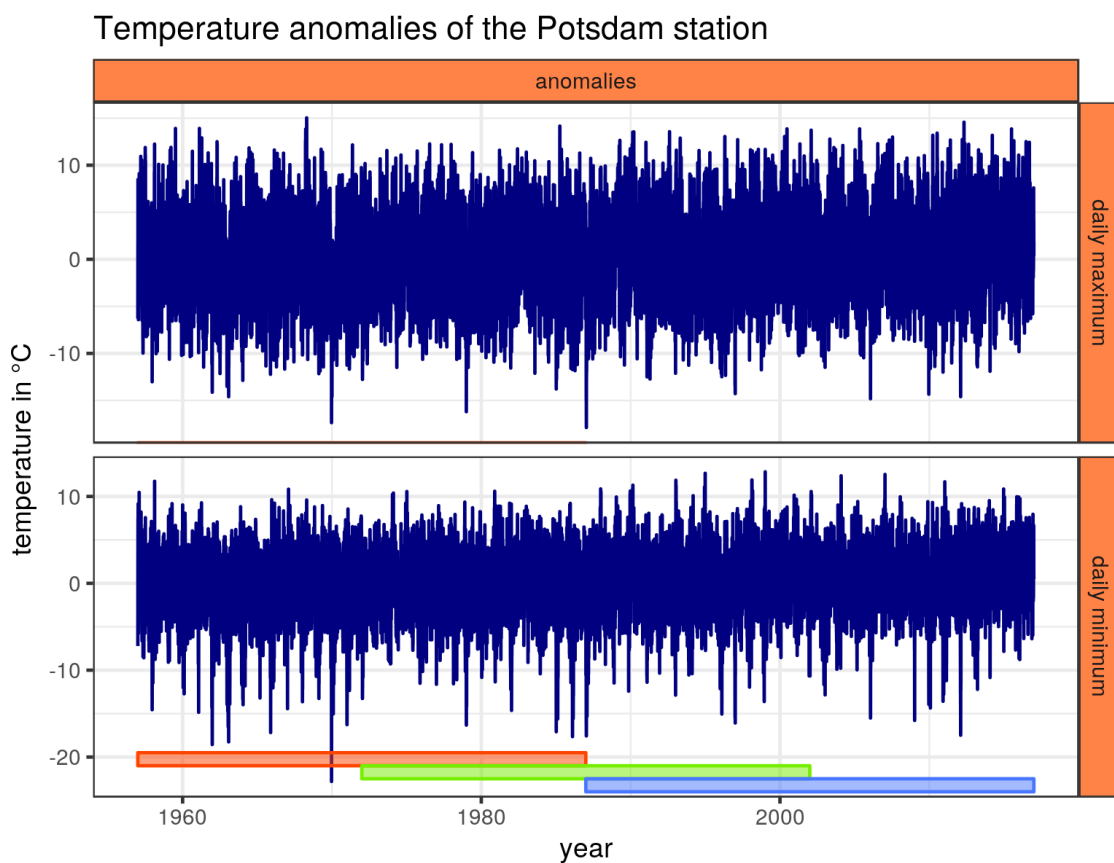


Figure 7.1.: Plot of the daily maximum and minimum temperature anomalies of the Potsdam station in Germany. Displayed are 60 years of data, from 1957 till 2016, which are required to be present in all time series analyzed within this chapter. In addition, the three time windows of the corresponding analysis are depicted using colored bars in the bottom of the figure. The first window (red) reaches from 1957 till 1986, the second one (green) from 1972 till 2001, and the third one (blue) from 1987 till 2016.

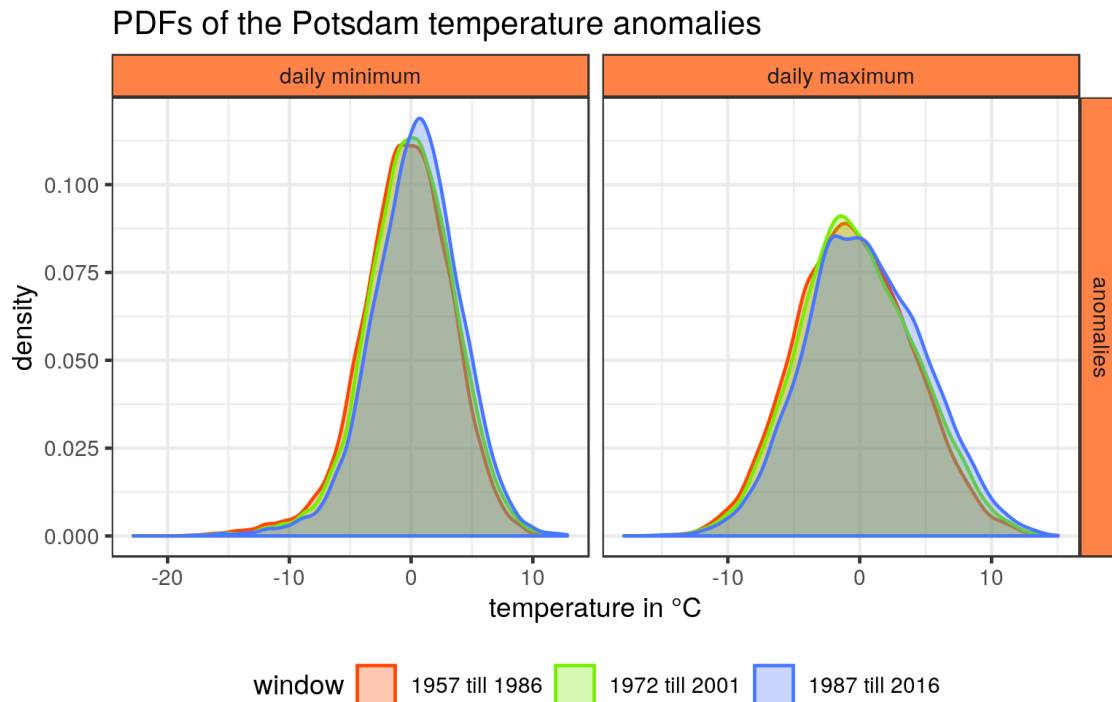


Figure 7.2.: Probability density functions (PDF) of the Potsdam temperature anomalies obtained using kernel density estimators. For the individual densities only measurements within the corresponding time windows of figure 7.1 have been used.

On the other hand, the daily minima do have a larger overall range of observable values and a pronounced asymmetry of the tails. The spikes towards negative temperatures are more than double the size of the ones towards positive temperatures. Upon closer inspection those large spikes towards negative temperatures all occur in winter months with three exceptions, one in mid November and two in the end of March. This is due to the heteroscedasticity of the temperature in the latitudes of Germany, where the variance in winter is larger than in summer.

In both the minima and maxima a small positive trend can be found. In addition to a mere shift of the time series the daily minima seem to have less spikes in negative and more in positive direction in the third time window (blue) compared to the first one (red).

To further investigate the temporal evolution of the series' behavior, its PDF and its first four moments are estimated within the three time windows represented by colored bar at the bottom of figure 7.1. Figure 7.2 shows the PDFs obtained using a kernel density estimator plotted on top of each other, with their colors corresponding to the ones of the particular time window, in figure 7.1. The numerical estimates of the first four moments can be found in table 7.1.

They revealed a consistent shift of the PDF towards higher temperatures. In total, the daily minimum temperature anomalies increased by 0.82°C and the maxima did show an even stronger increase by 0.88°C . A note of caution when interpreting these and the following numbers. They refer to e.g. changes in the mean temperatures of the first 30

Table 7.1.: Numerical estimation of the first four moments of the Potsdam temperature anomalies in the different time windows. The windows correspond to the colored bars at the bottom of figure 7.1.

window	mean	variance	skewness	kurtosis	type
1957 till 1986	-0.41	13.76	-0.51	4.14	daily minimum
1972 till 2001	-0.04	13.31	-0.34	3.88	daily minimum
1987 till 2016	0.41	13.11	-0.37	3.97	daily minimum
1957 till 1986	-0.44	19.15	0.10	2.81	daily maximum
1972 till 2001	-0.13	19.54	0.12	2.79	daily maximum
1987 till 2016	0.44	20.18	0.09	2.79	daily maximum

years compared to the ones of the last 30 years and thus underestimate the absolute increase between the first and last year of the series. An equal mean with opposite signs in the first and third time window (as in table 7.1) is a feature by construction when using anomalies.

In the variances of the densities a large asymmetry can be found. The variance of the maxima is almost twice the one of the minima. In addition, they do show opposite trends. While the variance of the minima is getting smaller, the one of the maxima is getting bigger. The daily minima can be used as proxies for temperatures at nighttime and the daily maxima for temperatures at daytime. This matches the perceived feeling that the variability of the climate does increase in time.

Apart from the variance the densities do also differ in their overall shape. The maxima do have a positive skewness and thus a slightly heavier tail towards higher temperatures. The minima, on the other hand, do have a negative and a much more pronounced skewness as well as a larger kurtosis. This results in a far stronger asymmetry and much heavier tails towards of the density towards lower temperatures. The latter corresponds to the major spikes towards very low temperatures occurring in winter. All these findings also indicate that the overall temperature PDF does have heavy tail towards both high and low values, with the one to the left being slightly heavier than the one to the right.

Considering the temporal evolution of the shape of the distributions, the density of the minima becomes much more sharp and focused. At the same time, the pronounced peak of the maxima obtains a bimodal structure. Then again, the skewness and kurtosis of both quantities do remain approximately constant, aside from a lower skewness in the first time window of the minima. In addition, the changes in the third and fourth moment do not appear to be consistent between the different time windows and might just be due to statistical uncertainties. The climate change thus seems to be most pronounced in the mean and variance of the temperature anomalies.

7.3.2. Non-stationary EVA and significance

After a discussion of the general changes in the daily Potsdam series we will now perform both the qualitative time window as well as the quantitative VGLM analysis. Using the obtained results we will identify if and how the extreme temperature anomalies did change over the past 60 years.

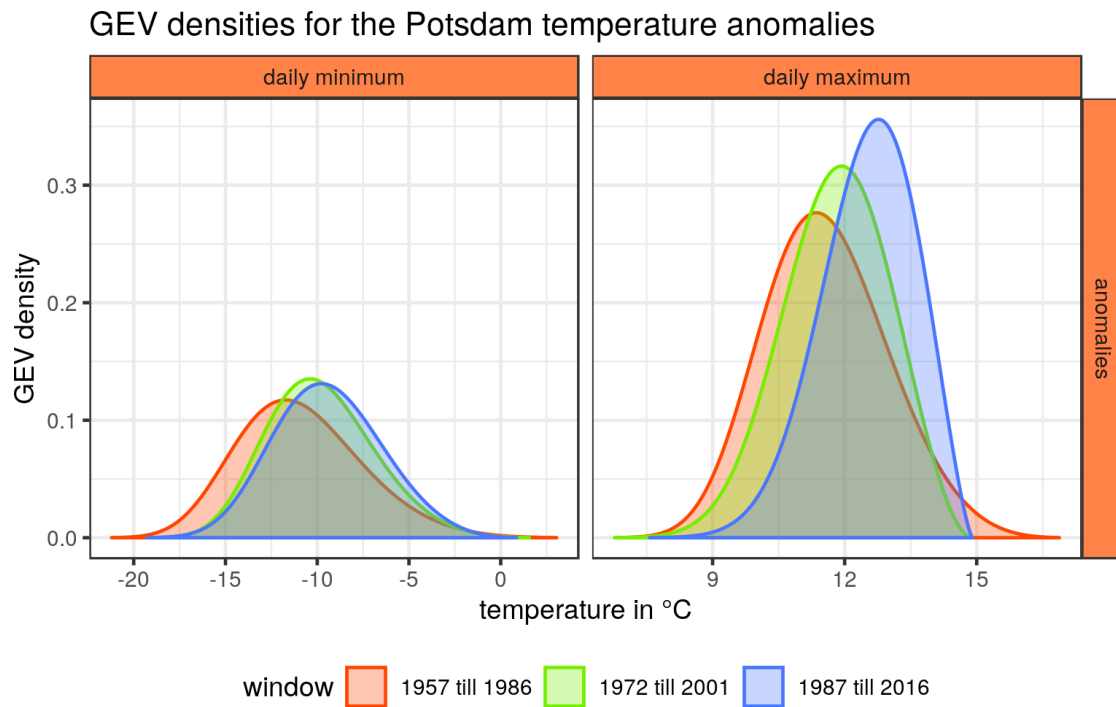


Figure 7.3.: GEV densities constituted by the parameters fitted within the different time windows. To generate this figures, firstly, the annual maxima/minima of the daily maximum/minimum temperature anomalies (see figure 7.1) were extracted. Afterwards, the series were split into three time windows and a stationary GEV distribution was fitted using all block maxima/minima within the individual windows.

Table 7.2.: GEV parameters and some return levels for the Potsdam temperature anomalies within the three time windows. All quantities except of the shape parameter are given in degree Celsius and all error estimates are calculated using the Monte Carlo method.

window	location	scale	shape	10 year ret	100 year ret	type
1957-1986	-12.28±0.59	3.19±0.48	-0.17±0.15	-18.26±1.07	-22.53±2.70	min
1972-2001	-10.95±0.62	2.77±0.41	-0.19±0.16	-16.04±0.96	-19.48±2.27	min
1987-2016	-10.51±0.68	2.89±0.41	-0.23±0.16	-15.58±1.28	-18.68±2.44	min
1957-1986	11.04±0.32	1.36±0.22	-0.21±0.15	13.48±0.42	15.04±0.82	max
1972-2001	11.41±0.27	1.25±0.17	-0.36±0.15	13.34±0.27	14.23±0.40	max
1987-2016	12.20±0.25	1.15±0.16	-0.43±0.15	13.87±0.22	14.52±0.28	max

We start with the time window approach and fit the stationary GEV distribution to the annual maxima of the daily maximum temperature anomalies or to the annual minima of the minima within the three time windows. The resulting distributions are displayed in figure 7.3 and the corresponding GEV parameters can be found in table 7.2. Note that we will fit GEV parameters to the block maxima/minima using the maximum likelihood method discussed in section 3.3. The distributions in figure 7.3 are not estimated from the histograms of the block maxima/minima but are the true GEV distributions corresponding to the estimated parameters, which are given in table 7.2.

Similar to the mean values of the overall PDFs, the location parameters do consistently shift towards higher temperatures too. For the block maxima of the daily maximum temperature anomalies, which will be abbreviated to *daily maxima* in the remainder of the EVA analysis, a shift of 0.76 °C was found. This is comparable to the 0.88 °C found for the mean value. The increase in location of the daily minima of about 1.77 °C is more than double the increase of the corresponding mean value. This already hints that the changes in the tails of the overall temperature distribution won't be summarized by mere shift towards higher values but instead a fairly complex and asymmetric one.

The scale parameters are getting smaller. For the daily minima a minor increase between the second and third time window can be found but taking the first time window into account there is still a general downward trend. This is in contrast to the findings for the overall variances of the anomaly series, which decreased for the daily minima but increased for the maxima. Therefore, the general variability of the daily maximum temperature anomalies is increasing while the variations in the large deviations from the mean climatology are getting smaller. These findings match the decrease in kurtosis found for the daily maxima. In addition to the temporal evolution also the relative size of the variability of the daily maxima compared to the minima differs between the extreme events and the bulk of the data. For daily minimum temperature anomalies the scale parameter is double the size of the maxima's one. For the variance of the overall densities the opposite behavior was found.

Both the daily minima and maxima are of Weibull type and their shape parameters are getting even more negative. This corresponds to a lightening of the tails. In addition, both of them are bounded and the biggest/lowest possible value is coming closer the lower the shape parameter gets (for a shape of zero the bound is moving towards infinity

and the distribution gets of Gumbel-type). But due to the findings in chapter 6, we should not trust the implications of this upper/lower bound, especially not when it is obtained from as few as 30 data points. That's why they won't be reported or discussed in the remained of the analysis anymore.

While the GEV parameters are fitted to the block maxima/minima it is the return levels we are most interested in the EVA. They represent the very large quantiles of the original distribution and provide insight into possible changes in the extreme weather events. For the daily minima a pronounced increase in return level at both the 10 and 100 year return period could be found. Thus, there will be less extreme deviations towards cold temperatures. The maxima show a slight increase in 10 year return level and first a decrease but afterwards a slight increase in 100 year return level. This rather complex behavior is most probably caused by several ingredients. The increase in location of both the daily maxima and minima results in a general gain in return levels. The decrease in scale, on the other hand, yields smaller return levels, and, finally, the more negative shape parameters cause the return levels to shrink in size as well. The difference in shape of the daily maxima between the first and second time window is larger than between the second and last window and the location parameter exhibits the opposite behavior. We can thus see a decrease in return levels at first and an increase thereafter. This behavior is more pronounced in the 100 than in the 10 year return level since the increase in location contributes a constant shift to the return level but the influence of the shape parameter increases as the return period becomes larger and larger (see equation (2.1.1)). In the 10 year return level it still manifests with first a small and then a larger increase in return levels. For the minima the picture is more simple since the increase in location and the decrease in scale and shape cause all return levels to decrease as well.

Following the qualitative time window analysis we perform the more quantitative VGLM approach. Here, we fit all of the block maxima/minima using a stationary GEV distribution and one with a linear trend in both the location and the scale parameter. Afterwards, we will use the so-called likelihood ratio test to check which model, the stationary or non-stationary, is more likely to describe the observed data. Our null hypothesis will be the one of the stationary model and we use a 5% significance level to test whether this hypothesis can be rejected in favor of the more complex non-stationary model. For a detailed description of the non-stationary analysis include the class of vector generalized linear models (VGLM) and the likelihood ratio test please see chapter 5. The return levels and the result of the hypothesis test are shown in figure 7.4. It resembles the findings of both the 10 and 100 year return levels obtained in the time window approach and proves the trends in the parameters to yield a significantly better description of the underlying data. This analysis therefore shows statistically sound evidence for a change in the extremely high and extremely low temperature anomalies. While for the former there is a slight increase in the moderate extreme events and a slight decrease in the largest events, the extreme deviations towards lower temperatures do all decrease at an even higher rate.

An underlying assumption for the VGLM approach to yield a good description of the block maxima/minima is that the changes occurring in the GEV parameters can be approximated by a first order linear model. If not, we would underfit the complex dynamics of the system and would strip the results of the VGLM approach off any meaning. As can be seen in table 7.2 all changes of the GEV parameters except for the

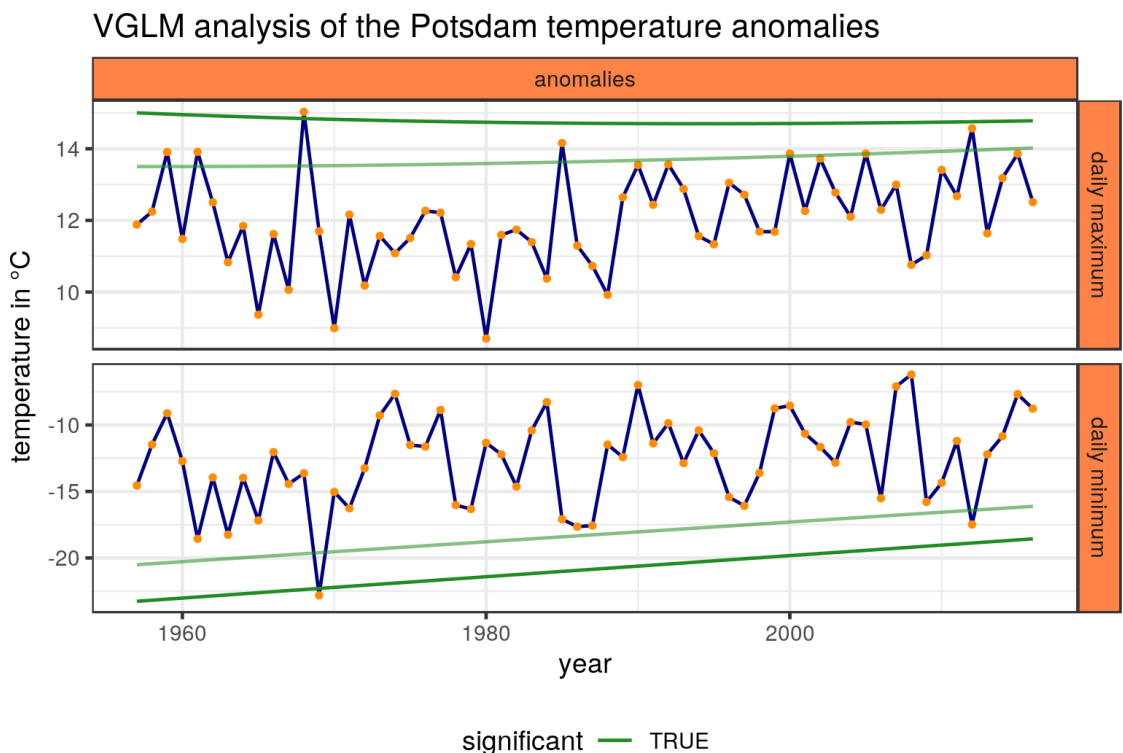


Figure 7.4.: This figure shows the results of the VGLM analysis of the Potsdam temperature anomalies (see figure 7.1). Firstly, the annual maxima of the daily maximum temperature anomalies (top) and the annual minima of the daily minima (bottom) were extracted. They are plotted using orange dots and connected by blue lines. In these block maxima/minima a non-stationary GEV distribution with a linear trend in time in both the location and scale parameter was fitted using the framework of the VGLM (see section 5.2.1). The 10 year (soft color) and the 100 year (dark color) return levels were extracted and added to the series. Finally, the fit of the non-stationary GEV distribution was compared to a stationary one using a likelihood ratio test (see section 5.4.2). If the null hypothesis of describing the block maxima/minima using the stationary model was rejected at a 5% significance level in favor of the non-stationary one, the return levels were plotted in green. If, on the other hand, the null hypothesis could not be rejected at this significance level, it was plotted in red. In this plot, however, the non-stationary model of both the block maxima and minima is significant and thus only green colored return levels are present. Note that while the trend in the location and scale parameter is indeed a linear one, the resulting return levels do not have to be a straight line (see equation (2.1.1)).

scale of the daily minima could be indeed described by a first order linear model. As a result I tried to use a more complex model featuring a first order linear model for the location parameter, a second order linear model for the scale parameter, and a constant shape parameter for the daily minima. But both the likelihood ratio test and various information criteria did not suggest this model to yield a significantly better description than the first order linear one. On the other hand, even the model featuring a linear trend in both the location and scale parameter might be already fairly complex and the description would be more appropriate with one containing a linear trend in only one of the parameters while keeping the other two constant. Such models showed to be on average worse performing than the one with two linear trends and a constant shape parameter.

7.4. Raw temperature series of the Potsdam station

Next, the raw temperature series of the Potsdam station will be analyzed. In contrast to the anomalies their changes do reflect the impact of the climate change on the society more directly. But since most of the results will be quite similar to the ones obtained for the Potsdam temperature anomalies in the previous section, only differing results will be reported. Note that the block maxima of the raw temperatures are distinctly different events than those of the anomalies and not just scaled versions. For one, all extremely high raw temperatures are located in summer while those of the anomalies can be found in any season throughout the year.

7.4.1. Time series and overall distribution

We, again, start with the plot of the time series in figure 7.5. Despite of the pronounced annual cycle we still find the variability of the daily maximum temperatures to be larger than the one of the minima.

The main difference between the estimated densities of the anomalies in figure 7.2 and of the raw series in figure 7.6 is the bimodal structure introduced by the annual cycle. The corresponding numerical estimates of the first four moments can be found in table 7.3. In general the bimodal structure remains constant in time. For the daily minimum temperature the left peak, which corresponds to the average temperature during winter, is nevertheless getting smeared out in the last time window. This indicates more mild winter months but also, on average, colder springs and autumns during nighttime since we are dealing with the daily minimum temperatures. The magnitude of the change in mean is the same as for the anomalies, the skewness and kurtosis remain, again, more or less constant, and the variance is approximately decreasing too. Upon closer inspection, the variance is getting smaller in the second time window for both the daily maximum and minimum temperatures but increases again in the third one. This might cause problems since we are fitting a linear model to the scale dependence. But as for the temperature anomalies, fitting a second order linear model resulted in overfitting and was not favored by neither the likelihood ratio test nor the information criteria.

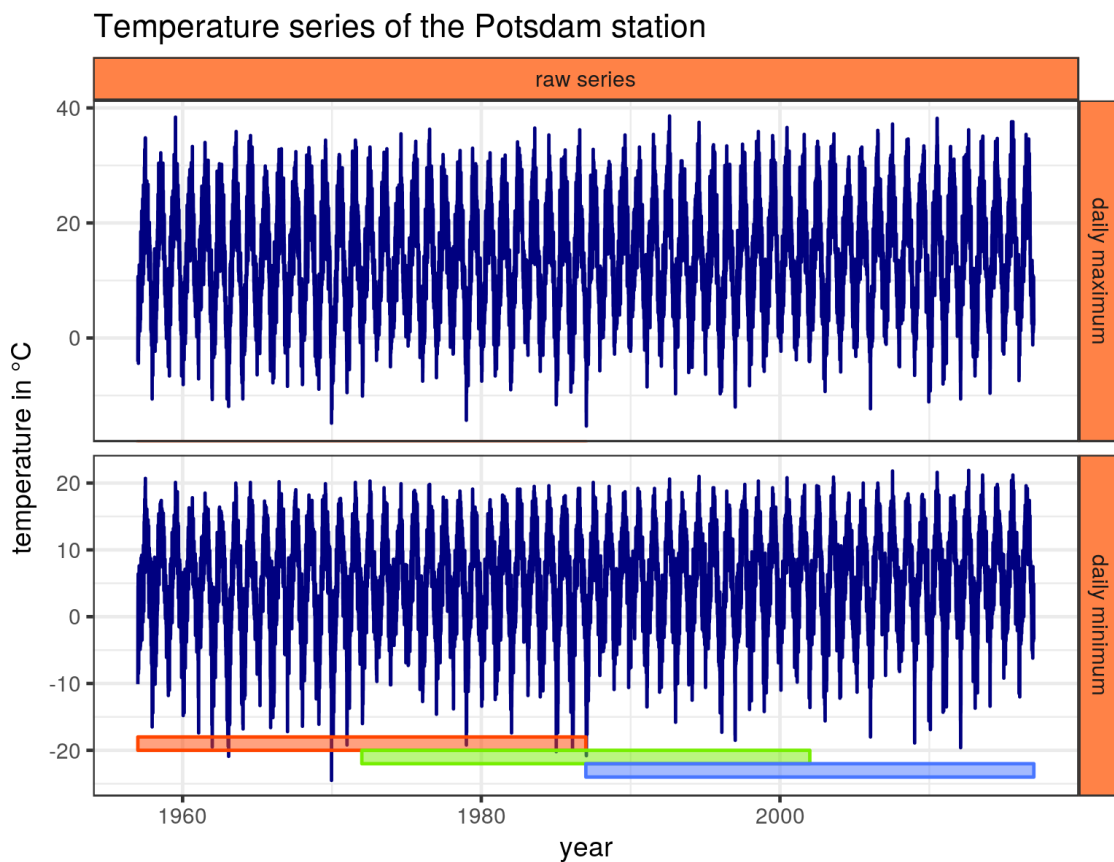


Figure 7.5.: Plot of the daily maximum and minimum temperature series of the Potsdam station. Compared to figure 7.1 the series display a pronounced annual cycle.

Table 7.3.: Moments of the raw Potsdam temperature series in the different time windows.

window	mean	variance	skewness	kurtosis	type
1957 till 1986	4.85	49.35	-0.39	2.81	daily minimum
1972 till 2001	5.22	46.50	-0.31	2.72	daily minimum
1987 till 2016	5.68	47.13	-0.30	2.67	daily minimum
1957 till 1986	13.19	84.67	-0.04	2.19	daily maximum
1972 till 2001	13.50	81.89	0.00	2.24	daily maximum
1987 till 2016	14.07	83.99	-0.03	2.23	daily maximum

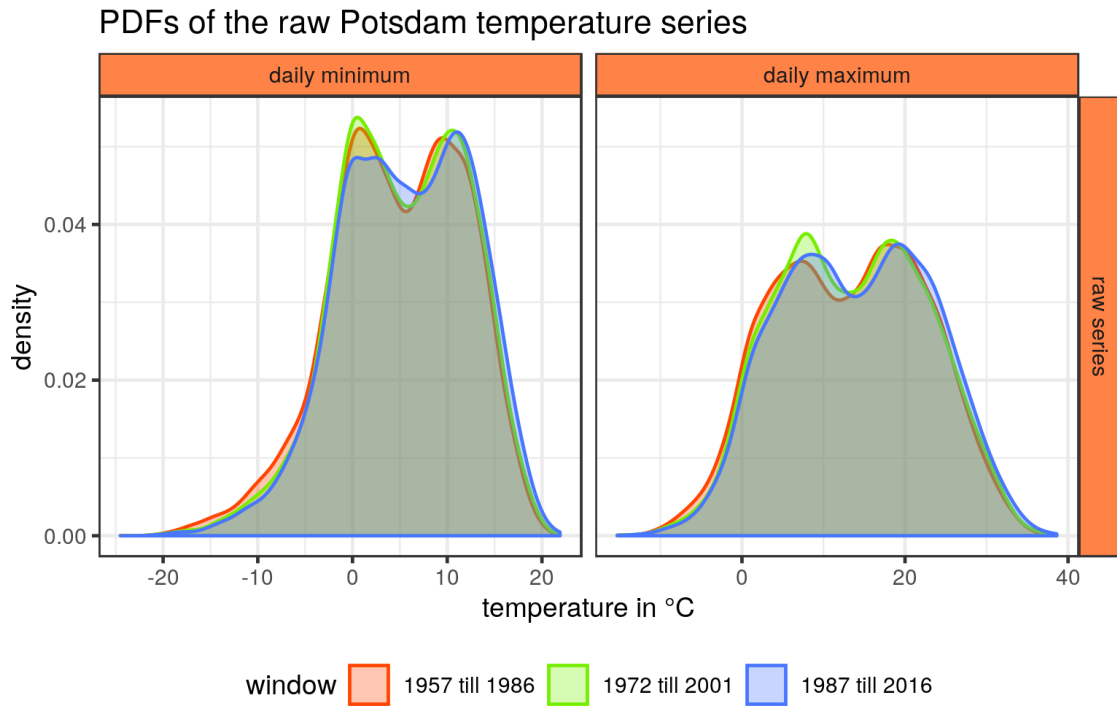


Figure 7.6.: This figure shows the PDFs of the raw temperature series of the Potsdam station in figures 7.5 similar to the one of the anomalies in figure 7.2.

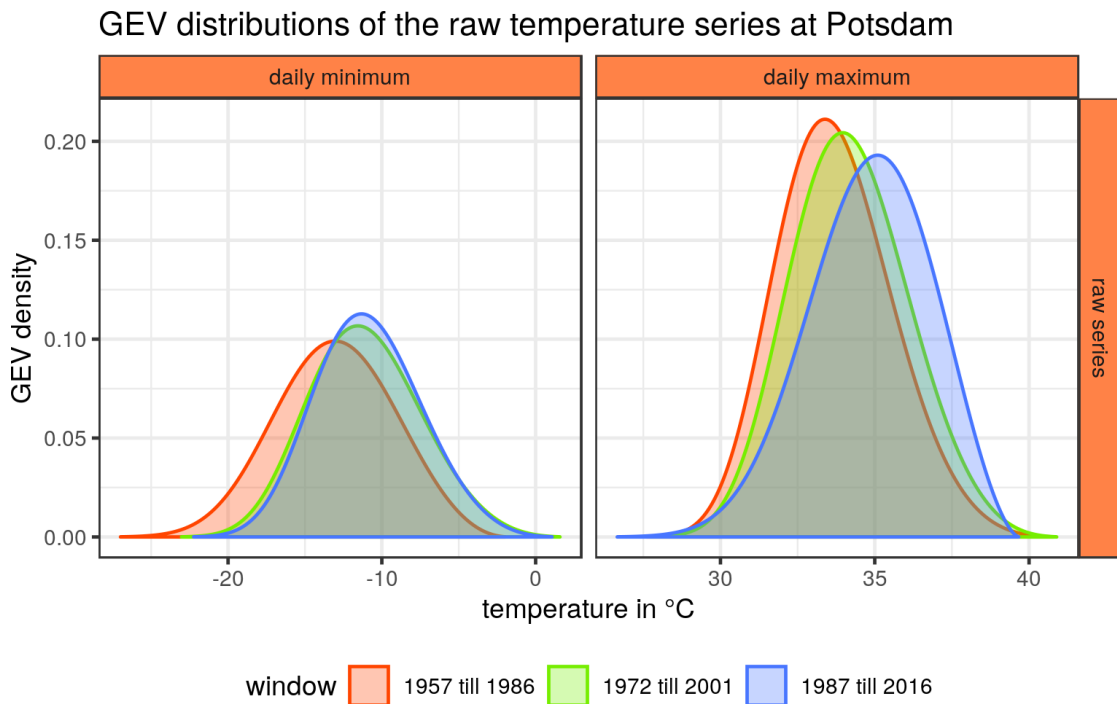


Figure 7.7.: GEV distributions constituted by the GEV parameters fitted to the block maxima/minima of the raw temperature series in figure 7.5.

Table 7.4.: GEV parameters and some return levels for the raw temperature series of the Potsdam stations in the three time windows. All quantities except of the shape parameter are given in degree Celsius and all error estimates are calculated using the Monte Carlo method.

window	location	scale	shape	10 year ret	100 year ret	type
1957-1986	-14.39 ± 0.85	3.92 ± 0.54	-0.31 ± 0.15	-20.76 ± 1.36	-24.06 ± 2.38	min
1972-2001	-12.46 ± 0.67	3.55 ± 0.57	-0.23 ± 0.13	-18.69 ± 1.06	-22.54 ± 2.09	min
1987-2016	-12.17 ± 0.62	3.36 ± 0.48	-0.23 ± 0.16	-18.07 ± 1.12	-21.70 ± 2.36	min
1957-1986	32.99 ± 0.40	1.78 ± 0.28	-0.20 ± 0.15	36.21 ± 0.52	38.31 ± 1.07	max
1972-2001	33.49 ± 0.37	1.85 ± 0.31	-0.23 ± 0.17	36.74 ± 0.52	38.73 ± 1.13	max
1987-2016	34.20 ± 0.45	2.07 ± 0.31	-0.38 ± 0.16	37.34 ± 0.42	38.72 ± 0.77	max

7.4.2. Non-stationary EVA and significance

The GEV distributions fitted into the raw daily temperature series can be found in figure 7.7 and the corresponding parameters including their fitting errors in table 7.4. One finds almost all fitting errors to be larger than the ones obtained for the temperature anomalies in the last section. While previously the changes were still bigger than the errors this relationship is now the opposite and it is indeed questionable what to learn from the results of the raw temperature series. But the time window approach is a more qualitative method, which enables us to access the overall change of the extreme events. To draw statistically sound conclusions, we will rely on the VGLM approach in combination with the likelihood ratio test. If there are indeed significant changes in the data, the hypothesis of a better description using a stationary model over a non-stationary one can be rejected. Whether or not the fitting errors in the 30 year windows are larger than the changes does not enter in this maximum likelihood-based approach.

The distributions in figure 7.7 show a decrease in scale for the daily minima but an increase for the maxima. This is in contrast to the findings of the anomalies, which showed a decrease for both types. The scale parameter undergoes a more consistent change than the variances of the overall daily series. The shape parameter, on the other hand, is shifting to smaller negative values for the minima and to higher negative values for the maxima. The increase of the shape parameters for the daily minima, which can not be found in the series of the anomalies, thus indicates that the tails of the overall temperature distribution are getting more fat towards low temperature and less fat towards high temperatures.

There is an overall increase in all return levels. The decrease in scale for the minima is compensated by the less negative value of the corresponding shape parameter. For the maxima the increase in the location and scale parameter causes the return levels to rise. At the same time influence of the more negative shape parameters results in smaller and smaller return levels for higher return periods. The extremely cold days in winter therefore become warmer, the moderate extreme hot days become more frequent, and the extremely hot days stays more or less the same. Albeit the slope of the return levels found in the VGLM analysis is almost the same compared to the anomalies, the trend in the extreme events of the daily minimum temperature is not significant. This is due to

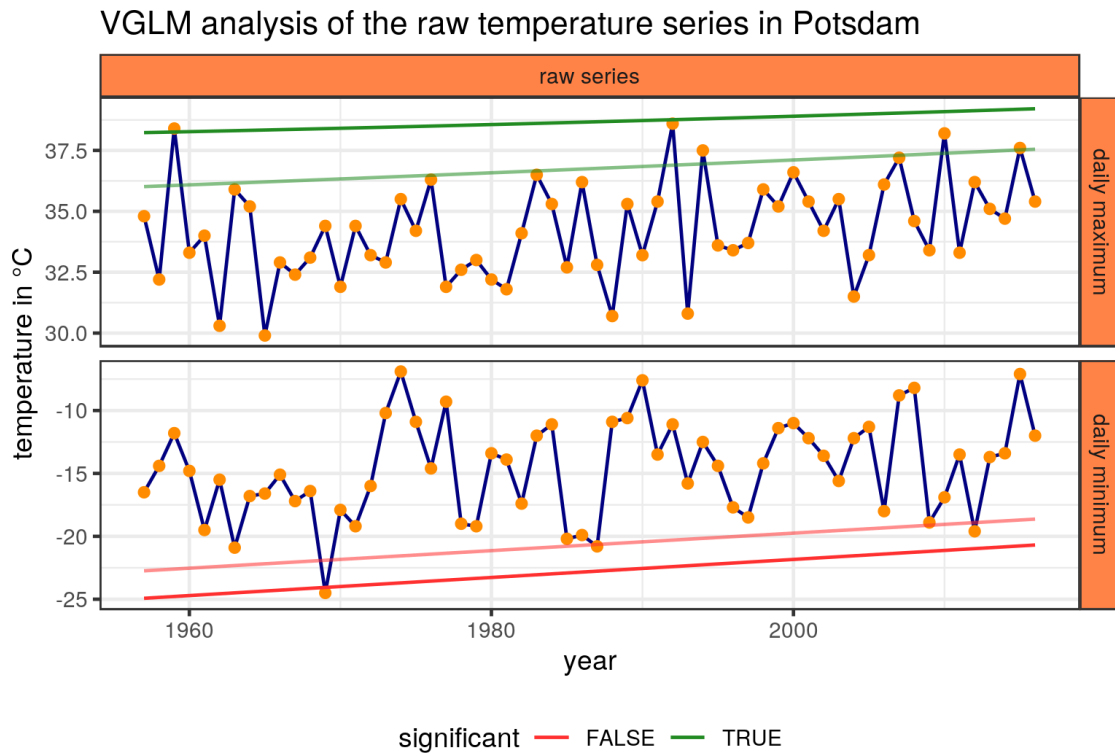


Figure 7.8.: The figure shows, similar to figure 7.4, the results of the VGLM analysis for the black maxima/minima of the daily maximum/minimum temperature series of figures 7.5. As indicated by the colors of the return levels (lighter color corresponds to the 10 year return level and darker color to the 100 year return level) the model describing the daily minimum temperatures is not significant. In contrast to the extreme events of the daily maximum temperature it is thus best described by a stationary model.

the larger variance in the block minima, which can be seen in figure 7.8, and is consistent with the larger error estimates obtained in the time window approach.

7.5. Analysis of the temperature series within Germany

The case study of the Potsdam temperature anomalies in section 7.3 provided us with some insights into possible temporal changes of the overall temperature distribution. But whether the obtained results for the local weather at Potsdam are reasonable proxies for the general climate within Germany is hard to tell without performing a similar analysis on large amount of stations within Germany. Since each GEV-fit is based on only 30 values, there are considerable statistical uncertainties, even when ignoring the possibility of local effects. Thus, we will analyze all series provided by the DWD, which meet our criteria in length and quality (see section 7.1), and compare the results with the ones of the Potsdam station.

Although the changes of the first four moments in time throughout Germany are quite interesting, the whole analysis will not be covered in this chapter. The figures and their interpretations are just too many and thus they are moved to the appendix (see A). Instead, only the results for the 10 and 100 year return levels will be reported in here. To fully conceive their temporal evolution on the basis of the changes of the individual parameters, studying the appendix is strongly recommended.

7.5.1. Temperature anomalies

One of the most important findings of the analysis is that the algorithm based on the augmented Lagrangian method (see section 3.2.1) is indeed working for all stations without any need of manual interactions in the time window approach. All other packages in R performing the stationary EVA do, at least at the point this document was written, throw errors for at least one of the stations.

Figures 7.9 and 7.10 show the distributions of the 10 and 100 year return levels calculated for all temperature anomalies. As before, the extracted block maxima/minima were split into three overlapping time windows, the stationary EVA was performed, and the resulting 10 and 100 year return levels were extracted from all of them. In order to better compare the distributions with the results of the Potsdam case study, the corresponding return levels are highlighted using vertical lines.

To be perfectly precise, not all of the 74 (daily maxima) or 75 (daily minima) stations have been used to create these distributions. The measurement station located at the mountain called Brocken in the very center of Germany has been removed to improve the visibility of the remaining results. As discussed in appendix A.1, we consider the parameters obtained for this series as artifacts. But wasn't just claimed the most important finding of this analysis is that there are no artifacts? Well, both is true since we unfortunately overloaded the term "artifacts". By introducing the constrained optimization based on the method of augmented Lagrangian multipliers we could indeed get rid of all *numerical* artifacts. In other words, the maximum likelihood fitting procedure did work for all stations, was always properly initialized, and found the global minimum of the negative log-likelihood function. But while this largely improves the extreme value analysis using the GEV or GP distribution, it is not a cure for all its weaknesses, just

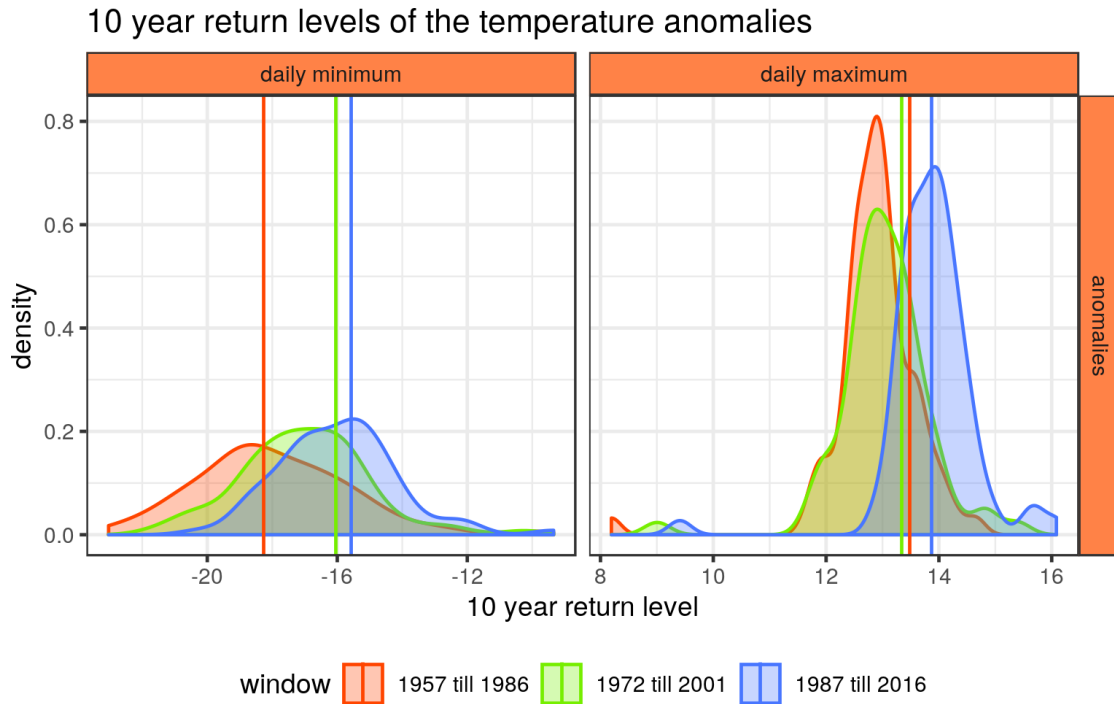


Figure 7.9.: Distribution of all 10 year return levels of the daily temperature anomalies series provided by the German weather service DWD. A detailed description of the preprocessing applied to the 73 (daily maxima) or 74 (daily minima) stations series can be found in section 7.1 and 7.2. This is one series less than the total amount contained in the data set. The missing one is the Brocken station, which was considered an artifact with respect to the issue described in chapter 6 and thus omitted in this analysis. A version of the analysis including this very station can be found in the appendix A.1 and in the corresponding figure A.13. To increase the comparability with the Potsdam case study the corresponding results are highlighted with a vertical line.

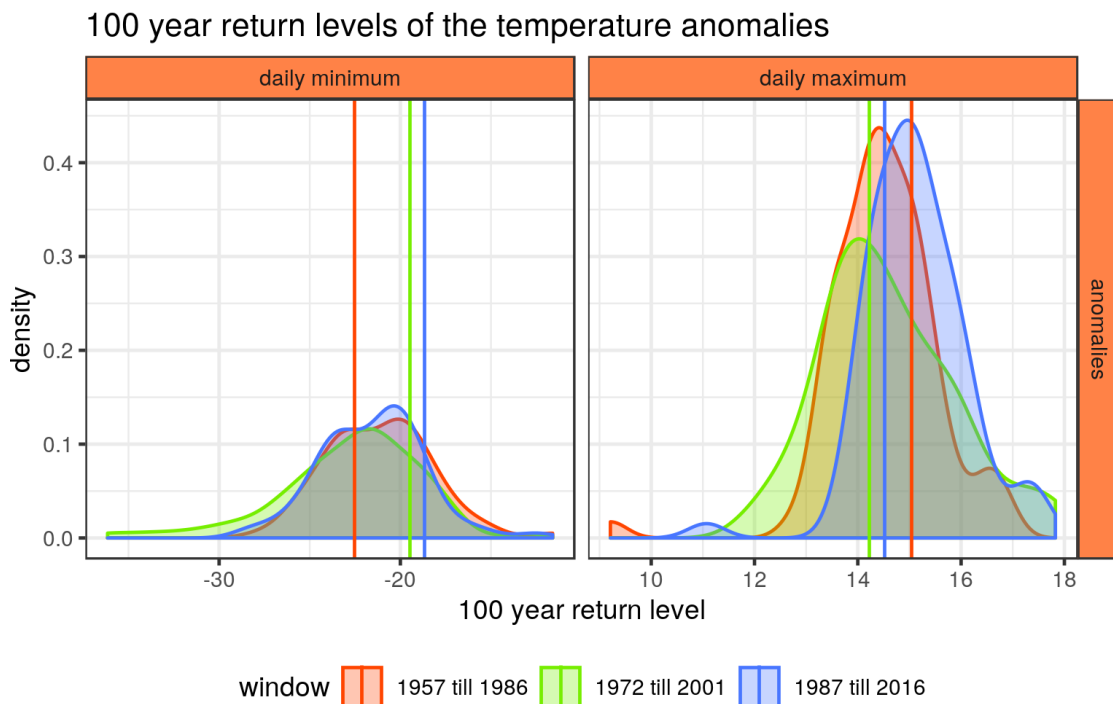


Figure 7.10.: Distribution of all 100 year return levels of the daily temperature anomalies throughout Germany. For details see figure 7.9 and section 7.1 and 7.2.

for the numerical ones. The problem encountered when fitting the Brocken series was the same as the one discussed in chapter 6. A very large event occurred within the first time window dominating its statistics and causing the fit to be of a Fréchet type with a very large shape parameter. In all other windows the series results in similar parameters as the remaining ones (see the appendix, where the results of the Brocken station were included in the figures). The time windows were therefore too short to fully sample the dynamics of the system. But shouldn't this be true for all stations and thus renders this whole analysis useless? I don't think so. As a fact, the Brocken station was the only one experiencing very big jumps in parameters and the only temperature series yielding such a large shape parameter (0.69 in the first window). In addition, the analysis of the whole ensemble provided by the DWD yielded another interesting result apart from the GEV parameter, which might give an explanation of the extraordinary behavior of the Brocken series. It turns out the kurtosis of the daily temperature anomalies is correlated with the altitude of the measurement station. The higher the elevation, the bigger the kurtosis, and thus the fatter the tails of its PDF. Since the elevation of the Brocken station is one of the largest among all considered time series, it might be constituted by a distribution with a fatter tail from the start and is therefore more prone to be dominated by a single very large event. We can easily check this hypothesis by reviewing the results of the series corresponding to the largest elevated measurement site, the mountain Zugspitze. As can be seen in the very south of Germany in figure 7.15 the highest station does indeed dominate the changes in the 100 year return levels. This does, of course, not prove the classification of the Brocken station as an artifact but given the limited amount of data

it should serve the purpose. Also it illustrates the intriguing details involved in handling observational data and the limitation of the full automation of the EVA.

Back to the actual figures. While the magnitude is differing between the individual time windows of the daily maximum and minimum temperature anomalies in 7.9 we can nevertheless see a consistent shift of the distributions towards higher values (temperatures). For the daily minima there is also a reduction of the overall width of the distributions causing an even larger increase in the 10 year return level. But in case of the 100 year return levels in figure 7.10 no such shift or consistent change in the distribution seems to be present, apart from the shift towards higher values in the last time window for the daily maxima. This means that the medium extreme events, those occurring on average every e.g. 10 years, of the daily minimum temperature anomalies decrease in time by a margin (roughly four degrees) while the corresponding large extremes, at a return period of 100 years, stay approximately the same. For the daily maxima both return levels do increase by approximately one degree. The overall change of the temperature anomaly distribution throughout Germany thus seems to be a fairly complex one.

Up to now, we just took the results and lumped them all together. By doing so we learned about the temporal changes of distributions constituted by all the individual return levels throughout Germany. But as we have already seen with the Brocken station, the series are not perfectly identical and their behavior might depend on both their geographic location and local elevation profile. To not miss such possible spatial information, the 10 and 100 year return levels are displayed on maps in figures 7.11 and 7.12.

When projecting the content of the distributions onto maps we have to further condense the information. Three different maps corresponding to the results obtained in the three time windows for each climatological quantity would be just confusing at best. Instead, only the numerical value of the return levels found in the last time window will be encoded in the color of each circle on the map. Their center correspond to the geographic location of the measurement station and their enveloping color to the difference between the value in the third and the one in the first time window. Thus, the inner color indicates the latest realization of the return level and the outer one its change in time. For a detailed explanation of the color scale please see section 7.2.

In those maps one can see that both the values and the temporal changes in the 10 and 100 year return levels do not occur uniformly throughout Germany. Stations located at coast turn out to be among the most mildest ones for both the daily minimum and maximum temperature anomalies. Also there is a tendency for more western stations (lesser longitude on the x-axis) to have less extreme deviations from the climatology towards low temperatures. Facing the temporal changes in the return levels the asymmetry in the very high quantiles of the overall distribution is quite pronounced again. In case of the daily minima a general increase in values and thus a decrease in magnitude of the 10 year return level can be found. The largest changes do occur in the south of Germany. The same is true for the daily maxima while they also feature an asymmetry between the stations in the east and the ones in west of Germany. Series corresponding to a larger longitude do have a tendency to experience a slower increase in the 10 year return level. Also note that the daily minima feature a wider range for the values and about two times the range for the differences compared to the daily maxima. For the 100 year return level of the maxima the patterns are a little bit more smudged. The increase does not occur

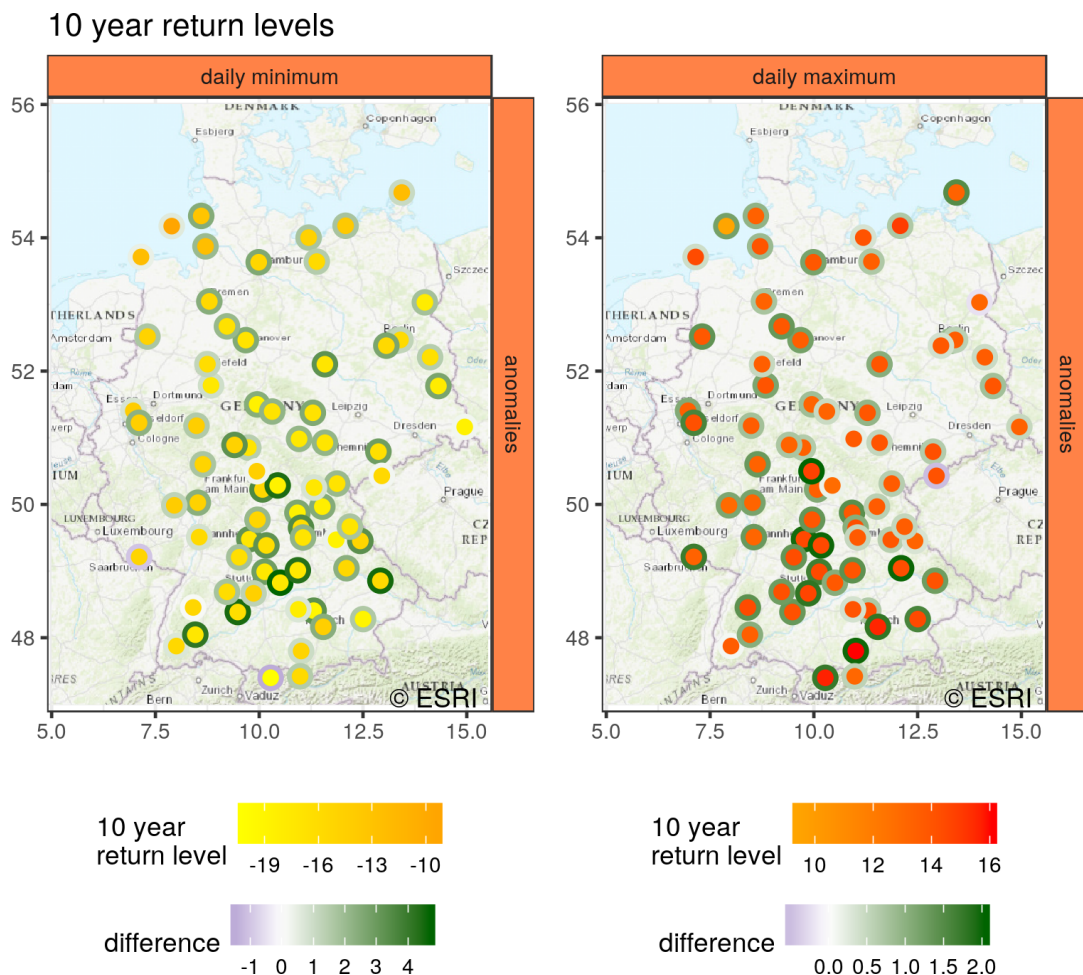


Figure 7.11.: 10 year return levels of the daily temperature anomalies throughout Germany. This map illustrates the spatial information missing in the distributions of the 10 year return levels in figure 7.9. The colors within the circles, ranging from yellow to red, display the value in the third and last time window at the measurement sites. The colors of their outer surrounding, on the other hand, show the magnitude of the temporal evolution calculated from the values of the third window minus the ones of the first. Note, to better compare the different maps of this analysis the color ranges were fixed as described in section 7.2.

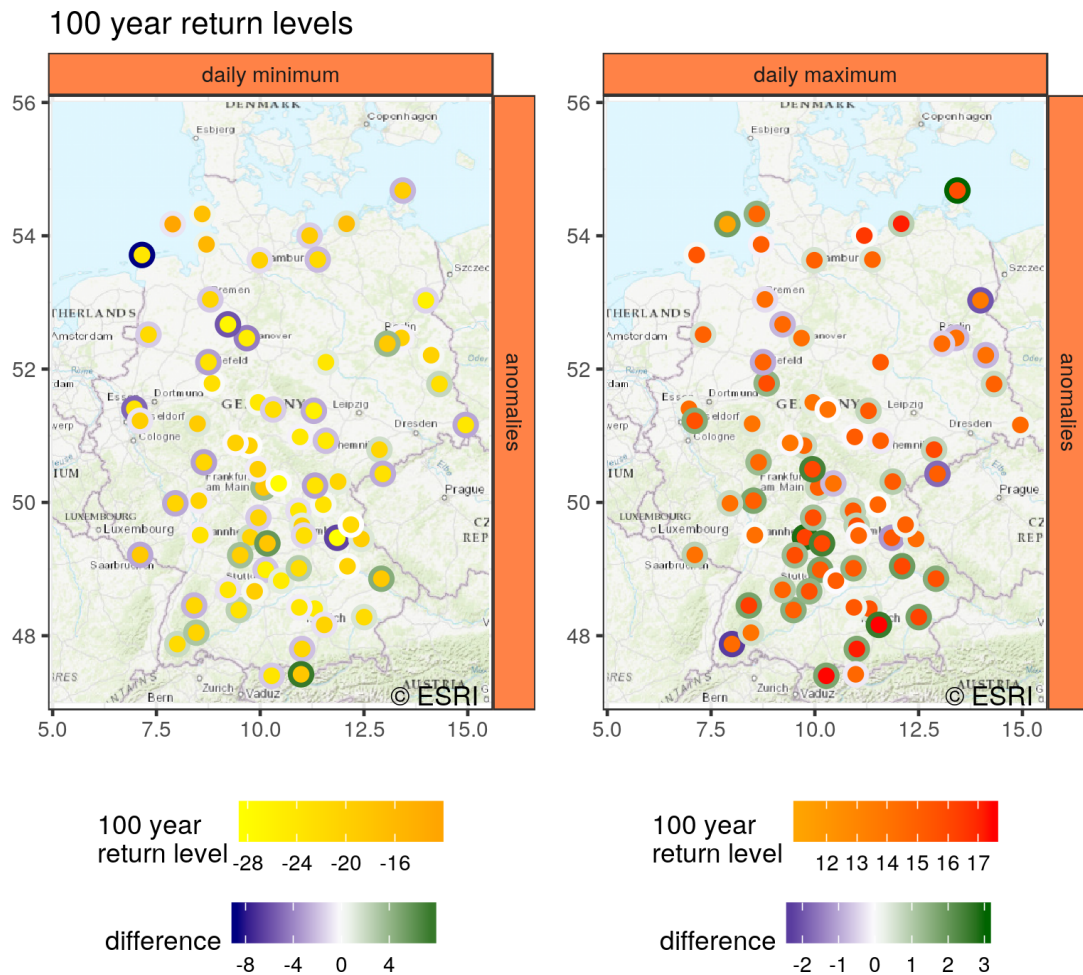


Figure 7.12.: 100 year return levels of the daily temperature anomalies throughout Germany. For a detailed description of the figure and colors please see figure 7.11 or section 7.1 and 7.2.

in all of Germany anymore but most prominently in the south, middle, and in coastal regions. Also now the decreases are bigger in number and showing a slight tendency to occur in the east. The changes in pattern are even larger for the daily minima. In contrast to the 10 year return levels the 100 year ones do show predominantly decrease. The largest ones do occur in the north and the south still mostly features increases. But since we are talking about the daily minimum temperature anomalies the extreme deviations from the mean climatology become less in the south and way larger in the middle and north of Germany.

Next, the results of the VGLM analysis are presented in figure 7.13. As a short reminder, in there we take the whole series, perform a non-stationary EVA by fitting a GEV distribution with a linear trend in both the location and scale parameter, and compare the result with a stationary model using the likelihood ratio test. This test checks whether we can reject the null hypothesis of describing the underlying data with a stationary model in favor for a non-stationary one at a 5% significance level.

The maps show two kinds information. The inner color corresponds to the slope of a straight line fitted into the curve of the 10 year return level. This fit is only an approximation of the non-linear dependence of the return levels on time but we will use it to determine the sign of the trends we will find to be significant. Note that label “significant trend” corresponds to the first order linear model used in the location and scale parameters. The return levels are just non-linear combinations of the GEV parameters and could very well be approximately constant while a trend is actually present. Also the slope will be only estimated for the 10 year return level since the non-linear curvature is even more pronounced for larger return periods and the resulting slope less reliable. The outer one represents the results of the hypothesis test. If it was rejected, it is colored in red. If not, a brown color will be used. For more details please see section 5.4.2. One can see in figure 7.13 that for the daily maxima indeed almost all stations show a significant trend while for the daily minima only for approximately half of the stations the hypothesis test could be rejected. Surprisingly, not all of them are heading towards higher values. Eight stations feature a significant trend in the GEV parameters and at the same time an overall increase in the 10 year return level for the daily maximum temperature anomalies. A more thorough discussion of the changes in the return levels, GEV parameters, and first moments of the underlying distribution as well as possible causes is given in the appendix A.1.

We could therefore show in a statistically sound way that the climate change is not just causing an overall shift in the temperature anomalies towards higher values but is also affecting the extreme events and thus the very tails of the temperature distribution. But the nature of those changes is far more complicated than a mere linear increase in value. Instead, we face both significant increases and decreases depending on the return period as well as the geographic location of the corresponding measurement station.

7.5.2. Raw temperature series

After discussing the changes in the distribution of the temperature anomalies throughout Germany we will have a look at the extreme events of the raw series. It is almost certain for the annual maxima to occur in summer and for the annual minima to occur in winter. We therefore only sample a series of a reduced effective length and loose some explanatory

VGLM analysis of the temperature anomalies

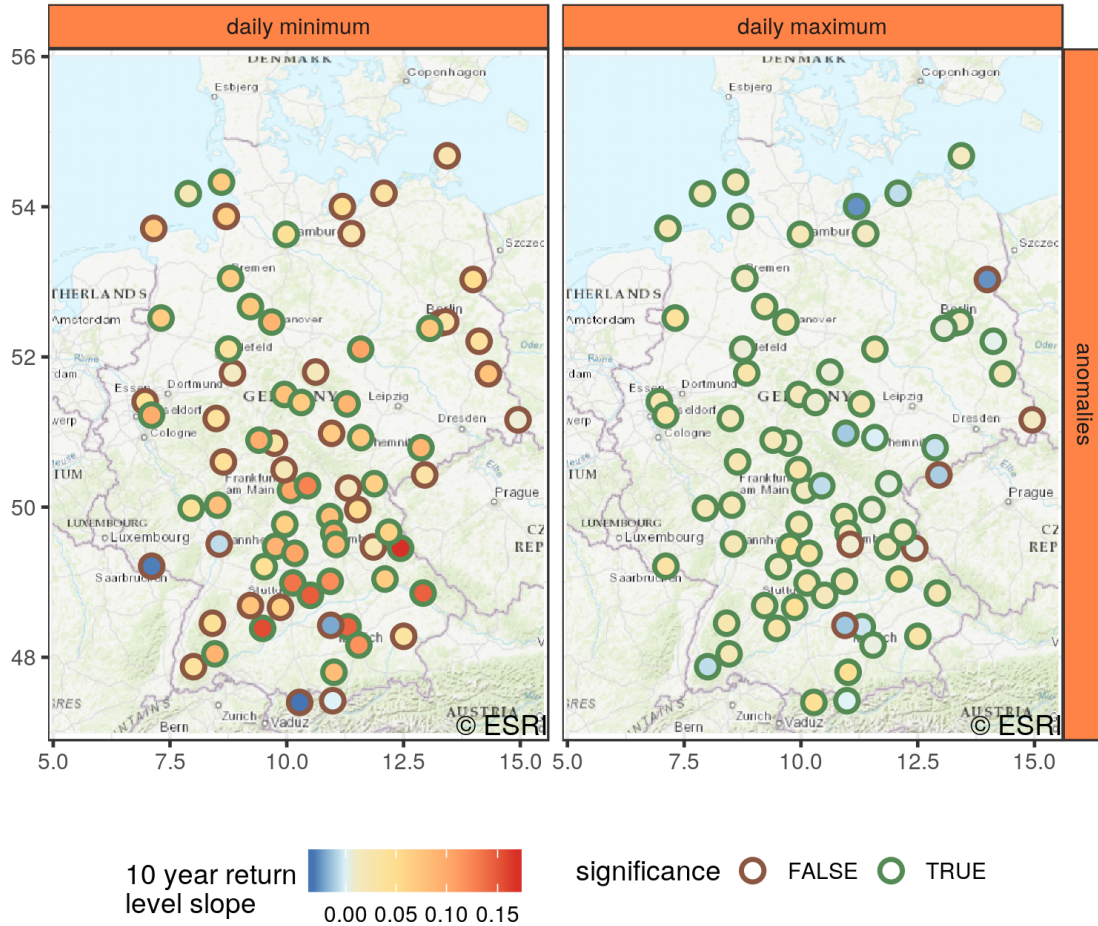


Figure 7.13.: Results of the likelihood ratio test whether to use a stationary model over a first order linear one. All stations, for which the null hypothesis of a better description with a stationary model was rejected, are enclosed by a green color. The circles are filled with the slope of the 10 year return levels fitted using the VGLM. Since it has a non-linear dependence in time the slope was estimated by fitting a straight line in its curve. This value might therefore not yield a perfect approximation of the actual temporal evolution but it is sufficient to determine its overall tendency. Note, the Brocken station, although still considered an artifact, is included in the figure again since it is not spoiling the color scales of this plot.

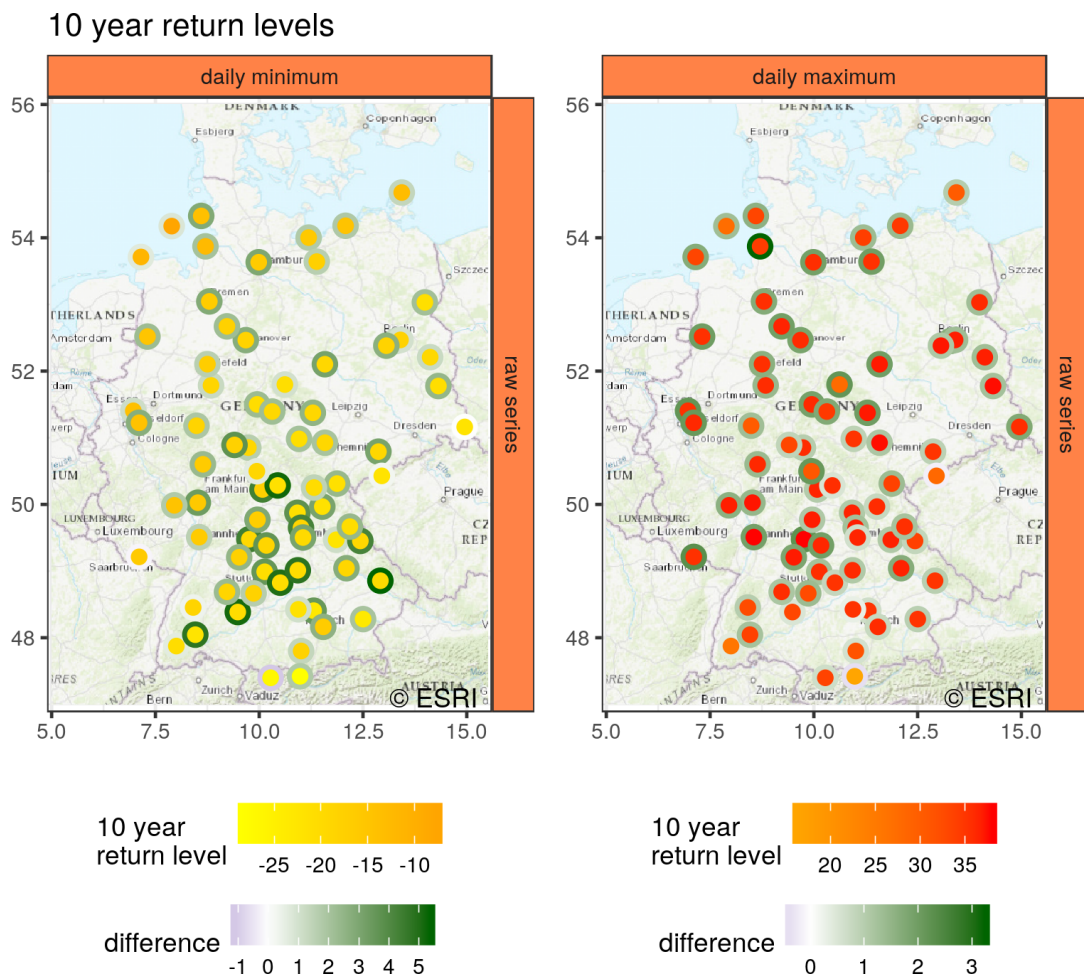


Figure 7.14.: 10 year return levels of the daily raw maximum temperature series throughout Germany. For a detailed description see figure 7.11 and section 7.1 and 7.2.

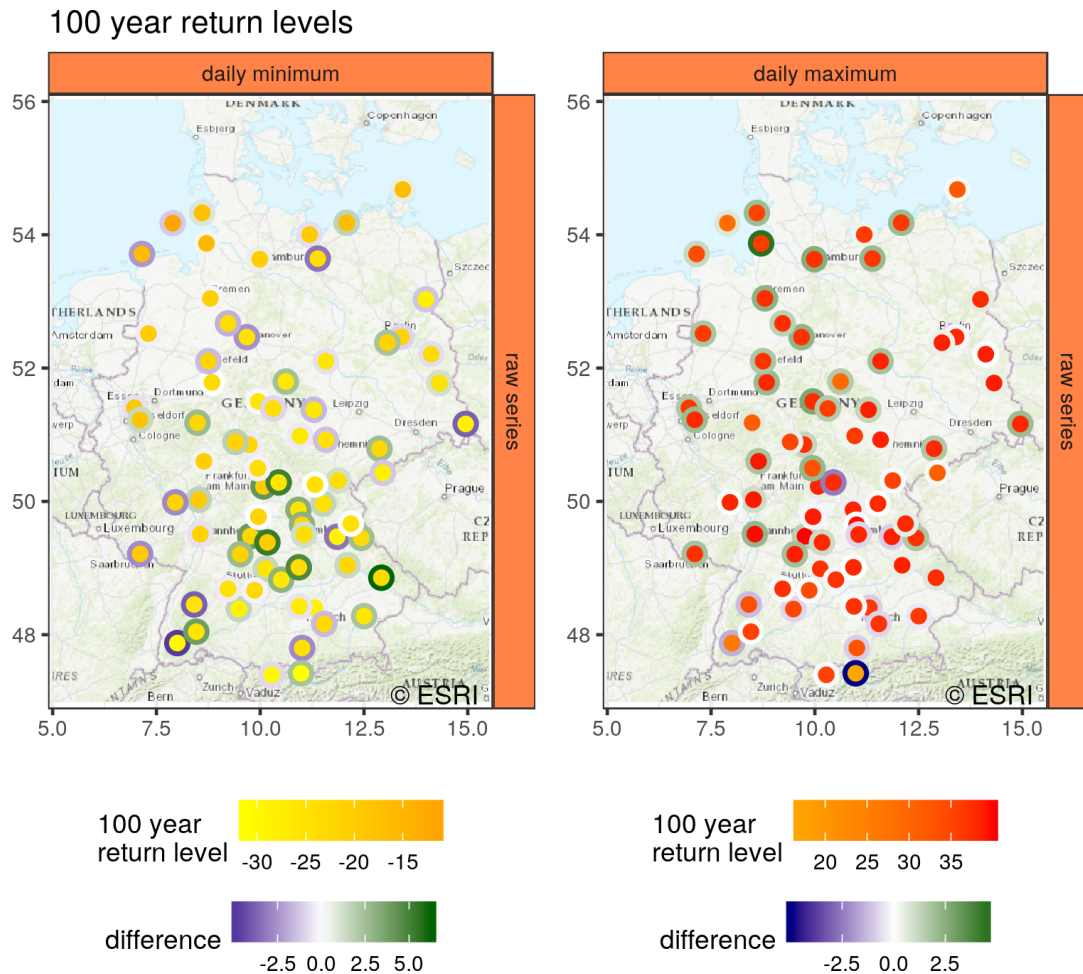


Figure 7.15.: 100 year return levels of the daily raw maximum temperature series throughout Germany. For a detailed description see figure 7.11 and section 7.1 and 7.2.

power when interpreting the results as the changes of the overall climate system. On the other hand, extreme events of the raw temperature series, like extremely hot days in summer or extremely cold days in winter, correspond to the ones having the most direct impact onto the everyday life of us humans. The results in this part of the analysis are thus not of primary interest for the investigation of the climate system but for the assessment of the consequences of its changes on society.

The temporal changes in the overall distributions of the return levels are very similar to the ones of the anomalies in figures 7.9 and 7.10. So, they were skipped to keep this chapter rather concise but the interested reader is encouraged to have a look at the whole analysis in A.2.

The spatial patterns of the changes are shown in figures 7.14 and 7.15. For the daily minima both the 10 and 100 year return levels look almost the same compared to the ones of the temperature anomalies, apart from the overall larger increase in case of the 100 year ones. The daily maxima, on the other hand, do show completely different patterns. While it was the south of Germany, which featured the strongest increase in

VGLM analysis of the raw temperature series

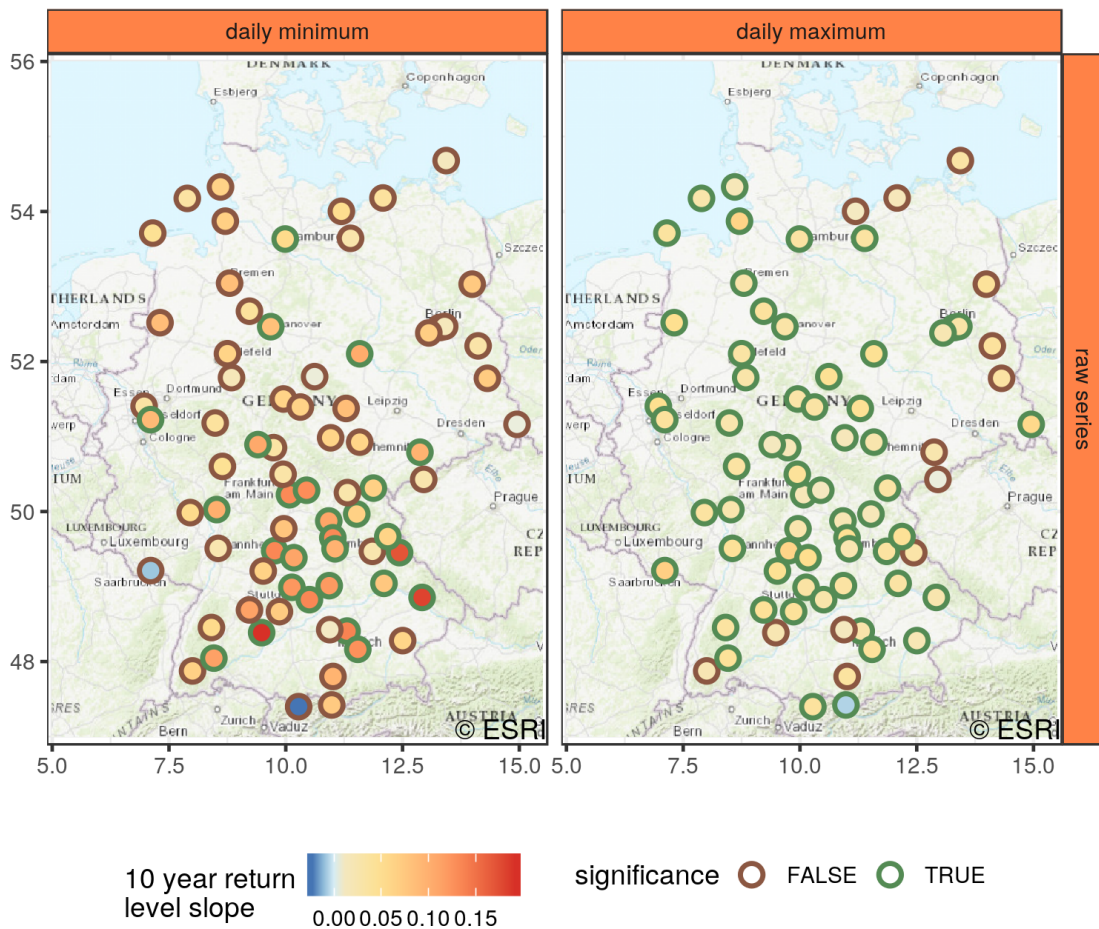


Figure 7.16.: Results of the likelihood ratio test whether to use a stationary model over a linear one for the daily raw temperature series. This plot is the counterpart of figure 7.13, which shows the results for the daily temperature anomalies and provides a more detailed description.

the extreme temperature anomalies, it's the north and west displaying the strongest increases in the raw temperature extremes. The changes in the east and south are only minor for the 10 year return levels and almost negligible or even negative for the 100 year return levels. This means, while the fluctuations in the temperature do increase for the south of Germany they do so during autumn, winter, or spring. Thus, these increasing fluctuations do not contribute the large deviations in summer sampled in the analysis of the raw temperature series. The actual changes of the overall temperature distribution does therefore not only seem to be highly non-linear and dependent on the particular spatial location, but also dependent on particular times within the year, like seasons or even more fine-grained temporal periods. This, unfortunately, renders our attempt to properly describe the changes using a non-stationary model almost impossible. But still we can check if there are at least changes present, which can be described by a first order linear model in the location and scale parameter of the GEV distribution.

The results can be seen in figure 7.16. Compared to the corresponding ones for the temperature anomalies in figure 7.13 a few less hypotheses could be rejected. Only 61 trends for the daily maximum temperatures (anomalies: 68) and 27 for the daily minima (anomalies: 41) out of the 74 individual series were found to be significant. But the general picture remains the same. What does it mean that for one series we have a significant trend in the temperature anomalies but none in the raw data including the annual cycle? The cause of this discrepancy can be found by looking at the ranges of values in the previous maps. The difference between the parameters in the first and last time window, and thus the slope of the linear trend, is approximately the same for the anomalies and raw series. At the same time the range in values is larger in case of the raw series, especially for the daily maxima. Since an increase in variance will worsen the fitting results in terms of the global minimum of the negative log-likelihood, the deviance statistic will decrease. This increase in variance can thus already push the statistic below the threshold set by the Chi-squared distribution and prevent the hypothesis from being reject. But this does not mean there are no trends anymore in the raw temperatures. We just would need more data to make statistically sound statements about its existence, which is covered by the annual cycle.

7.6. Analysis of the Potsdam precipitation data

After analyzing the daily temperature in the previous section, we will now turn to the daily accumulated precipitation. Since there is just one value measured per day, the integrated amount of rainfall during one day, no separation as for the temperatures in daily maxima and minima will be done. Also the annual cycle in the precipitation data is far less pronounced and almost absent. That's why we will drop the calculation of the anomalies and just handle the raw data itself. So, it will be way less figures than in the previous section.

The baseline for the results we are expecting is the analysis in (Trenberth, Fasulo, and Smith 2005; Trenberth 2011). They found the overall amount of moisture to rise in the atmosphere around the globe. This leads to an increase of the precipitation in most regions worldwide. But especially in Europe no trend could be found in the nodes of the grid data and only the average over all rainfall in Europe revealed a positive trend. Let's check this result by analyzing all stations provided by the DWD, which meet the requirements discussed in section 7.1. Since these measurements yield more direct and pure observations of the rainfall data projected onto a rather coarse grid, the analysis might reveal some insights not found in the grid data yet.

Again, we will start our analysis with the case study of the rainfall at the Potsdam station.

7.6.1. Time series and distribution

The time series of the daily accumulated precipitation in figure 7.17 has some major differences compared to the daily temperature series. First of all, there are only positive values and zeros corresponding to the total amount of rainfall occurring during one day given in millimeters. It is not at all visible in the plot but more than half of the entries are actually zero for the Potsdam station (11540 days out of 21915). In other words,

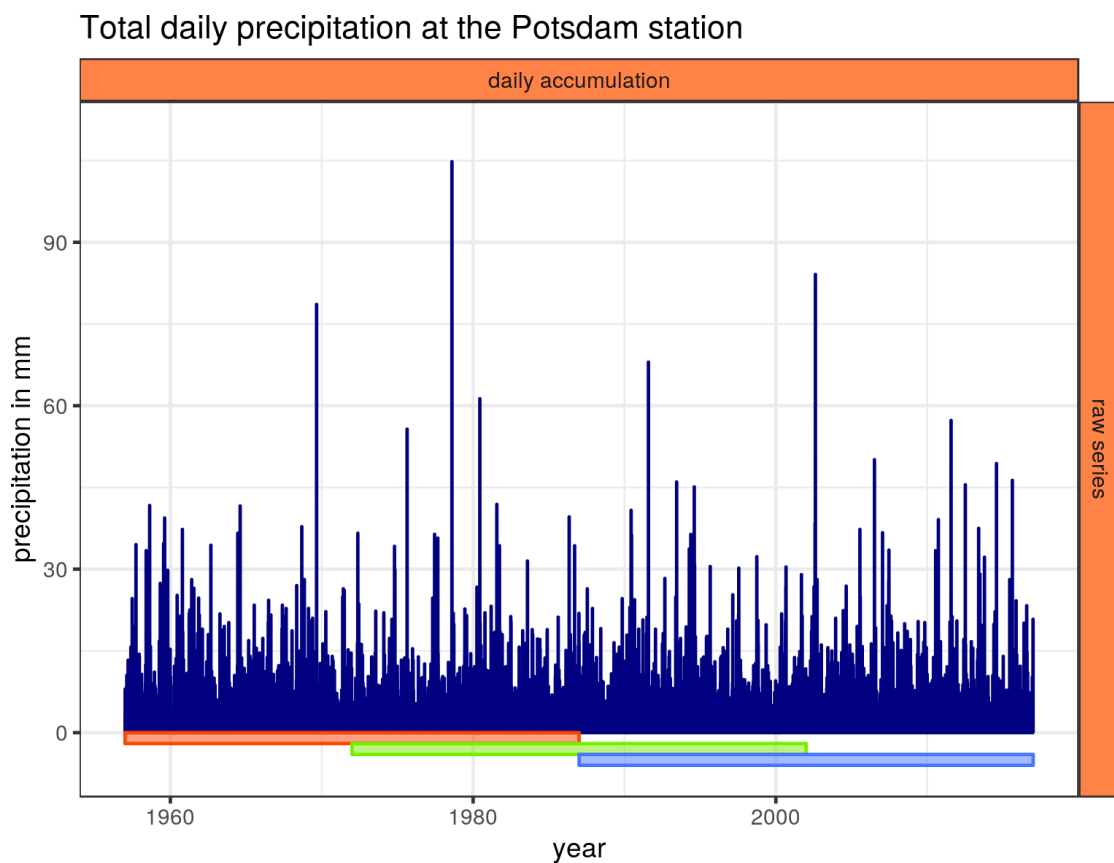


Figure 7.17.: Plot of the daily accumulated precipitation of the Potsdam station in Germany. Displayed are the 60 years of data, starting in 1957 and ending in 2016, required to be present in all time series analyzed within this chapter. In addition to the rainfall series, the time periods spanned by the three different time windows in the corresponding approach are highlighted. The first window (red) reaches from 1957 till 1986, the second one (green) from 1972 till 2001, and the third one (blue) from 1987 till 2016.

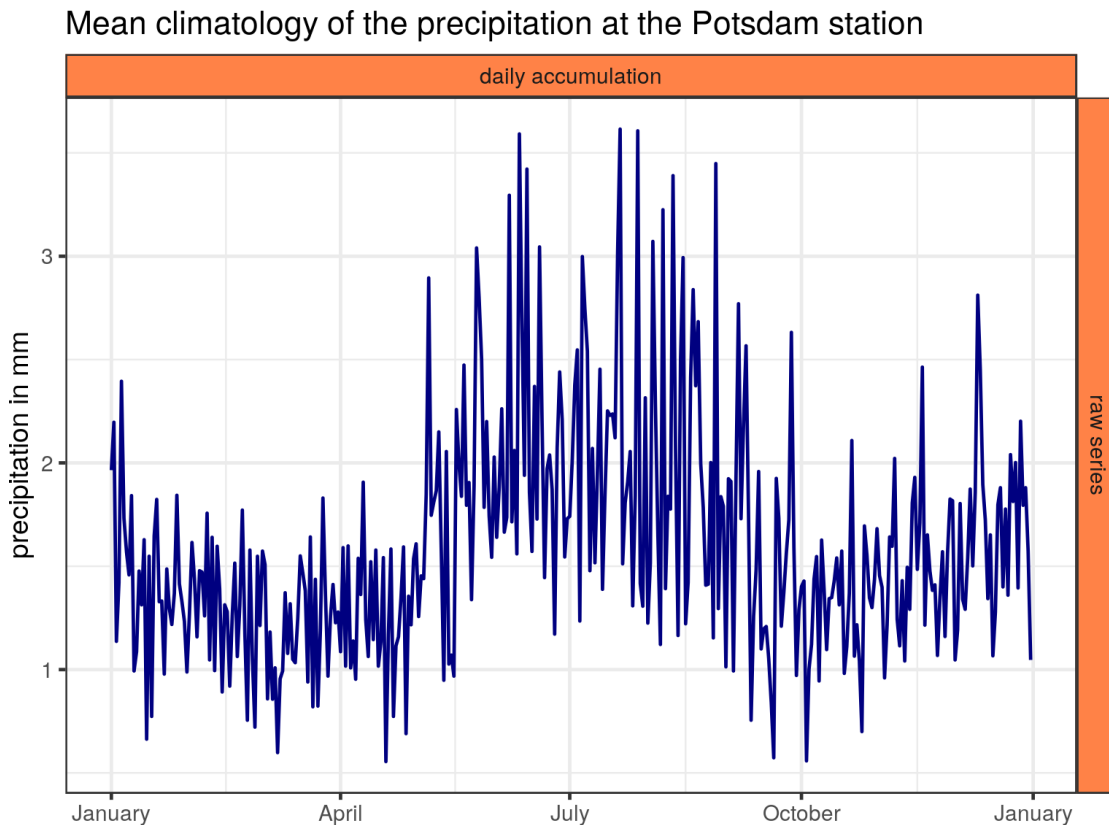


Figure 7.18.: Mean climatology of the daily accumulated precipitation of the Potsdam station in figure 7.17. The mean climatology is obtained as an intermediate step when calculating the anomalies of a time series (see section 4.3.1). There the mean values of all individual days of the year are calculated using only those particular days throughout the years. E.g. all February 2nd of a time series are averaged. The mean climatology thus looks exactly the same for all years in the time series and is only plotted for one specific year.

there are more days in Potsdam without any precipitation at all than days featuring rainfall. Compared to the temperature series the precipitation shows no pronounced annual cycle. Also the direct inspection of the mean climatology, the average of the individual days throughout all the years in the series, in figure 7.18 does not require us to investigate the anomalies of the rainfall (raw series minus climatology) as well. An annual cycle is indeed present but it is very noisy and small compared to the original data and its extreme events. Furthermore, no obvious trend can be seen in precipitation data in figure 7.17.

The estimated PDF of the rainfall data within Potsdam can be found in figure 7.19. Due to the logarithmic transformation of the x-axis, a constant offset of 0.1 mm was added to the whole series to move the peak at zero slightly to higher values. When reviewing the distributions of the daily accumulated precipitation within the different time windows no relevant changes at all can be found. The same is true for the estimated first four moments in table 7.5. Apart from a decrease in skewness and kurtosis in the

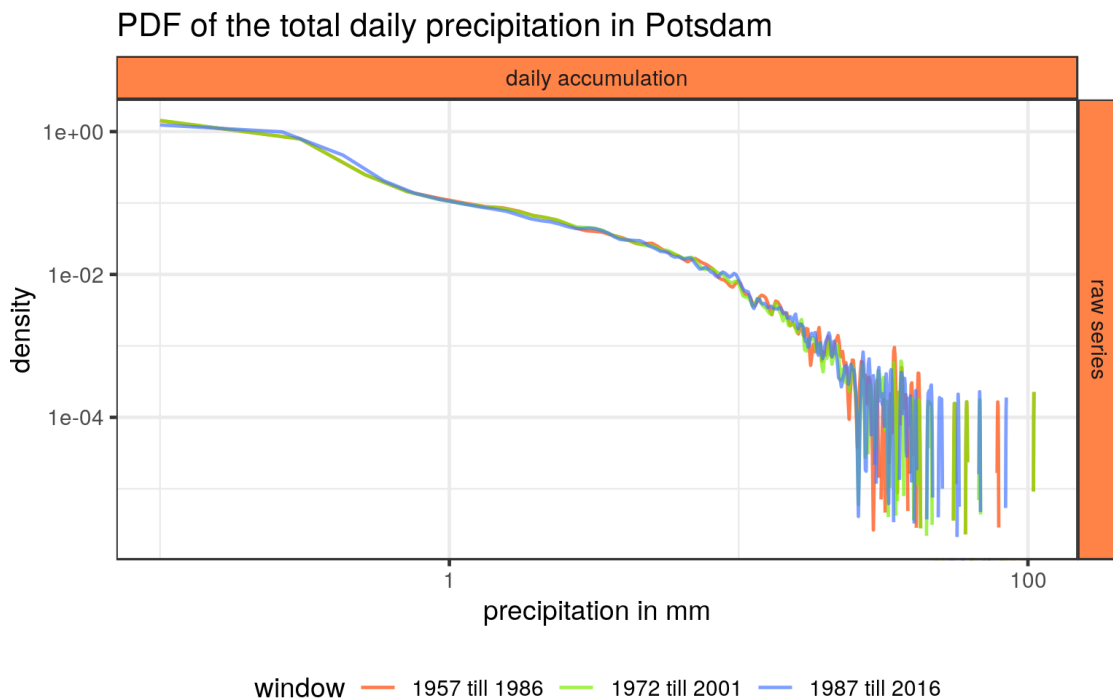


Figure 7.19.: Probability density function (PDF) of the daily accumulated precipitation at the Potsdam station obtained using a kernel density estimator. For the three densities only measurements within the corresponding time windows, highlighted in figure 7.17, have been used. In order to improve the visibility of this heavy-tailed curve, both the x- and y-axis have been transformed to logarithmic scale. Since the transformation of the x-axis would have removed the peak of the distribution, which resides at zero rainfall, a constant offset of 0.1 mm has been added to the whole series. This way, the peak is retained but shifted towards higher values. Please note that the flickering of the PDF at very high values is an artifact of the kernel density estimation. Since Gaussian distributions of fixed bandwidth are positioned at all observed precipitation, they only add up to a smooth curve at lower precipitation values. For very higher ones the bandwidth is too small and the probability to reside in two neighbouring Gaussians gets marginal. Since the tails of those solitary Gaussians would easily dominate the y-axis on a logarithmic scale, it was truncated at a density of 2×10^{-6} . Larger values of the bandwidth, on the other hand, would yield a smooth curve for the whole range of precipitation but distort the result at lower values.

Table 7.5.: Numerical estimation of the first four moments of the Potsdam daily accumulated precipitation series within the different time windows. The windows correspond to the colored bars at the bottom of figure 7.48.

window	mean	variance	skewness	kurtosis	type
1957 till 1986	1.61	14.82	6.51	92.58	daily accumulation
1972 till 2001	1.56	13.75	6.67	98.48	daily accumulation
1987 till 2016	1.60	14.56	5.71	62.18	daily accumulation

Table 7.6.: GEV parameters and some return levels for the Potsdam daily accumulated rainfall within three time windows. All quantities except of the shape parameter are given in mm and all error estimates are calculated using the Monte Carlo method.

window	location	scale	shape	10 year ret	100 year ret
1957 till 1986	27.0±1.7	9.4±1.4	0.29±0.21	57.0± 8.6	119± 65
1972 till 2001	26.2±1.8	9.2±1.6	0.31±0.19	56.3±11.3	121±162
1987 till 2016	28.3±2.1	9.6±1.8	0.23±0.20	56.4± 8.0	106± 46

third time window, everything seems to remain stationary. But the decrease is just occurring in one time window preceded by a slight increase in both the skewness and kurtosis. Therefore, no conclusion can be drawn for the temporal evolution of those parameters without analyzing the whole set of precipitation series throughout Germany.

7.6.2. Non-stationary EVA and significance

Figure 7.20 and table 7.6 show the results of the time window analysis for the daily accumulated precipitation data. Similar to the first four moments, the resulting GEV parameters and the derived return levels remain more or less constant. Only in the shape parameter in the third time window a major decrease does occurs, which is probably due to the corresponding decrease of both the skewness and kurtosis of the underlying series. As a result the 100 year return levels drop by more than 10 %. But for both quantities the error estimates are larger than the actual changes. Therefore, no conclusion can be drawn despite of that more data would be required to analyze the extreme events of the daily accumulated rainfall.

Figure 7.21 shows the resulting 10 (bright red) and 100 year return levels (dark red) obtained in the VGLM analysis. Surprisingly, both of them feature a slope, which is opposite compared to the one of the time window approach. While for the VGLM approach we find a general increase in the return levels, being even more pronounced the larger the return period, the corresponding results in the time window approach do decrease on a similar rate. How is this possible? The extreme value analysis, especially the estimation of the shape parameter, is very sensitive to the largest block maxima/minima. The lesser data is provided to the fitting procedure, the more difficult it is to trust the results of the analysis. With only 30 points at hand the stationary EVA performed in the individual time windows does see a stationary time series and is dominated by

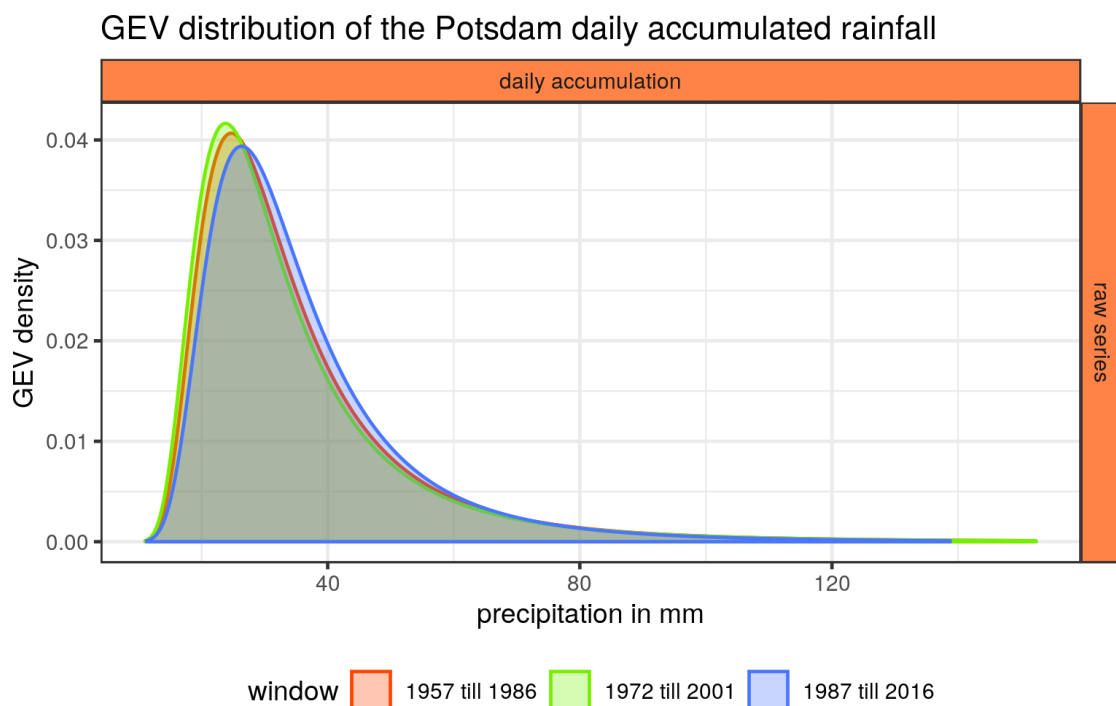


Figure 7.20.: GEV distributions constituted by the parameters fitted within the different time windows. To generate this figures, firstly, the annual maxima of the daily accumulated precipitation (see figure 7.17) were extracted. Afterwards, the series were split into three time windows and to all block maxima within one window a stationary GEV distribution was fitted.

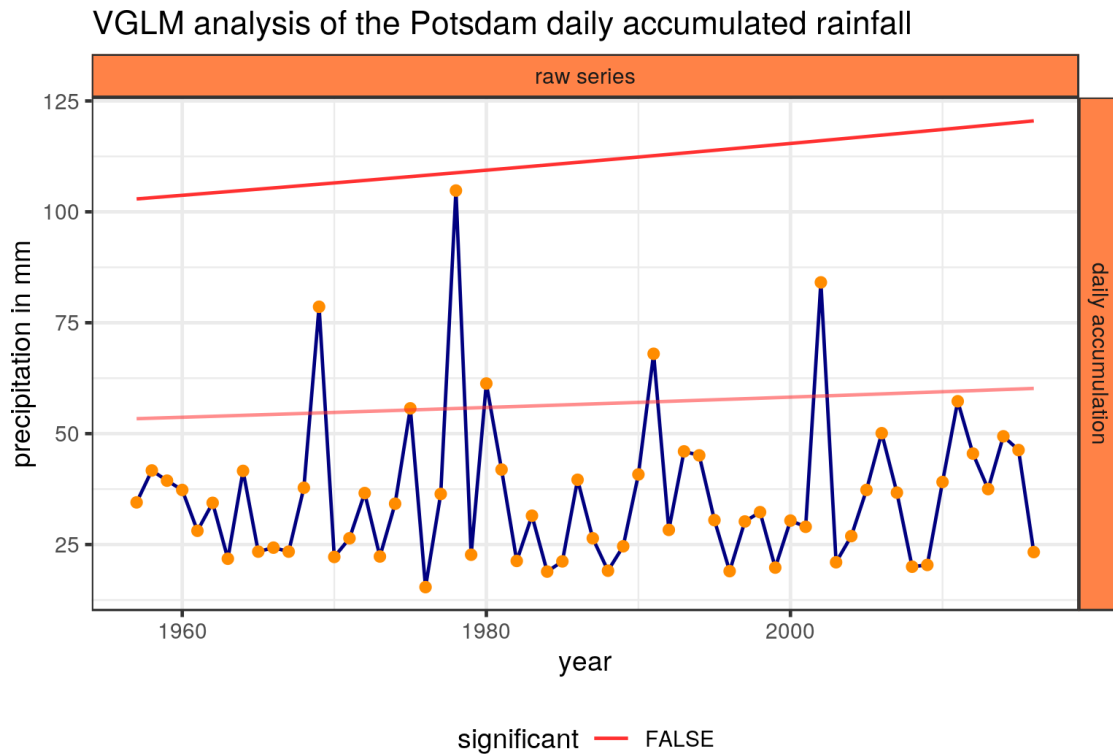


Figure 7.21.: This figure shows the results of the VGLM analysis for the Potsdam daily accumulated precipitation. At first, the annual maxima of the raw rainfall series (see figure 7.17) were extracted. They are plotted as orange dots and connected by blue lines. Into these block maxima a non-stationary GEV distribution with a first order linear trend in time in both the location and scale parameter was fitted using the framework of the VGLM (see section 5.2.1). The 10 (soft color) and 100 year (dark color) return levels were extracted and added to the series. Finally, the fit of the non-stationary GEV distribution was compared to a stationary one using the likelihood ratio test (see section 5.4.2). If the null hypothesis of describing the block maxima using the stationary model was rejected at a 5% significance level in favor of the non-stationary one, the lines of the return levels were plotted in green. If, on the other hand, the null hypothesis could not be rejected at this significance level, they were plotted in red. Note that while the trend in both the location and scale parameter is indeed a linear one, the resulting return levels do not have to be a straight line (see equation (2.1.1)).

the largest value in the data set. Since the largest one is occurring in 1978 and thus is present in both the first and the second time window, their results are almost identical. The reduction of the shape only occurs due to the absence of the 1978 event and the parameter is now dominated by the second largest one. This highlights the shortcomings of the time window approach since in a stationary picture both the largest and second largest event should be able to occur in almost all possible combinations without affecting the results of the analysis. The VGLM approach, on the other hand, is more robust since it incorporates all 60 annual block maxima of the time series. Therefore, it is able to fit the overall trend (the mean value of the block maxima between 1957 and 1986 is 35.96 mm and increases to 36.15 mm between 1987 and 2016) without being affected too much by the largest events in the series. Note that we are still using a first order linear model in both the location and scale and no trend in the shape parameter. But this linear model, responsible for the trend in the return levels in figure 7.21, does not appear to yield a more appropriate model of the block maxima than the stationary one. The null hypothesis could not be rejected and the likelihood ratio test was not passed. Or in short, the VGLM analysis finds the time series to be stationary, which is consistent with the large fitting errors found in the time window approach.

7.7. Temporal evolution of the precipitation throughout Germany

When investigating the precipitation series of the Potsdam station we couldn't draw any conclusions about their temporal evolution except that we do not have enough data to draw statistically sound conclusions. What does change when we incorporate all other stations and plot the resulting return levels altogether? Please note that in contrast to the distributions of the temperatures shown earlier in this chapter the ones of the precipitation data in figure 7.23 do not feature different climatological quantities. We do not distinguish between the maximum and minimum daily values or apply different preprocessing. Instead, only the results for the accumulated daily precipitation are shown; the 10 year return levels to the left and the 100 year ones to the right.

Compared to the findings for the Potsdam series (marked using vertical lines) one does indeed see some overall changes. There is a general tendency at both return periods for the distributions to acquire more mass at lesser values and in return to obtain lighter tails. Also the spread of the distributions is decreasing, especially for the 10 year return levels. Thus, the extreme precipitation events are getting smaller in magnitude due to the climate change. This reflects the findings in the time window analysis of the Potsdam series but is in contrast to the results of the corresponding VGLM analysis. The plot of the actual changes of the individual series throughout Germany will shed some light into this matter.

Figure 7.23 shows the spatial distribution of both the return levels in the last time window and the differences between the third and first time window. On average the decreases in the return levels are more prominent than the increases. But while there is a tendency for the increases to occur in the middle and north of Germany these patterns are quite noisy and probably due to more complex dependencies, like the local elevation profile. The whole analysis can be found in A.3. Worth mentioning is the finding that the

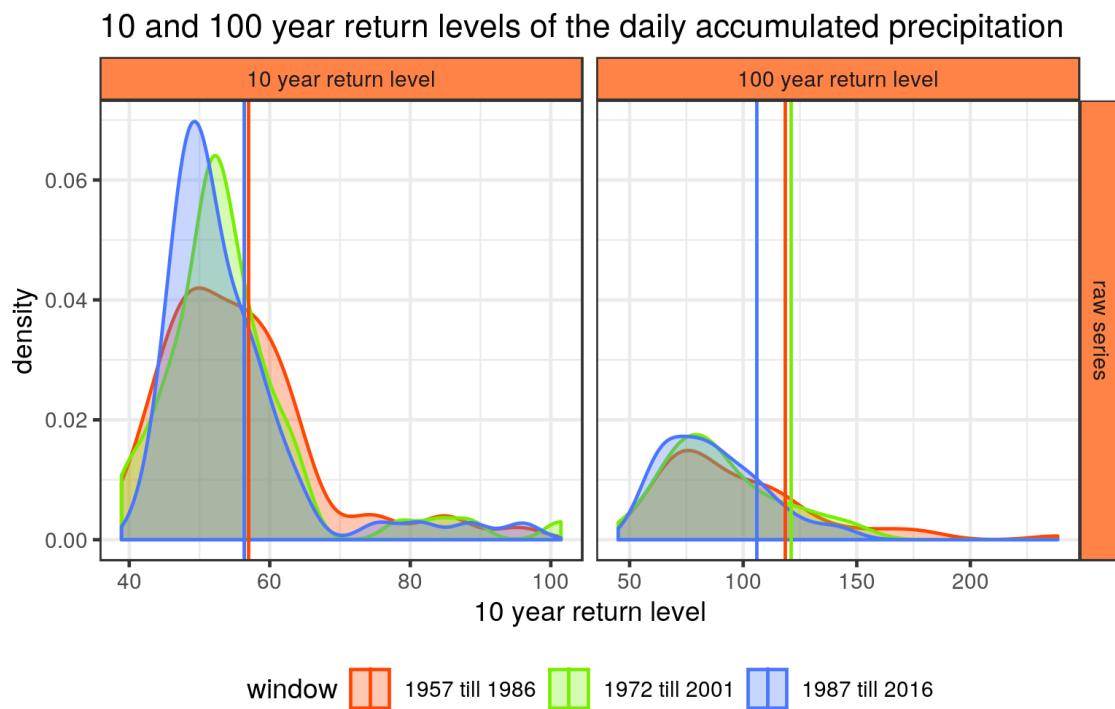


Figure 7.22.: Distribution of all 10 year return levels for the daily accumulated precipitation throughout Germany. For details of the preprocessing applied to the 57 different contributing series please see section 7.1 and 7.2.

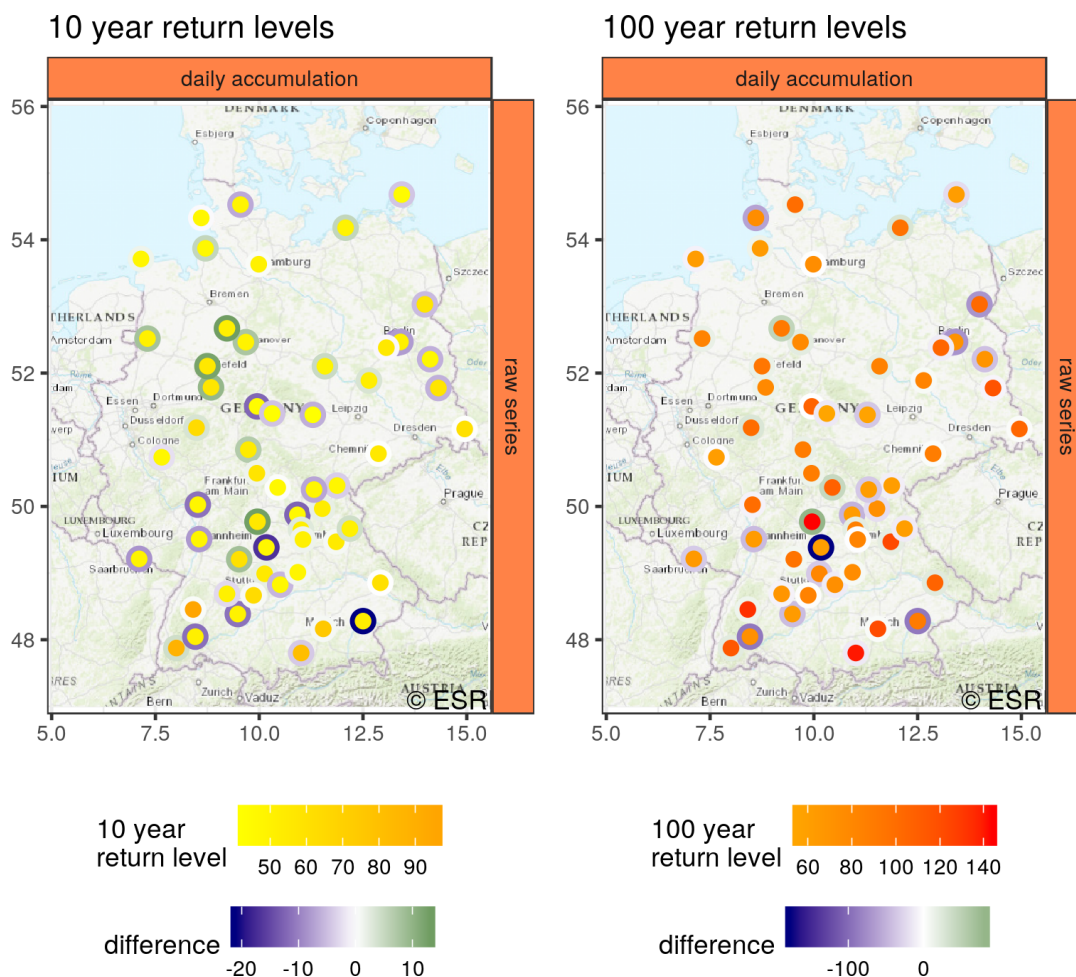


Figure 7.23.: Map of all the individual 10 year return levels of the daily accumulated precipitation contributing to figure 7.22. For a detailed description of the figure and colors please see section 7.1 and 7.2.

first and second, as well as the third and fourth moments of the individual precipitation series are highly correlated. A fact, which might be especially interesting for people modelling the distribution of rainfall in order to feed simulations.

In contrast to the findings of the temperatures only four series fail the likelihood ratio test in the VGML analysis displayed in figure 7.24, one showing an increase in the extremes and three showing a decrease. But here it is not at all obvious that we should expect trends in the precipitation. After all, the rainfall over Europe is supposed to remain constant. So, why should the corresponding extreme events do? Well, for one we have to reject the hypothesis tests of four different stations and there is an even bigger number of stations with a deviance statistic quite close to the threshold corresponding to the 5% significance level. In addition, the variance of the precipitation data is larger than the one of the temperatures and we would thus need longer time series to detect a trend of the same magnitude. But in the end we have no way to tell whether this is the case or if the precipitation series are indeed stationary. To conclude, we do find four stations featuring a statistically sound trend in their extreme values and need more data to verify the presence or absence of temporal changes in the remaining series.

VGLM analysis for the daily accumulated precipitation

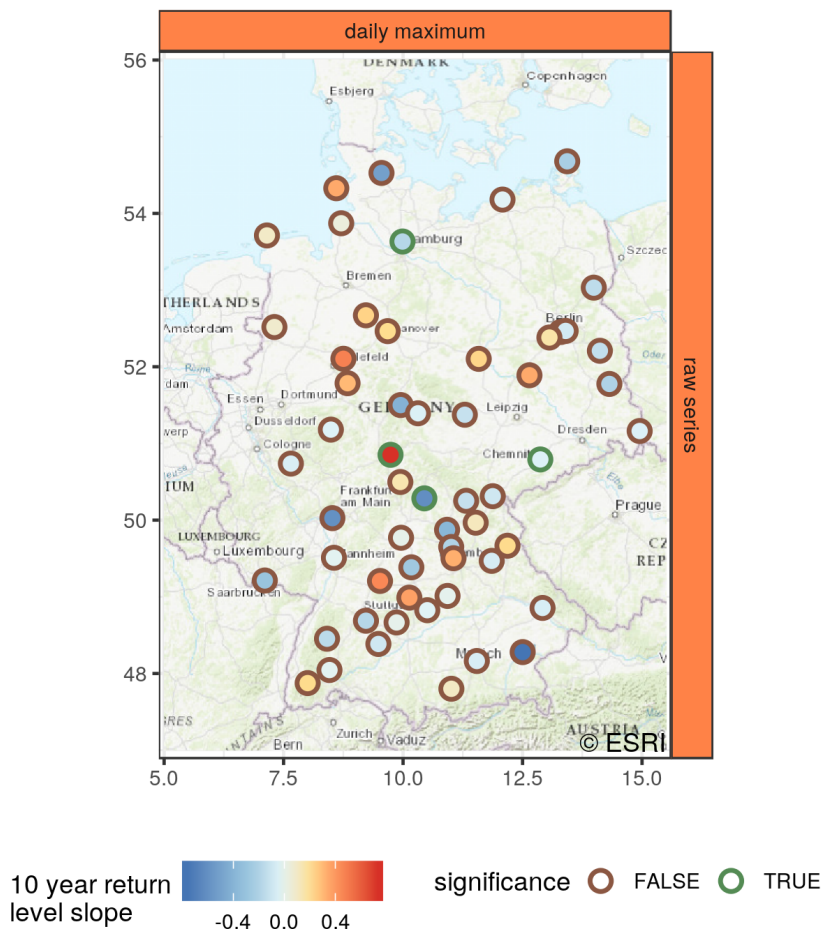


Figure 7.24.: Results of the likelihood ratio test whether to use a stationary model over a first order linear one for the daily accumulated precipitation. All stations, for which the null hypothesis of a better description with a stationary model was rejected, are enclosed by a green color. The circles are filled with the slope of the 10 year return levels fitted using the VGLM. Since it has a non-linear dependence in time the slope was estimated by fitting a straight line in its curve. This value might therefore not yield a perfect approximation of the actual temporal evolution but it is sufficient to determine its overall tendency.

8. Application of the non-stationary EVA to the ERA-Interim data set

In the previous chapter we analyzed the station data provided by the German weather service (DWD) in order to draw conclusions about the temporal evolution of the extreme events in the daily temperature and precipitation and to check the performance of our modified algorithm on a large data set. But both aspects we didn't cover exhaustively yet. The climate system spans the whole globe and picking a detail as small as Germany might yield misleading results. Our world is connected not only climatologically but also on an environmental, social, economic, and political basis. To properly access the changes on regional scales, we have to embed them into the global one. Moreover, the well-oiled analysis of the up to 74 stations provided by the DWD was a nice result and isn't matched by any package providing the error-prone default optimization. But the number of involved stations is far too small to accurately show the robustness of our improved software numerically. Both problems do have the same solution: the analysis of the ERA-Interim reanalysis data set.

In this chapter I will first introduce the overall concept of a reanalysis data set and the peculiarities of ERA-Interim one. Afterwards, we compare the findings of the DWD analysis with the ones of the ERA-Interim data set and highlight advantages and shortcomings of both methods. Finally, we will have a look at some selected results of the time window and VGLM analysis of the global data set. For a thorough treatment of the preprocessing and steps of the analysis please see section 7.1 and 7.2.

8.1. The concept of reanalysis

The overall idea of the reanalysis is to use all sorts of measurements throughout the world, project them onto a global grid, and interpolate between the measurement steps using a state-of-the-art weather model. It forces the simulated weather to obey the laws of physics and at the same time to match the observational data as close as possible. This enables the user to not just work with the atmospheric variables incorporated as measurements but also to access additional information about the state of the atmosphere. If e.g. only sea-surface temperature and pressure measurements are present and used to feed the weather model, the resulting reanalysis data set will feature temperatures and pressure values at a large number of different heights and at places not covered by the measurement itself. Also additional so-called derived quantities, like the cloud coverage or the daily accumulated precipitation, can be accessed.

But the reanalysis data is of course no free lunch. The dynamics of the weather model are intrinsically chaotic. Even when forced to match the global observations one day and the next it is free to do in the 24 hours in between whatever obeys the laws of physics. In addition, the description of the climate system, although improving in time, are far from

perfect. The dynamics of the atmosphere do cover a very large set of different spatial scales and a lot of them are smaller than the resolution of the underlying grid. The probably most prominent example is the formation of clouds. Some quantities, like the daily accumulated rainfall, do therefore only present very vague approximation of the true state of the climate system and are not meant for direct statistical analysis. Others, like the two metre temperature used in the remaining part of the chapter, do yield a very plausible approximation since they were part of the measurement too.

To be able to perform the extreme value analysis (EVA) or other statistical methods on the global data, we need the series at the individual grid points to be as long as possible. But since the state-of-the-art weather models are constantly under development inconsistencies would arise from the change in model. In order to meet both requirements and provide long series of best possible quality, big climate centers do use all their historical data in the reanalysis run of a new version of their weather model. In time, they will produce a large set of consistent and comprehensive data sets of the past atmospheric state referenced to a specific version of a weather model. One of such centers is the European Centre for Medium-Range Weather Forecast (ECMWF), which provides, among others, the ERA-Interim reanalysis data set (see Berrisford et al. (2011) for information about the data retrieval).

8.2. The ERA-Interim data set

The two most important ingredients for producing the ERA-Interim data set is the data assimilation scheme, which is used to project all the different sources of measurements onto a global grid, and the atmospheric weather model. Both components are highly complex and their thorough treatment would go beyond the scope of this chapter. Instead, only an appetizer will be served. The assimilation scheme uses, among other, clear-sky radiance measurements of different satellites and surface pressure measurements obtained in direction observations. It is called 4D-Var since it is not only able to assimilate all these different sources onto a global grid but also to adjust the correction biases in an completely automated scheme. The correction biases are required to make the different measurements as consistent as possible. The atmospheric model is of version Cy31r2 and incorporates e.g. surface drag, ocean waves, the geometry of land surface, and cloud convection. For more details on both components please see Dee et al. (2011).

The data set starts in 1979, when the global observing system TIROS-N was launched, and is constantly updated to also cover recent times. Since we are only interested in complete years of data, the resulting series will have a length of 39 years in total. They will be split in two overlapping time windows of a length of 25 years. As already discussed in the previous chapter 25 data points, representing the annual maxima or minima, are not that much when fitting a distribution constituted by three parameters. But, on the other hand, windows of 30 years would result in a way too large overlap cancelling most of the changes within the system. Having at least 25 points and a modest overlap seems to be an appropriate trade-off. Unfortunately, the time windows do not coincide with the ones used in the DWD analysis. But the second one of the DWD from 1972 till 2001 and the first one of the ERA-Interim analysis from 1979 till 2003 as well as the last ones from 1987 till 2016 (DWD) and 1993 till 2017 (ERA-Interim) match reasonably well.

The spatial resolution of the underlying grid is approximately 80 km and composed of 115680 grid points on 480 different longitudes and 241 different latitudes.

We will use the temperature of the air at two meters above the surface of land or water. It is provided for four time points during the day, midnight, 6am, noon, and 6pm. From those four different values the daily maximum and minimum temperature will be approximated by selected the highest respectively the lowest value. This is of course only a crude approximation of the true daily temperature maximum and minima measured at the stations of the DWD but it will serve the purpose. For both daily series the temperature anomalies will be calculated (see section 4.3.1) and, finally, the block maxima of the daily maximum temperature and temperature anomalies and the block minima for both quantities were extracted. For all four series per grid point the time window and VGLM analysis discussed in chapter 7 will be performed and some selected results will be reported.

One open question still remains. Why did we choose the ERA-Interim data set above all other products provided by the ECWMF? To not digress too much from the main subject of this chapter, I will just answer the question with respect to two data sets, which sound even more promising for the application of the EVA: the ERA-20C and the CERA-20C data sets. The model driving the latter one does not only describe the dynamics of the atmosphere but also those of the ocean and couples both systems (see Laloyaux et al. (2016)). This results in a more complete approximate of the overall climate system and more realistic reanalysis data. In addition, both the ERA-20C and CERA-20C do span the whole 20th century and thus provide a lot of data we could use in our non-stationary EVA. But this vast amount of data is not of the same quality for all points in time. It features an assimilation scheme (Poli et al. 2015), which is able to smoothly incorporate observational data covering only a limited subset of the overall length of the data set. For the first years of data thus only intake ship and buoy measurements are used to feed the model. With no observations within Germany used at all the data will of course feature an error in the resulting temperature series rendering our analysis useless. Therefore, we will use the shorter but more accurate ERA-Interim data set over the longer ones. In principle the usage of CERA-20C data set restricted to the years 1979 till 2017 should yield the most appropriate data set. But this would result in a lot of confusion for people being triggered by the CERA-20C buzzword. So, again, we will use the ERA-Interim reanalysis data.

8.3. Comparability between the DWD and ERA-Interim data sets

In terms of accuracy and quality the time series provided by the DWD are second to none and the ideal starting point to gain insights into the distribution of the climate in Germany. Then again, the weather in Germany is only a small part of the global climate system. It is therefore important to get a general idea about the changes in the temperature distribution throughout the world, which will be done using the ERA-Interim reanalysis data set. But beforehand we have to check the quality of its results by comparing them to the ones of the DWD series. After all, the underlying grid is quite coarse-grained and the atmospheric model might spoil the statistics of the temperature

series.

Figure 8.1 shows the results of the time window analysis of the both ERA-Interim (background) and the DWD (circles) series, which share the same color scale. For the DWD data the values of the shape parameter correspond to the last time window from 1987 till 2017 and for the ERA-Interim data set the last window spanning from 1993 till 2017 was used as well. Both plots show the results for the daily maximum temperatures with the ones for the raw series on the left and the ones for the temperature anomalies on the right. After starring at the results for the raw series for quite a while ones starts to see some agreements. For both data sets the highest values can be found in the south of Germany and the values in the north and in a small spot at the very south tend to be lesser than the average. Apart from these tendencies the results of the ERA-Interim analysis can not properly reproduce the ones of the DWD and there is quite some noise in the patterns. But not as much as for the daily maximum temperature anomalies. In there one has to believe in patterns in the first place to actually find them in the figure.

The results of the shape parameters are not the end of the story yet. From experience, which I obtained in implementing the Monte Carlo error estimates presented in section 2.2.3.2, the GEV parameters of similar time series are able to compensate each other to still yield almost the same return levels for medium size return periods, like 10 years. So let's compare the 10 year return levels for the daily maximum temperatures in the last time window presented in figure 8.2. While the results obtained for the ERA-Interim data set do not match the ones of the DWD perfectly, the overall character is preserved and the fluctuations are smaller than the ones for the shape parameters. The station located in the German Ocean in the north west does feature very small values for both data sets. But the remaining stations at the coast are less affected by the water than the corresponding regions in the ERA-Interim data set. Other general patterns are not matched by the reanalysis data. Apart from the overall correspondence of the results for the raw series, there are six stations in the middle and south of Germany, which differ from the ERA-Interim results by a margin. All of them are located at higher altitudes and thus their smaller return levels are due to the local geographic peculiarities of the station. Since the reanalysis data set is based on a grid of approximately 80 km, the local elevation profile is smoothed out and measurements obtained at the top of the highest mountain within the corresponding region are bound to yield different results. As a conclusion, the findings of the time window analysis obtained for the ERA-Interim data set do match the ones of the DWD data reasonably well when restricting the discussion on moderate return levels. What about the VGLM analysis?

Figure 8.3 shows the results of the VGLM analysis. The color scale represents the deviance statistic, a measure for the appropriateness of the linear model with respect to the stationary one. It is arranged in such a way that blue and green correspond to values where the null hypothesis of using a stationary model instead of one with a first order linear model in both the location and scale parameter could not be rejected. The colors yellow, orange, and red indicate a rejection and thus showing the existence of a trend. The circles, representing the results for the DWD stations, do have an outer color of red if the test could be rejected and a blue one if this was not the case. See section 5.4.2 for a more detailed description of the so-called likelihood ratio test.

Before we start interpreting the figure an important difference between the two data sets has to be highlighted. The VGLM analysis uses the full length of the series, which

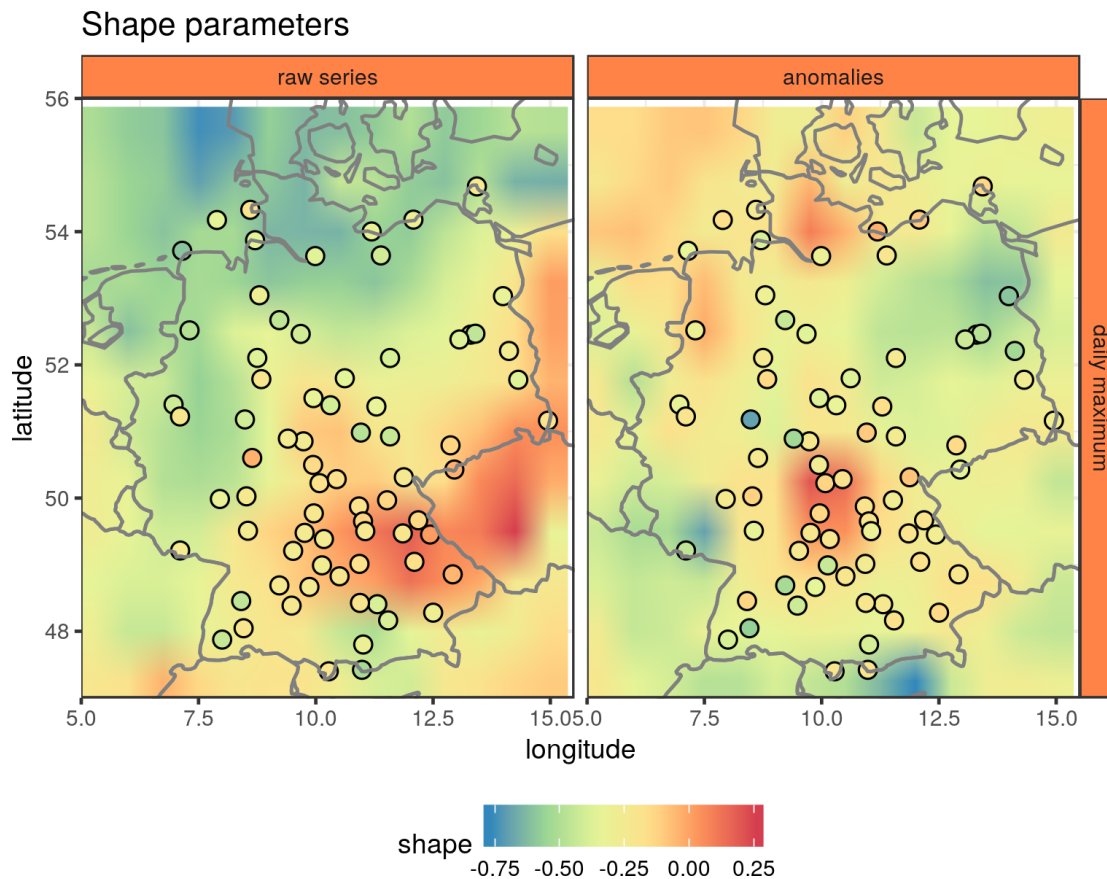


Figure 8.1.: A comparison between the shape parameters obtained in the analysis of the DWD and ERA-Interim data set. For both the raw series (left) and anomalies (right) of the daily maximum temperature were used. Plotted are the shape parameters of the last time window. This spans from 1993 till 2017 for the ERA-Interim set and from 1987 till 2016 for the DWD series. The results of the reanalysis data set is displayed using colored tiles and constitutes the background of the image with the borders of Germany and its neighbours plotted on top. The shape parameters for the stations provided by the DWD are plotted in circles located at its geographic position. The same color scale was used for the results of both data sets. For details about the preprocessing of the DWD data see section 7.1 or section 8.2 for the ERA-Interim data set.

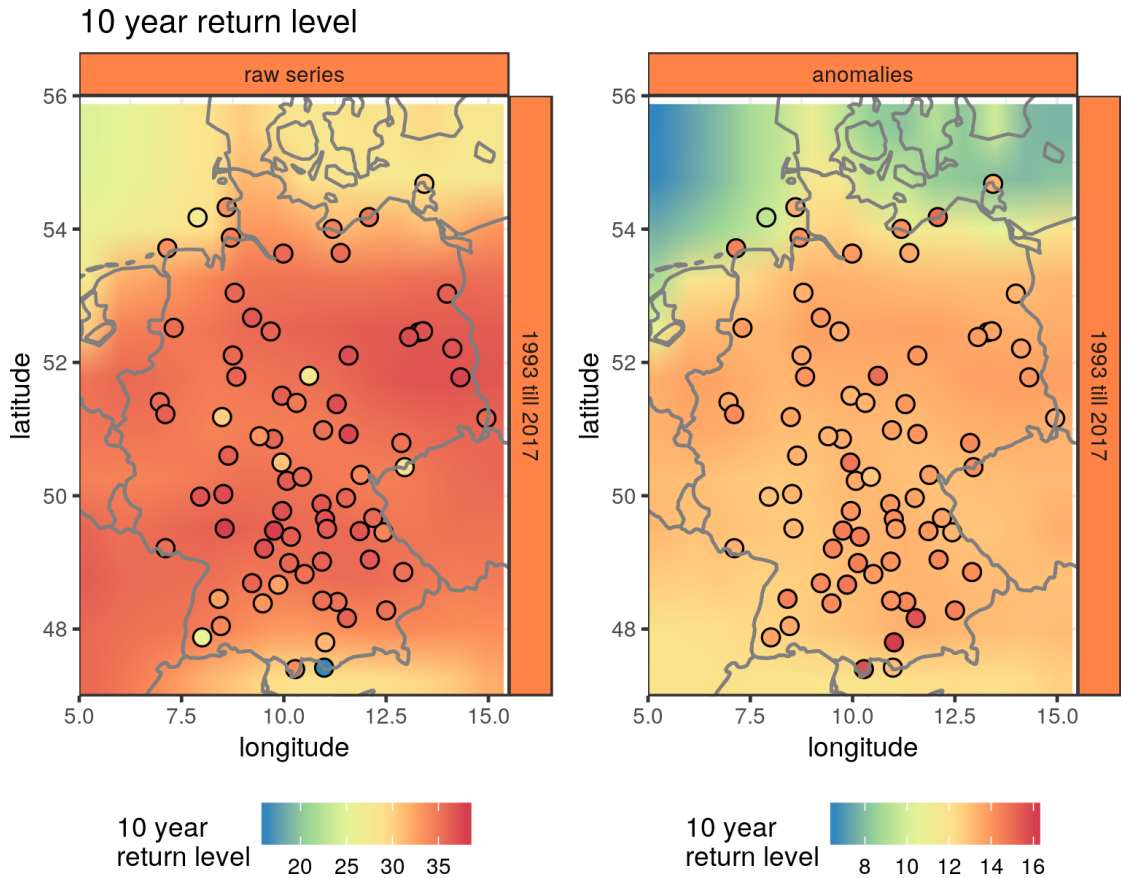


Figure 8.2.: Comparison between the 10 year return levels obtained for the the daily maximum temperatures of the DWD station series and the ERA-Interim reanalysis data set. For details about the structure of the figure please see figure 8.1 and for more information on the preprocessing section 7.1 (DWD) or 8.2 (ERA-Interim).

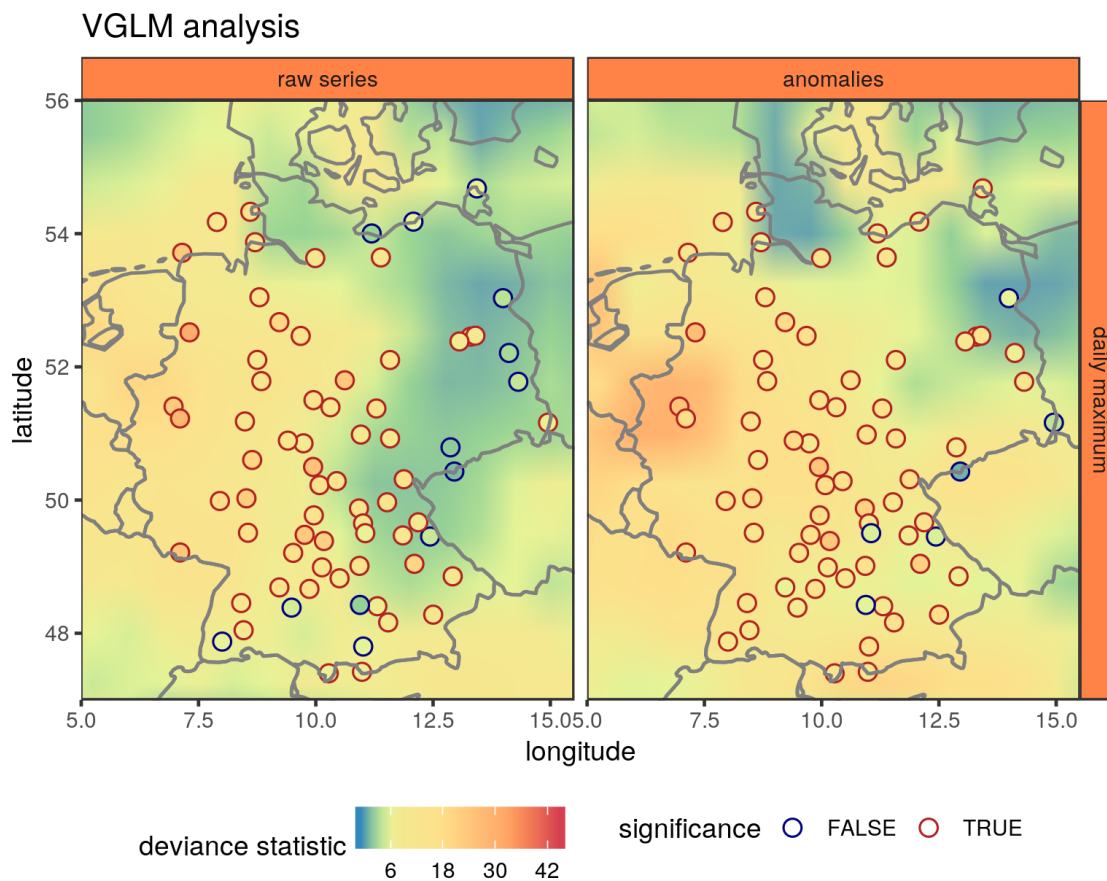


Figure 8.3.: Comparison of the results obtained in the VGLM analysis of the daily maximum temperatures DWD station series and the ERA-Interim reanalysis data set. The structure is quite similar compared to figure 8.1. But here the color scale is not equivalently spaced but stretched in such a way color blue and green correspond to a deviance statistic the likelihood ratio test could not be rejected with. For the colors yellow, orange, and red a significant trend in the extreme temperatures could be found. To ease the distinction between both groups in case of the DWD station series, the corresponding circles do feature a specific outer color. If blue, the hypothesis could not be rejected. If red, it could. For more information on the preprocessing please see section 7.1 (DWD) or 8.2 (ERA-Interim)

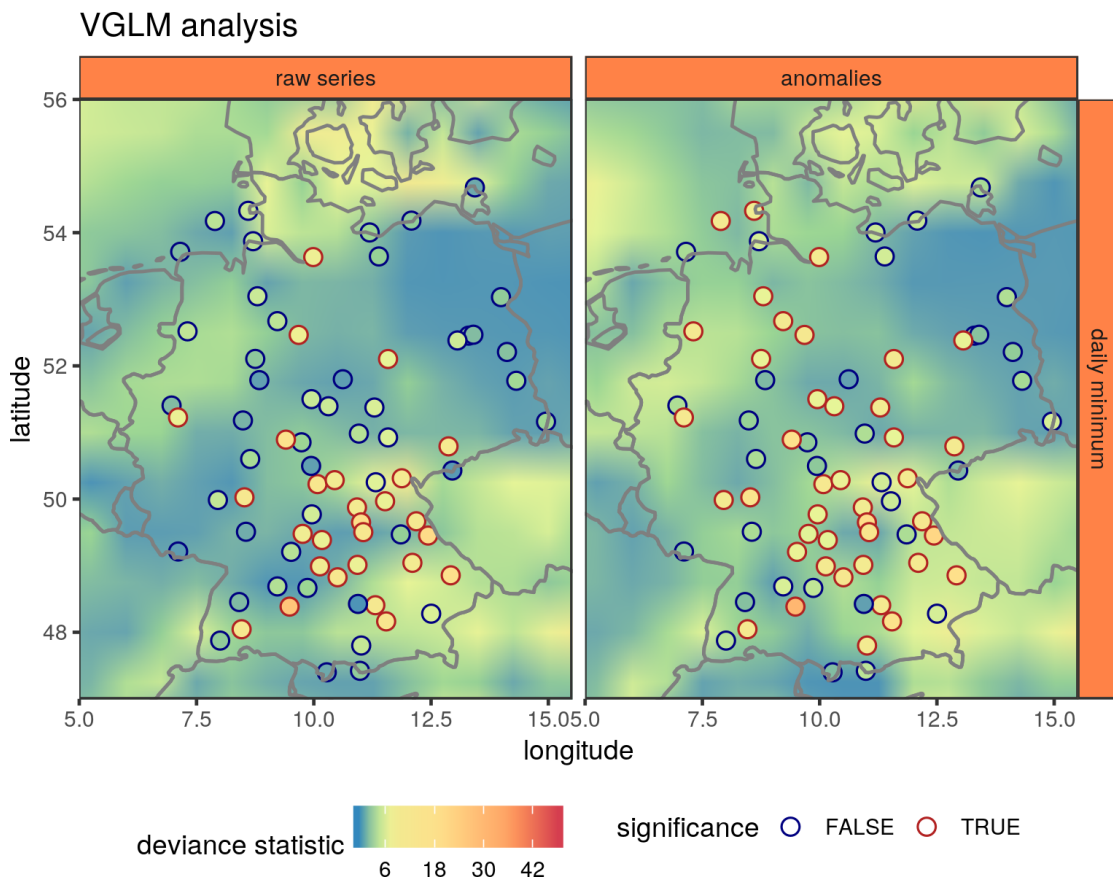


Figure 8.4.: Comparison of the results obtained in the VGLM analysis of the daily minimum temperatures DWD station series and the ERA-Interim reanalysis data set. For more information please see figure 8.3 or sections 7.1 (DWD) or 8.2 (ERA-Interim).

spans 60 years in the DWD data set but only 39 in the ERA-Interim one. Even if both sets would feature the same trends in the extreme temperatures and the same level of fluctuations, the results might still differ since the reanalysis series are too short. The longer a time series, the easier it is to find trends in the presence of noise. The results for the raw series are in almost perfect agreement. For both data sets the majority of the hypothesis tests could be rejected and only in the east, the very north, and the outmost south west corner of Germany no significant trend could be found in the extreme temperatures. In addition, a number of stations, especially in the east, do feature significant trends but the corresponding region in the ERA-Interim data set does not. This is probably due to the shorter length of the reanalysis data set as already discussed. For the daily maximum temperature anomalies the consistency is not that evident. Five out of the six stations, for which the hypothesis could not be rejected, are located in a region the results of the ERA-Interim analysis do show a significant trend in. On the other hand, for almost all stations a significant trend could be found and this overall feature was very nicely reproduced.

The reader might wonder why all the figures above exclusively cover the results for

the daily maximum and not for the daily minimum temperature series. For one, this is to make the chapter more concise. A thorough analysis of climate data is presented in appendix A and the purpose of this chapter is solely to embed the findings of the DWD analysis in a global scale. But in addition the results for the daily minima are much more noisy as can be seen in figure 8.4. Most significant stations are located in regions of brighter color but the hypotheses of stationarity of the ERA-interim series could barely be rejected. Therefore, only the results of the daily maximum temperatures will be discussed in the remainder of this chapter.

8.4. Time window analysis of the ERA-Interim data set

In figure 8.5 we can see the 10 year return levels of the daily maximum temperature anomalies obtained in the time window spanning the years between 1993 and 2017. It is composed 115680 results obtained for the individual series in the ERA-Interim reanalysis data set. The stationary fit of the GEV distribution in the time window approach could be performed for all stations without a single error thrown by the improved algorithm. The same holds true for the VGLM analysis in the next section. But since the latter relies on a different algorithm one might wonder if the constrained optimization was actually necessary. First of all, the fit of the VGLM model itself already features the linear constraints described in section 3.3. Without them a lot of fits would have failed. But even in their presence the routine only works for the full series as used in the VGLM analysis. When applying the VGLM fit on the individual time windows some will throw an error. This can only be prevented by the adding the non-linear constraint.

The actual patterns generated by the 10 year return levels do look quite smooth and very plausible. In most parts of the sea the extreme fluctuations from the climatology are comparably small. This is very reasonable since the oceans are supposed to act as a heat reservoir and temperature series of the sea do show a smaller variance than on land. The second lowest return levels can be found on land within the torrid zone, between 23.4° south and 23.4° north. The largest values do occur in the north frigid (66.5° till 90° north) and in the south frigid (66.5° till 90° south). Since the variability of the climate varies significantly between those zones, these findings do make a lot of sense.

The differences of the 10 year return levels for the temperature anomalies between the first and second time window can be found in figure 8.6. The largest positive one can be found in Siberia and the largest negative one in Antarctica. But apart from this no clear dependence on the latitude can be found, at least not symmetric with respect to the equator. The extreme deviations from the temperature climatology do decrease over most parts of the sea. On land one can see a general increase except in Canada, Russia and most parts of Africa.

In figure 8.7 the 10 year return levels for the raw series in the ERA-Interim data set can be found. The most extreme and hottest temperatures are found in the north of Africa and in Australia. Also one can find a gradual decrease of the return levels at larger latitudes. The lowest values are located in Antarctica and they are incredibly low. Remember that these values of less than -20°C correspond to the hottest temperature, which occurs on average only once every 10 years! In addition, both the Andes and Tibet are quite pronounced since they still feature a high elevation profile when over a grid

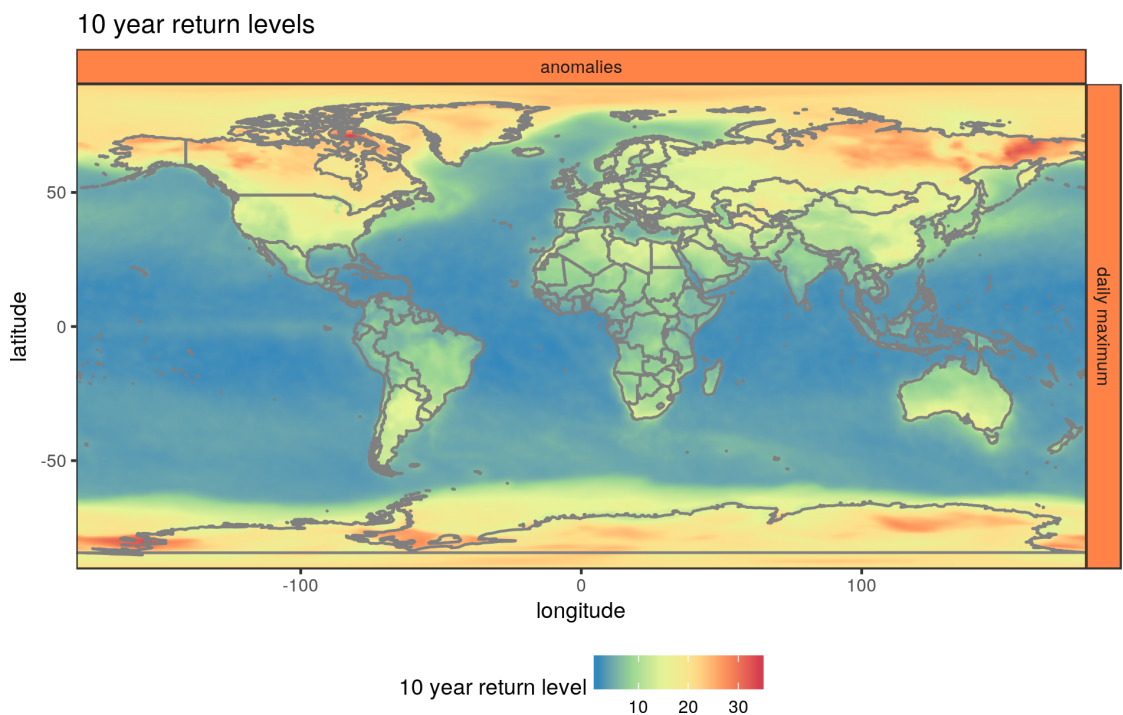


Figure 8.5.: 10 year return levels of the daily maximum temperature anomalies obtained for the ERA-Interim data set. The plotted values correspond to the ones found in the last time window spanning the years from 1993 till 2017. The individual pixels correspond to the results of the 115680 different grid points in the data set. To better grasp the spatial information, although the coastal lines are already quite prominent, the borders of all countries were added on top. For more information about the preprocessing please see section 8.2.

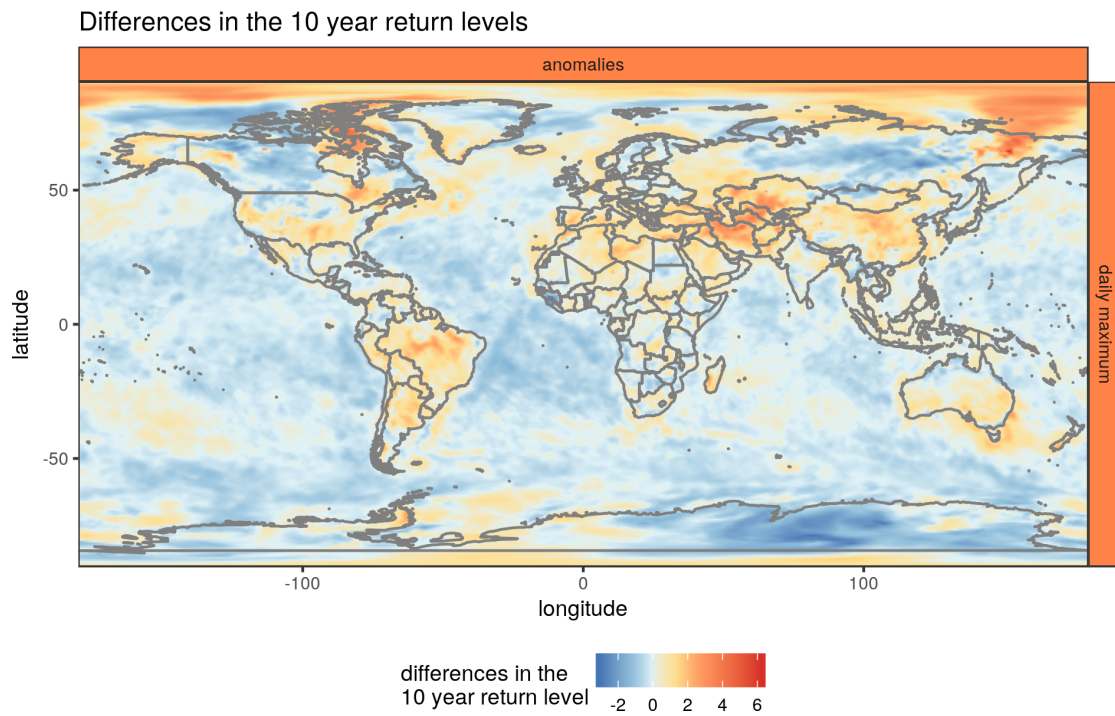


Figure 8.6.: Differences in the 10 year return levels of the daily maximum temperature anomalies obtained for the ERA-Interim data set. The plotted values correspond to the ones found in the last time window spanning the years from 1993 till 2017 minus the ones of the first window ranging from 1979 till 2003. For more information about the preprocessing please see section 8.2.

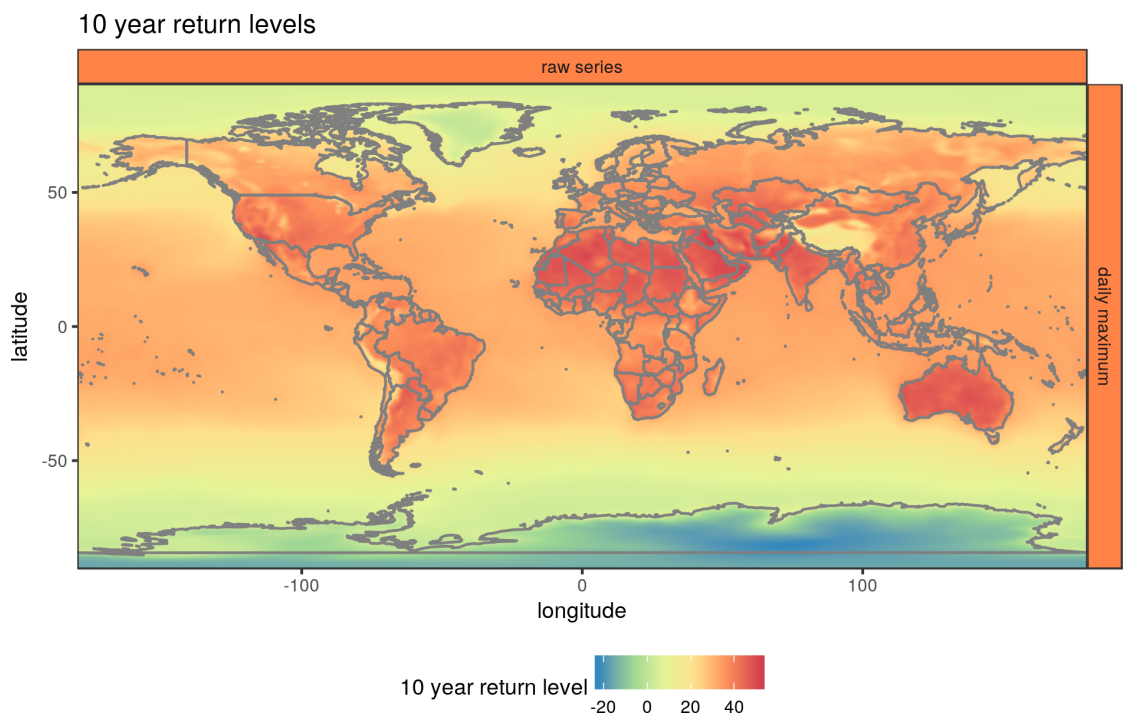


Figure 8.7.: 10 year return levels of the raw daily maximum temperature series obtained for the ERA-Interim data set. For more information about the preprocessing please see section 8.2.

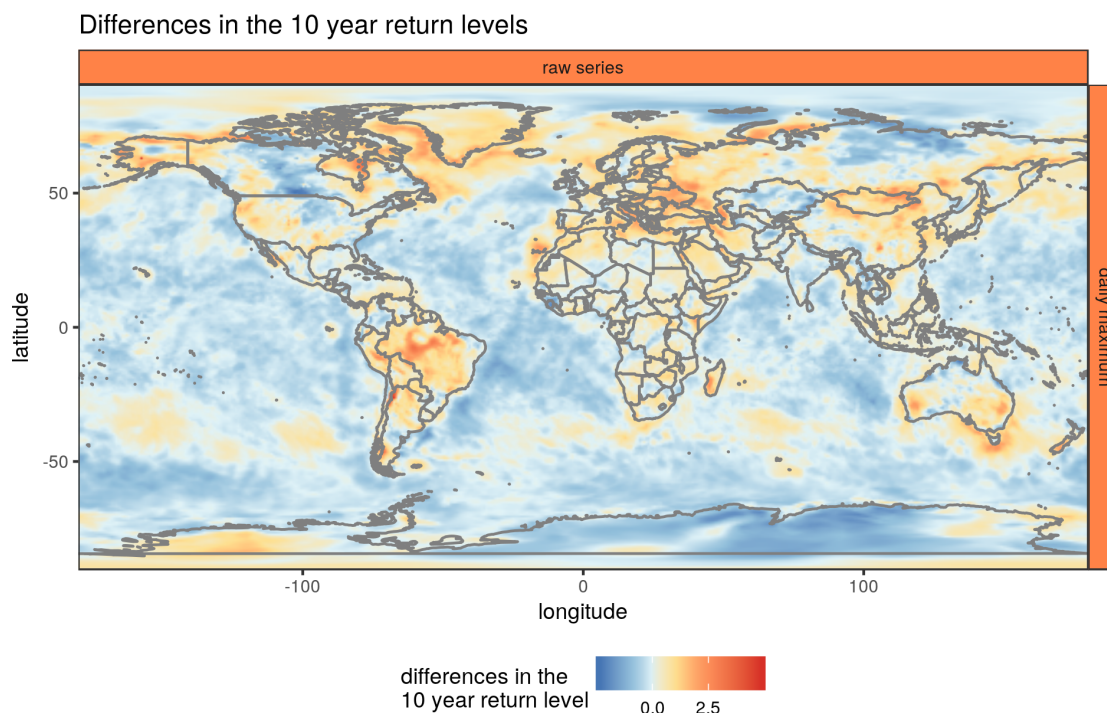


Figure 8.8.: Differences in the 10 year return levels of the raw daily maximum temperature series obtained for the ERA-Interim data set. The plotted values correspond to the ones found in the last time window spanning the years from 1993 till 2017 minus the ones of the first window ranging from 1979 till 2003. For more information about the preprocessing please see section 8.2.

length of approximately 80 km.

The changes in the return levels are shown in figure 8.8 and are almost identically compared to the ones of the temperature anomalies. Only now the region of Turkmenistan, Kazakhstan, and Afghanistan does feature a decrease instead of a large increase. In addition, the Mediterranean region and the east of Europe show an even more pronounced increase in the return level. It is, however, remarkable that on a large fraction of the earth's surface the return levels are decreasing.

8.4.1. VGLM analysis of the ERA-Interim data set

Surprisingly, the results of the VGLM analysis of the daily maximum temperature anomalies in figure 8.9 show a lot of structure. The color represent the deviance statistic and its scale is the same as in figure 8.3, which contains the results of the DWD analysis as well. The larger the value, the better the series can be described using a first order linear model compared to a stationary one. For a certain range of values, indicated by the colors blue and green, the likelihood ratio test of describing the data more appropriately using a stationary instead of a linear model can not be rejected. For higher values, plotted using the colors yellow, orange, and red, the hypotheses can be rejected and a statistically sound trend in the extreme temperature (anomalies) could be found. Most of

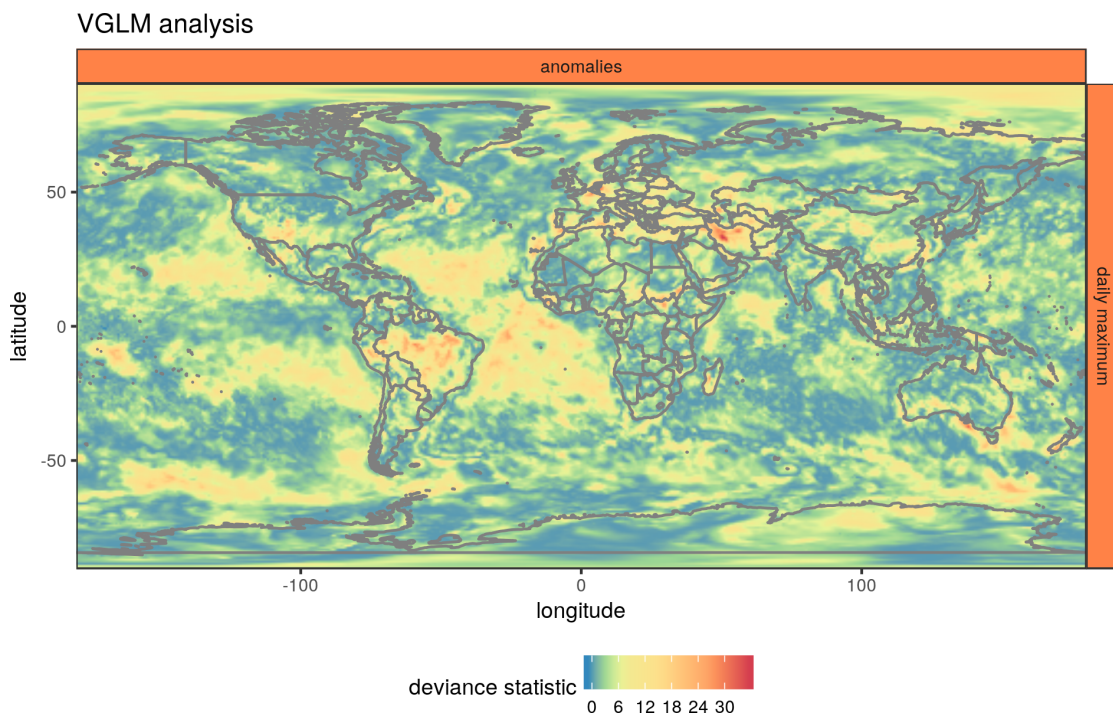


Figure 8.9.: Results of the VGLM analysis performed on the daily maximum temperature anomalies of the ERA-Interim reanalysis data set. As in the previous figures the obtained values are depicted in colored tiles with the borders of the countries plotted on top of them. Similar to figure 8.3 the color scale is stretched in such a way color yellow, orange, and red correspond to a rejection of the likelihood ratio test. For the color blue and green, on the other hand, no statistically sound trend could be found in the series. Apart from this binary interpretation of the results there is also a continuous one, the higher the deviance statistics the better the data is described with a first order linear model instead of the stationary one. For more information about the preprocessing see section 8.2 and for details about the likelihood ratio test see section 5.4.2.

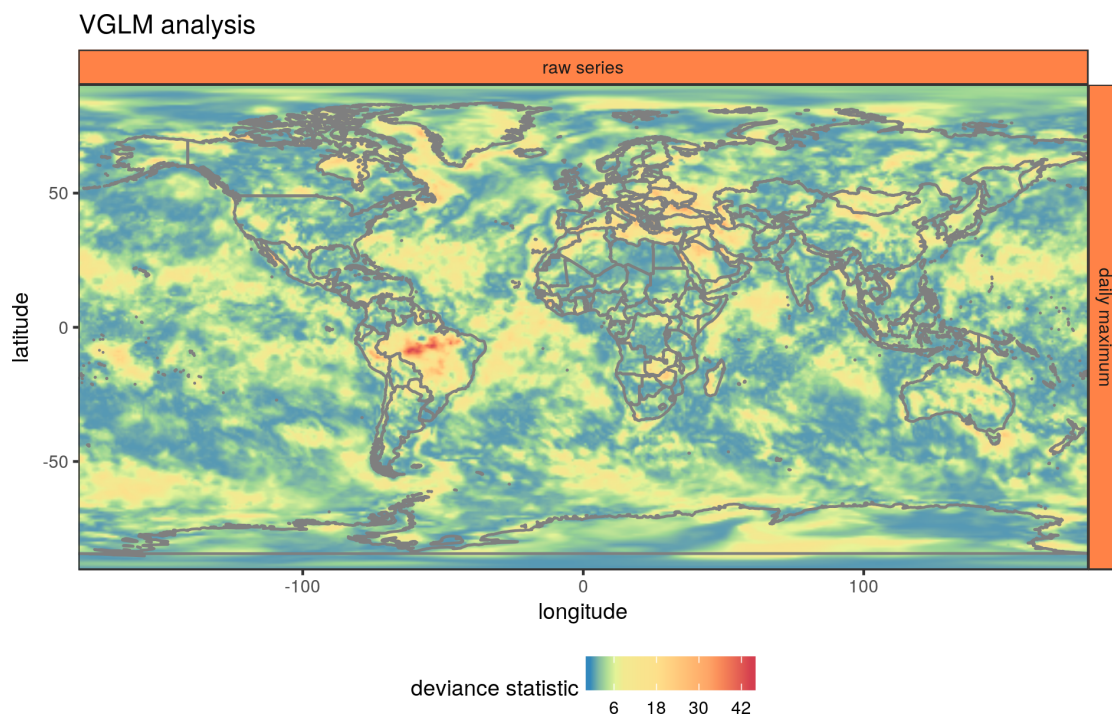


Figure 8.10.: Results of the VGLM analysis performed on the raw daily maximum temperature series of the ERA-Interim reanalysis data set. For more information about the preprocessing see section 8.2 or figure 8.9 and for details about the likelihood ratio test see section 5.4.2.

the globe does not exhibit such significant trends. But the majority of the found ones are located in the Pacific sea near Mexico or Peru and Chile or in most parts of the Atlantic sea. Also on land in the Mediterranean region, the middle east, and Brazil significant trends could be found.

The results of the VGLM analysis of the raw temperature series are, again, quite similar to the ones of the anomalies. They can be found in figure 8.10. The patterns of significant trends over the ocean are less pronounced while the ones on land appear more clearly.

9. Summary

Throughout this thesis I first understood and solved the problems of the available extreme values analysis (EVA) software packages, implemented a fixed version of the stationary fit for the generalized extreme value (GEV) and generalized Pareto (GP) distribution, and, finally, performed a large scale analysis on extreme temperatures and precipitation time series.

But let's start at the beginning. The main subject of investigation in this thesis is the temporal evolution of the extreme events in temperature and precipitation over Germany. It is well established that there is a climate change and earth is getting warmer. But apart from difficulties in pinpointing a single mean value to represent the state of the global climate system and in handling different local characteristics, most treatments of the climate change focus on average statistics. It is of course important to access the change in the first moments of e.g. the temperature distribution. But then again, it would be more than careless to not investigate the evolution of its tails. After all, it is the extreme events, like massive rainfall, extremely hot summers, or incredibly cold winters, causing the most harm to society. I therefore strove to perform a non-stationary EVA on a larger collection of station data provided by the German weather service (DWD).

It didn't matter which software package I was using to do analysis. For at least one of the stations the fit was failing and, as a result, rendering a massive parallel application impossible. After digging through a lot of source code I realized the problem was a conceptual one. The negative log-likelihoods of the GEV and GP distribution, which are used by the maximum likelihood (ML) fit in the EVA, contain several logarithms and thus give rise to a region in parameter space inaccessible to the algorithm. This itself wouldn't spoil the optimization yet, but, unfortunately, the boundary to this forbidden region features a steep slope towards smaller values before hitting the undefined domain. This causes the algorithm to occasionally get trapped and to not reach the global minimum of the negative log-likelihood. Either this boundary effect or the initialization of a parameter combination inside the forbidden region, apart from unrelated bugs, causes the software packages to sometimes throw an error and to abort the optimization.

To solve the numerical issues of the EVA, I introduced the constrained optimization using the augmented Lagrangian method. In order to allow other researchers to benefit from the findings in this thesis as much as possible, I provided an implementation of the fixed optimization routines via my software package `climex`¹, which is written in the statistical programming language R and published under GPL-3 license. Its well-set default arguments and collection of auxiliary functions will ease the heavy scripting duties associated with massive parallel analyses, like those covered in this thesis. In addition, I also introduced estimates for the fitting errors of the calculated parameters and return levels. They come in three different flavors: the well-known estimate based on the delta

¹The source code can be found in <https://gitlab.com/theGreatWhiteShark/climex> and a live example of the included web application can be accessed via climex.pks.mpg.de

method, which is used by almost all other packages to access the errors of the GEV/GP parameters and performs very poorly, and two new ones based on either the Monte Carlo method or bootstrap.

Only after fixing the numerical issues of the maximum likelihood optimization the analysis of the two climatological data sets, the aforementioned station data of the DWD and the ERA-Interim reanalysis data set embedding the findings of the first part of the analysis in a global context, could be performed. While for the former one it would be still feasible to adjust the initial parameter combinations of the failing fits by hand, the analysis of the latter would not be possible without the introduction of the constrained optimization. The analysis itself was done in two different ways. Firstly, in the so-called time window approach, I separated the individual series into several overlapping time windows and performed a stationary EVA in each of them. This means, I segmented the series within the time windows into annual blocks, extracted their maximal or minimal values, and fitted the GEV distribution to the resulting events. Next, the GEV parameters and the corresponding return levels were collected from the different time windows and compared in order to find possible patterns of changes in space and time. This more qualitative method was completed with the more quantitative VGLM analysis. In there, the whole time series was used and a non-stationary GEV distribution containing a first order linear trend in both the location and scale parameter was fitted. The result was compared to a baseline stationary model and a hypothesis test was performed to determine whether to discard the stationary model and to use the non-stationary one to describe the data instead. A linear change in the GEV parameters is of course just a crude approximation but this approach allowed us to search for trends in the temperature and precipitation extremes in a statistically sound way.

Due to the climate change a general shift of the temperature towards higher values and thus more hot and less cold extremes would be expect. Indeed, we could find the location parameters of the GEV distributions, which can be thought of as the mean event size at a return period (or quantile) of approximately the block size of one year, fitted to both the daily maximum and minimum temperatures to increase. But the overall changes are far more complex and dependent on the geographical location as well as the considered return period, which is quite unexpected. For the 100 year return levels of the daily maximum temperatures a decrease was found in the east and the center of Germany for both the raw series and the anomalies, as well as a quite strong reduction for the raw series in the very south of Germany. For the daily minima about half the stations did show an increase in the 100 year return levels in the north of Germany for the anomalies and uniformly distributed one throughout all of Germany for the raw series.

This behavior is due to the non-linear dependence of the return levels on the fitted GEV parameters. But since their estimates are not free of errors, we can not extrapolate to arbitrarily high return levels. To access whether the obtained return levels are still meaningful, we have to calculate their fitting errors. This task, which can not be handled with any other software package, was done using the Monte Carlo-based approach I develop in this thesis. In case of the 100 year return levels of the raw temperatures the changes are of the same order as the fitting errors and therefore the observed behavior might be arbitrary and simply due to fluctuations. For the temperature anomalies, on the other hand, the changes exceeded the errors and can thus be interpreted as manifestations of the dynamics of the underlying climate systems. Since many of the

results are quite similar to the ones obtained using the raw series, both might represent the actual changes of the system. But, again, this qualitative time window approach was done to provide the reader with some detailed insights into the temporal evolution of the climate system. Statistically sound tests for the presence of trends in the extreme events of the temperature and precipitation will be done using the quantitative VGLM analysis.

This non-stationary EVA resulted in significant trends in the GEV parameters for the daily maximum temperatures of almost all stations and for about half the stations in case of the daily minima. So, there is statistically sound evidence for a change in the extreme temperatures and, surprisingly, it is not exclusively towards higher values. The VGLM analysis yielded several rejected hypotheses featuring a negative trend in the 10 year return levels. One for the raw daily maximum temperatures, eight for the daily maximum temperature anomalies, but none for the daily minima. Note that the time window and the VGLM analysis did not always yield the same slopes in the return levels for all stations. With only 60 data events at hand, representing the annual maxima or minima, both methods are driven with a number of points, which can be considered the absolute minimum. Thus, the results might show some degree of overfitting but this, unfortunately, is a general problem of the EVA. In case of the precipitation data the situation is even more grave since only four hypothesis tests could be rejected. One of these stations showed an increase in the extreme precipitation, the other three a decrease. The remaining 53 stations featured both increases and decreases on equal terms spread throughout Germany without any clear patterns. Apart from these four stations, the analysis of the precipitation data thus showed neither any evidence for a change in the extreme events nor any clear tendencies. If there is indeed a temporal evolution going on in the tails of the daily accumulated rainfall, more data is required to find it. For the length at hand the EVA does not seem to be an appropriate tool of investigation.

The analysis of the temperature data of the ERA-Interim reanalysis data set yielded quite surprising results too. While in some parts of the globe, especially on land, the 10 year return levels were found to increase, they do in general decrease in most parts of the earth and almost entirely over the sea. But since we found a huge discrepancy between the results of the analysis using the station data within Germany and the results obtained for the corresponding grid points of the reanalysis data set, we can not be sure whether the patterns in the return levels of the ERA-Interim data are trustworthy. It is likely that the statistics of the reanalysis data does not properly resembles nature at larger quantiles in which case the results obtained using the EVA have no meaning.

The change of the climate system is far from a simple shift towards higher temperature or precipitation values. The evolution of the extreme precipitation is very hard to access and the changes in the tails of the asymmetric temperature distribution are fairly complex. To nevertheless make reliable statements about the climate change, long measurements of high quality, a sound theoretical foundation, and well-oiled numerical routines are key. With my contribution to the latter requirement I hope I could advance climatological studies and contribute to a more solid numerical basis future researchers can rely on.

A. Analysis of the temperature and precipitation series provided by the DWD

In here the full analysis of the temperature and precipitation series provided by the German weather service (DWD) will be presented. A thorough treatment of the preprocessing and structure of the analysis including an explanation of the figures and colors can be found in sections 7.1 and 7.2.

A.1. Temperature anomalies within Germany

The baseline of this analysis will be the case study of the Potsdam temperature anomalies in section 7.3. But as already mentioned in the description of the structure of the analysis in section 7.2, we can neither compare the estimated PDFs of all stations nor estimate a PDF composed of all their data. Instead, the first four moments of every station will be calculated and the resulting distributions of the moments, including a highlight of the result obtained for the Potsdam station, are displayed for the three different time windows.

The resulting distributions of the mean values and variances can be found in figure A.1 and A.2. As for the Potsdam station we can see a consistent shift in the mean values of all examined stations. Also the variances of all daily maximum temperature anomalies are getting bigger and the ones of the minima do decrease, while the values of the latter are on average smaller than for the maxima. These findings are in perfect agreement with the results of the Potsdam station. It indicates the same changes for the anomalies within Germany: a general shift of the distribution function of the temperature anomalies towards higher temperatures, a decrease in variability of the temperatures at nighttime, and an increase in variability of the daily maximum temperatures.

Apart from the findings of the case study, we can extract a lot more information about the temporal change in skewness and kurtosis of the temperature anomalies from figure A.3 and A.4. While for the Potsdam station these changes were too small to be distinguishable from noise introduced by numerical artifacts, the skewness and kurtosis for all stations in Germany do yield a consistent and interesting picture. The daily maxima do have on average a positive skewness while the minima do have a negative one, which is about as twice as big. The skewness of the daily maxima is furthermore slightly increasing and thus shows the opposite behavior of the Potsdam station. At the same time the skewness of the minima is decreasing by almost 50%. This picture matches the one we obtained in the GEV analysis of the Potsdam station. The extreme events of the minima do change a lot faster than the ones of the maxima and the kurtosis is, again, larger for the minima. This matches the findings from the Potsdam series, which showed

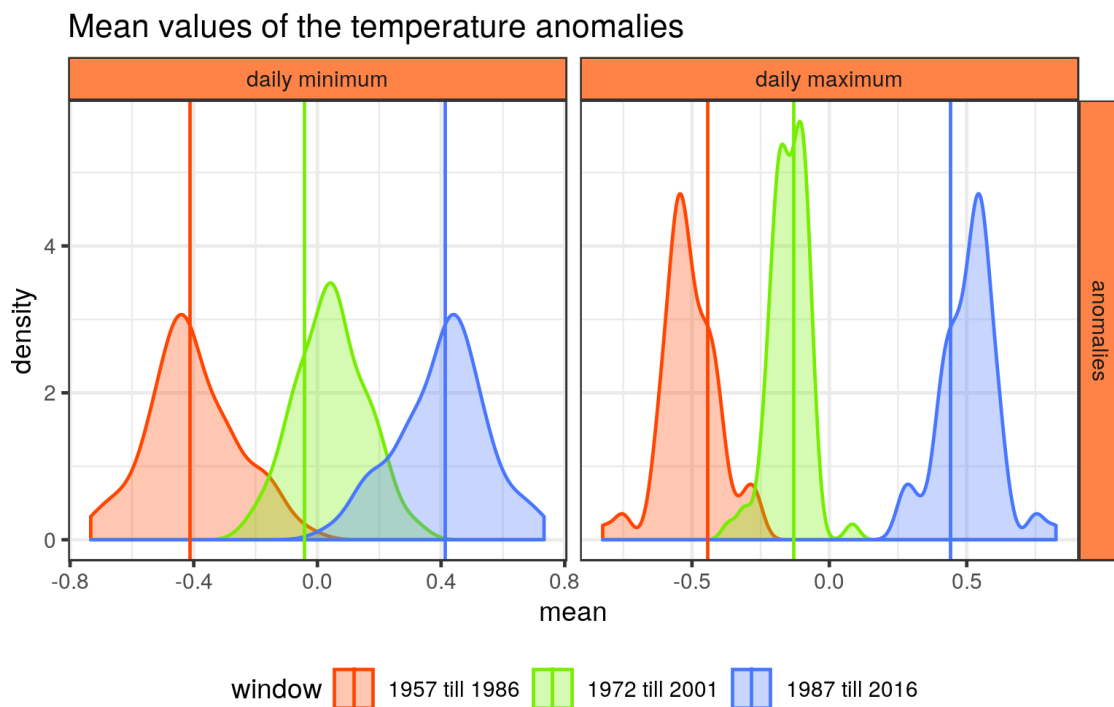


Figure A.1.: Distributions of all mean values calculated from the daily temperature anomaly series provided by the German weather service. A detailed description of the preprocessing applied to the 74 (daily maxima) or 75 (daily minima) stations series can be found in section 7.1. To increase the comparability with the Potsdam case study the corresponding results are highlighted with a vertical line.

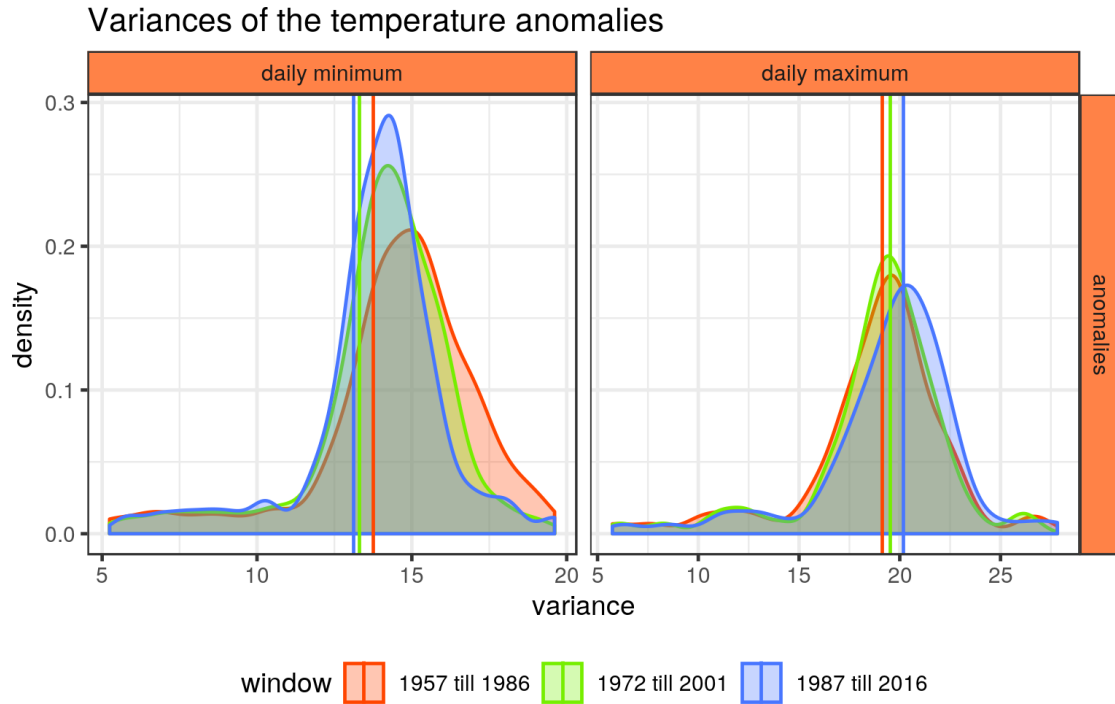


Figure A.2.: Distribution of all variances of the daily temperature anomalies within Germany. For a more detailed description see figure A.1 and section 7.1.

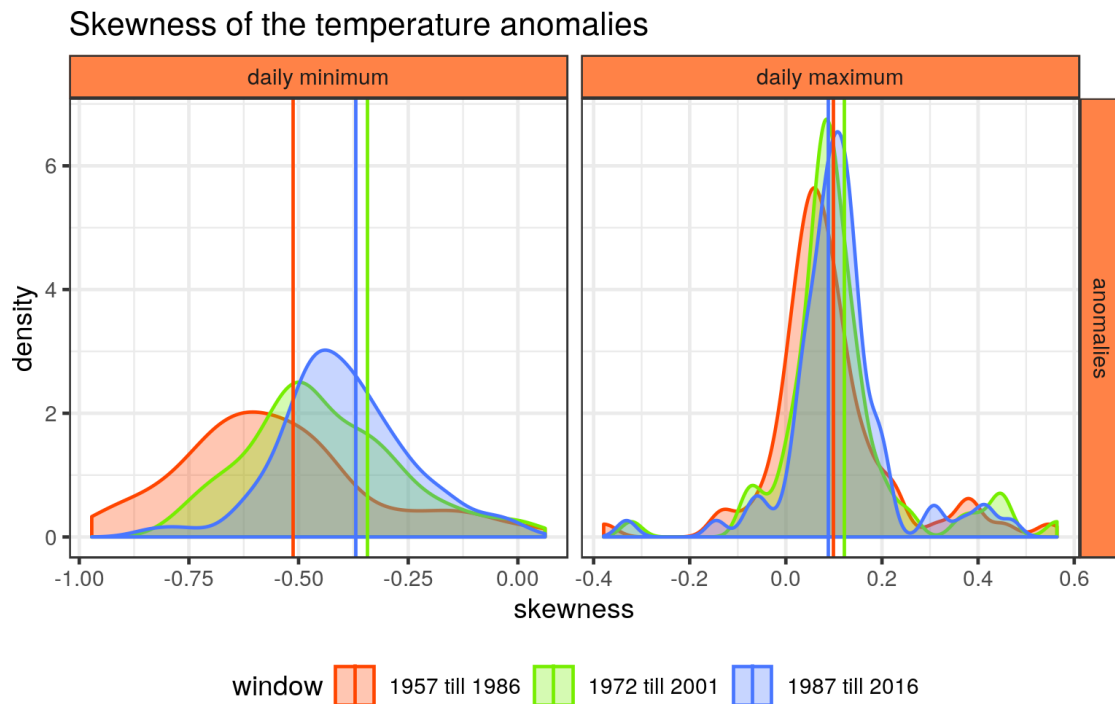


Figure A.3.: Distribution of all skewness values of the daily temperature anomalies within Germany. For a more detailed description see figure A.1 and section 7.1.

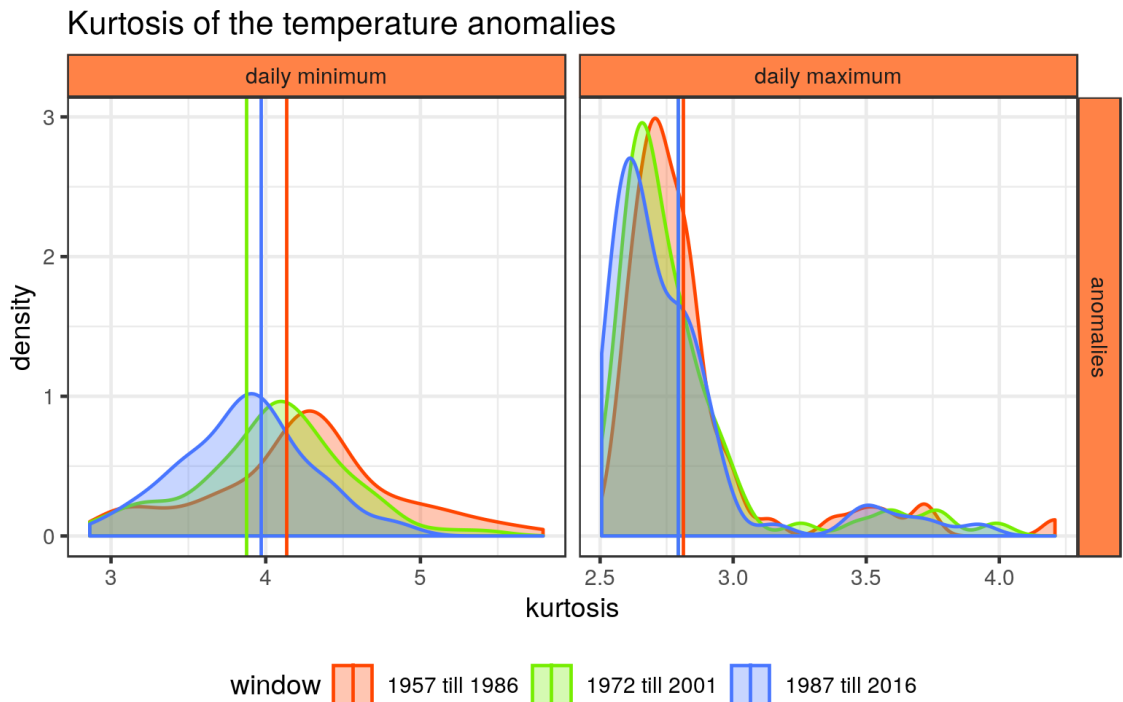


Figure A.4.: Distribution of all kurtosis values of the daily temperature anomalies within Germany. For a more detailed description see figure A.1 and section 7.1.

larger variability and a higher number of spikes towards low temperatures. The temporal change in fourth moment does yield a uniform picture. One can find a general decrease suggesting less heavy tails at both ends of the overall temperature distribution and thus a general decrease in extreme events (apart from the shift towards higher temperatures).

To search the temporal evolution of the first four moments of the daily temperature anomalies for spatial patterns, the values obtained in the last time window as well as the difference between the first and third window are shown in the following figures.

When searching for spatial patterns in the mean values of the daily temperature anomalies in figure A.5 one can find nothing but slightly larger changes for the maxima compared to the minima. It's counter-intuitive but with its chilly and rainy summers and its vicinity to the ocean the mean temperature in the southern UK is actually higher than in the center of Germany. Although another project involving a principal component analysis of reanalysis data over Germany indicated the presence of such an effect as well, a higher mean in the vicinity of the sea, no such pattern can be found in the figure. This could be due to the selection of the stations involved into the analysis. Since we are concerned with the *change* of the climate, we do not care that much about different offsets in the series. Stations were therefore not picked to show a similar degree of the so-called *heat island effect*, where measurement stations located in cities do have higher values on average due to traffic, heat emission generators and electronic devices etc. Or it maybe is just not a dominant pattern and barely visible in figure A.5.

The variance in figure A.6 is smallest at the measurement stations near the coast or on islands and is getting larger the further the measurement station is located away from

Mean values of the temperature anomalies

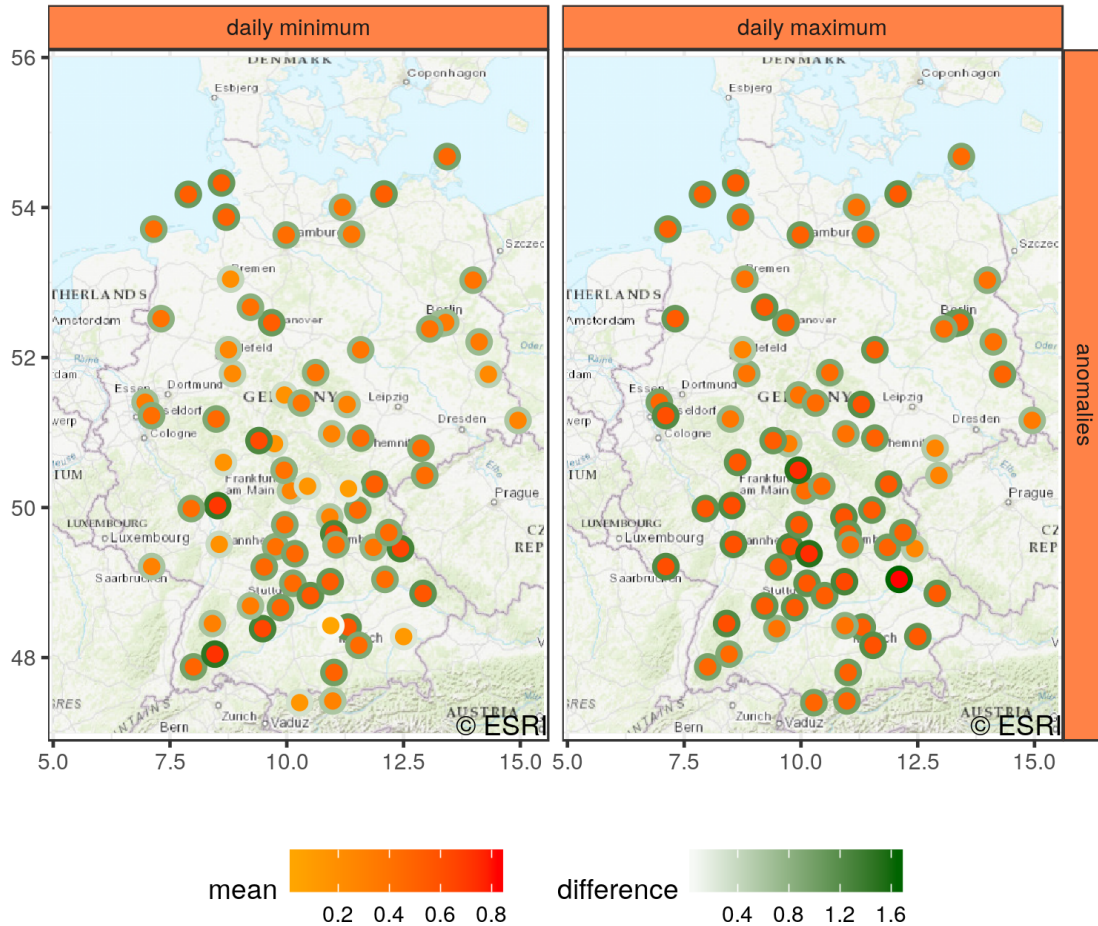


Figure A.5.: Mean values of the daily temperature anomalies throughout Germany. This map illustrates the spatial information missing in figure A.1. The colors within the circles, ranging from yellow to red, display the value of the mean in the third and last time window at the measurement site. The colors of their outer surrounding, on the other hand, show the magnitude of the temporal evolution calculated from the values of the third window minus the ones of the first. Note, to better compare the different maps of this analysis the color ranges were fixed as described in section 7.2

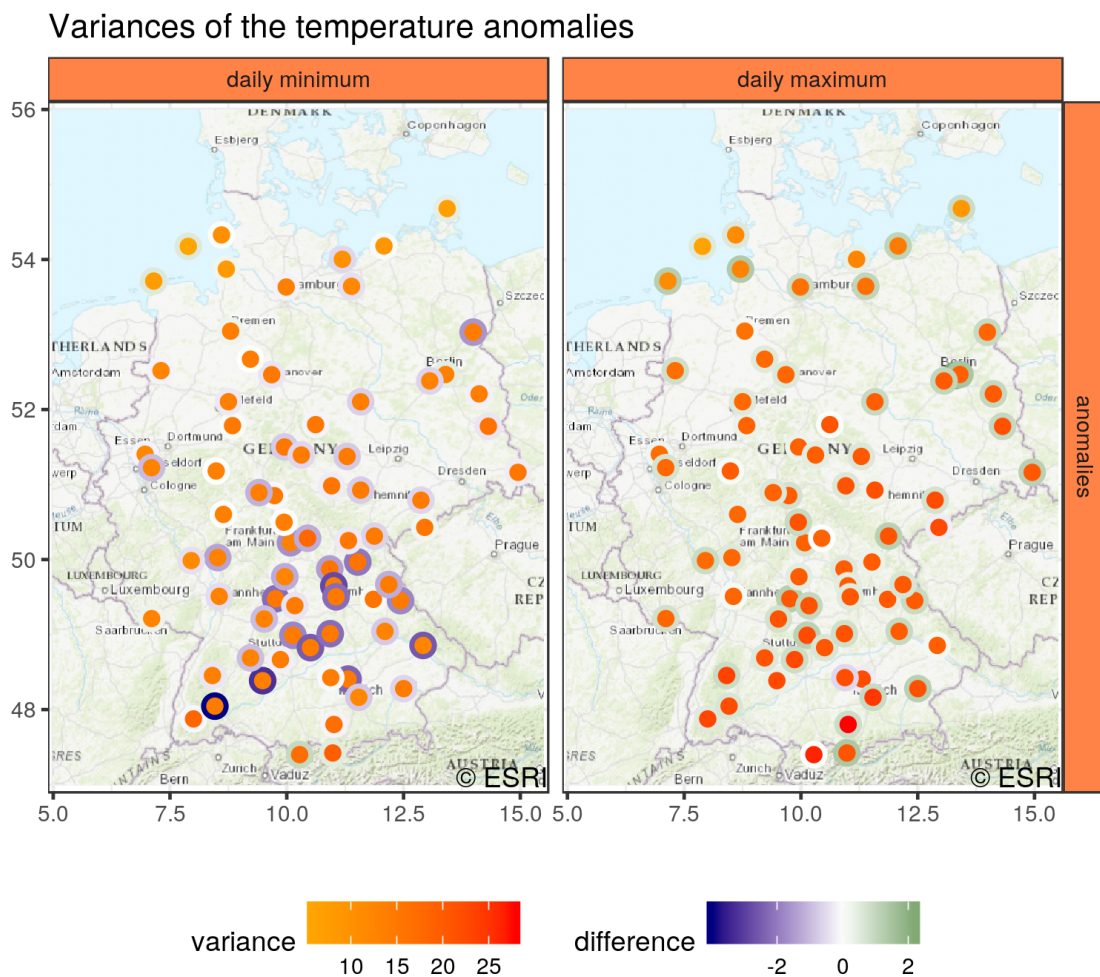


Figure A.6.: Variations of the temperature anomalies throughout Germany. See figure A.5 and section 7.2 for a detailed description.

Skewness of the temperature anomalies

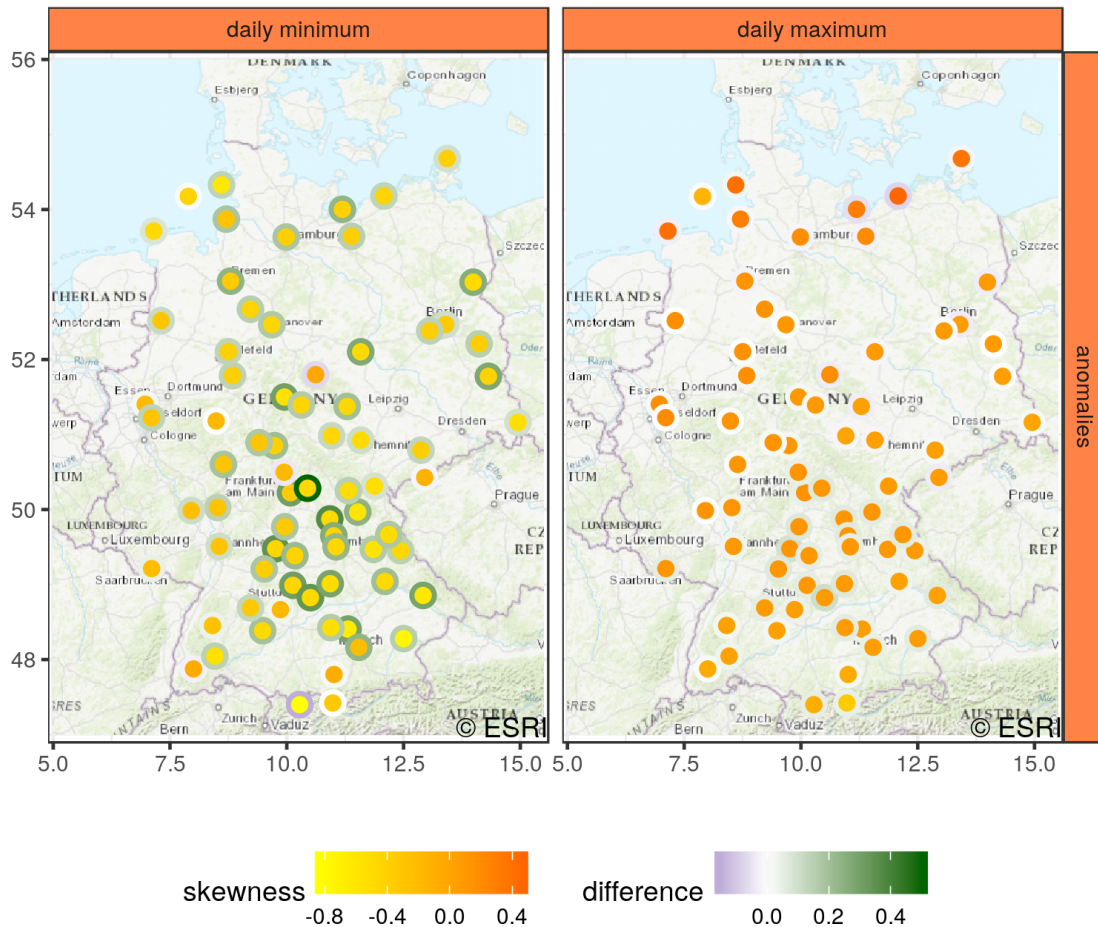


Figure A.7.: Skewness values of the temperature anomalies throughout Germany. See figure A.5 and section 7.2 for a detailed description.

the sea. This was expected since water has a large heat capacity and oceans are assumed to act as the main heat reservoir for our climate system (Randall et al. 2007). While no clear patterns for the changes in variance can be found for the maximum temperatures, the largest ones for the minima do occur in the south of Germany. Since these changes are rather strong reductions in the temperature variability and there is an addition increase in mean values of the minima towards, the temperatures during nighttime have become a lot more mild especially in Southern Germany.

The skewness of the maxima is largest in the coastal areas and exhibits a slight decrease towards the south. For the minima, however, no such pattern could be found. Small changes in the skewness towards higher values can be present in the region of Bavaria for the maxima and most stations in the minima. In addition, there an interesting pattern can be found. Stations more eastwards than a longitude of 10 degree exhibit a larger change in skewness than the western ones. This might indicate a change in frequency or path of Großwetterlagen but further research is required to check this hypothesis. A notable exception for the minima is a city called Oberstdorf located in the very south of

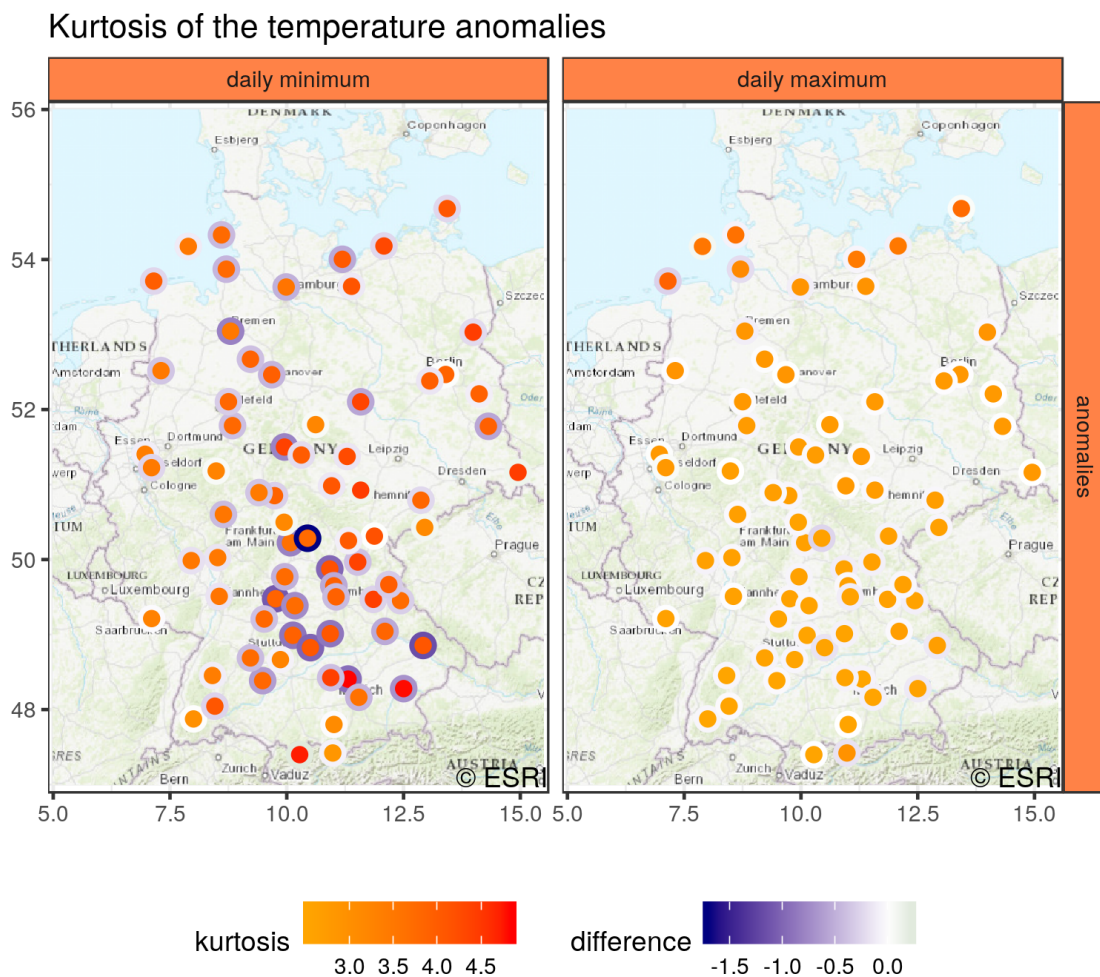


Figure A.8.: Kurtosis values of the temperature anomalies throughout Germany. See figure A.5 and section 7.2 for a detailed description.

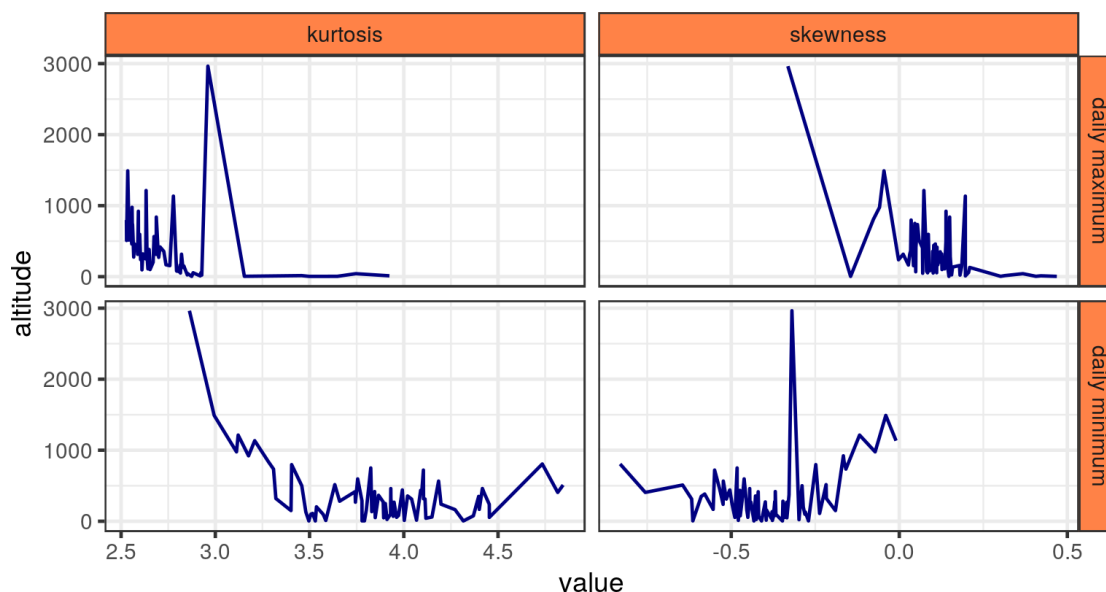


Figure A.9.: A plot of the skewness and kurtosis values of the daily temperature anomalies series against the altitude of the corresponding measurement station. The figures on the left correspond to the kurtosis of the daily maxima (top) and minima (bottom) and the ones on the right to the skewness, as indicated by the orange annotation boxes. Only for the kurtosis of the daily minima a pronounced dependence on the altitude can be found.

Germany and elevated 806 meters above zero. While in almost all stations the skewness of the minima does increase for Oberstdorf it is actually decreasing. That's why it yields the lowest value in all of Germany. In the corresponding time series the number spikes towards low temperatures is highest in the third time window and one may see something like an overall downwards trend in the daily minimum temperature. This shows quite nicely that climate is an average quantity and its local manifestations can differ quite a lot.

The changes in the kurtosis of the minima in figure A.8 seem to be strongly coupled to the ones of the skewness since both exhibit almost identical patterns. For the maxima the kurtosis is largest in the coastal region but no obvious tendency for a gradient towards the south can be seen. In addition, an interesting pattern in the kurtosis of the minima can be found. Especially stations, which seem to be located at mountainsides, exhibit a smaller value. A plot of the kurtosis against the altitude in figure A.9 revealed that there truly is such a correlation. For the maxima such a pattern could still be present with a much larger background noise if the skewness of the mountain Zugspitze, which is located at the largest altitude of 2964 m, would be considered an outlier. For the skewness, on the other hand, no such correlation is obvious, except that higher elevated stations do show a smaller negative skewness in their minima than lower ones. This leads to the hypothesis that higher elevated stations might have less extreme deviations of the mean climatology towards low temperatures than lower ones.

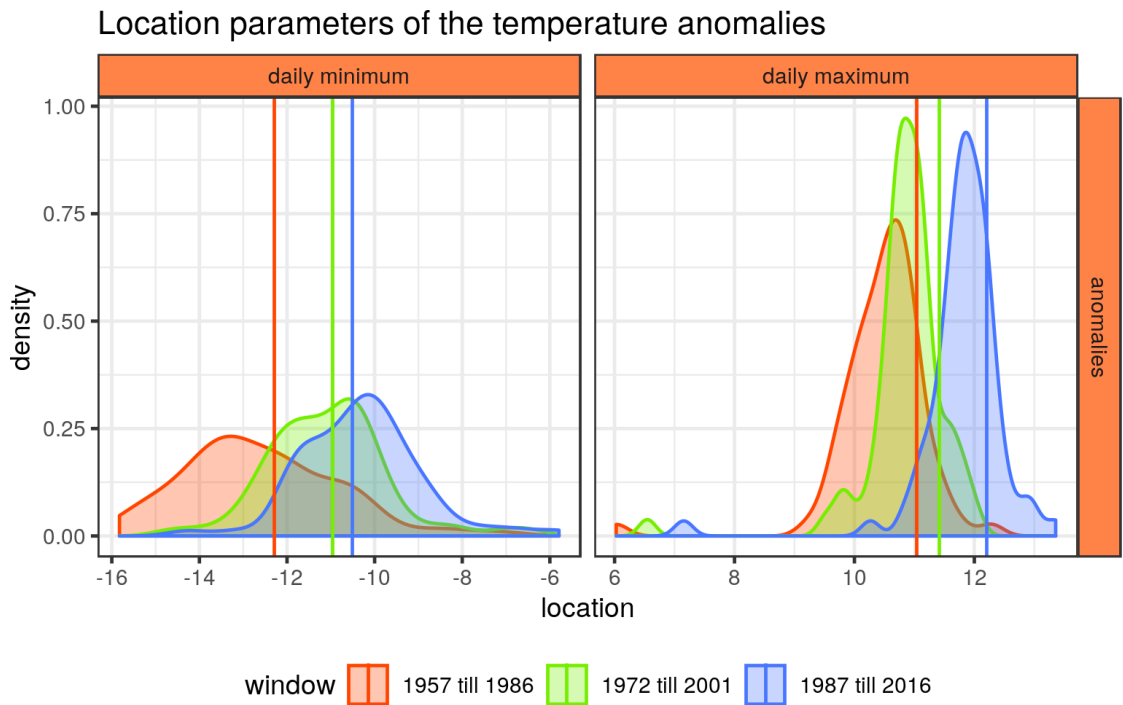


Figure A.10.: Distribution of all location parameters calculated for the different stations provided by the DWD. For a distribution of a specific color only the annual block maxima, for the daily maximum temperature anomalies, or block minima, for the daily minimum temperature anomalies, of the corresponding time window were used. In order compare the results with the case study of the Potsdam station, its results are marked with vertical lines. For detailed description see section 7.1 and 7.2.

A.1.1. Non-stationary EVA and significance

Now, let’s come to one of the main problems of this analysis, the question of “how do extreme weather events change in time within Germany?”.

For the location parameters in figure A.10 we can find the same behavior as for the mean values of the DWD series. Although there is a general trend towards warmer temperatures, the rate of change is not the same for the minima and maxima. While the former exhibit a large change between the first and second time window, the maxima show a larger step between the second and third window. The outlier of the maxima at very low temperatures in the location parameters is a measurement station on top of a mountain called Zugspitze in the very south of Germany. In the location parameter we can also see a tendency of the distribution to become more narrow starting in the second time window, while the spread of the daily minima is approximately double the size the value of the maxima. This means the picture of the extreme weather within Germany is getting more uniform.

The change in scale for the maxima is far less pronounced than expected from the analysis of the Potsdam station. Between the first and second time window in figure A.11 there is both a reduction in the overall scale value and the skewness of its distribution. Between the second and third window the average scale increases a tiny bit and gets

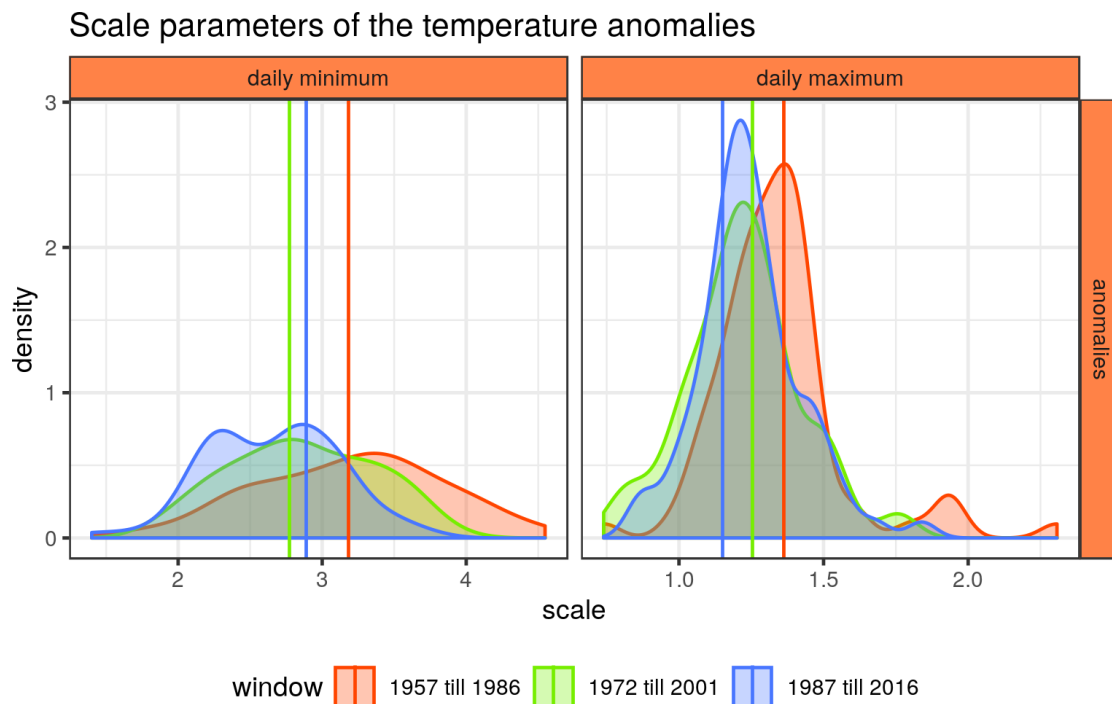


Figure A.11.: Distribution of all scale parameters of the daily temperature anomalies throughout Germany. For details see figure A.10 and section 7.1 and 7.2.

more narrow. For the minima the scale decreases throughout all three windows with the difference of the first and second window being the largest one. In addition, the spread of the distributions decreases in time as well. Interestingly a pronounced bi-modal structure emerges in the third window. Therefore, the variability with respect to the very large deviations from the climatology decreases throughout Germany.

The results of the shape parameter in figure A.12 are the most interesting ones. The distribution for the maxima spreads in the second time window and becomes narrow again in the third one featuring a shift to more negative values. This will cause both a reduction of the magnitude of the extreme events as well as a more uniform picture of the GEV distributions. While there was still a small fraction of maxima exhibiting positive shape parameters in the first and second window, there is almost none in the third. The general tendency of these findings do match the results of the Potsdam station but they are far less pronounced. In case of the minima the distribution of the shape within the first time window has almost its entire weight at negative shape values. But for the second window it shifts towards higher values by a margin and obtains a positive skewness. In the third window the position of the distribution is roughly the same with now a negative skewness and a lesser spread. This causes the ratio of negative to positive shape parameters to increase to approximately 2 : 1. As a result the return levels of the minima should increase and thus show the opposite behavior compared to the Potsdam station and might lead to an overall increase in extreme daily minimum temperatures in some parts of Germany.

Facing the rather large changes in the shape parameter of the block maxima obtained

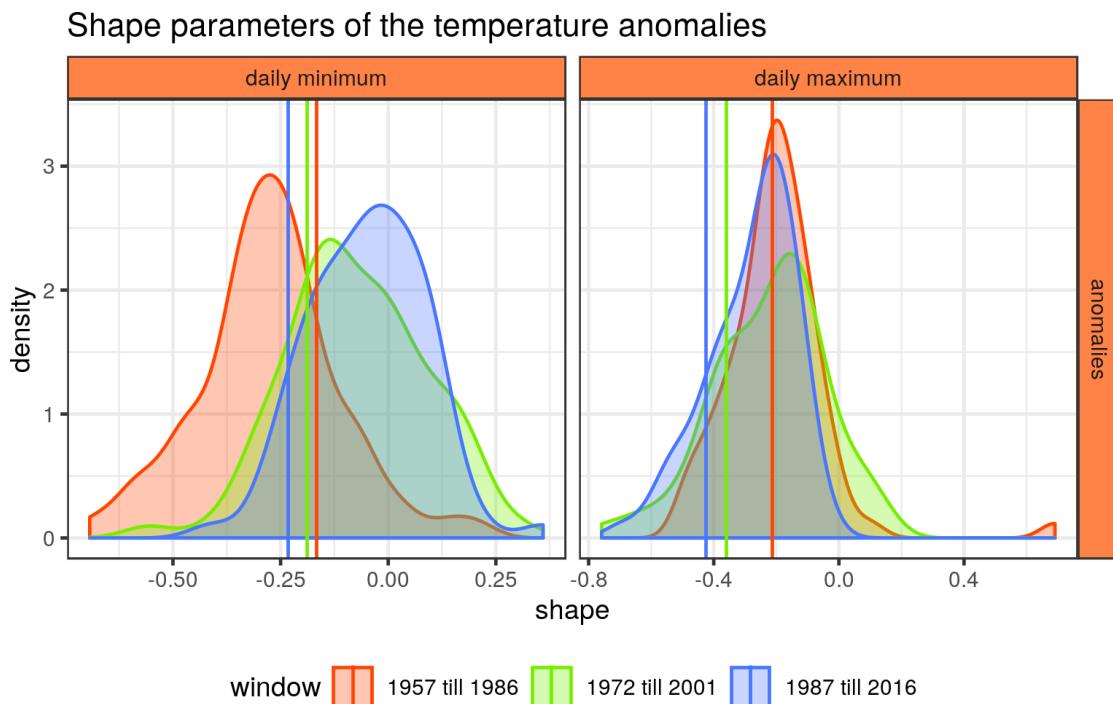


Figure A.12.: Distribution of all shape parameters of the daily temperature anomalies throughout Germany. For details see figure A.10 and section 7.1 and 7.2.

for the daily minimum temperature anomalies one might ask again, do we really have a proper description of the data when using a first order linear model with a constant shape? Both merely all suggestions in the literature (e.g. Gilleland and Katz 2014, @Yee2007, @ColesBook) and personal experience advise against the usage of a temporal dependency in the shape parameter. It needs a lot of data to be approximated appropriately since it is very sensitive to outliers and the most extreme block maxima/minima (see chapter 6). In addition, the 60 years of data are just not enough to trust a six parameter fit including a linear trend in the shape and since we seek to automate the fitting routine in order to apply it on all stations within Germany (and all nodes of a reanalysis data set in chapter 8), trust is what we need. A manual inspection of the fitting results of each station would not be feasible. We therefore keep the assumption of a constant shape parameter in the following application of the VGLM analysis.

In the 10 year return levels in figure A.13 we can see the behavior we expected from investigating the GEV parameters. For the maxima the decrease in scale and shape compensates the slight increase in the location parameters. This results in a distribution of the return levels with a tiny shift to the right, a wider spread in the second time window, and a more pronounced shift in the third one. The increase in spread is likely to be introduced by the broadening of the distribution of the shape parameters. For the 100 year return levels in figure A.14 the decrease in shape and scale becomes dominant and causes the distribution of the second window to shift to lower temperatures and to increase in spread. For the third window the larger shift of the location parameters, again, pushes the distribution to higher temperatures and thus yields larger return levels

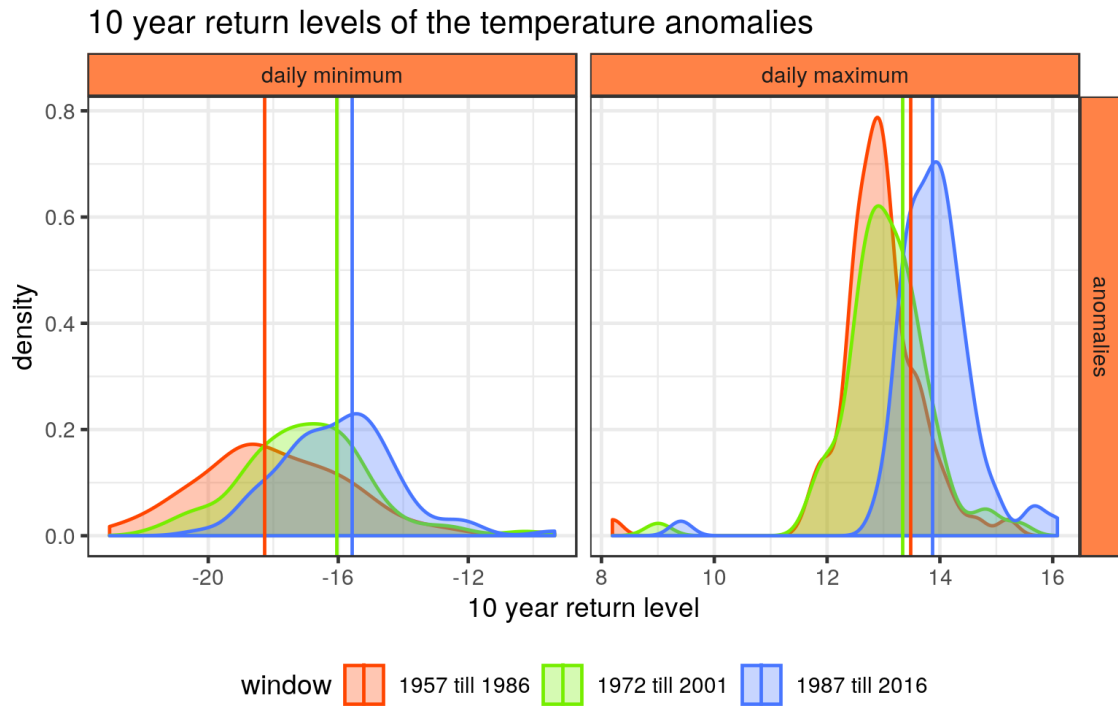


Figure A.13.: Distribution of all 10 year return levels of the daily temperature anomalies throughout Germany. For details see figure A.10 and section 7.1 and 7.2.

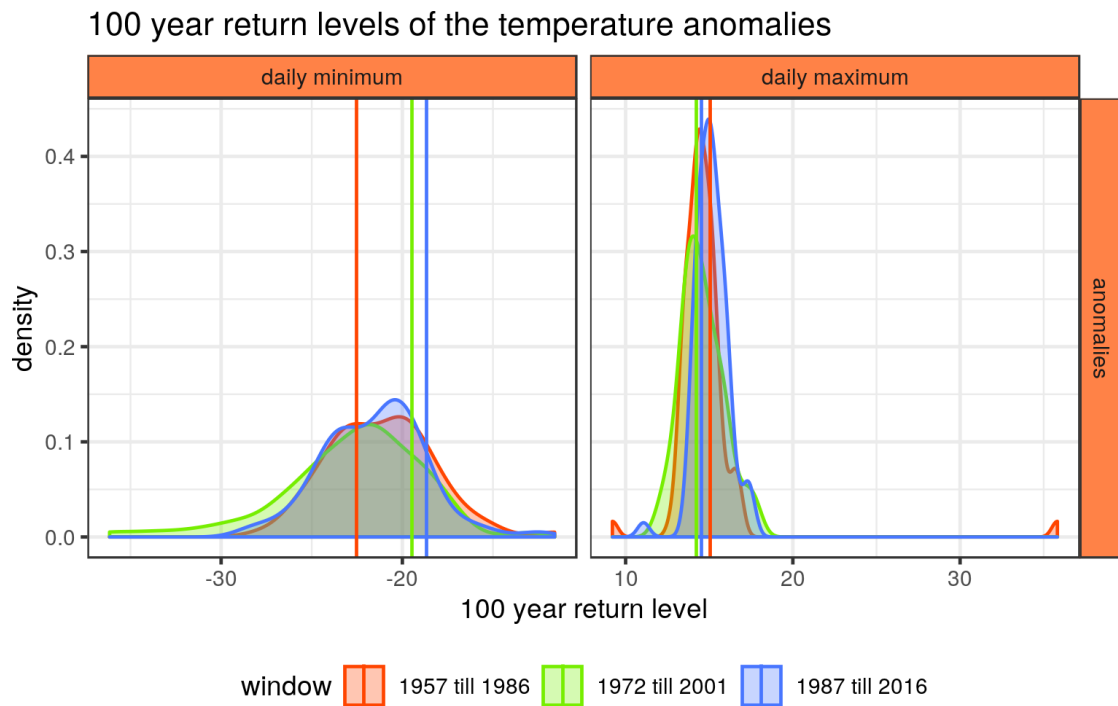


Figure A.14.: Distribution of all 100 year return levels of the daily temperature anomalies throughout Germany. For details see figure A.10 and section 7.1 and 7.2.

than in the first or second window.

In the 10 year return levels of the minima one can see first a large then a smaller shift towards higher temperatures and an overall narrowing of the distribution. Both is probably caused by the temporal evolution of the location parameters. For the 100 year return levels the distribution remains more or less constant with an increase in spread in the second window, which will decrease again in the third. This behavior is totally different compared to the results of the Potsdam station and the ones for the 10 year return levels. It is most likely caused by the strong increase in shape cancelling the contributions of both the decrease in scale as well as the increase in location. The outlier at a very large daily maximum temperature in the first time window of the 100 year return levels corresponds to the Brocken station, which is a mountain in the center of Germany. In the first time window the fit results in a location parameter of 11.21, a scale of 0.74, and a shape of 0.69. It is thus responsible for the largest 100 year return level, the largest shape, and the smallest scale parameter for all maxima. But what looks like an artifact of the analysis corresponds to a valid fit. The problem encounter here is similar to the one discussed in the chapter 6. The series contains very large events in its first years and thus the shape parameter gets overestimated when restricting our view only on the first time window. But still, it is an artifact and won't serve as a proof for the massive decrease of the extreme events.

The spatial distribution of the location parameters and their changes in figures A.15 are almost identical to the ones found for the mean values of the series. The patterns of the scale parameters in figure A.16, on the other hand, do not match the ones of the variances in figure A.6. For the maxima there is a small tendency for larger decreases to happen in the north and middle of Germany and for increases to occur in the south. For the minima no patterns can be found. At first sight, one might expect a similar dependence on the amplitude for the positive changes in the scale like the one found for the kurtosis. But upon closer inspection no such dependence could be found.

The shape parameters of the daily maxima in figure A.17 are, on average, shifting towards lower values. But for coastal stations an increase can be found instead. In addition, there is a larger increase near the ocean for the maxima too. Considering only the contribution of the shape parameters to extreme weather events this indicates that stations near the sea will suffer more extreme temperature anomalies during both day and nighttime. For stations further away from the coast only the extreme fluctuations around the climatology during nighttime are getting stronger. The outlier of the maxima with the massive difference in shape and 100 year return level in the very center of Germany is the Brocken station we already declared an artifact in the previous part.

The biggest increase towards higher temperatures in both the 10 year return levels in figure A.18 and the 100 year return levels in figure A.19 for the maxima and minima occurs in the south of Germany. For the maxima this is caused by a slight increase in scale in the south of Germany and its decrease in the center and north. For the minima the location and scale parameters would promote a more uniform picture of the changes in return levels. But since the shape parameters becomes larger in both the center and north of Germany, the shift of the 10 year return levels towards higher temperatures is compensated to some extends. For the 100 year return levels, on the other hand, the change in shape becomes the dominant driver and thus causes a shift towards lower temperatures in the center and north. A small shift towards higher temperatures in

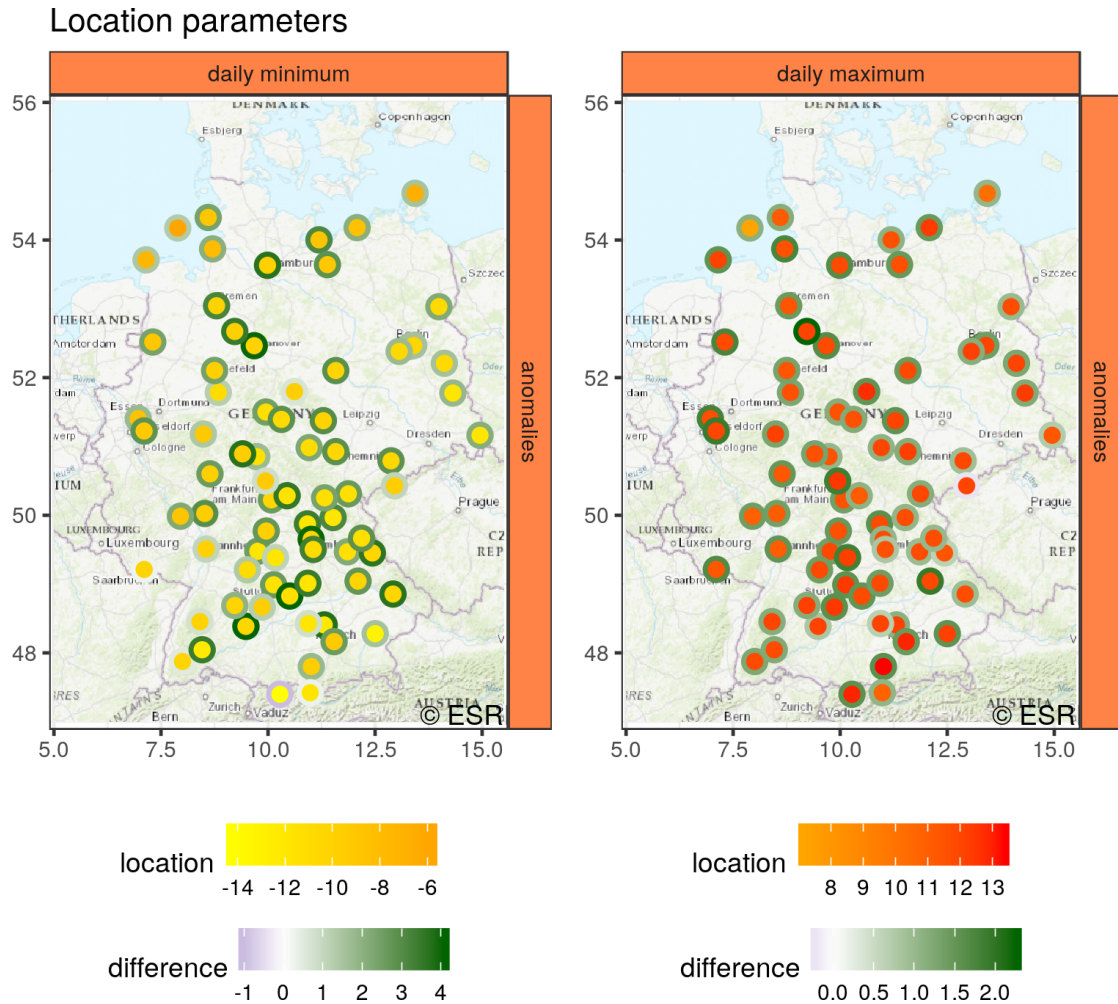


Figure A.15.: Location parameters of the daily temperature anomalies throughout Germany. In contrast to figure A.5 the results of the time window analysis of both the daily minimum and maximum temperature anomalies are colored in two separate scales. Else, the intricate details of the figure would be barely visible. This also means that color orange is overloaded and corresponds to a value of approximately -5°C in the left figure and 6°C in the right figure. For a detailed description please see section 7.1 and 7.2.

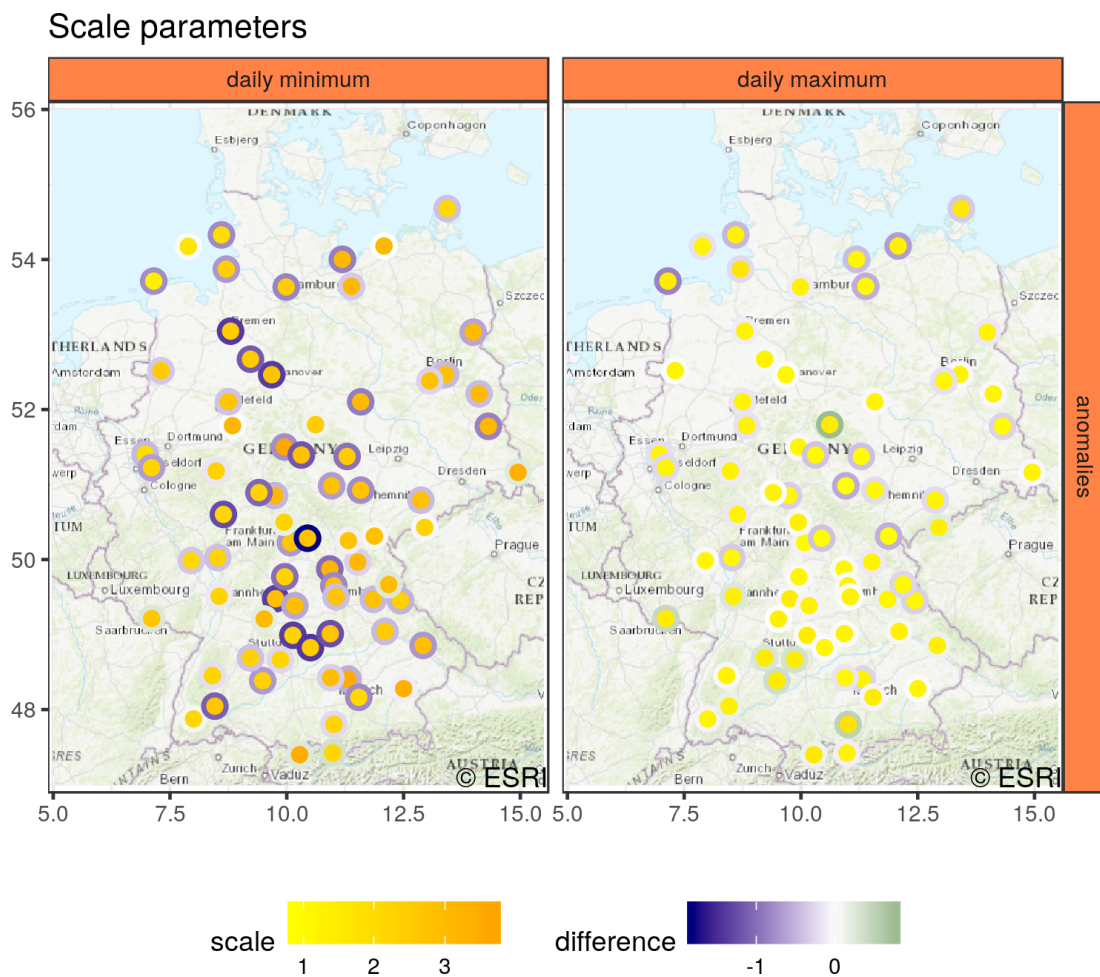


Figure A.16.: Scale parameters of the daily temperature anomalies throughout Germany. For a detailed description of the figure and colors please see section 7.1 and 7.2.

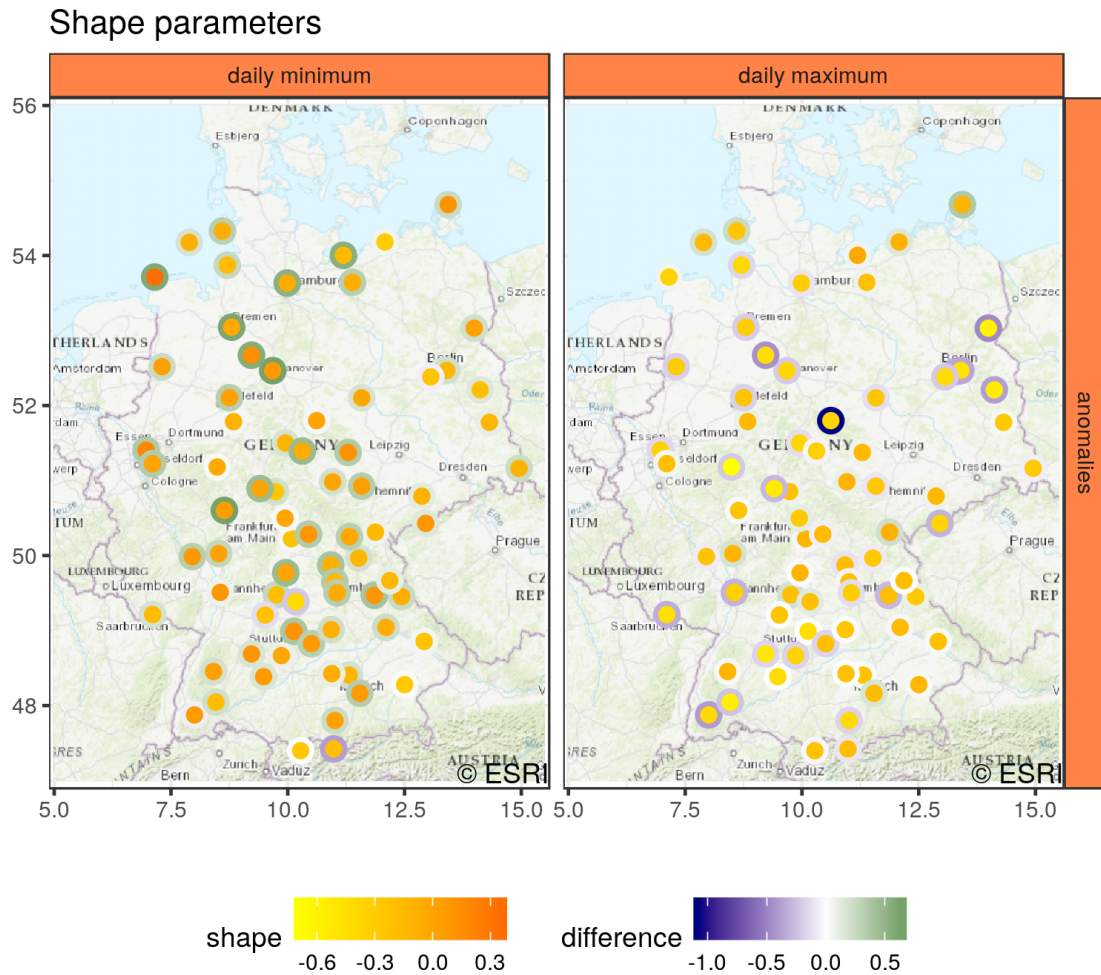


Figure A.17.: Shape parameters of the daily temperature anomalies throughout Germany. For a detailed description of the figure and colors please see section 7.1 and 7.2.

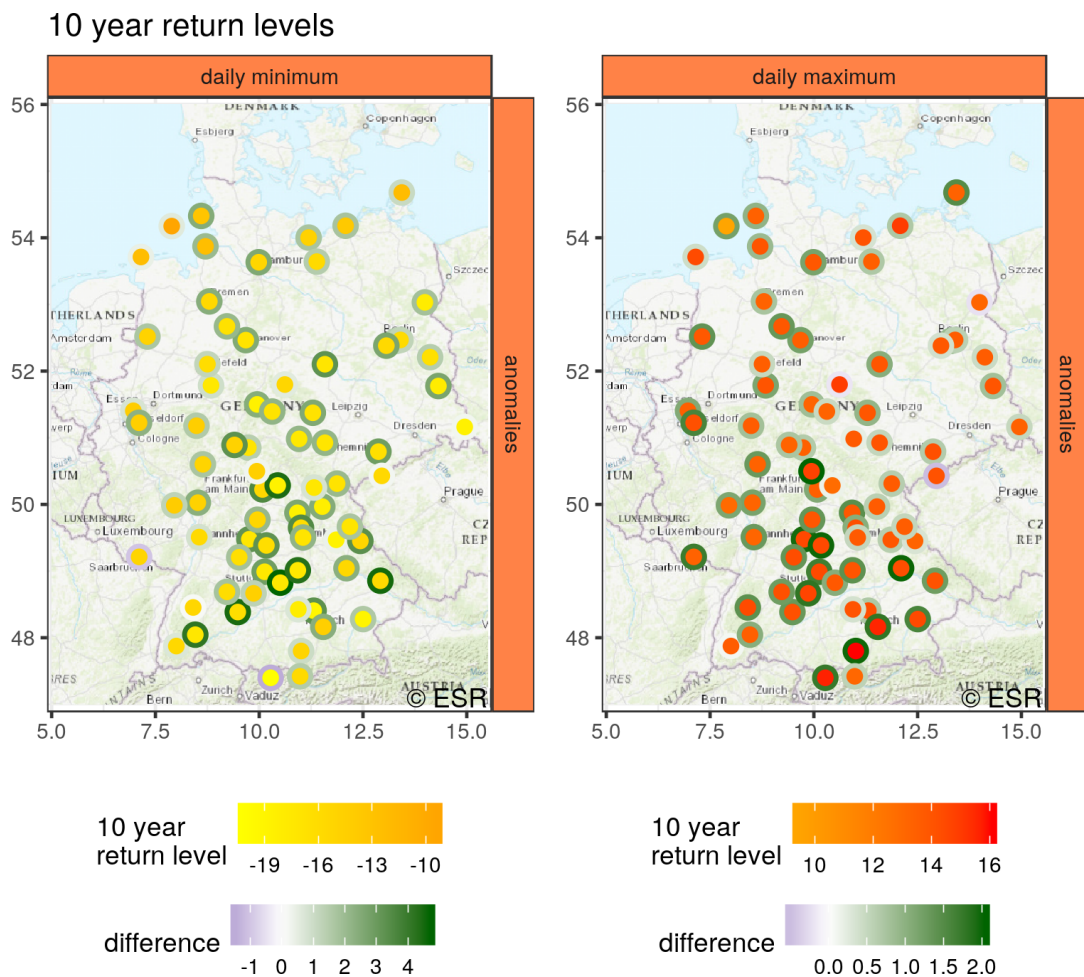


Figure A.18.: 10 year return levels of the daily temperature anomalies throughout Germany. For a detailed description of the figure and colors please see figure A.15 or section 7.1 and 7.2.

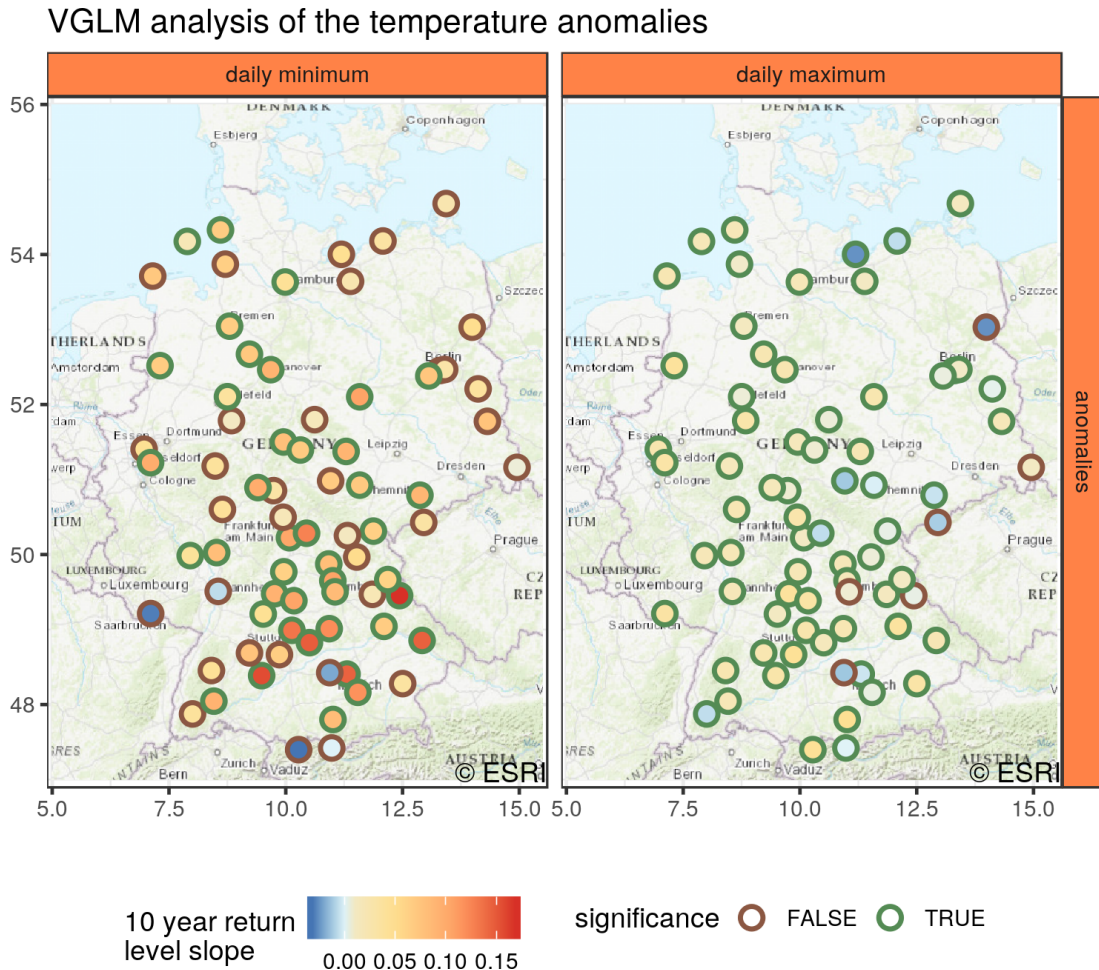


Figure A.20.: Results of the likelihood ratio test whether to use a stationary model over a first order linear one. All stations, for which the null hypothesis of a better description with a stationary model was rejected, are enclosed by a green color. The circles are filled with the slope of the 10 year return levels fitted using the VGLM. Since it has a non-linear dependence on time the slope was estimated by fitting a straight line in its curve. This value might therefore not yield a perfect approximation of the actual temporal evolution but it is sufficient to determine its overall tendency.

the south still remains. This means deviations from the climatology towards higher temperatures during daytime become more pronounced. At the same time the extreme deviations towards lower temperatures during nighttime become smaller for a return period of around 10 years. But for very large extremes, e.g with return periods of 100 years, those deviations actually become larger in many parts of Germany. Mostly in the south there will be a further decrease instead.

Finally, we also perform the VGLM analysis on the whole set of daily temperature anomalies of the DWD stations. As shown in figure A.20 for most of the daily maxima the null hypothesis of using a stationary GEV distribution have been rejected at a 5% significance level. This means that there is indeed a very high probability for the large fluctuations around the temperature climatology to change in time, possibly due to the climate change. The nature of the changes is in agreement with the findings of the time window analysis. Most of the 10 year return levels do show a positive trend but for eight stations a decrease in combination with a significant trend in the GEV parameters could be found. In case of the daily minima around half of the hypotheses had to be rejected and none of them showed a decrease in the return levels. This asymmetry is caused by the different variances of the block maxima and block minima series. The latter, featuring almost double the variance of the block maxima, has bigger noise levels and the resulting trends have to be larger in order to not pass the likelihood ratio test. In other words, just because there are more stations of the daily minima passing the hypothesis test does not necessarily mean the extreme minima are “more” stationary than the maxima. It is just harder to be verified in a statistical sense and since the slopes in the block minima are often even larger than for the corresponding block maxima it is most probably both sides of the overall temperature anomaly distribution being non-stationary.

A.2. Raw temperature series throughout Germany

The analysis of the raw temperature series will be kept rather concise. Only those results differing from the ones of the temperature anomalies in the last section will be reported.

Figures A.21 shows the temporal evolution of the mean values of the raw temperature series throughout Germany. Their shift is much less pronounced as in the anomalies but still present. In addition, for the maxima the shift is a lot bigger than for the minima. The outlier at very low temperatures of both the maxima and minima is the measurement station at the Zugspitze, the highest mountain within Germany. Since the measurement takes place at the top of the mountain, the mean values are much smaller but the series does nevertheless not represent an artifact and will be included in the remainder of the analysis.

The variances depicted in figure A.22 show a different behavior compared to the ones of the anomalies in A.2. For the maxima the distribution in the second window underwent a small shift but in the third window it, again, looks almost exactly like in the first one. In case of the anomalies we did find a slight shift towards higher values instead. For the minima the behavior of the variances is more similar to the anomalies. It features a shift towards smaller values for the second and third window. But now the third window has a larger spread than the second one. Why is the behavior different? The annual cycle present in the raw series does not simply increase the overall variance to

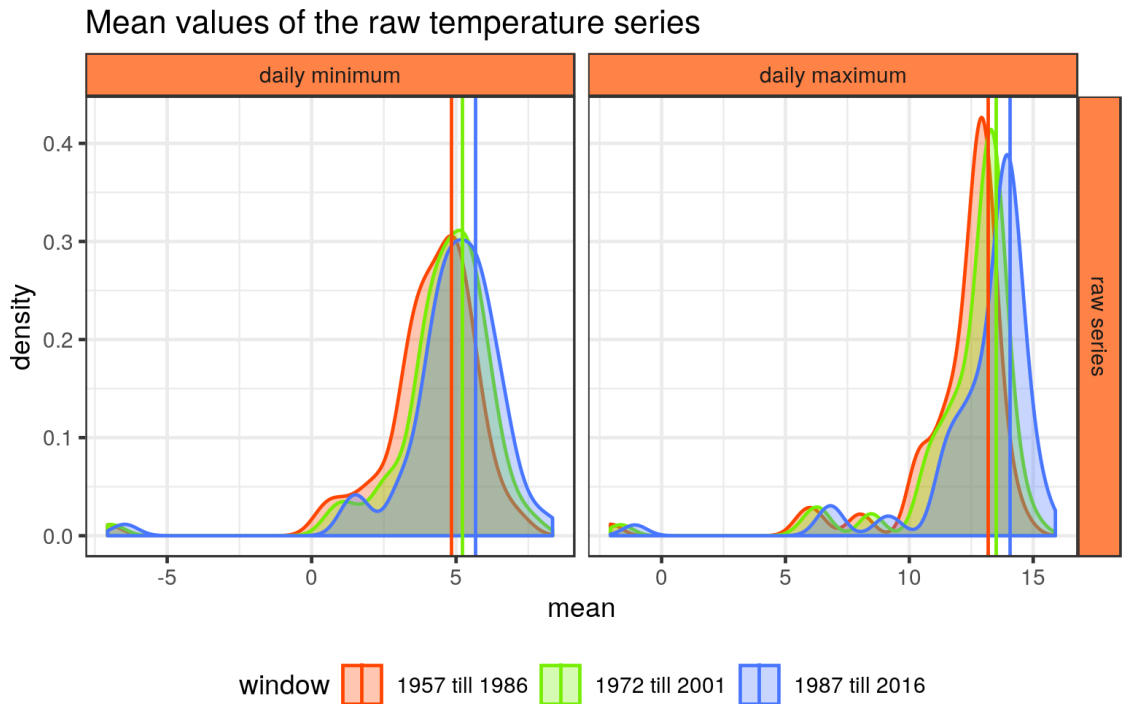


Figure A.21.: Distribution of the mean values of the raw temperature series within Germany. For more details see section 7.1.

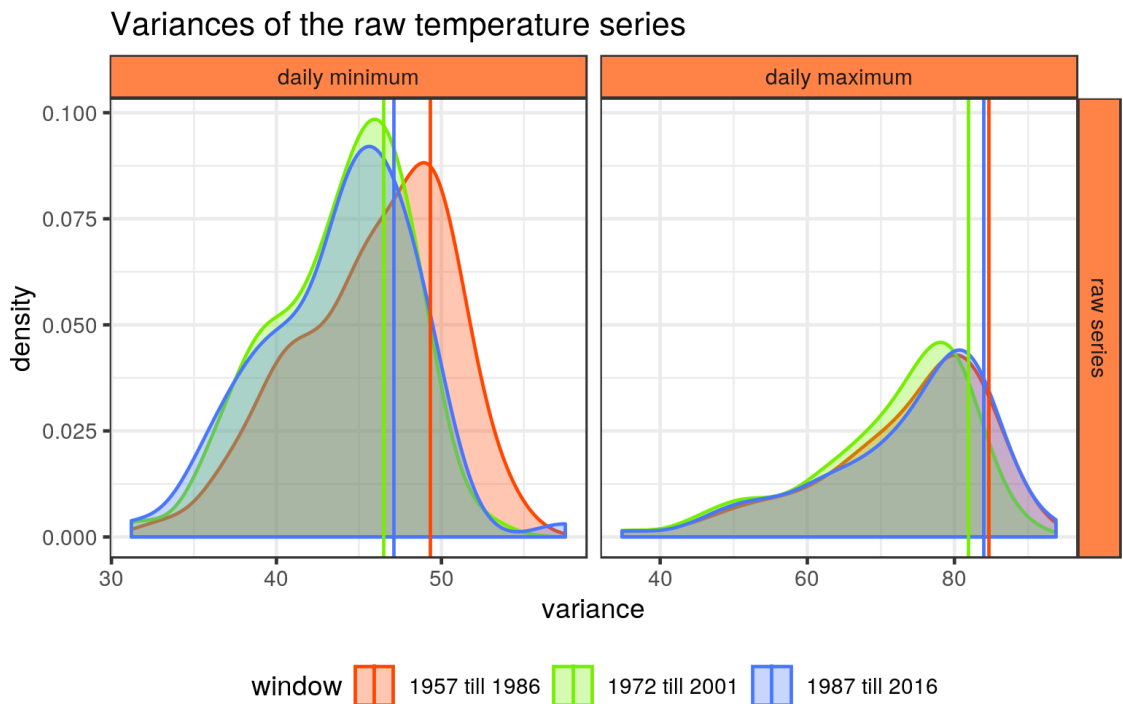


Figure A.22.: Distribution of the variances of the raw temperature series within Germany. For more details see section 7.1.

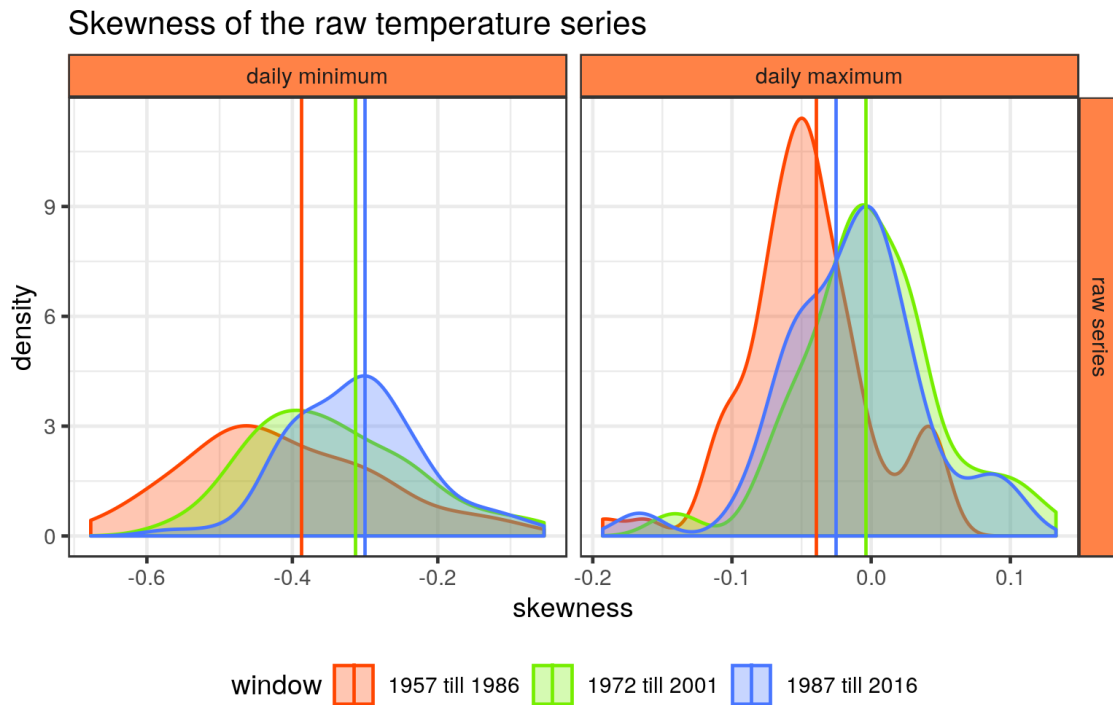


Figure A.23.: Distribution of the skewness values of the raw temperature series within Germany. For more details see section 7.1.

values about four times as high as the ones for the anomalies, it also changes the way fluctuations contribute to the estimates. For the anomalies all fluctuations, regardless of their direction, do increase the variance. But when dealing with raw series only those fluctuations do increase the variance, which result in a larger absolute values after adding both the raw temperature and the fluctuation itself. Therefore, most fluctuations towards lower values in summer or towards higher values during winter will cause the variances to decrease. Due to heteroscedasticity of the temperature distribution the overall behavior of the raw series compared to the temperature anomalies does change.

While the skewness of the daily maximum temperature anomalies was predominantly positive, it is mostly of negative value for the raw series (see figure A.23). In the second and third time window the distribution shifts towards higher values, obtains a larger spread, and is now approximately centered around zero. For the daily minima the picture is qualitatively the same as for the temperature anomalies.

The kurtosis is in general of smaller values than for the anomalies. It also shows a similar behavior compared to the anomalies for the daily minima in figure A.24. This includes a shift towards lesser values and a broader spread. But there are important differences between the two. In case of the daily minima of the raw series the kurtosis features a bimodal structure in the distribution of the third time window. For the daily maxima the behavior of the raw series is opposite compared to the anomalies. Here, the distribution does not shift towards lesser values but gains more spread and more mass at higher kurtosis values.

The spatial distribution of the mean values in figure A.25 is almost the same as for

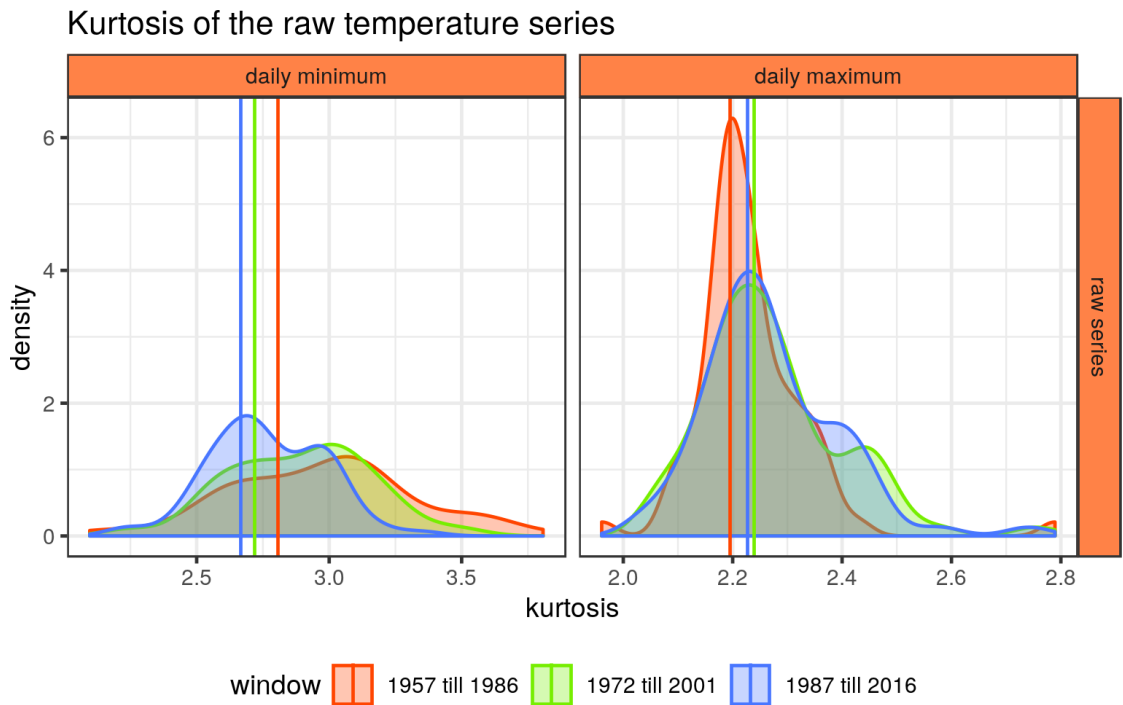


Figure A.24.: Distribution of the kurtosis values of the raw temperature series within Germany. For more details see section 7.1.

the anomalies. The variance in figure A.26, on the other hand, feature a less pronounced trend in the north-south direction. Most of the increases for the maxima occur in the south of Germany and most of the decreases in the north, in contrast to the anomalies, which yielded a more uniform increase throughout all of Germany. For the minima there is, again, a tendency for larger decreases to occur in the south of Germany. But compared to the anomalies this pattern is also less pronounced.

The skewness in figure A.27 is very similar to the anomalies. A slightly stronger shift towards positive values can be found in both the daily maxima and minima. In addition, the increase in skewness in the south of Germany is a lot more pronounced.

The differences between figure A.28, depicting the kurtosis for the raw temperatures series throughout Germany, and figure A.8, the corresponding one for the temperature anomalies, are subtle but many. For the daily minima the values of the kurtosis seem to be more or less uniformly distributed with a small trend between north and south with lesser values in the north. While no such pattern was found for the daily minimum temperature anomalies, the exact opposite pattern was discovered for the daily maximum temperature anomalies. For the daily maxima of the raw series, on the other hand, no trend is visible and the values seem to be uniformly distributed. In addition, the largest changes towards higher kurtosis in the daily maxima do occur in the north close to the sea. For the minima the decrease in kurtosis in the south of Germany is even more pronounced than for the anomalies. But, again, the distributions of the raw temperature series are bimodal and the results for the moments of such series might be misleading.

Mean of the raw temperature series

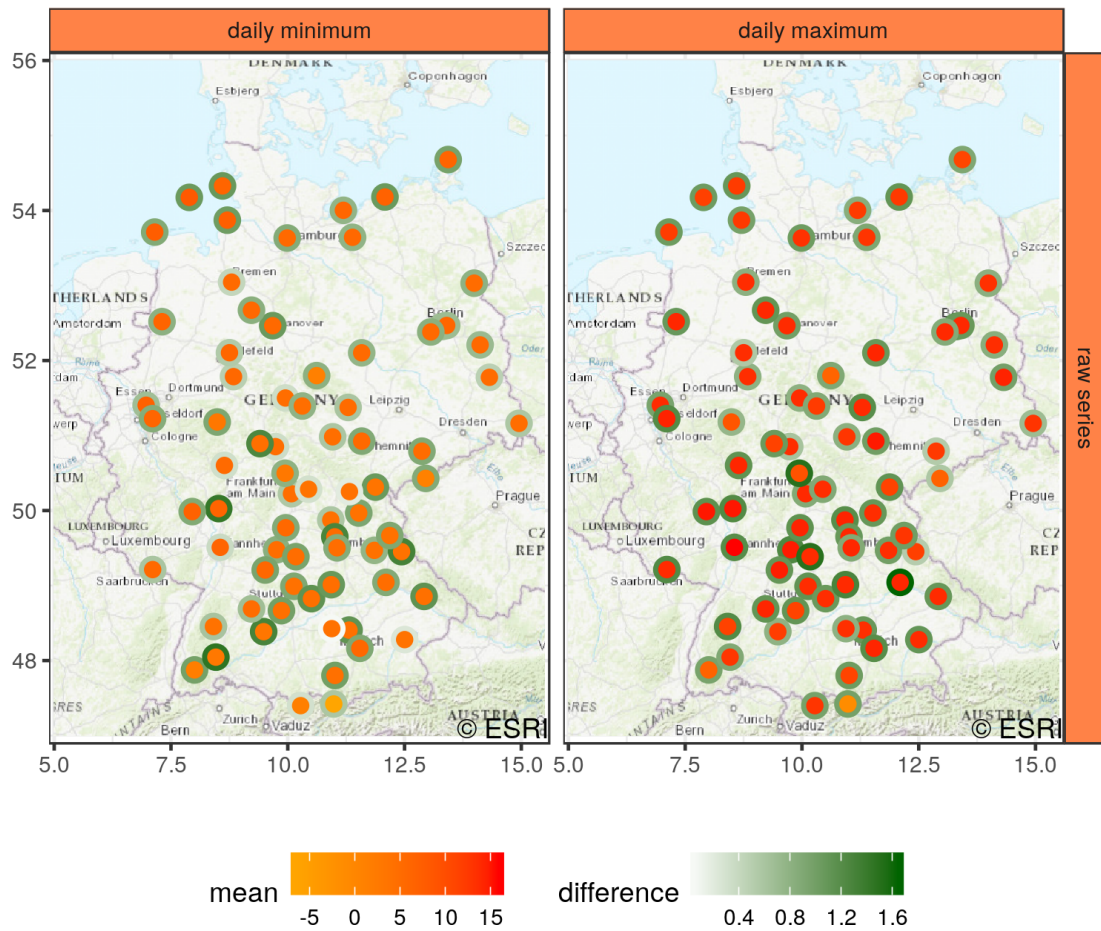


Figure A.25.: Mean values of the raw temperature series throughout Germany. The figure is analog to A.5, depicting the mean values for the daily temperature anomalies instead. See section 7.1 and 7.2 for a detailed description.

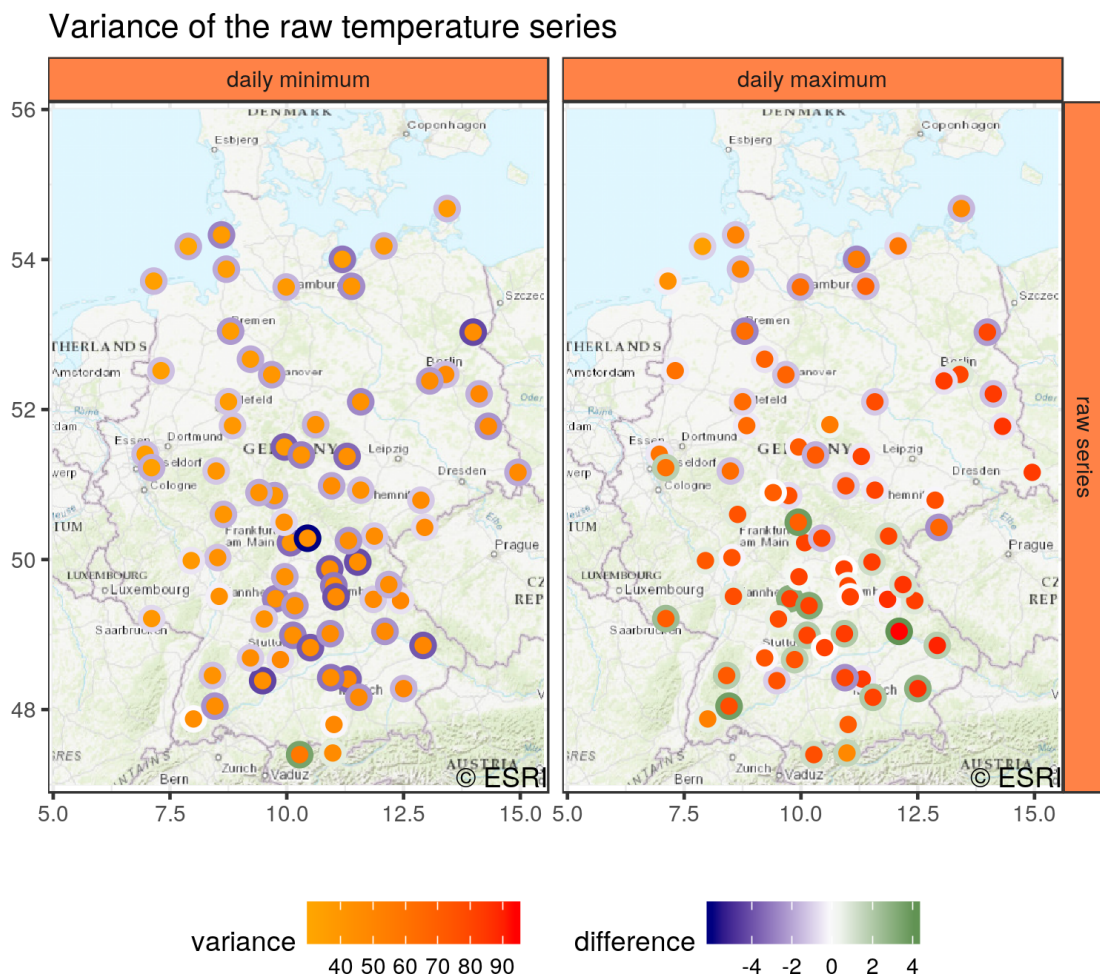


Figure A.26.: Variances of the raw temperature series throughout Germany. See figure A.25 and section 7.1 and 7.2 for a detailed description.

Skewness of the raw temperature series

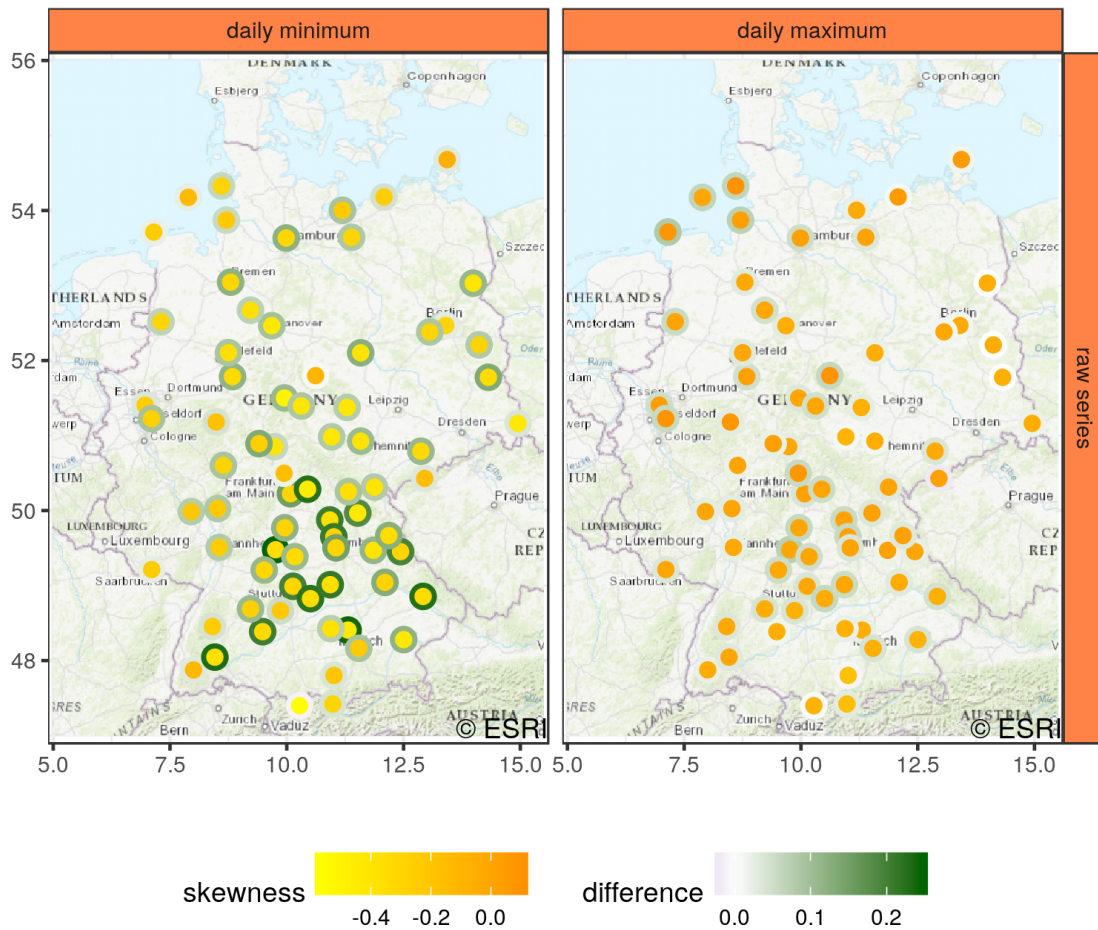


Figure A.27.: Skewness values of the raw temperature series throughout Germany. See figure A.25 and section 7.1 and 7.2 for a detailed description.

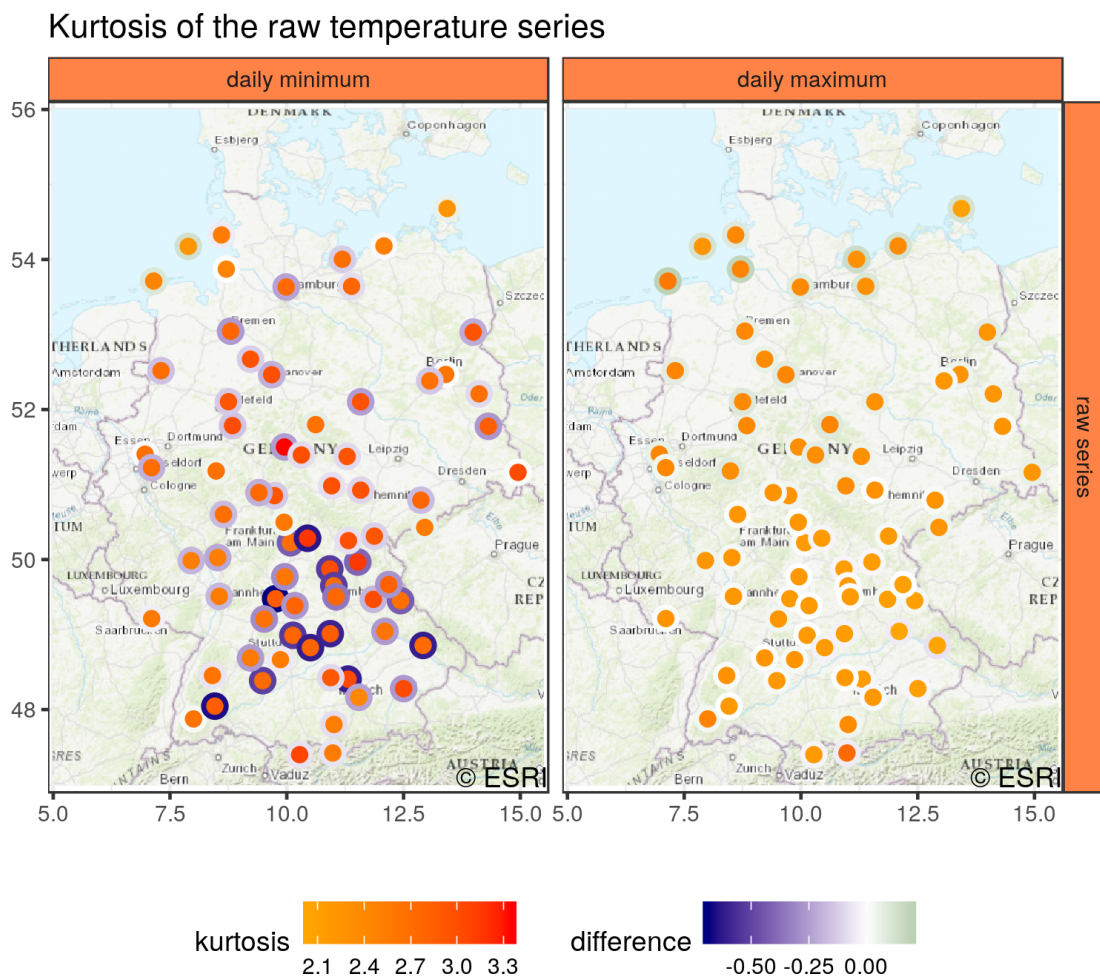


Figure A.28.: Kurtosis values of the raw temperature series throughout Germany. See figure A.25 and section 7.1 and 7.2 for a detailed description.

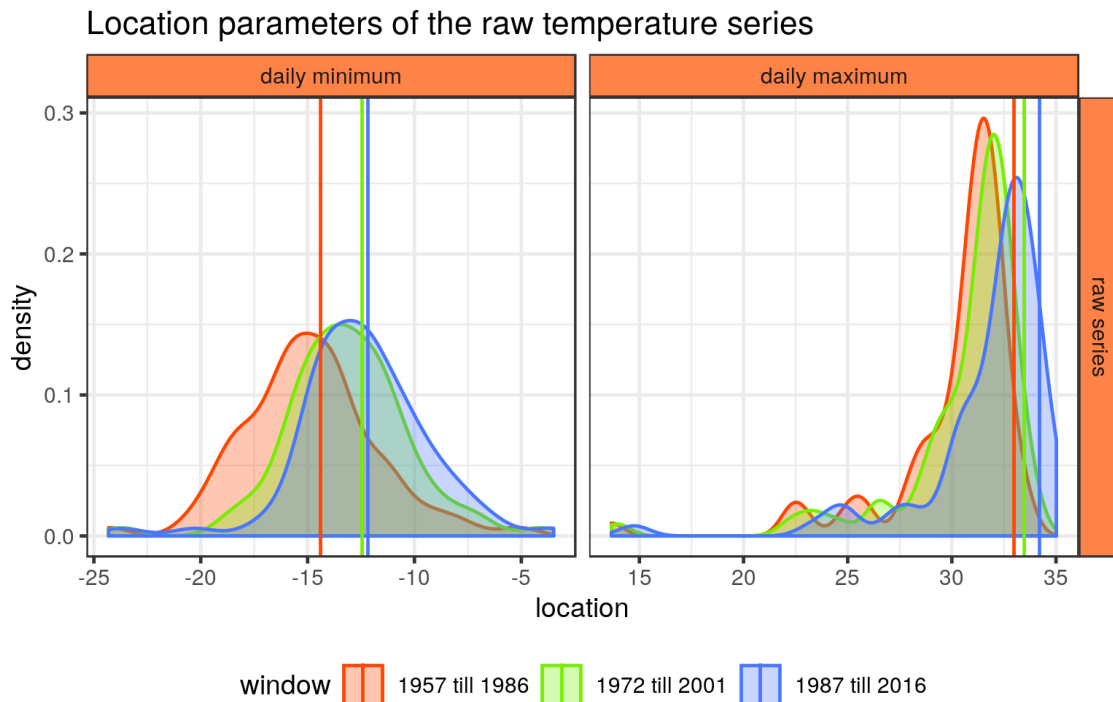


Figure A.29.: Distribution of the location parameters for the daily raw temperature series. This plot is the counterpart of figure A.10 and a more detailed description of the underlying data and analysis can be found in there and in section 7.1 and 7.2.

A.2.1. Non-stationary EVA and significance

Now let's turn to the last part of the temperature analysis. The change in location in figure A.29 is approximately the same as for the anomalies.

The scale parameters in figure A.30, on the other hand, show a different behavior. Apart from a shift for the daily minima in the third time window towards lower values, the distributions remain at roughly the same positions. The scale of the daily maxima, which is on average of higher values than for the anomalies, has more mass at larger values in the third time window than in all others. On closer inspection the usage of a first order linear model to describe those changes does not seem very plausible. But as mentioned beforehand, since a more complex model would not pass a likelihood ratio test, we will keep it as a baseline to probe an overall change of the distribution. Even if we find the test to reject the null hypothesis of a stationary model we will not claim the changes to be perfectly linear. Instead, we learn that there are indeed changes in the distribution function. But their actual nature can be better accessed using the time window approach.

Figure A.31 shows the changes in the shape parameters. The ones of the daily minima do look quite similar to their counterparts of the temperature anomalies. But then again, although a shift towards higher values takes place, only a fraction of the mass is covering positive shape values for the minima of the raw temperature series. In addition, the shift of the maxima towards lower shapes is stronger than for the anomalies.

The resulting return levels in figure A.32 and A.33 match the results of the anomalies

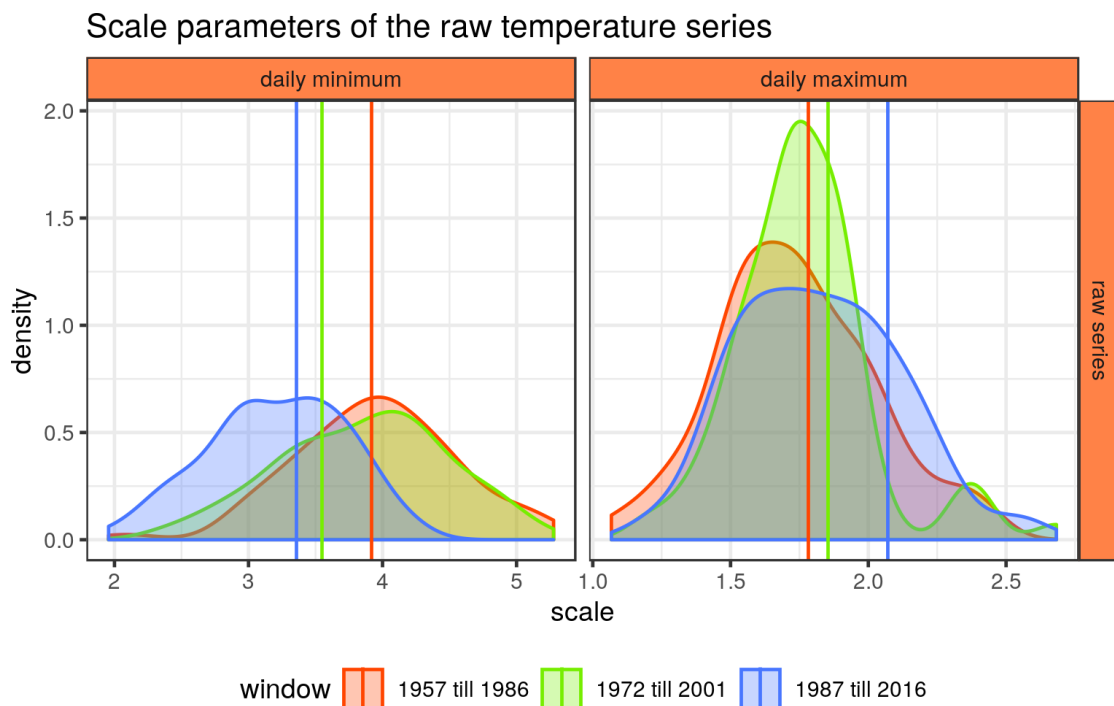


Figure A.30.: Distribution of the scale parameters for the daily raw temperature series. A more detailed description of the underlying data and analysis can be found in section 7.1 and 7.2.

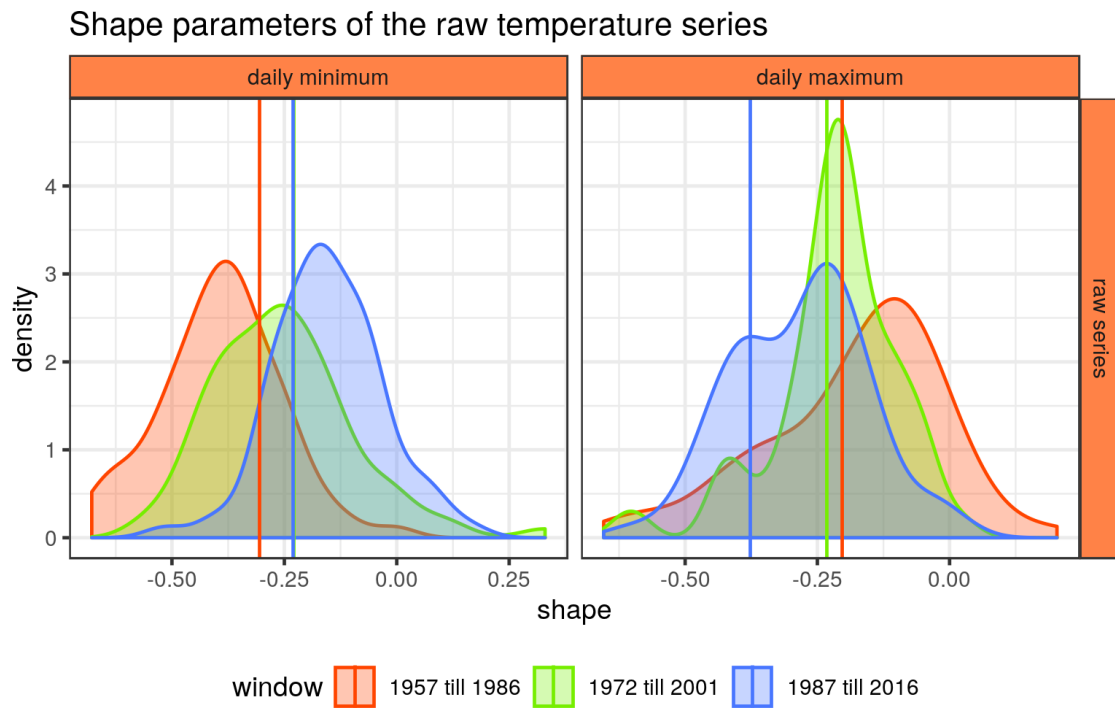


Figure A.31.: Distribution of the shape parameters for the daily raw temperature series. A more detailed description of the underlying data and analysis can be found in section 7.1 and 7.2.

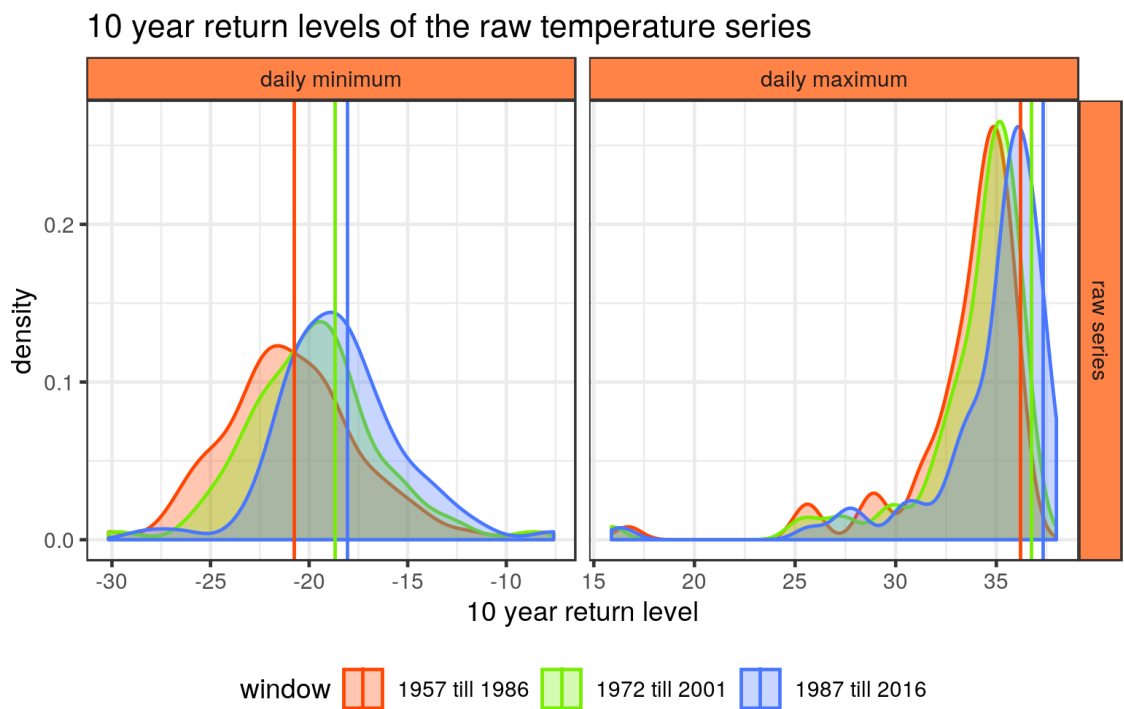


Figure A.32.: Distribution of the 10 year return levels for the daily raw temperature series. A more detailed description of the underlying data and analysis can be found in section 7.1 and 7.2.

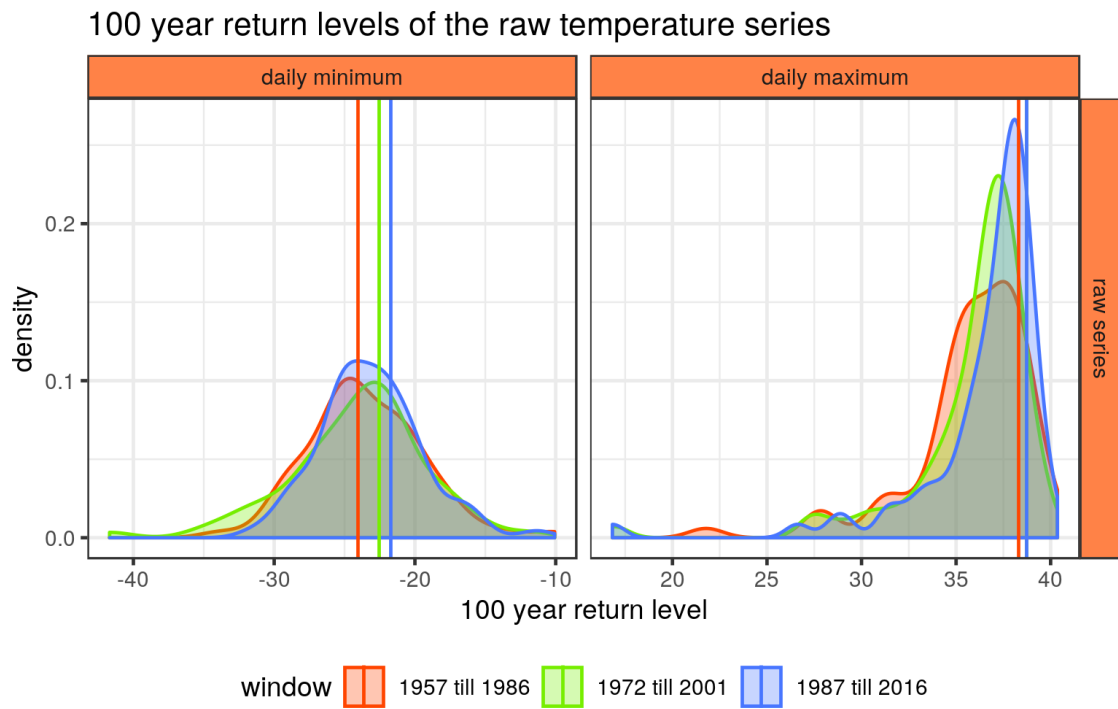


Figure A.33.: Distribution of the 100 year return levels for the daily raw temperature series. A more detailed description of the underlying data and analysis can be found in section 7.1 and 7.2.

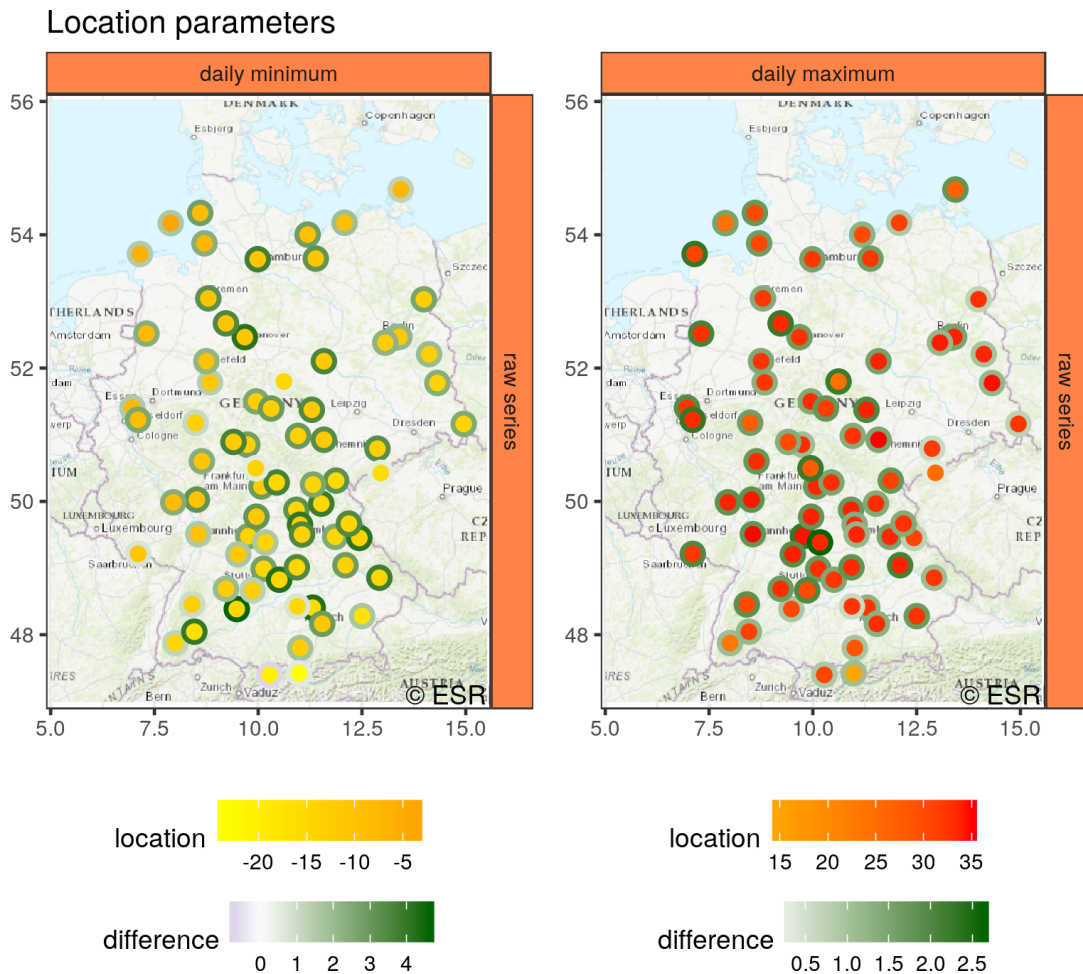


Figure A.34.: Location parameters of the daily raw maximum temperature series throughout Germany. For a detailed description see figure A.15 and section 7.1 and 7.2.

almost perfectly. The only differences are, that there is no change in spread of the 10 year return level distribution for the daily maxima and the spread is getting more less and less for the 100 year return levels. In addition, the Brocken station is no outlier anymore.

The spatial distribution of the location parameters of the daily raw temperature series in figures A.34 matches the one for the anomalies in figure A.15 in most parts. In case of the scale parameters in figure A.35 the decrease of the daily maxima now mostly occurs in the center of Germany while there is the tendency to an increase in the remaining regions. The decrease in scale for the daily minimum temperature series is even more pronounced compared to the anomalies.

The shape parameters in figure A.36 feature their strongest decrease for the daily maxima in the south and most of their increases in the north of Germany. But still, the highest shape values can be found for both the daily minima and maxima in the south of Germany. The increase in shape of the minima is more uniformly spread compared to the anomalies.

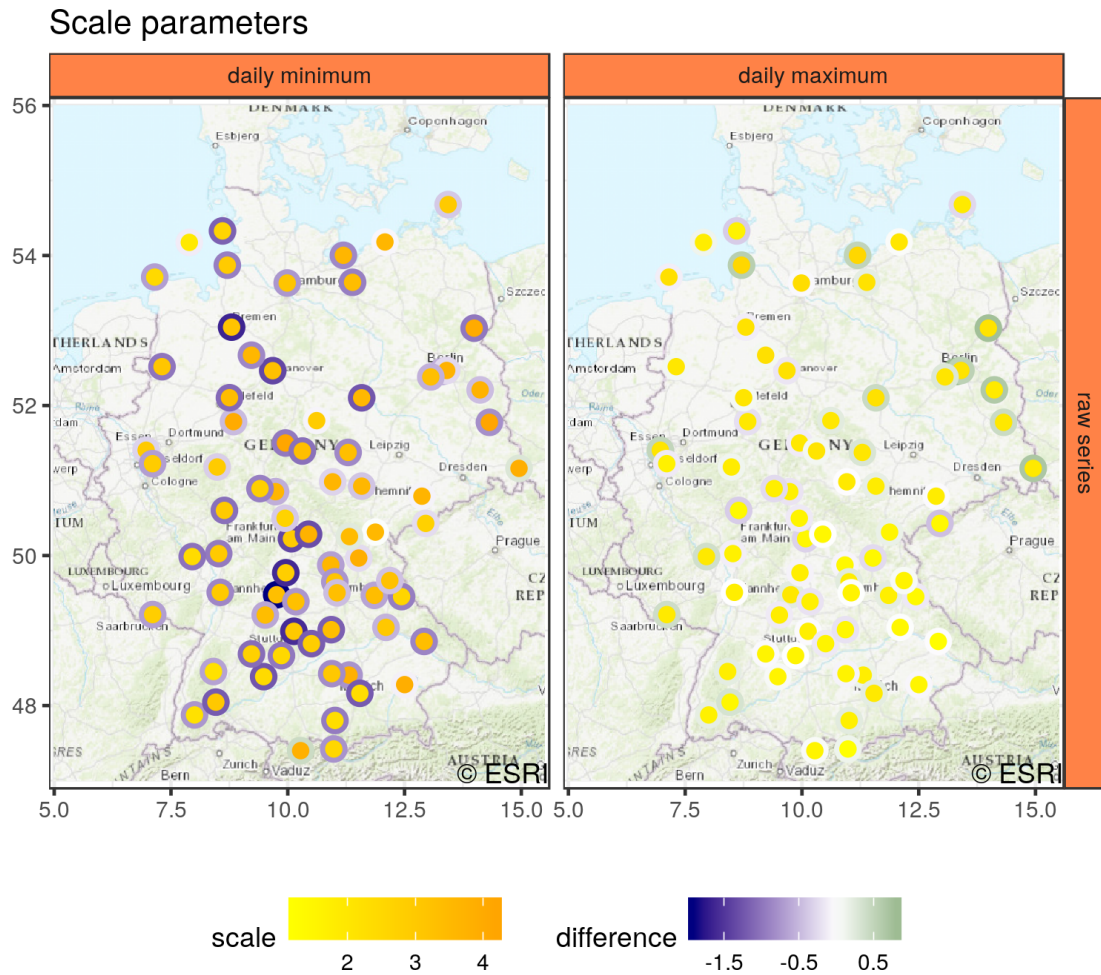


Figure A.35.: Scale parameters of the daily raw maximum temperature series throughout Germany. For a detailed description see section 7.1 and 7.2.

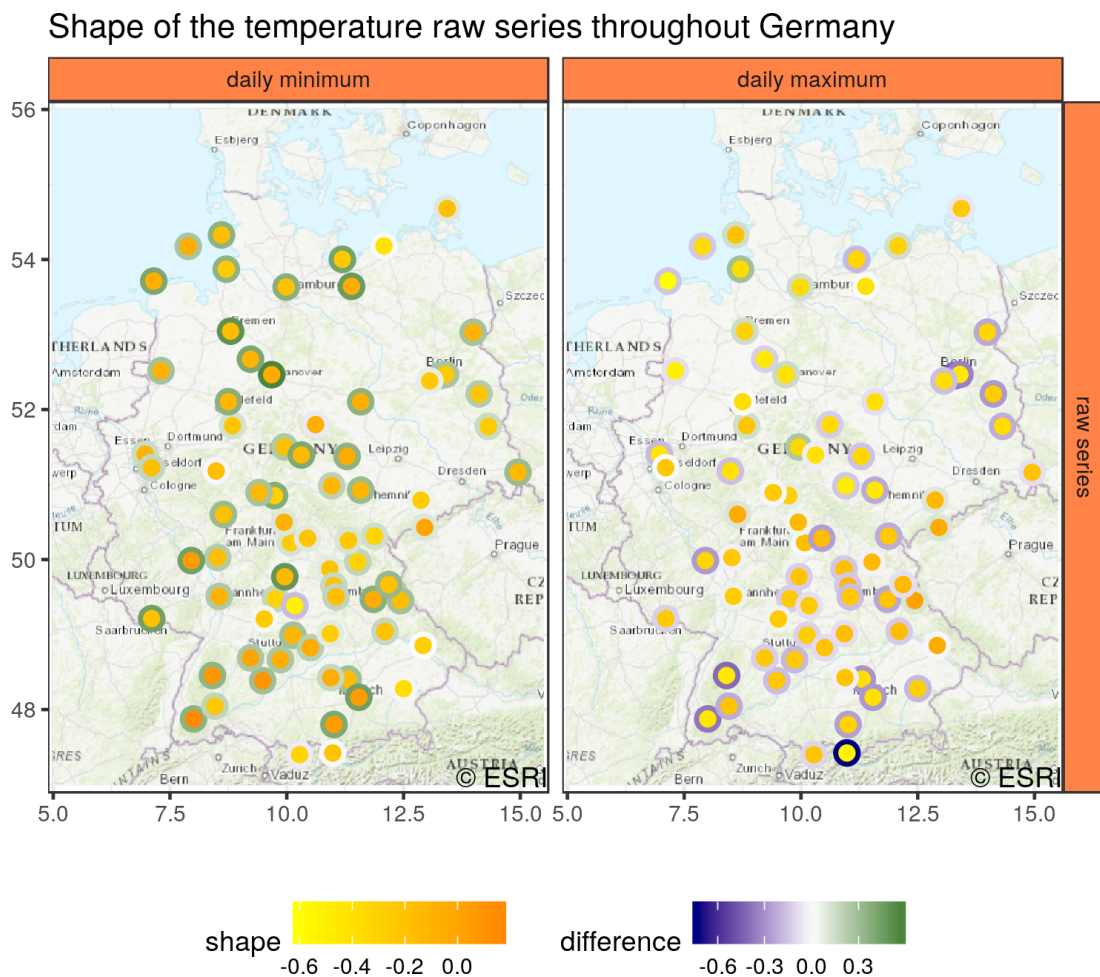


Figure A.36.: Shape parameters of the daily raw maximum temperature series throughout Germany. For a detailed description see section 7.1 and 7.2.

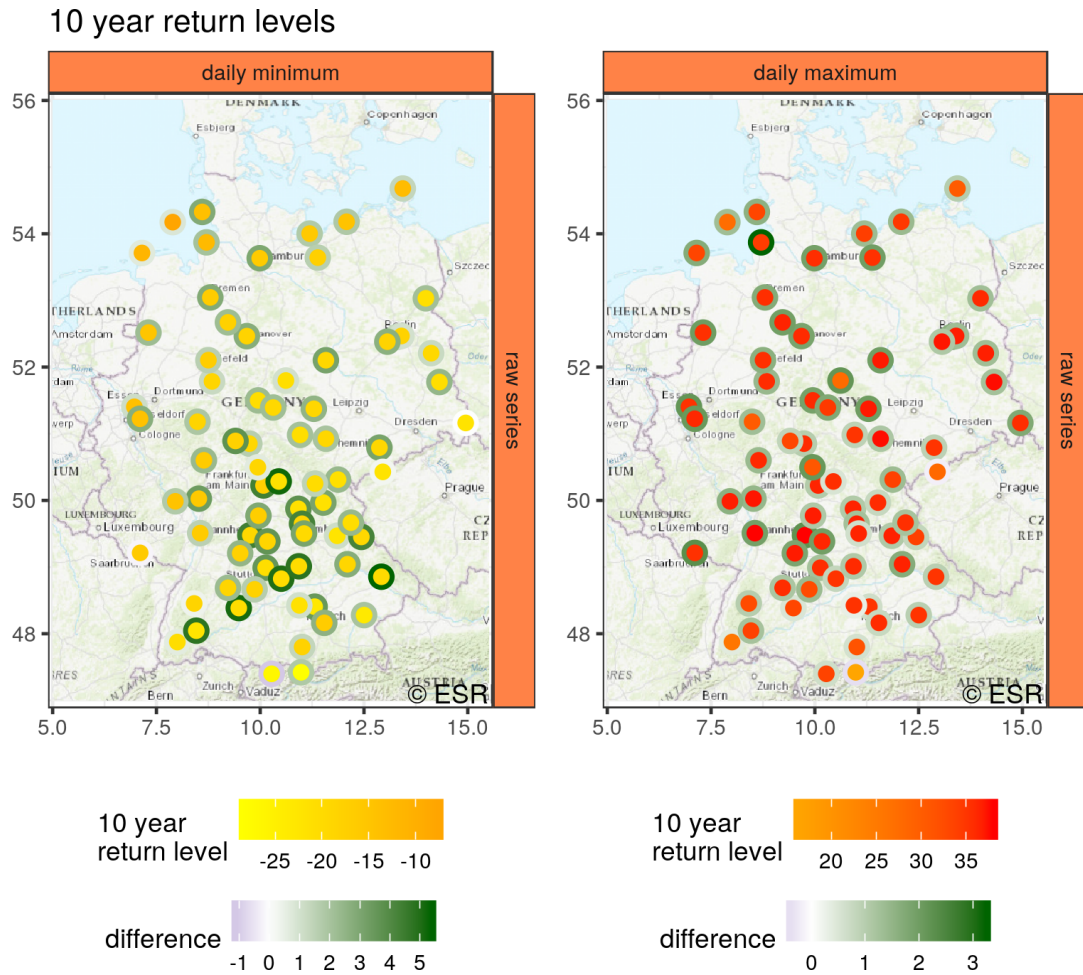


Figure A.37.: 10 year return levels of the daily raw maximum temperature series throughout Germany. For a detailed description see figure A.15 and section 7.1 and 7.2.

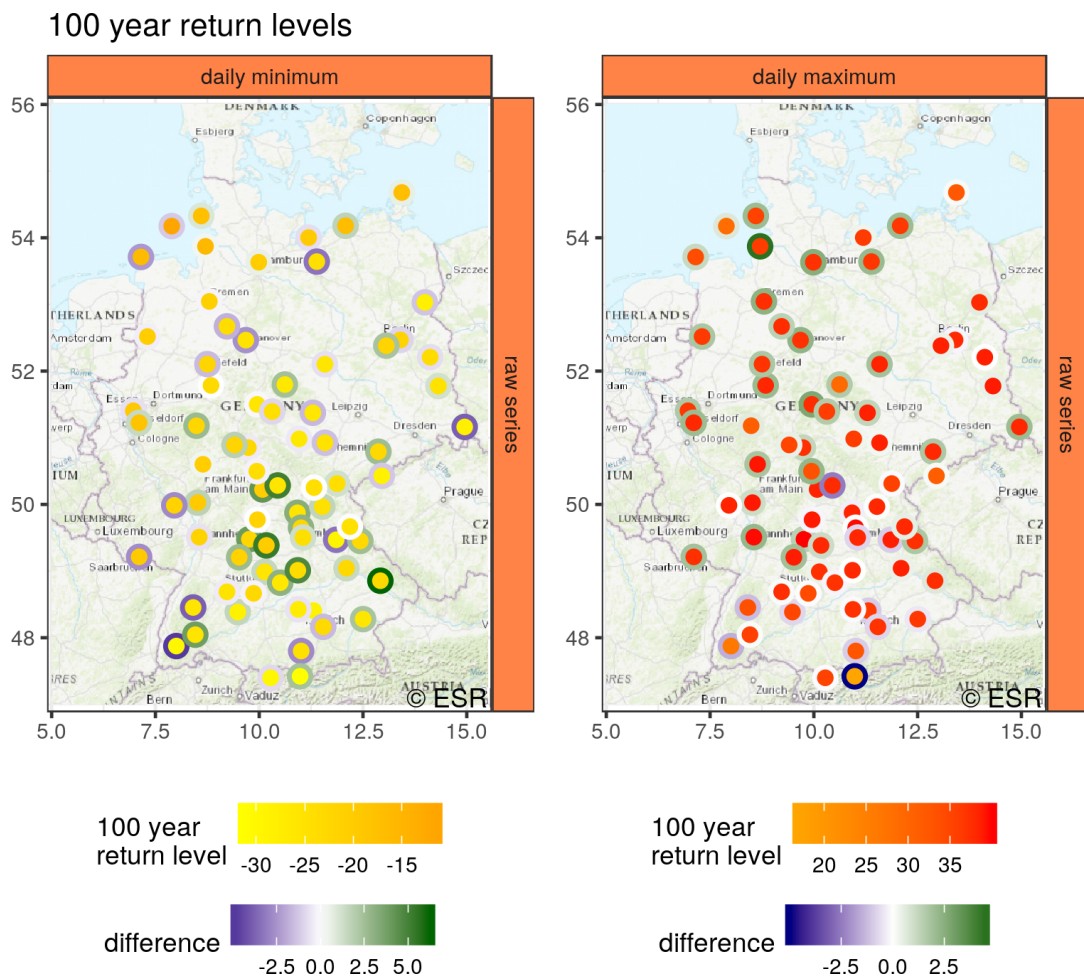


Figure A.38.: 100 year return levels of the daily raw maximum temperature series throughout Germany. For a detailed description see figure A.15 and section 7.1 and 7.2.

VGLM analysis of the raw temperature series

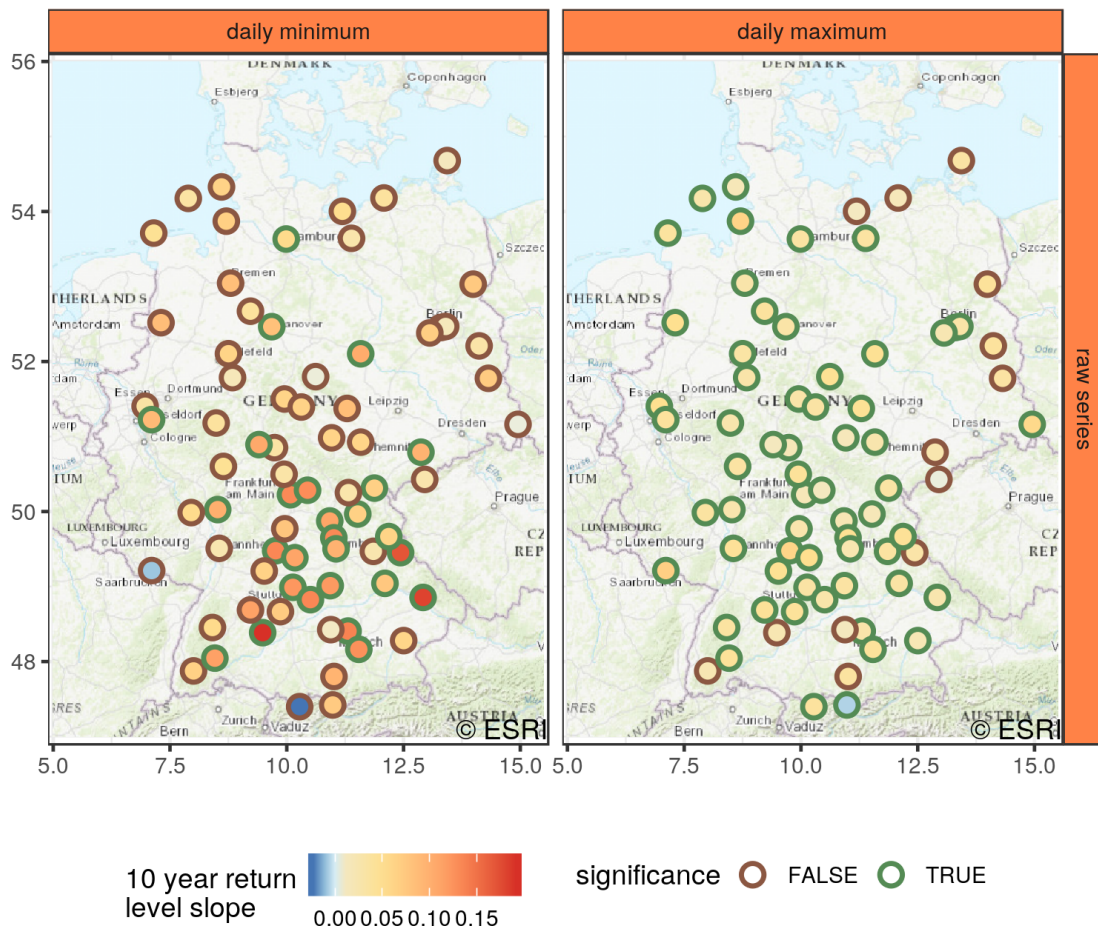


Figure A.39.: Results of the likelihood ratio test whether to use a stationary model over a linear one for the daily raw temperature series. This plot is the counterpart of figure A.20, which shows the results for the daily temperature anomalies and provides a more detailed description.

Figures A.37 and A.38 show the resulting return levels of the daily raw temperature series. As a result of the changes in the shape parameters the increase in return levels of the maxima is strongest in the north while there is even a decrease in the 100 year return levels in the south of Germany. For the minima, on the other hand, the results are quite similar to the anomalies. Thus, the extremely hot and cold days become smaller in magnitude in the south of Germany and larger in the north. The medium extreme hot days become larger in all of Germany but at a faster pace in the north. The medium extreme cold days become lesser in all of Germany but more fast in the south.

The result of the VGLM analysis in figure A.39 are quite similar to the corresponding ones for the daily anomalies. In general less null hypothesis could be rejected for both the daily minima and maxima. This is probably due to the increase in variance for the block maxima, which we already experienced in the case study of the Potsdam series in section 7.4.

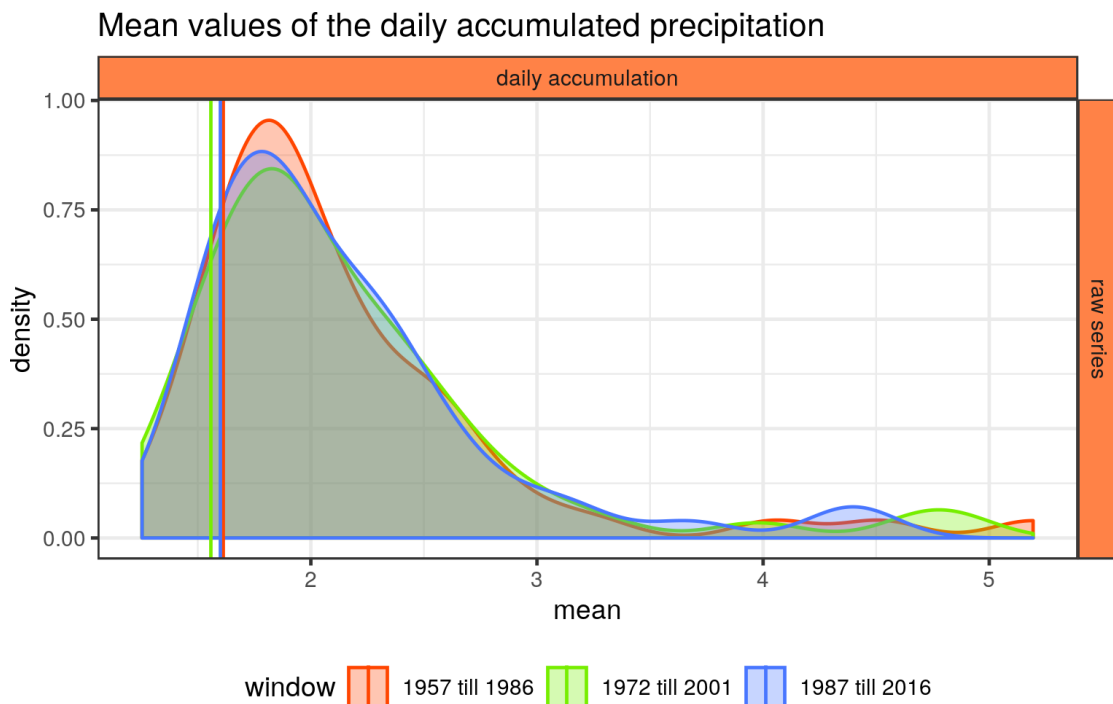


Figure A.40.: Distributions of all mean values calculated from the daily accumulated precipitation throughout Germany. A detailed description of the preprocessing of the 57 stations can be found in section 7.1. To increase the comparability with the case study of the Potsdam station the corresponding results were marked with a vertical line.

A.3. Precipitation throughout Germany

For a general discussion of the precipitation data and how they do differ from the temperature data, please see section 7.6.

The mean values of all 57 precipitation series provided by the DWD can be found in figure A.40. A slight shift in the overall mass of the distribution towards higher values can be seen but in general the distribution remains constant in time.

More pronounced changes can be found for the variances in figure A.41. The bulk of the distribution obtains more mass at higher values in both the second and the third time window. At the same time, the largest variances in the ensemble are getting smaller. The skewness in figure A.42 shows the opposite behavior compared to the findings of the Potsdam case study in section 7.6.

The distribution of the kurtosis in figure A.43, on the other hand, gains more mass at lower values. The very large values in the first and second time window are getting smaller by a factor of approximately three. Thus, while constant in their mean values, the variability of the rain becomes larger and medium sized precipitation events become more frequent in Germany. But the magnitude of extreme precipitation events will possibly decrease. We will check this assumption in the next section.

The maps for the mean values in figure A.44 and the variances in A.45 do show the very similar patterns. The largest decreases do happen in the south west of Germany

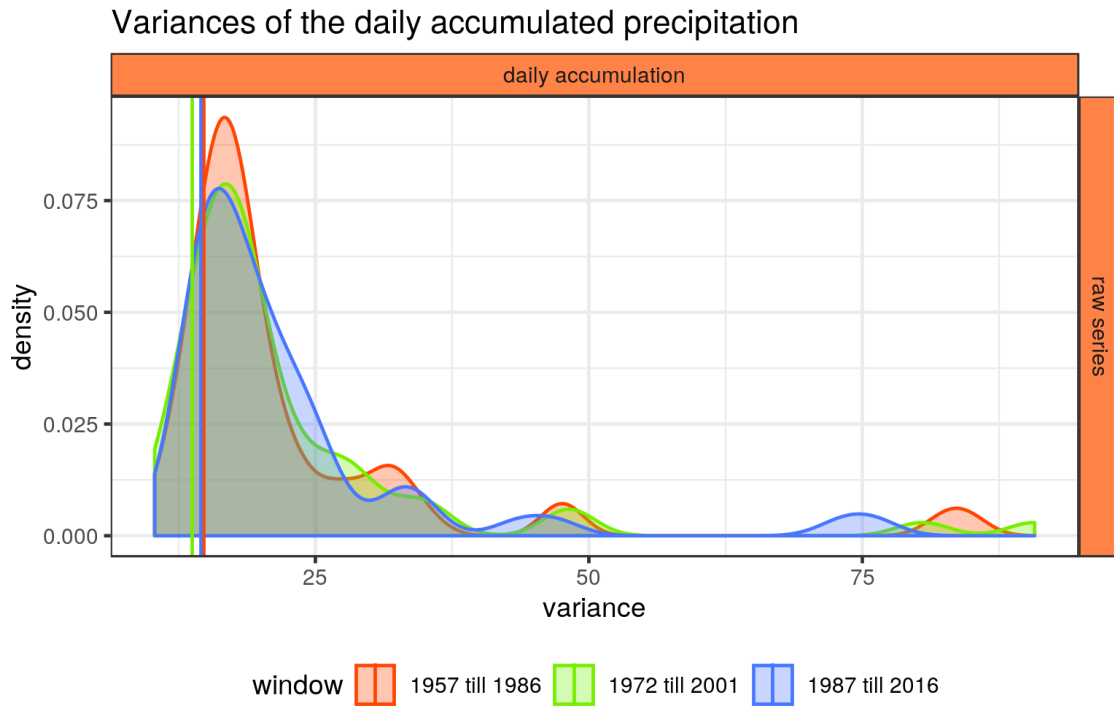


Figure A.41.: Distribution of all variances of the daily accumulated precipitation within Germany. For a more detailed description see figure A.40 and section 7.1.

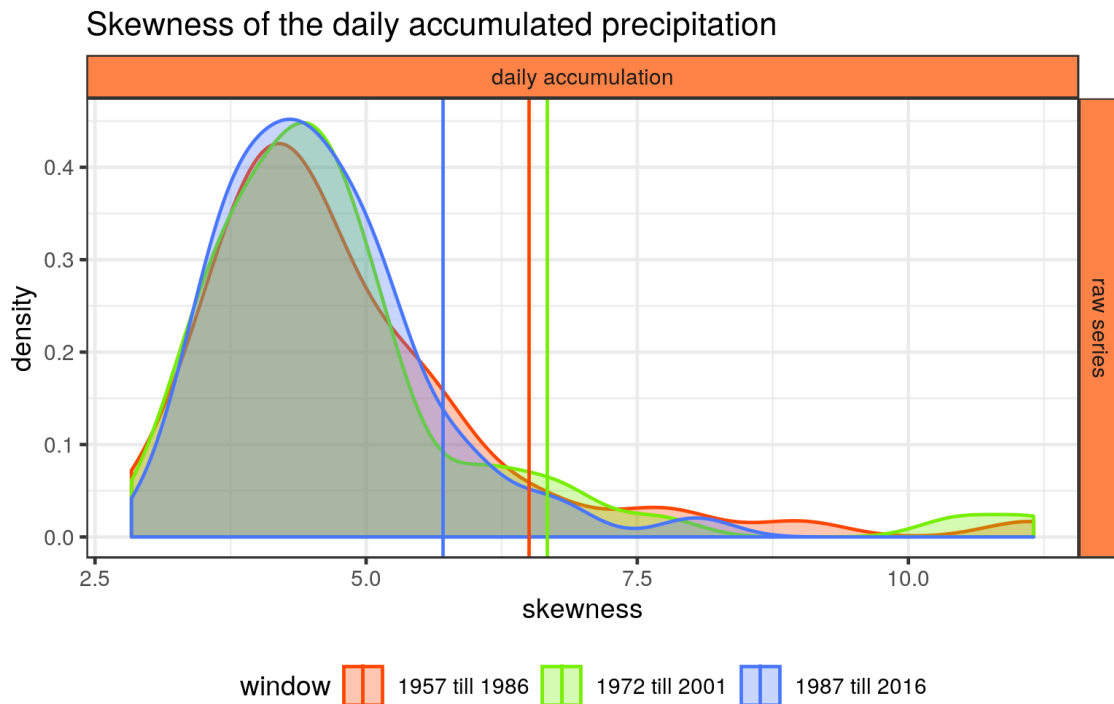


Figure A.42.: Distribution of all skewness values of the daily accumulated precipitation within Germany. For a more detailed description see figure A.40 and section 7.1.

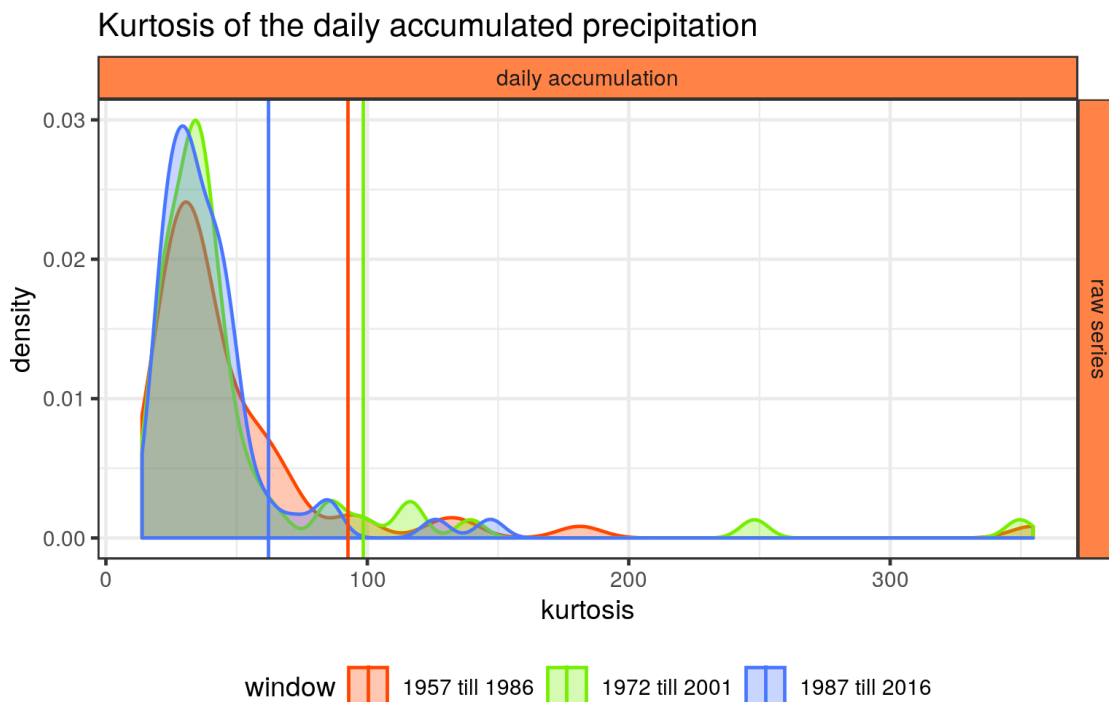


Figure A.43.: Distribution of all kurtosis values of the daily accumulated precipitation within Germany. For a more detailed description see figure A.40 and section 7.1.

while the biggest increases do occur in the north and north west.

For the skewness in figure A.46 and the kurtosis in A.47, again, the patterns appear to be quite similar. Now, both the largest values and the largest decreases can be found in the east of Germany. The dominance in color for the differences between the time windows is due to the heavy tailed nature of the precipitation data, as can be seen for the Potsdam series in figure 7.19. The biggest negative difference is caused by a series featuring its largest events, more than three times the size of the second largest one, in the first time window. For the biggest positive difference in both kurtosis and skewness the largest event, now more than double the size of the second largest, is located in the third time window. In both cases the overall series itself looks pretty stationary.

But overall the maps seem to be as coherent and uniform as the corresponding ones for the temperature series. This result is unexpected since the correlation length in temperature data is said to be larger than for precipitation and the latter is stronger affected by the local elevation profile.

To further investigate the spatial correlations between the moments of the distributions, all estimates were plotted against both the longitude and latitude of the corresponding station in figure A.48. One can see a pronounced trend in the skewness and kurtosis to have both higher values and a larger variability for higher longitudes. The mean, and to some extends also the variance, does decrease with higher longitudes. The main reason for this correlation can be found in atmospheric dynamics over Europe. In general, air saturates with water over the Atlantic Ocean and moves eastwards to central Europe. The probability of rain to occur thus is higher in the west of Germany, which the humid

Mean values of the daily accumulated rainfall

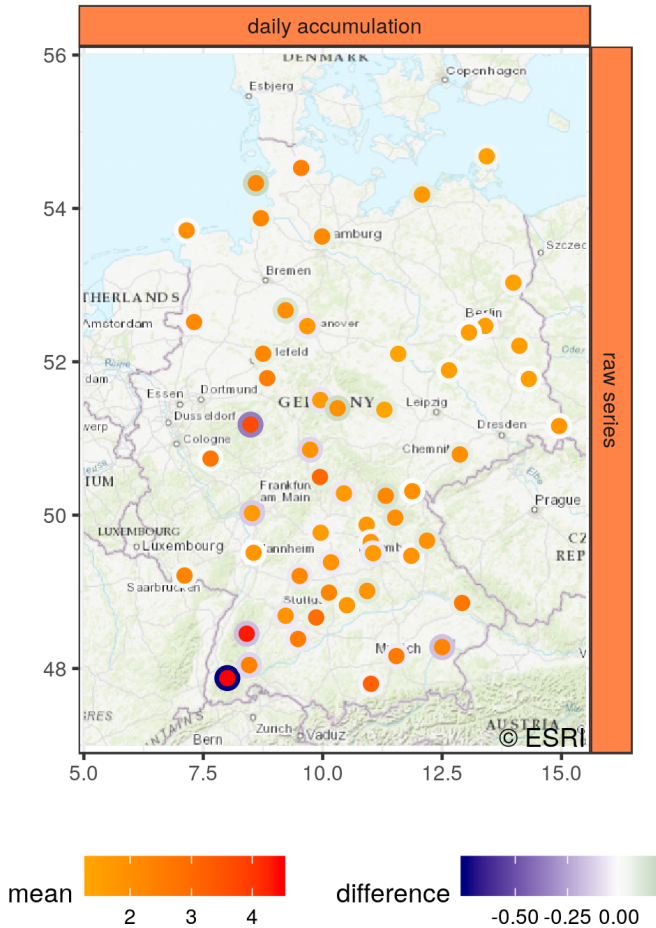


Figure A.44.: Mean values of the daily accumulated precipitation throughout Germany. This map illustrates the spatial information missing in figure A.40. The colors within the circles, ranging from orange to red, display the values of the mean within the third and last time window. The colors of their outer surrounding, on the other hand, show the magnitude of the temporal evolution calculated from the values of the third window minus the ones of the first. Note that for better comparability of the different maps in this analysis the color ranges were fixed as described in section 7.2

Variiances of the daily accumulated rainfall

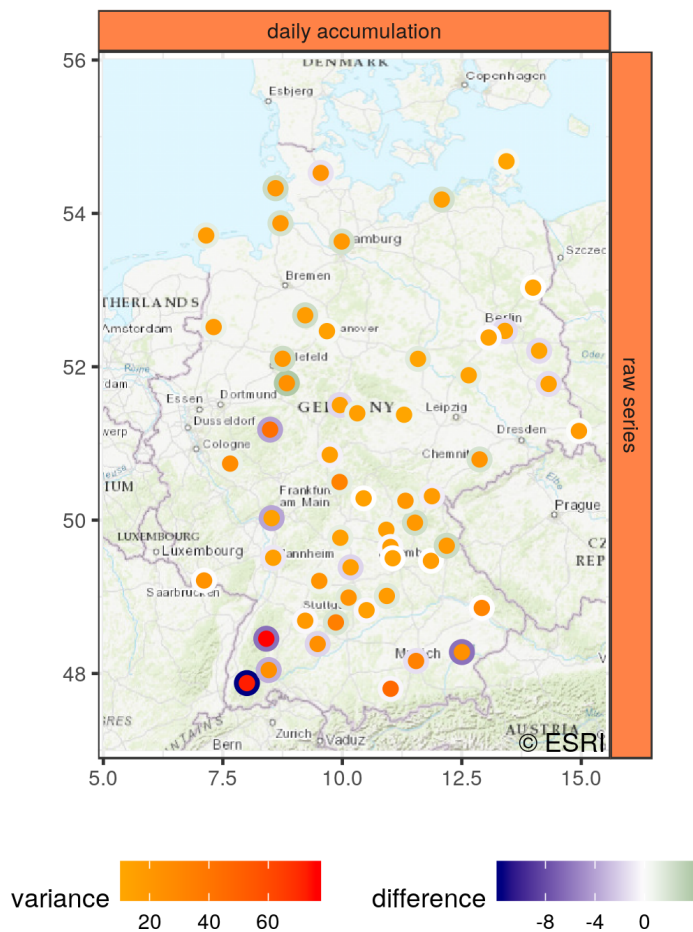


Figure A.45.: Variiances of the daily accumulated precipitation throughout Germany. See figure A.44 and section 7.2 for a detailed description.

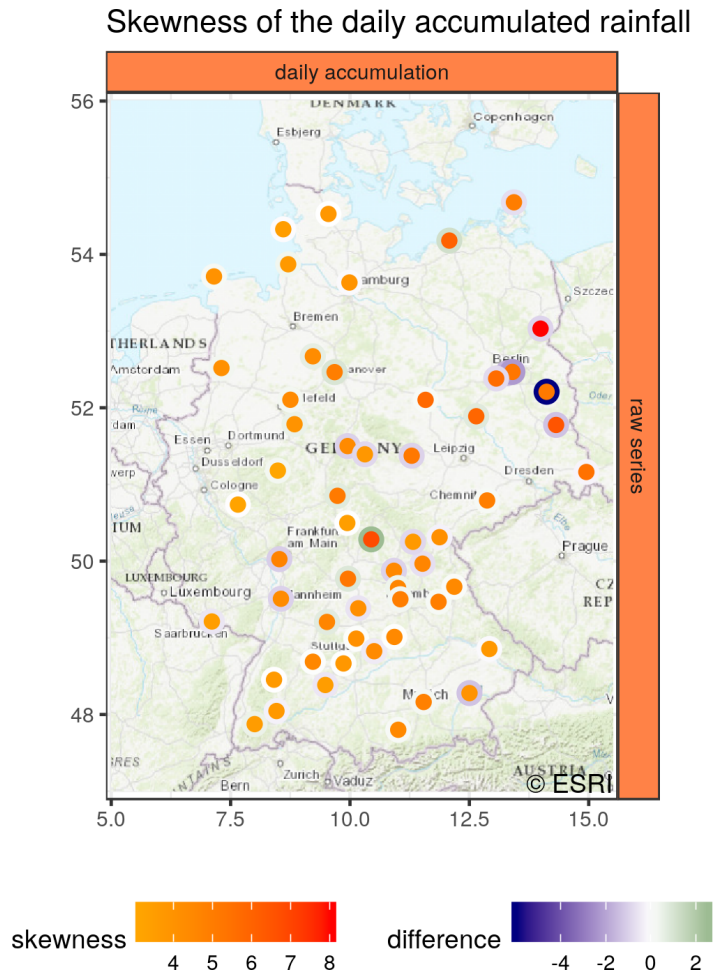


Figure A.46.: Skewness of the daily accumulated precipitation throughout Germany. See figure A.44 and section 7.2 for a detailed description.

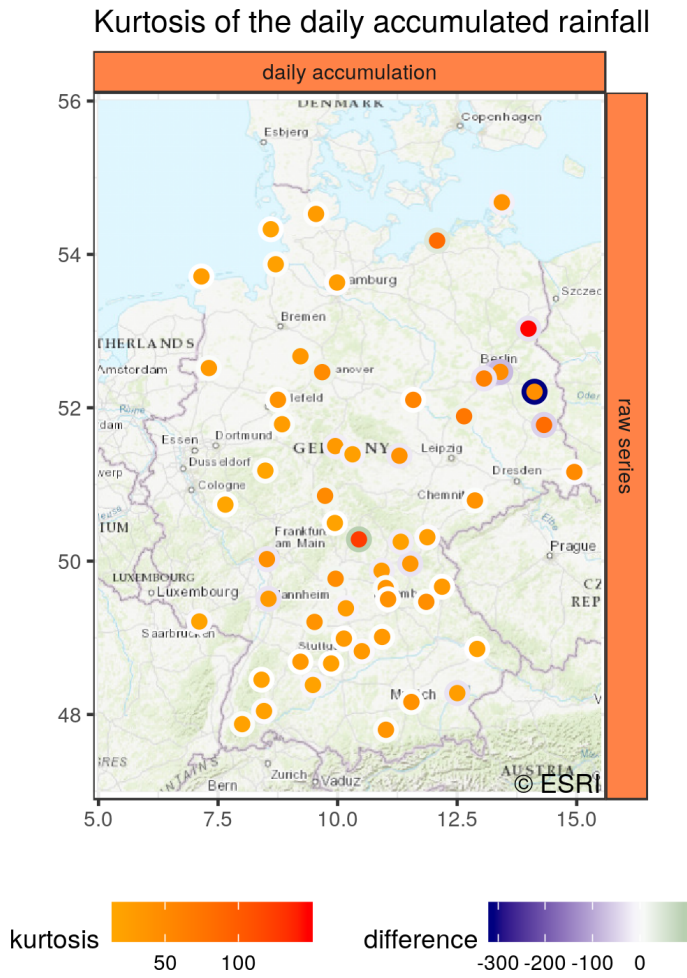


Figure A.47.: Kurtosis of the daily accumulated precipitation throughout Germany. See figure A.44 and section 7.2 for a detailed description.

Dependence of the precipitation on the geographic location

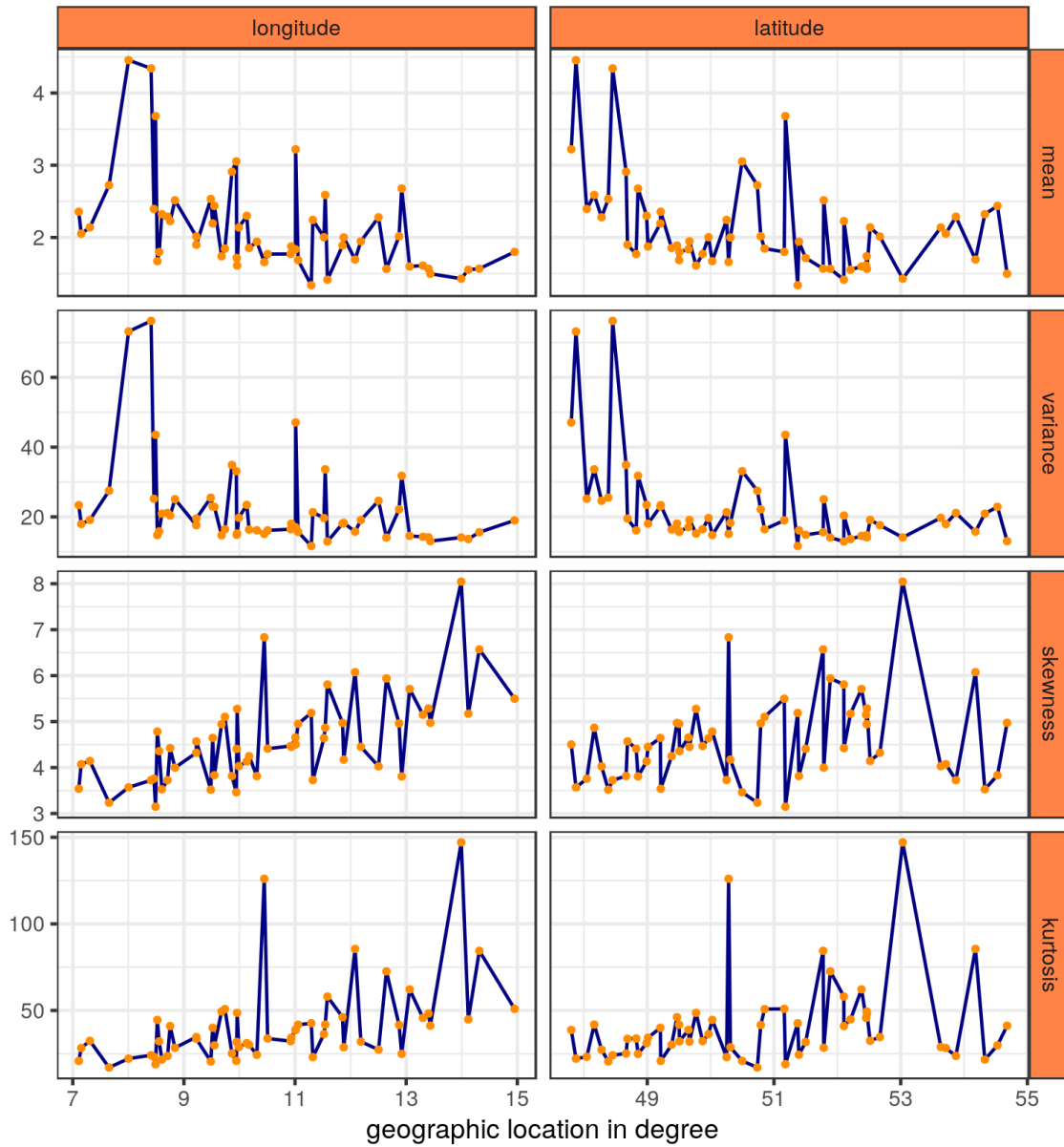


Figure A.48.: Plot of the dependence of the first four moments of the daily accumulated precipitation series on the geographic location of the corresponding measurement station. The rows do display the different moments while the columns represent the longitudes and latitudes of the stations.

air passes first. Similar patterns can be found for the latitudes. Both the mean and variance are larger and more variable in the south of Germany and the kurtosis and skewness do increase in both value and variability with higher latitudes. The increase of the mean precipitation in the south is probably due to presence of several larger mountainsides, where the humid air has to drop a lot of its water content in order to pass the hills and mountains. The increase in skewness and kurtosis, on the other hand, might be attributed to a general correlation between lesser values and variability in mean and variance and higher values and variability in skewness and kurtosis judging from figure A.48. But upon closer inspection of the corresponding map plots this hypothesis is high improbable. The correlation between skewness or kurtosis and latitude is probably an artifact because of the overall shape of Germany. This causes all stations in the east to also be located at relative high latitudes. Therefore, the correlation between the third and fourth moment and the latitude is probably due to the correlation between both quantities and the longitudes of their measurement stations.

Speaking of correlations, the similarities in the maps of the mean and variance as well as skewness and kurtosis also do appear in figure A.48. Both the first and second and the third and fourth moments seem to be highly correlated. And indeed, the Pearson correlation is 0.958 between the mean and variance and 0.943 between the skewness and kurtosis. This finding might be very useful in determining the right distribution for numerical simulations of precipitation events.

A.3.1. Non-stationary EVA and significance

The changes in the location parameters in figure A.49 are more pronounced than the ones in the mean values of the raw precipitation series. The bulk of the distribution, again, gets more and more mass at higher values while at the same time the largest location parameters become smaller. The scale parameters in figure A.50 do shift towards smaller values.

The same is true for the shape parameters in figure A.51. Therefore, more and more mass of the distribution function is located at negative shape values. From all we have discussed so far it may seem counterintuitive to have a precipitation series with a negative shape parameter. But not every time series and not every time window features a very dominant precipitation event, which surpasses all others. For some stations one can find block maxima, which look instead similar to the ones of raw daily temperature series of the Potsdam in figure 7.8.

Compared to the findings of the case study, the overall changes in the return levels throughout Germany in figure A.52 and A.53 are much more pronounced. For both return periods the distributions shift towards lower values and become much more narrow. For the 100 year return levels the distribution loses a lot of weight at its right tail. This means, while the overall distribution of the precipitation remains approximately constant throughout Germany, the extreme rainfall events do occur more seldom and do have a smaller magnitude.

The maps of the mean values in figure A.44 and the location parameter in figure A.54 are very similar and a stronger tendency for the occurrence of decreases in the 10 year return levels in figure A.57 can be found. Apart from this, all other maps of the time window approach do not show any spatial patterns at all. In contrast to the four

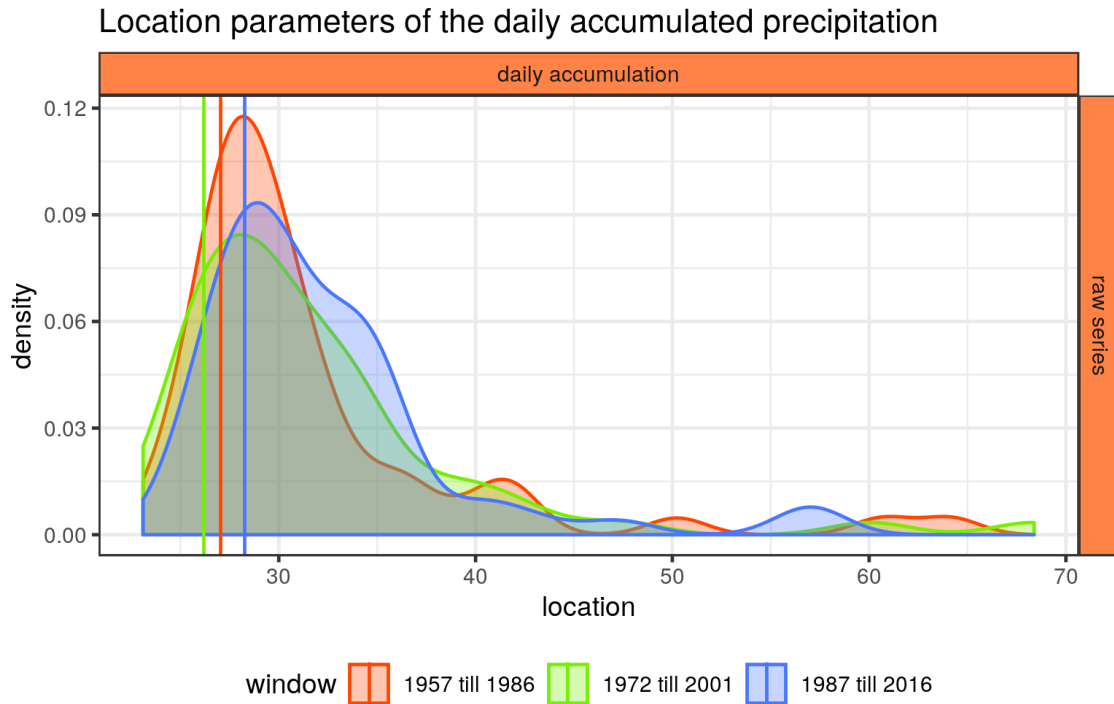


Figure A.49.: Distributions of all location parameters calculated for the daily accumulated precipitation of the 57 different stations provided by the DWD. For a distribution of a specific color only the annual block maxima of the corresponding time window were used. In order compare to the results with the case study of the Potsdam station, its results are marked with vertical lines. For detailed description see section 7.1 and 7.2.

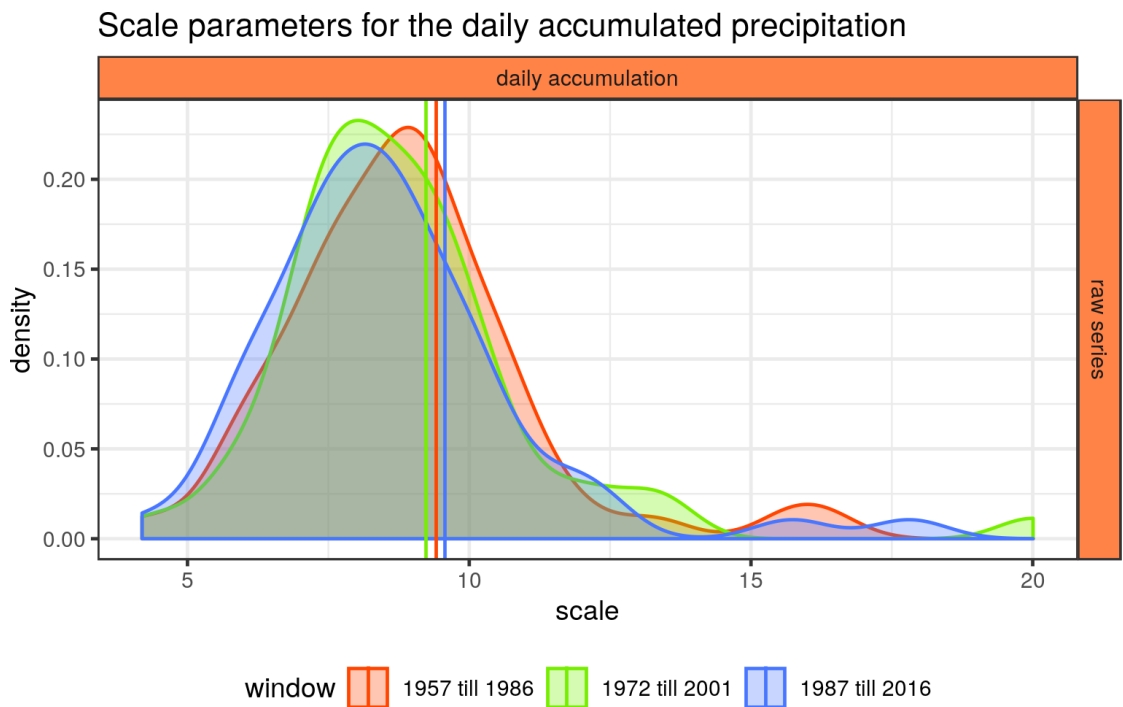


Figure A.50.: Distribution of all scale parameters for the daily accumulated precipitation throughout Germany. For details see figure A.49 and section 7.1 and 7.2.

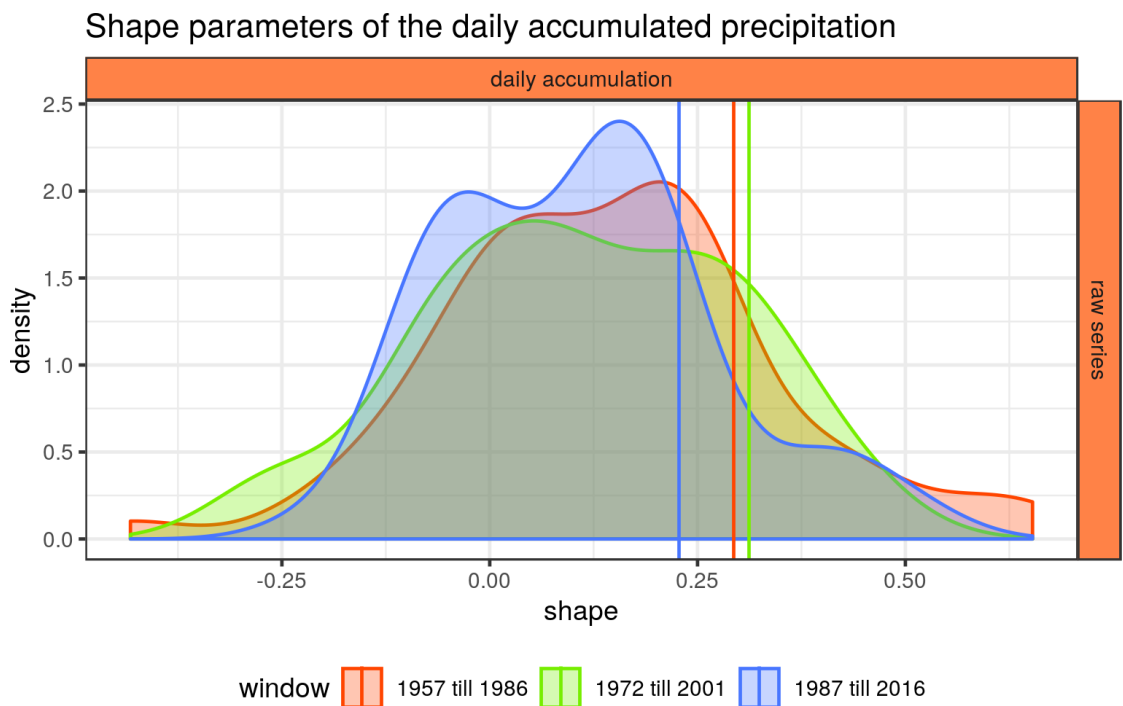


Figure A.51.: Distribution of all shape parameters for the daily accumulated precipitation throughout Germany. For details see figure A.49 and section 7.1 and 7.2.

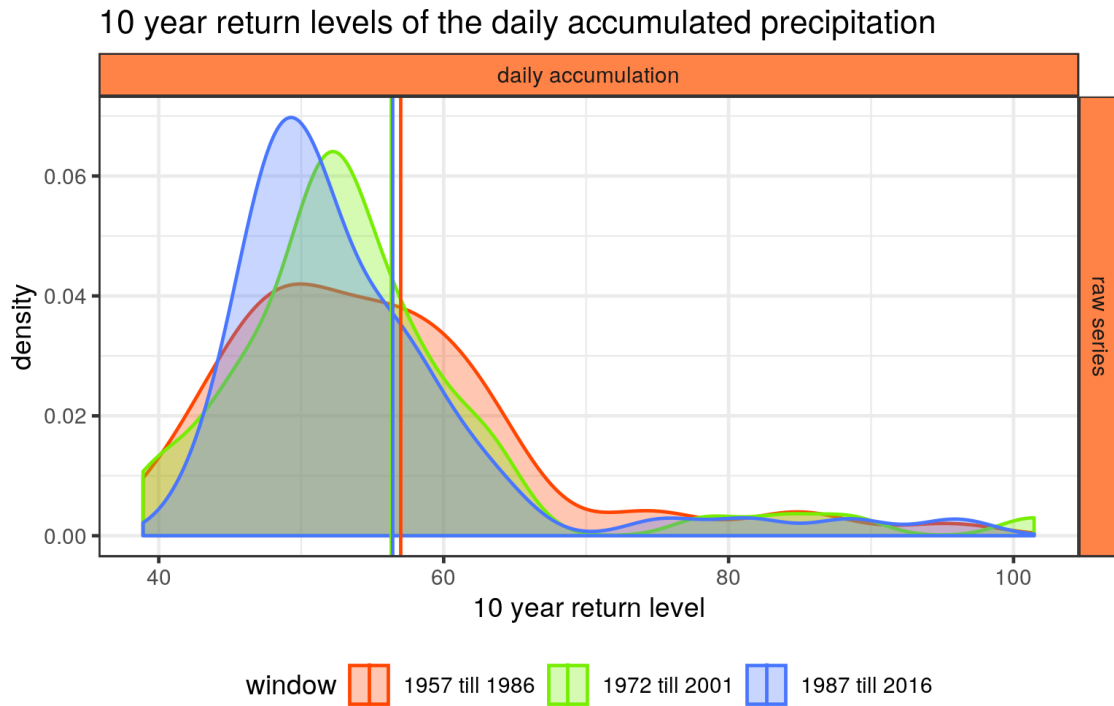


Figure A.52.: Distribution of all 10 year return levels for the daily accumulated precipitation throughout Germany. For details see figure A.49 and section 7.1 and 7.2.

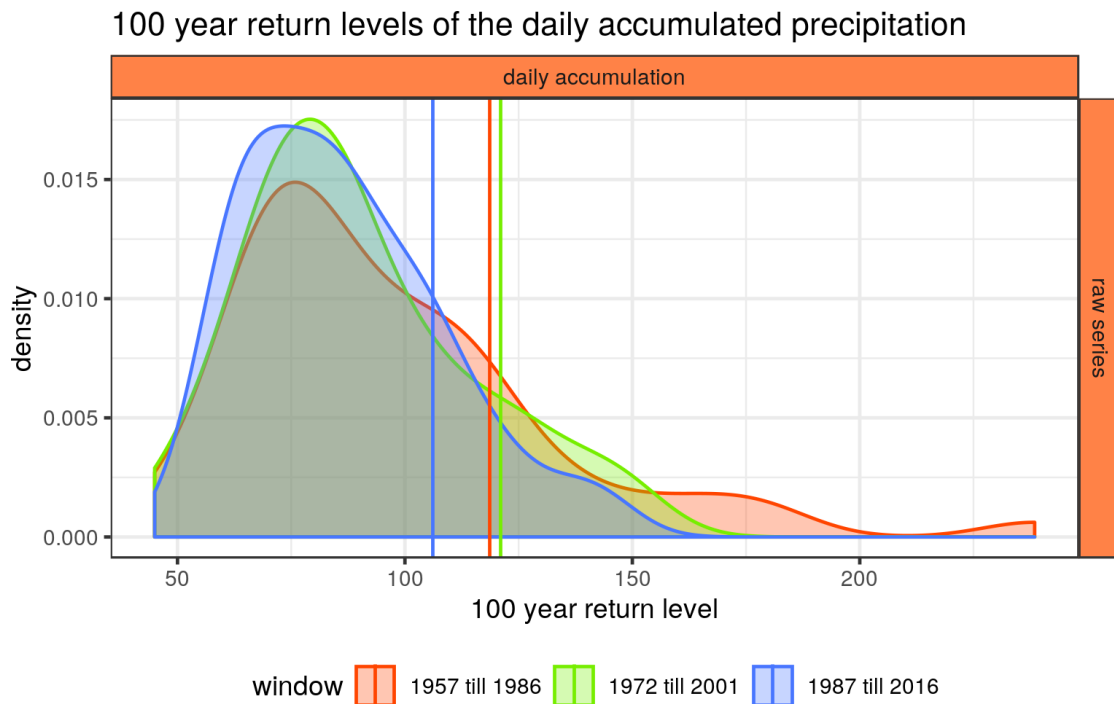


Figure A.53.: Distribution of all 100 year return levels for the daily accumulated precipitation throughout Germany. For details see figure A.49 and section 7.1 and 7.2.

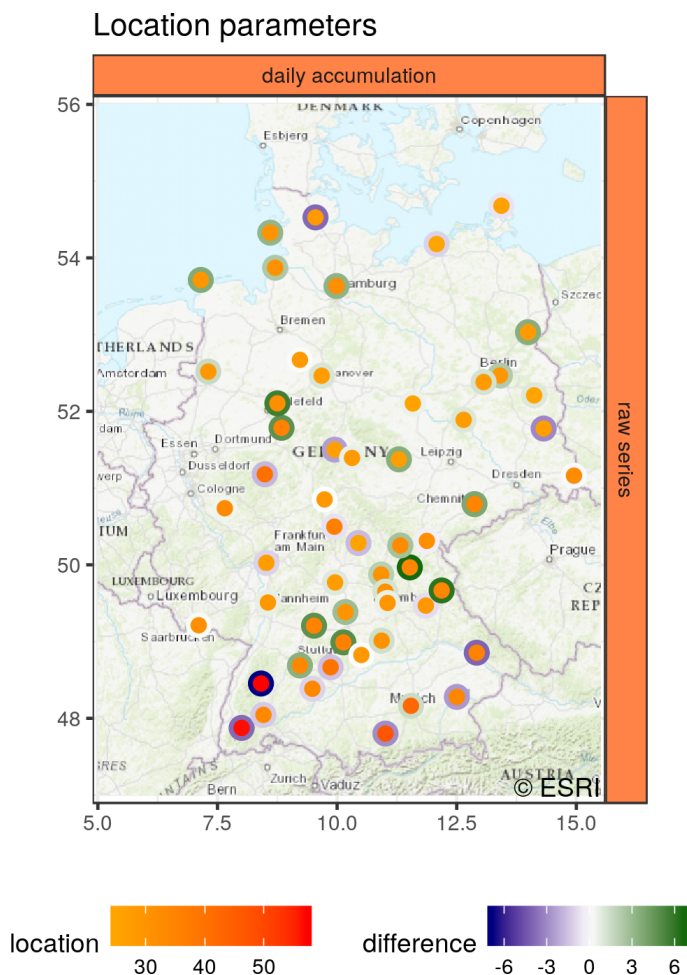


Figure A.54.: Location parameters of the daily accumulated precipitation throughout Germany. For a detailed description of the figure and colors please see section 7.1 and 7.2.

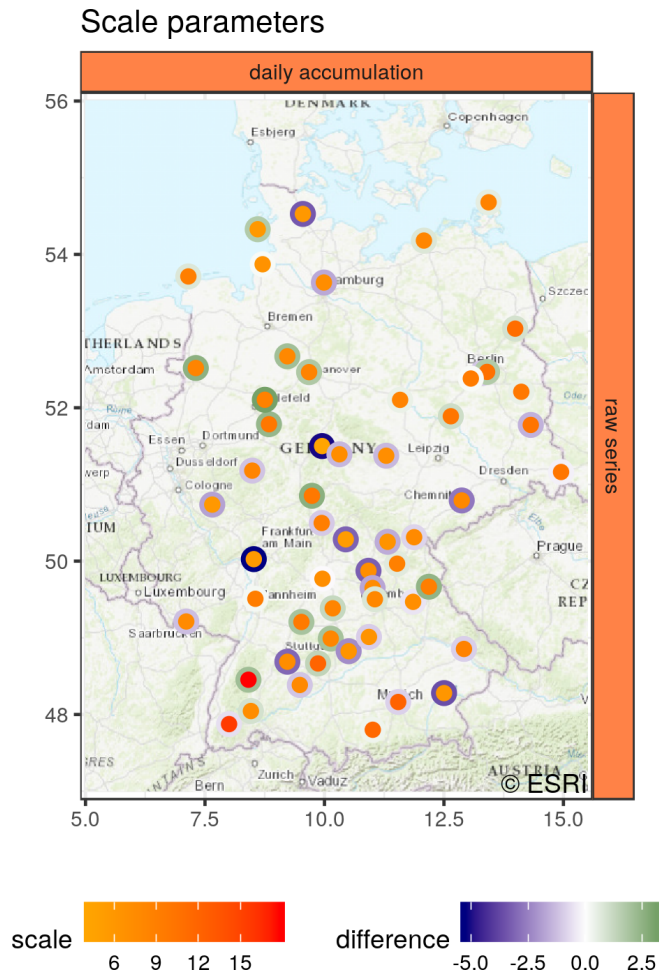


Figure A.55.: Scale parameters of the daily accumulated precipitation throughout Germany. For a detailed description of the figure and colors please see section 7.1 and 7.2.

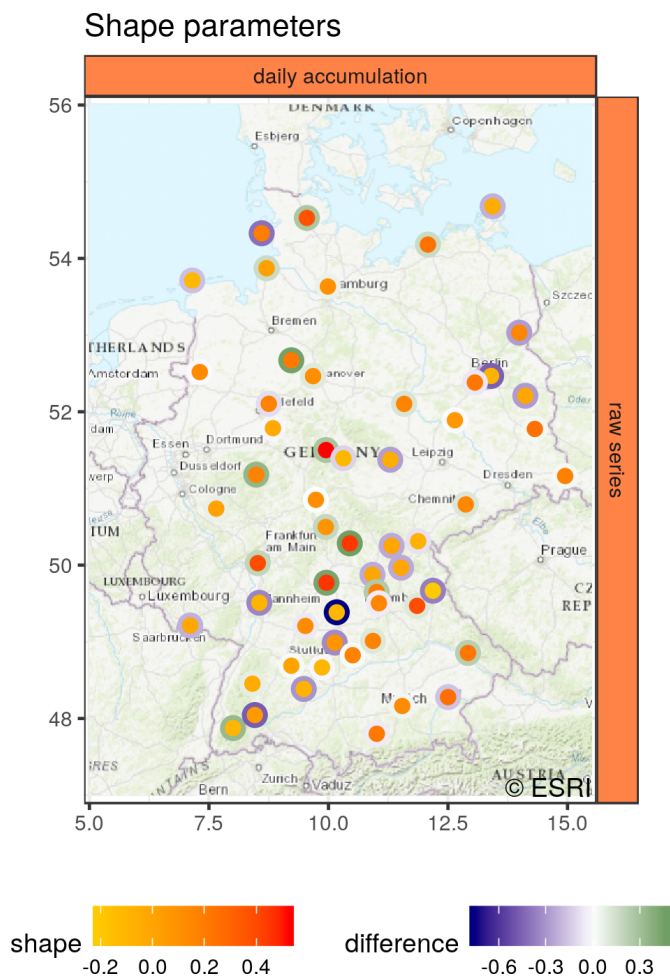


Figure A.56.: Shape parameters of the daily accumulated precipitation throughout Germany. For a detailed description of the figure and colors please see section 7.1 and 7.2.

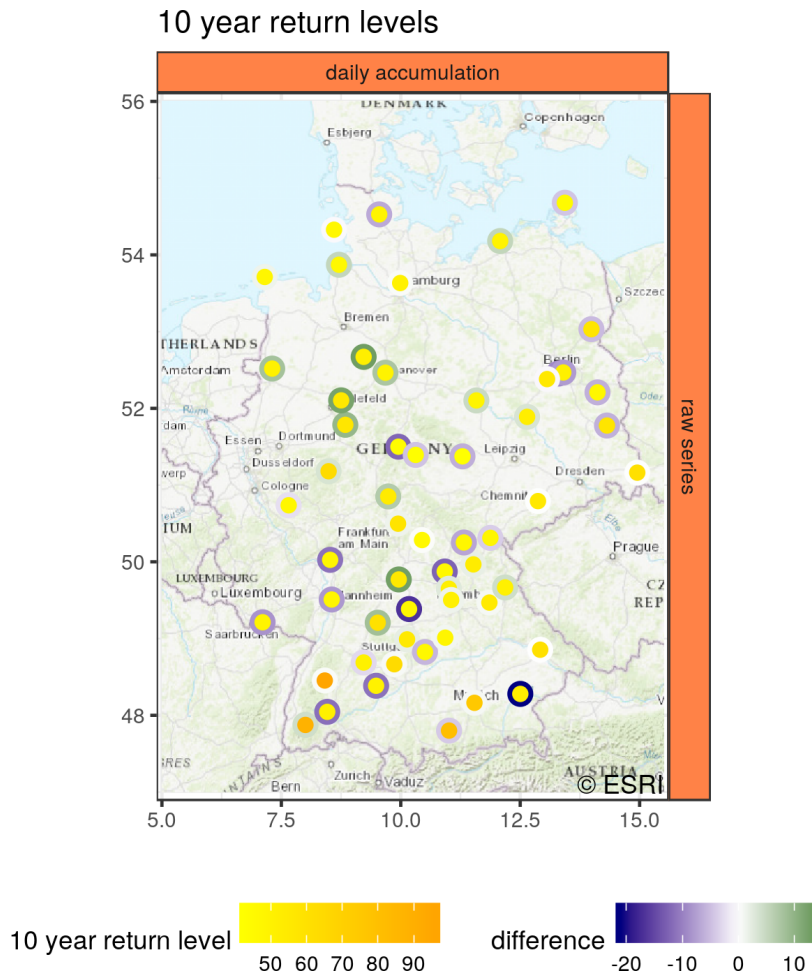


Figure A.57.: 10 year return levels of the daily accumulated precipitation throughout Germany. For a detailed description of the figure and colors please see section 7.1 and 7.2.

moments of the series their extreme events do appear to be indeed almost independent of each other. At most at regional scales on a distance of a few hundreds of kilometers correlations between the results can be found.

Finally, let's review the results of the VGLM analysis of the precipitation data throughout Germany in figure A.59. For almost none of the stations the null hypothesis could be rejected. Therefore, statistically speaking, no trend in the extreme precipitation events in Germany could be found. But this doesn't match the results obtained in the time window analysis where the distributions of the return levels were found to shift towards lower values, especially for the 100 year return levels. But this change is either not of the a nature, which can be well described by a first order linear model, or it is not pronounced enough to pass the likelihood ratio test. Thus, as a final results of the precipitation analysis, we indeed find a general tendency towards less extreme precipitation events but actual nature of the changes in the extremes do depend on the geographic location and

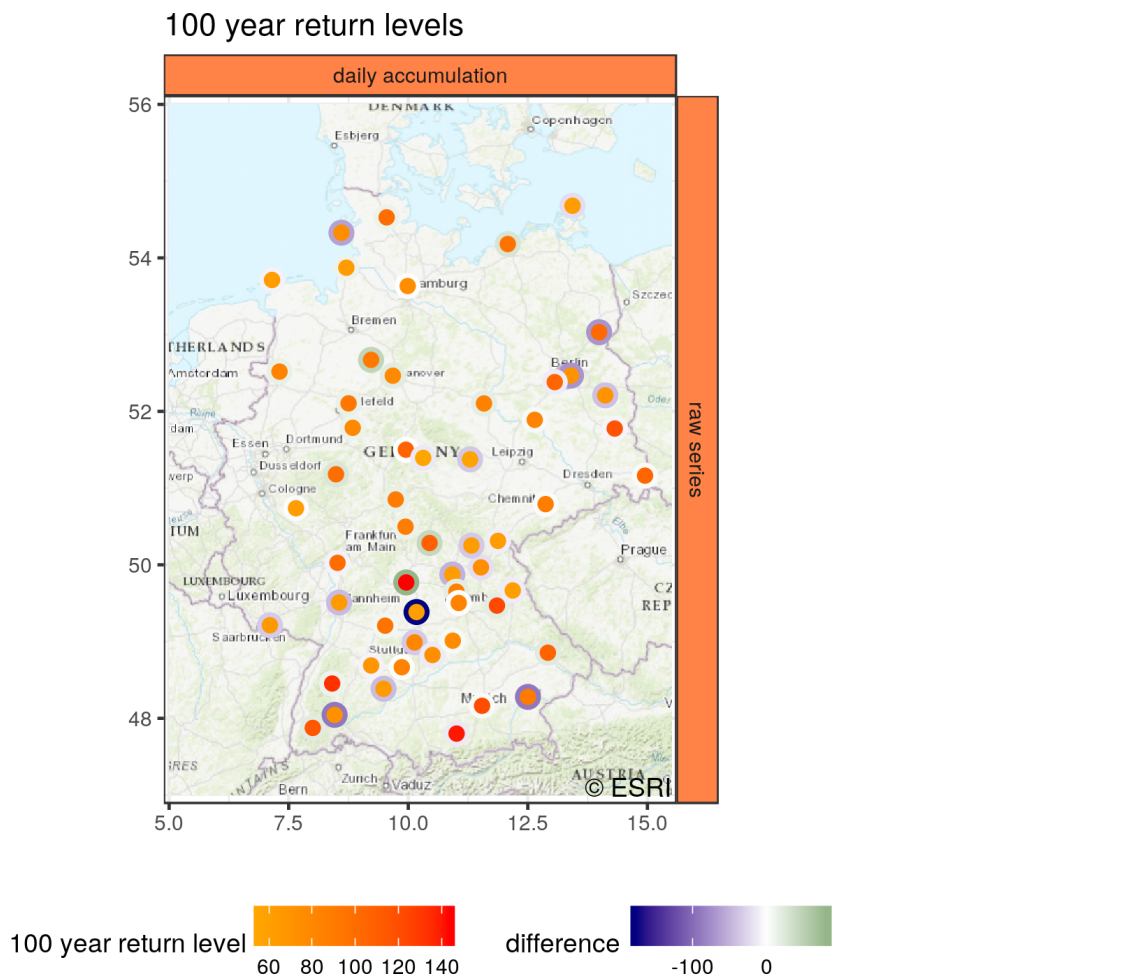


Figure A.58.: 100 year return levels of the daily accumulated precipitation throughout Germany. For a detailed description of the figure and colors please see section 7.1 and 7.2.

VGLM analysis for the daily accumulated precipitation

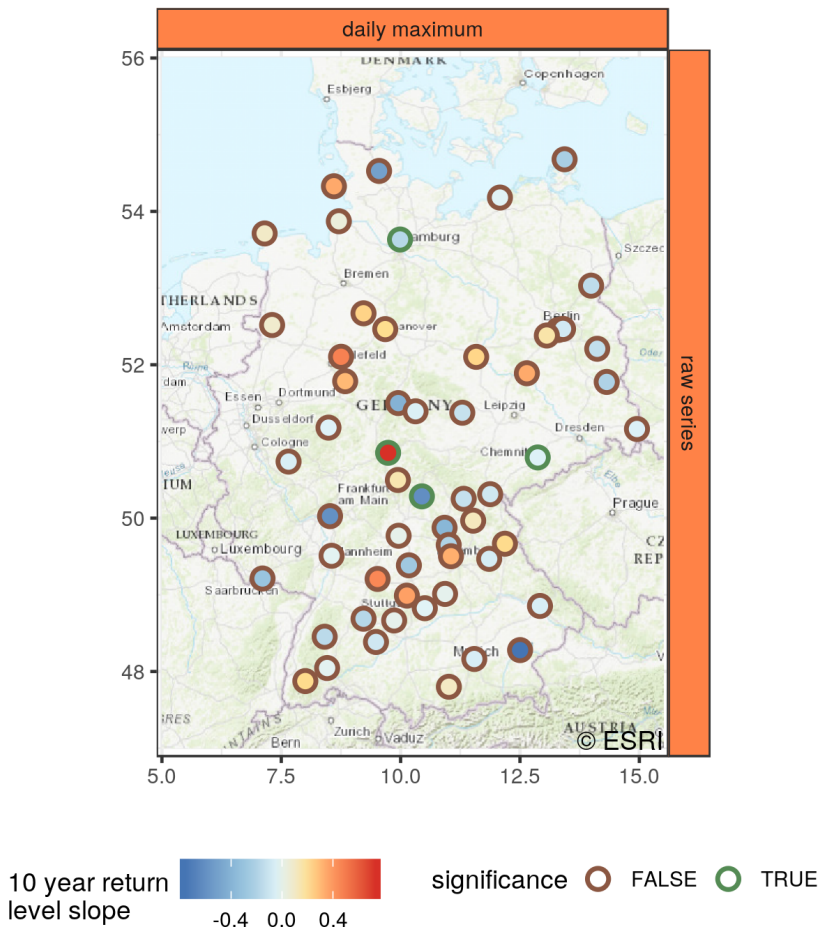


Figure A.59.: Results of the likelihood ratio test whether to use a stationary model over a first order linear one for the daily accumulated precipitation. All stations, for which the null hypothesis of a better description with a stationary model was rejected, are enclosed by a green color. The circles are filled with the slope of the 10 year return levels fitted using the VGLM. Since it has a non-linear dependence on time the slope was estimated by fitting a straight line in its curve. This value might therefore not yield a perfect approximation of the actual temporal evolution but it is sufficient to determine its overall tendency.

more data is required to statistically validate this result.

Abu-Mostafa, Yaser S., Malik Magdon-Ismael, and Hsuan-Tien Lin. 2012. *Learning from Data*. AMLBook.

Akaike, Hirotogu. 1973. "Information Theory and an Extension of the Maximum Likelihood Principle." 2nd International Symposium on Information Theory.

Ashkenazy, Yosef, Hezi Gildor, Martin Losch, Francis A Macdonald, Daniel P Schrag, and Eli Tziperman. 2013. "Dynamics of a Snowball Earth ocean." *Nature* 495 (7439). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 90–93.

Benzi, Roberto, Giorio Parisi, Alfonso Sutera, and Angelo Vulpiani. 1982. "Stochastic resonance in climatic change." *Tellus* 34 (1): 10–16.

Berrisford, Paul, Dick Dee, Paul Poli, Roger Brugge, Keith Fielding, Manuel Fuentes, Per Kållberg, Shinya Kobayashi, Sakari Uppala, and Adrian Simmons. 2011. "The Era-Interim Archive - Version 2.0." European Centre for Medium Range Weather Forecasts.

Box, George E. P., Gwilym M. Jenkins, and Gregory C. Reinsel. 2008. *Time Series Analysis: Forecasting and Control*. 4th ed. John Wiley & Sons.

Budyko, M. I. 1969. "The effect of solar radiation variations on the climate of the Earth." *Tellus* 21 (5): 611–19.

Castillo, Enrique, and Ali S. Hadi. 1994. "Parameter and quantile estimation for the generalized extreme-value distribution." *Environmetrics* 5 (4): 417–32.

Chavez-Demoulin, V., and A. C. Davison. 2005. "Generalized additive modelling of sample extremes." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54 (1): 207–22.

Christidis, Nikolaos, Gareth S. Jones, and Peter A. Stott. 2014. "Dramatically increasing chance of extremely hot summers since the 2003 European heatwave." *Nature Climate Change* advance on (December). Nature Publishing Group.

Cleveland, Robert B., William S. Cleveland, Jean E. McRae, and Irma Terpenning. 1990. "STL: A Seasonal-Trend Decomposition Procedure Based on Loess." *Journal of Official Statistics* 6 (1): 3–73.

Coles, Stuart G. 2004. *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Coles, Stuart G., and Mark J. Dixon. 1999. "Likelihood-Based Inference for Extreme Value Models." *Extremes* 2 (1). Kluwer Academic Publishers: 5–23.

Coles, Stuart G., and J. A. Tawn. 1996. "A Bayesian analysis of extreme rainfall data." *Applied Statistics* 45 (4). Wiley-Blackwell: 463–78.

Dansgaard, W., S. J. Johnsen, H. B. Clausen, D. Dahl-Jensen, N. S. Gundestrup, C. U. Hammer, C. S. Hvidberg, et al. 1993. "Evidence for general instability of past climate from a 250-kyr ice-core record." *Nature* 364 (6434): 218–20.

Davidon, William C. 1991. "Variable Metric Method for Minimization." *SIAM Journal of Optimization* 1 (1).

Davison, A. C., and Richard L. Smith. 1990. "Models for Exceedances over High Thresholds." *Journal of the Royal Statistical Society, B* 52: 393–442.

Dee, DP, SM Uppala, AJ Simmons, P Berrisford, P Poli, S Kobayashi, U Andrae, et al. 2011. "The Era-Interim Reanalysis: Configuration and Performance of the Data." *Quarterly Journal of the Royal Meteorological Society* 137: 553–97.

Deutscher Wetterdienst. 2018. <ftp://ftp-cdc.dwd.de/pub/CDC/>.

- Diebolt, Jean, Armelle Guillou, Philippe Naveau, and Pierre Ribereau. 2008. “Improving probability-weighted moment methods for the generalized extreme value distribution.” *Revstat - Statistical Journal* 6 (1).
- Edwards, Paul N. 2011. “History of climate modeling.” *Wiley Interdisciplinary Reviews: Climate Change* 2 (1): 128–39.
- Esper, Jan, Edward R Cook, and Fritz H. Schweingruber. 2002. “Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability.” *Science (New York, N.Y.)* 295 (5563): 2250–3.
- Falk, Michael, Jürg Hüsler, and Rolf-Dieter Reiss. 2011. *Laws of Small Numbers: Extremes and Rare Events*. 3rd ed. Birkhäuser.
- Fawcett, Lee, and David Walshaw. 2007. “Improved estimation for temporally clustered extremes.” *Environmetrics* 18 (2): 173–88.
- Ferro, Christopher A. T., and Johan Segers. 2003. “Inference for Clusters of Extreme Values.” *J. R. Statist. Soc. B*, no. 65: 545–56.
- Fisher, R. A., and L. H. C. Tippett. 1928. “Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample.” *Cambridge Philosophical Society* 24 (2): 180–90.
- Franzke, Christian L. E., Terence J. O’Kane, Judith Berner, Paul D. Williams, and Valerio Lucarini. 2015. “Stochastic climate theory and modeling.” *Wiley Interdisciplinary Reviews: Climate Change* 6 (1): 63–78.
- Frechét, Maurice. 1927. “Sur La Loi de Probabilité de L’écart Maximum.” *Annales de La Société Polonaise de Mathématique* 6: 93–116.
- Gilleland, Eric, and Richard W. Katz. 2014. “extRemes 2.0: an extreme value analysis package in R.” *Journal of Statistical Software*.
- . 2016. “extRemes 2.0: An Extreme Value Analysis Package in R.” *Journal of Statistical Software* 72 (8): 1–39.
- Gilleland, Eric, Mathieu Ribatet, and Alec G. Stephenson. 2013. “A software review for extreme value analysis.” *Extremes* 16 (1): 103–19.
- Gumbel, E.J. 1958. *Statistics of Extremes*. Columbia University Press.
- Haan, Laurens de, and Ana Ferreira. 2006. *Extreme Value Theory - an Introduction*. Springer.
- Hansen, James, Makiko Sato, and Reto Ruedy. 2012. “Perception of climate change.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (37): E2415–23.
- Harvey, A. C. 1978. “Linear Regression in the Frequency Domain.” *International Economic Review* 19 (2): 507–12.
- Hasselmann, K. 1976. “Stochastic climate models Part I. Theory.” *Tellus* 28 (6): 473–85.
- Hastie, Trevor J., and Robert Tibshirani. 1986. “Generalized Additive Models.” *Statistical Science* 1 (3): 297–318.
- Hastie, Trevor J., Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning*. 2nd ed. Springer.
- Hipel, K.W., and A. Ian McLeod. 1994. *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier.
- Hirabayashi, Yukiko, Roobavannan Mahendran, Sujjan Koirala, Lisako Konoshima, Dai Yamazaki, Satoshi Watanabe, Hyungjun Kim, and Shinjiro Kanae. 2013. “Global flood

risk under climate change.” *Nature Climate Change* 3 (9). Nature Research: 816–21.

Hosking, Jonathan R.M. 1990. “L-Moment - Analysis and Estimation of Distributions Using Linear-Combinations of Order-Statistics.” *Journal of the Royal Statistical Society. Series B: Methodological* 52 (1).

Hosking, Jonathan R. M., J. R. Wallis, and E. F. Wood. 1985. “Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments.” *Technometrics* 27 (3). Taylor & Francis Group: 251–61.

IPCC. 2014. “AR5 - Impacts, Adaptation and Vulnerability.”

Jennrich, R. I., and P.F. Sampson. 1976. “Newton-Raphson and Related Algorithms for Maximum Likelihood Variance Component Estimation.” *Technometrics* 18 (1): 11–17.

Jones, Philip D., Keith R. Briffa, T.P. Barnett, and S.F.B. Tett. 1998. “High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with General Circulation Model control-run temperatures.” *The Holocene* 8 (4): 455–71.

Kamae, Youichi, Hideo Shiogama, Masahiro Watanabe, and Masahide Kimoto. 2014. “Attributing the increase in Northern Hemisphere hot summers since the late 20th century.” *Geophysical Research Letters* 41 (14): 5192–9.

Kelley, C. T. 1999. *Iterative Methods for Optimization*. SIAM.

Laloyaux, Patrick, Magdalena Balmaseda, Dick Dee, Kristian Mogensen, and Peter Janssen. 2016. “A Coupled Data Assimilation System for Climate Reanalysis.” *Quarterly Journal of the Royal Meteorological Society* 142: 65–78.

Lange, Kenneth. 2013. *Optimization*. 2nd ed. Springer.

Leadbetter, M.R., and S. Nandagopalan. 1987. “On Exceedance Point Processes for Stationary Sequences Under Mild Oscillation Restrictions.” Edited by Rolf-Dieter Reiss and Jürg Hüsler. Springer.

Martins, Eduardo S., and Jerry R. Stedinger. 2000. “Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data.” *Water Resources Research* 36 (3): 737–44.

Massah, M., and H. Kantz. 2016. “Confidence intervals for time averages in the presence of long-range correlations, a case study on Earth surface temperature anomalies.” *Geophysical Research Letters*, no. 43.

McLeod, A. Ian, and Hyukjun Gweon. 2013. “Optimal Deseasonalization for Monthly and Daily Geophysical Time Series.” *Journal of Environmental Statistics* 4 (11).

Müller, Philipp, and Holger Kantz. 2018. “The climex package: Resolving flaws in the maximum likelihood fit of extreme value distribution.” *To Be Submitted*.

Nelder, J. A., and R. Mead. 1965. “A Simplex Method for Function Minimization.” *The Computer Journal* 7 (4): 308–13.

Nelder, J. A., and R. W. M. Wedderburn. 1972. “Generalized Linear Models.” *Journal of the Royal Statistical Society A* 135 (3): 370–84.

Nocedal, Jorge, and Stephen J. Wright. 2012. *Numerical Optimization*. Springer.

Pauli, F., and Stuart G. Coles. 2001. “Penalized Likelihood Inference in Extreme Value Analysis.” *Journal of Applied Statistics* 28: 547–60.

Poli, Paul, Hans Hersbach, David Tan, Dick Dee, Jean-Noël Thépaut, Adrian Simmons, Carole Peubey, et al. 2015. “ERA Report Series: The Data Assimilation System and Initial Performance Evaluation of the Ecmwf Pilot Reanalysis of the 20th-Century Assimilating Surfaces Observations Only (Era-20C).” European Centre for Medium Range Weather Forecasts.

- Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P Flannery. 2007. *Numerical Recipes*. Third. Cambridge University Press.
- Randall, D.A., R.A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, et al. 2007. "Climate Models and Their Evaluation. In: Climate Change 2007: The Physical Science Basis." In, edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller, 591–662. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom; New York, NY, USA.
- R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reiss, Rolf-Dieter., and Michael Thomas. 2007. *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. 3rd ed. Birkhäuser.
- Sarkar, A, S Singh, and D Mitra. 2011. "Wind climate modeling using Weibull and extreme value distribution." *International Journal of Engineering, Science and Technology* 3 (5). MultiCraft Limited.
- Schär, Christoph, and Gerd Jendritzky. 2004. "Climate change: hot news from summer 2003." *Nature* 432 (7017): 559–60.
- Schär, Christoph, Pier Luigi Vidale, Daniel Lüthi, Christoph Frei, Christian Häberli, Mark A Liniger, and Christof Appenzeller. 2004. "The role of increasing temperature variability in European summer heatwaves." *Nature* 427 (6972). Macmillian Magazines Ltd.: 332–6.
- Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." *The Annals of Statistics* 6 (2): 461–64.
- Smith, Richard L. 1985. "Maximum likelihood estimation in a class of nonregular cases." *Biometrika* 72 (1): 67–90.
- . 1986. "Extreme Value Theory Based on the R Largest Annual Events." *Journal of Hydrology* 86: 27–43.
- Steenbergen, Raphaël D.J.M., Tessa Koster, and Chris P.W. Geurts. 2012. "The effect of climate change and natural variability on wind loading values for buildings." *Building and Environment* 55: 178–86.
- Storch, Hans von, and Francis W. Zwiers. 1999. *Statistical Analysis in Climate Research*. Cambridge University Press.
- Sura, Philip. 2011. "A general perspective of extreme events in weather and climate." *Atmospheric Research* 101 (1-2): 1–21.
- Trenberth, Kevin E. 2011. "Changes in Precipitation with Climate Change." *Climate Research* 47: 123–38.
- Trenberth, Kevin E., John Fasulo, and Lesley Smith. 2005. "Trends and Variability in Column-Integrated Atmospheric Water Vapor." *Climate Dynamics* 24: 741–58.
- Unkašević, M., D. Vujović, and I. Tošić. 2005. "Trends in extreme summer temperatures at Belgrade." *Theoretical and Applied Climatology* 82 (3-4): 199–205.
- Valor, A., F. Caleyó, D. Rivas, and J.M. Hallen. 2010. "Stochastic approach to pitting-corrosion-extreme modelling in low-carbon steel." *Corrosion Science* 52 (3): 910–15.
- Wallace, John M., and Peter V. Hobbs. 2006. *Atmospheric Science: An Introductory Survey*.
- Weissman, I. 1978. "Estimation of Parameters and Large Quantiles Based on the K

Largest Observations.” *Journal of the American Statistical Association* 73 (364): 812–15.

Yee, Thomas W. 2015. *Vector Generalized Linear and Additive Models*. Springer.

Yee, Thomas W., and Trevor J. Hastie. 2003. “Reduced-rank vector generalized linear models.” *Statistical Modelling* 3 (1): 15–41.

Yee, Thomas W., and Alec G. Stephenson. 2007. “Vector Generalized Linear and Additive Extreme Value Models.” *Extremes*, no. 10: 1–19.

Yee, Thomas W., and C. J. Wild. 1996. “Vector Generalized Additive Models.” *Journal of the Royal Statistical Society. Series B (Methodological)* 58 (3): 481–93.

Yu, Zhou, Marc Hoell, Holger Kantz, and Philipp Müller. 2018. “In Preparation.”