An investigation of present and future climate extremes using climate model data

DISSERTATION

Submitted to the University of Bremen
in partial fulfillment of the requirements
for the academic degree of
doctor rerum naturalium (Dr. rer. nat.)
Faculty 01 – Physics / Electrical Engineering

Submitted by:	Justus Contzen
Reviewers:	Prof. Dr. rer. nat. Gerrit Lohmann
	Prof. Dr. rer. nat. Thorsten Dickhaus
	Prof. Dr. rer. nat. Jürgen Kurths
Date of Submission:	July 6, 2023
Date of Defense:	November 30, 2023

Abstract

This work deals with the investigation of climate extremes using data from global circulation models. These models are one of the main tools that are used by climate scientists to study and quantitatively assess the future effects of climate change. Climate extremes like droughts, heat waves or heavy rain can potentially have severe consequences on people and economies worldwide, and changes in climate extremes can cause these effects to become more severe in the future, so a quantitative investigation of changes in climate extremes is of high relevance.

After a general introduction to the climate system, to climate extremes and to the mathematical investigation of extremes we discuss in the first part of the work the validation of climate models in terms of extremes. When validating a climate model, data from a historical run of it are compared to observed data. Validation is often done with respect to mean values and covariances, validation with respect to extremes is less common. We present a method for model validation in terms of extreme values based on classical extreme value theory. We also compare the spatial dependencies of extremes between model and observational data by using a clustering algorithm. We apply these methods to data from the Alfred Wegener Institute Earth System Model and from other models of the Coupled Model Intercomparison Project CMIP6. We find that both the distribution of climate extremes at a fixed location and the spatial structure of extremes are in general well met by the climate models. However, a systematic overestimation or underestimation of climate extremes can be detected in some regions, in particular in mountainous areas.

The second part of the work investigates how climate models predict climate extremes to change in the future. To this end, we design statistical models to globally describe changes in extreme temperatures over time. This enables us to investigate how large the changes in the extremes will be and how quickly they will take place, depending on the location. It turns out that there is a tendency for stronger and quicker changes over land, while changes over the ocean are slower and more gradual. We investigate different statistical models and analyze their performance. We also find that in the Arctic region the changes in extremes are characterized by a rapid increase in variance followed by a slower change in mean values, which differs from the behavior in other regions. This can

be explained by the melting of sea taking place in the Arctic that influences the local climate.

In the third part of the work, we discuss some methods to investigate spatial dependencies of extremes in more detail. If an extreme event affects a large area simultaneously, it can have particularly severe consequences. Max-stable processes are one possible approach for the statistical investigation of climate extremes and their spatial dependencies. Most of the existing models, however, are only of limited applicability if the data investigated cover a large and heterogeneous area. We discuss a clustering approach that has recently been proposed to address this issue. We show that there are cases in which that approach is not well applicable, propose an alternative method and use computer-simulated data to evaluate the two approaches. We obtain that our proposed approach can lead to better results in some cases. The work on this algorithm is intended to be a first step towards applying similar methods also to climate model data.

Acknowledgments

First of all, I would like to thank my supervisors Prof. Dr. Thorsten Dickhaus and Prof. Dr. Gerrit Lohmann for their guidance and support throughout the PhD project and for the constructive and insightful discussions. I am grateful to Manfred Mudelsee for his helpful suggestions and to Enno Prigge and the MarDATA Graduate School for the assistance and for the opportunities provided in the PhD project. I would like to thank my friends and colleagues at the AWI and the University of Bremen for a lot of pleasant talks and activities. Finally, I would like to thank my parents Angela and Rüdiger for their continued and unwavering support also in difficult times.

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

a.s.	almost surely
AWI	Alfred Wegener Institute
AWQD	Average Weighted Quantile Difference
BIC	Bayesian Information Criterion
BM	Block Maxima
cdf	cumulative distribution function
CMIP	Coupled Model Intercomparison Project
EDC	Extremal Dependence Clustering
ESM	Earth System Model
ETCCDI	Expert Team on Climate Change Detection and Indices
GCM	Global Circulation Model
GEV	Generalized Extreme Value
GP	Generalized Pareto
iid	independent and identically distributed
IPCC	Intergovernmental Panel on Climate Change

KS	Kolmogorov-Smirnov
LEC	Local Estimates Clustering
MLE	Maximum Likelihood Estimation
pdf	probability density function
PoT	Peaks over Threshold
PWME	Probability-Weighted Moment Estimator
QQ	Quantile-Quantile
RCP	Representative Concentration Pathway
RMSE	Root Mean Squared Error
SSP	Shared Socioeconomic Pathway

Chapter 1

Introduction

1.1 Motivation

In Quechua, the language of the Inca, the word for ”in front of” also has the meaning ”in the past”, and the word for ”behind” can also mean ”in the future”. The idea behind this choice of words is a notion of going backwards into the future, into the unknown that is not visible, while keeping in sight experiences and events lying in the past.

Currently, global climate change causes mankind to face a future that is maybe more uncertain than ever, with possibly drastic changes awaiting us (IPCC, 2021). Using information about and experiences from the past, it is to some extent possible to reduce our uncertainty of what the future will bring, so that we do not move toward the future blindly, but use our knowledge in the best possible way to keep negative consequences under control or adapt to them.

To predict future developments of the Earth's climate, one of the main assets used by climate scientists are global circulation models (GCMs). These are large-scale computer programs based on the physical equations and properties governing the climate system, which are calibrated based on the known or estimated climatic conditions of historic and prehistoric past times (Stocker, 2013). Using different scenarios of how future greenhouse gas emissions will develop, climate models are used to create simulations of the future climate and study their predictions about the effects of climate change. Climate models necessarily have to simplify the immensely complex climate system of the Earth and some parts of the climate system are still not understood well, therefore the accuracy of the

simulations is inevitably limited (Parker, 2013).

One of the concerns associated with climate change is a possible increase in frequency and magnitude of climate extremes. Climate extremes like heat waves, droughts and heavy rain can have devastating effects on human societies, ecosystems and enonomy, and therefore a precise understanding of how they will develop in the future is crucial (Seneviratne et al., 2012). This understanding can be increased also by using climate models, keeping in mind the limitations in the simulation accuracy mentioned above.

1.2 Aims and scope of this thesis

In this dissertation, which is based on an interdisciplinary project combining mathematics and climate science, it will be investigated how well climate models are able to describe climate extremes, and what changes to climate extremes are predicted by climate models. In addition to that, a possible improvement of a method to analyze the spatial structure of climate extremes is presented. Given the seriousness and urgency of climate change and in particular of changes in climate extremes, research on these topics is of high importance. In particular, the following three research questions are addressed:

1. *Using historical runs of climate models, how well do the statistical properties of climate extremes in the simulated data match with those of observed climate extremes?*

Addressing this question sheds light on how well climate models are able to simulate extremes under current climatic conditions. Comparing historical runs of a climate model to observed data can be used to validate the climate model: The reliability of a model that is not able to simulate the current climate accurately is probably limited also when it comes to future simulations. Validation of climate models is often done by comparing mean values and covariances (Tapiador et al., 2012), and less attention is given to validation in terms of climate extremes. Our investigation puts focus specifically on validation in terms of extremes, and uses different mathematical approaches to investigate the accuracy of climate models in that regard. When comparing climate model and observed data, we will focus on the distribution of extremes at fixed locations and on the location of regions in which extremes occur simultaneously.

It needs to be noted that even if a climate model simulates climate extremes well under current conditions, it is not guaranteed that is does so also in future scenarios. Nevertheless, an investigation of the accuracy of a climate model under historic conditions will give some insight on how much confidence can be put in its predictions regarding the future development of climate extremes.

2. *How do climate models predict climate extremes to change in the future?*

As already stated, one of the main purposes of climate models is to conduct simulations to describe the future effects of climate change. Analyzing the output of future simulations with respect to climate extremes will help us obtain a better understanding of this aspect of climate change. While similar investigations have already been done regionally and for the next few decades (for example Panagoulia et al., 2014; Sarhadi and Soulis, 2017), the development of extremes in the more distant future and on a global level is still less well understood (Rummukainen, 2013; IPCC, 2021). To address this, we extend existing statistical models for time-dependent extremes to make them more flexible and to make them applicable specifically to changes in climate extremes under some of the scenarios of the future development of climate. We will apply these statistical models to data from different climate models that cover a time span of more than three centuries. We will apply different statistical models and discuss and compare their results.

Changes in climate extremes will be investigated in terms of their timing and of the magnitude of the changes, as both of these parameters are important for assessing the possible consequences of the changes. In addition to that, also the regional variability of the changes will be investigated, as it is well-known that changes in extremes can vary considerably depending on the region (Xie et al., 2015).

3. *How can statistical methods to describe the spatial structure of extremes be improved?*

Extreme climate events can occur simultaneously throughout large areas, or also in separate remote regions of the world, and this may cause particularly high damage. It is therefore of interest to investigate climate extremes not only at one fixed location, but also in a spatial context, and to understand how extremes at different locations influence each other.

There exist some statistical models that can be used for this task, but most of them require to assume a certain kind of spatial homogeneity of the data. That assumption is not reasonable if the data cover a large and heterogeneous area, restricting the applications of said models to smaller areas like Switzerland (Ribatet et al., 2015). One recently developed approach to overcome this issue is to divide the area of investigation into different smaller regions on which the statistical models can then be applied (Saunders et al., 2021). We will discuss this approach and demonstrate that there are cases in which it is of limited applicability because the resulting regions do not necessarily fulfill the desired requirements. We then propose an alternative approach and we conduct a simulation study which points out that it can lead to improved results. These results are a first step to conducting a spatial investigation of extremes also for data from global climate models.

These three questions are each the topic of a published or submitted article. Historical runs of climate models are investigated in "Variability and extremes: statistical validation of the Alfred Wegener Institute Earth System Model (AWI-ESM)" (Contzen et al., 2022). The projected future developments of climate extremes are discussed in "Long-term development of temperature extremes in a warming Earth" (Contzen et al., 2023). Methods to investigate the spatial structure of extremes are investigated in "Regionalization approaches for the spatial analysis of extremal dependence" (Contzen et al., in press).

The structure of this thesis will follow the research questions outlined above. In the next chapter, an introduction to climate models and to the Earth's climate system in general is given. It is followed by a chapter exploring the mathematical foundations of extreme value theory that will be used throughout the work. The three research questions will each be addressed in one chapter based on the corresponding article. Parts of the introduction and methods sections of the articles are incorporated in the introductory chapters of the thesis. In the final chapter of the thesis, the conclusions that have been drawn in the previous chapters are summarized and discussed.

1.3 Own contributions to this work

All three papers on which this work is based were written in collaboration with Thorsten Dickhaus and Gerrit Lohmann. For each of them, I wrote the manuscript and conducted the statistical analyses. Thorsten Dickhaus and Gerrit Lohmann supervised the work, reviewed the manuscripts and made corrections.

For the first paper, Thorsten Dickhaus and Gerrit Lohmann were responsible for the initial concept and the design of the study. I conducted the analyses and evaluated the results, and I also designed the new dissimilarity measure that was used for the clustering. Thorsten Dickhaus contributed to the statistical methodology used in the study, and Gerrit Lohmann helped with the interpretation of the results and the writing of the introduction and the conclusion.

For the second paper, the idea to use non-stationary GEV distributions to investigate future changes of extremes was developed jointly by the three of us. I suggested to use logistic functions to describe the changes in the parameters. Thorsten Dickhaus suggested using a simulation study as a proof of concept on the applicability of the models, and Gerrit Lohmann contributed to the analysis of the results and to the conclusions that were drawn from them.

The concept of the third paper was designed by myself, and I also designed the simulation study and the examples that were used to discuss the theoretical attributes of the clusterings. Thorsten Dickhaus helped with the statistical analysis and contributed much to the work by pointing out areas in which improvements were necessary in previous versions of the manuscript.

Chapter 2

The climate system and climate models

2.1 Climate and climate change

The weather conditions at a certain time and place, like temperature, wind speed or precipitation, are determined by a complex interplay of numerous factors. Due to the highly non-linear nature of the atmosphere and ocean dynamics and due to their chaotic behavior (a small change in an initial state can accumulate over time and lead to a highly different outcome), an accurate weather forecast is not feasible for more than two weeks in advance even if the initial state were known completely (Holton and Hakim, 2013). The investigation of the statistical properties of the weather at a certain location, however, is possible on time scales of years, decades and even longer. These statistical properties are subsumed under the term climate, and while the study of climate does not allow for the prediction of the exact weather at a certain time and place, information like the average weather conditions and their variability, as well as probabilities for extreme weather events still provide invaluable information for the agricultural sector, healthcare, urban planning and many other areas (Ruddiman, 2008).

Instrumental measurements for climate variables started in the second half of the 19th century with initially only a few observation stations in selected regions of the world. Over time, weather stations have become much more frequent and measurement methods have improved (Jones, 2001). Technological advances like satellites (starting in the 1970s) allow for a fine-gridded measurement of many climate variables, including

some which were previously inaccessible (like the temperature in the troposphere for instance). For the investigation of the climate further back in the past, it is necessary to rely on historical documents or on proxy data like tree rings or pollen records, giving a much less accurate picture than instrumental measurements (Helle and Schleser, 2004; Soon and Baliunas, 2003). The information that can be gained from those sources is usually more sparse and more inaccurate for time periods that are further in the past. One important method that provides rather detailed data about the climate in the last few hundred thousands of years is the investigation of ice cores from Greenland and Antarctica. The thickness and chemical composition of the yearly layers of the ice can be used to draw conclusions about past temperatures and other climatic conditions, and air enclosed in the ice provide information of the atmospheric composition (Delmas, 1992). However, this method is limited to regions near the poles. In other regions, analysis of sediments can be used to gain information about the long-term past (He et al., 2020).

The data collected from all those methods show that the climate is a permanent process of transition over time, with changes that can range from relatively small variations on an annual or decadal time-scale to large-scale long-term trends over the course of thousands or even millions of years. External factors like the strength of the Sun or the characteristics of the Earth's rotation around it are called "forcings" and provide the most fundamental influences on the climate system. As a prominent example, the start and end of the ice ages was heavily influenced by periodic changes in the Earth's orbit, known as Milanković cycles (Ruddiman, 2008). The climate system itself can be internally divided into the components atmosphere, hydrosphere (oceans, rivers and lakes), biosphere (vegetation), land surface and cryosphere (ice and snow), and all of these components influence themselves and each other in a complex set of interactions, parts of which are still insufficiently understood. Fig. 2.1 depicts some of these interaction between different parts of the climate system. The random and chaotic nature of some of these internal processes in the climate system result in a natural internal variability of the climate (Rohli and Vega, 2018).

In addition, human actions have been influencing Earth's climate, with the earliest detectable effects maybe as far back as $6,000$ years ago, when the cultivation of rice possibly led to an increase of methane in the atmosphere, preventing a natural cooling that would otherwise have occurred due to orbital changes (Ruddiman, 2008). In the recent 200 years, the industrialization and the large-scale emission of greenhouse gases like CO_2

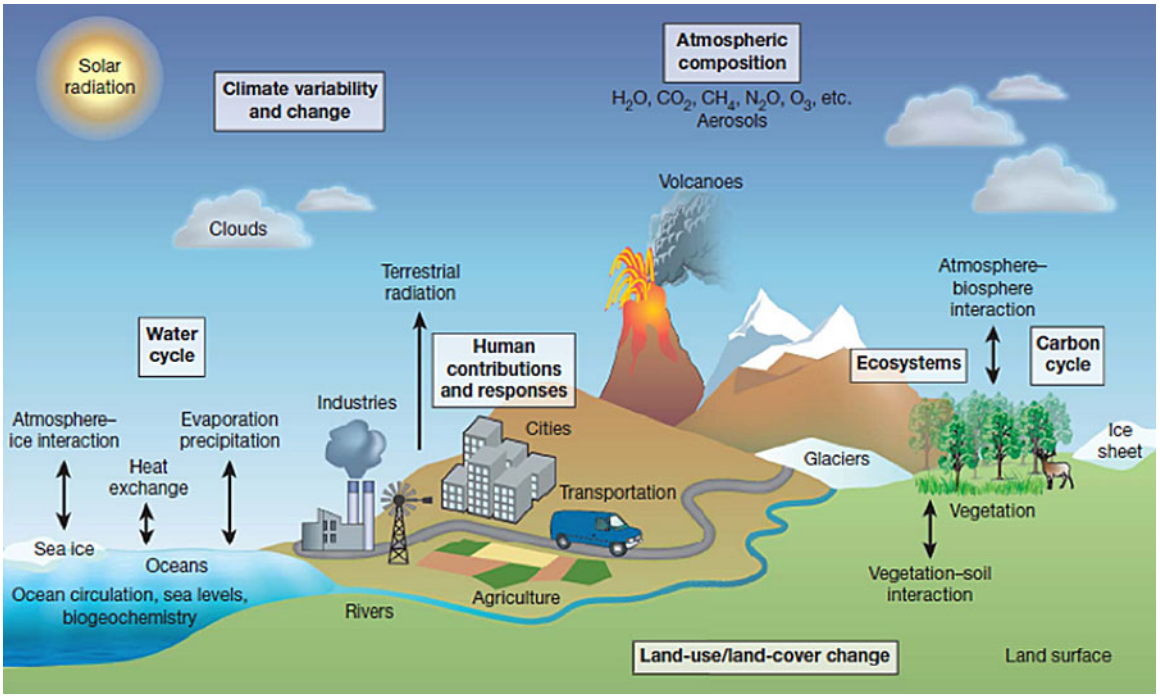

Figure 2.1: Simplified schematic of major components of the climate system, including human contributions, and their influences on each other (U.S. Climate Change Program, 2003, Fig. 2.5).

into the atmosphere have started to cause even stronger changes in the climate system (Seneviratne et al., 2012). When investigating the climate, the human (or anthropogenic) contributions are usually considered as an external forcing, as opposed to the natural trends and variability discussed above.

Earth is currently experiencing a warming of the land surfaces, the ocean and the atmosphere on a global scale, and there is a broad consensus that this is predominantly caused by humans (IPCC, 2012). However, the study of observational data to quantify these changes poses several challenges. First, the data coverage very much depends on the region investigated. In some regions, like sub-Saharan Africa, the amount of available data is limited because systematic measurement does not exist or has been started only recently (Dinku, 2019). But also long-term observations can exhibit missing values, or quality issues due to changed measurement methods over time. Long-term records may also be distorted at some measurement stations because of changes of the local environment, like the growth of cities (Koedel et al., 2022). All of these need to be taken into account when analyzing observational records. Furthermore, due to the short time span of observational records, it is often difficult to distinguish precisely between anthropogenic changes and those that are part of a natural variability on a decadal scale. This applies in particular to variables for which a broad data coverage relies on satellite measurements, for instance the strength of tropical cyclones (Sobel et al., 2016).

Regardless of their causes, changes in climate have always had a strong effect on human societies. In Zhang et al. (2007), it is shown that in historic times, periods of cooling were often followed by an increased instability of the society and the outbreak of wars. As an example, a cooling period in the fourth and fifth century AD, the Late Antique Little Ice Age, coincided with a period of extensive warfare in regions as far apart as Europe and Northern China (Büntgen et al., 2016; Yang et al., 2021), and a period of prolonged drought might have contributed to the collapse of the Maya civilization in the eighth to tenth century (Douglas et al., 2015). A review summary by Carleton and Hsiang (2016) collects recent empirical studies that find correlations between climate variables and various social factors like mortality, crime rate, or productivity all around the globe. For these reasons, it is evident that a large-scale global climate change as the one we are currently experiencing can have tremendous consequences on mankind.

2.2 Climate extremes

The Intergovernmental Panel on Climate Change (IPCC) defines a climate extreme as "the occurrence of a value of a weather or climate variable above (or below) a threshold value near the upper (or lower) ends of the range of observed values of the variable" (Seneviratne et al., 2012, page 116). Some possible types of extremes are heatwaves, droughts, heavy rainfall, storms, and hurricanes or cyclones. All of these can cause high damage to natural environments (Zscheischler et al., 2014), agriculture (Toreti et al., 2019), human health, societies and economies (Jongman et al., 2014; Ciais et al., 2005; Kovats and Kristie, 2006), which is the reason why possible future changes in climate extremes are particularly concerning. The damage of an extreme event depends not only on the strength of the event itself, but also on a variety of other factors, in particular the exposure and vulnerability of the area where it occurs (Cardona et al., 2012). Extremes that occur over a long time or over a large-scale area may have especially severe consequences, as the possibilities to mitigate their effects, for instance by providing humanitarian aid, are limited. The same applies if two different types of extremes occur at the same time and location, for example heat and drought (compound extremes). Extremes can also strengthen each other in a positive feedback loop (Cooney, 2012).

To quantify the strength of extremes, different measures, called extreme indices, exist. Commonly used are for example the maximum value of maximum daily temperature over a given time period or the number of consecutive summer days (days that exceed $25°$C) as indices for heat. The number of days exceeding a certain precipitation threshold (10mm or 20mm) is one of the possible indices for heavy rain (Karl et al., 1999). The Expert Team on Climate Change Detection and Indices (ETCCDI) has defined a list of 27 extreme indices that are often used for the investigation of extremes (Table 2.1).

There are numerous ways in which climate extremes can change over time: They can change in terms of their magnitude, their frequency, their duration, and their spatial distribution. A thorough investigation of the future changes in climate extremes is important to assess the risks and potential damages that can be caused by them. To explain how extremes can change in more detail, we take as an example those extremes that are defined as the events when a certain climate variable exceeds a fixed threshold value. The threshold value is defined as a certain percentile based on the (empirical) distribution of the variable in a reference period. Examples for this type of extremes include the ETCCDI indices TX10p or TN90p (see Table 2.1). Assume that the climate variable follows a normal

Table 2.1: The 27 core indices of the ETCCDI (ETCCDI, 2009). Abbreviations used: TN – daily minimum temperature. TX – daily maximum temperature. TG – daily average temperature. RR – daily amount of precipitation. Percentiles are based on the reference period $1961 - 1990$.

ID	Name
FD	Number of frost days ($\mathrm{TN} < 0°C$)
SU	Number of summer days ($\mathrm{TX} > 25°C$)
ID	Number of icing days ($\mathrm{TX} < 0°C$)
TR	Number of tropical nights ($\mathrm{TN} > 20°C$)
GSL	Growing season length (based on a consecutive period of days with $\mathrm{TG} > 5°C$)
TX_x	Monthly maximum value of TX
TN_x	Monthly maximum value of TN
TX_n	Monthly minimum value of TX
TN_n	Monthly minimum value of TN
TX10p	Percentage of days with $\mathrm{TX} < 10^{\mathrm{th}}$ percentile
TN90p	Percentage of days with $\mathrm{TN} > 90^{\mathrm{th}}$ percentile
TX90p	Percentage of days with $\mathrm{TX} > 90^{\mathrm{th}}$ percentile
WSDI	Warm spell duration index (consecutive days with $\mathrm{TX} > 90^{\mathrm{th}}$ percentile)
CSDI	Cold spell duration index (consecutive days with $\mathrm{TN} < 10^{\mathrm{th}}$ percentile)
DTR	Daily temperature range (Monthly mean difference between TX and TN)
Rx1day	Monthly maximum 1-day precipitation
Rx5day	Monthly maximum consecutive 5-day precipitation
SDII	Average amount of precipitation on wet days ($\mathrm{RR} \geq 1\mathrm{mm}$)
R10mm	Annual number of days with $\mathrm{RR} > 10\mathrm{mm}$
R20mm	Annual number of days with $\mathrm{RR} > 20\mathrm{mm}$
Rnnmm	Annual number of days with RR $>$ nn mm, nn a user-defined threshold
CDD	Maximum length of dry spell (consecutive days with $\mathrm{RR} < 1\mathrm{mm}$)
CWD	Maximum length of wet spell (consecutive days with $\mathrm{RR} \geq 1\mathrm{mm}$)
R95pTOT	Annual total precipitation on days when $\mathrm{RR} \geq 95^{\mathrm{th}}$ percentile
R99pTOT	Annual total precipitation on days when $\mathrm{RR} \geq 99^{\mathrm{th}}$ percentile
PRCPTOT	Annual total precipitation

distribution. If the mean of the distribution increases over time while the variance stays the same, the probability that the threshold value is exceeded will increase. The same

is true if the mean of the distribution stays the same but its variance increases. Both types of changes can also be combined, which can lead to a stronger effect. In Fig. 2.2, taken from IPCC (2001), these changes and their influence on extremes are illustrated. For non-normal distributions, another possible effect is that the heavy-tailedness of the distribution increases, meaning that "outliers" occur more frequently, which would also influence the frequency of extremes. This very simplified example already shows that changes in extremes are complex and in order to quantify them, an investigation of changes in the mean value of a climate variable is not enough.

Another important fact to keep in mind is the regionality of extremes. It was already mentioned that the impact of an extreme depends on the exposure and vulnerability of the affected region. In addition to that, the changes in extremes also vary spatially due to regional changes in the climate system. Climate change can lead to changes in wind patterns and in the strength and position of ocean currents, and these changed patterns can have an influence on extremes and can cause their changes to differ widely across regions (Xie et al., 2015).

2.3 Climate models

In order to improve the understanding of the climate system in general, to answer questions about the current climate change and to draw predictions on how that change will develop in the future, climate models are an important tool. In the following introduction we will follow Stocker (2013). Climate models aim to simulate the Earth's climate by performing calculations based on the known physical principles and equations that govern the climate system. These equations include basic physical conservation laws like the conservation of mass and of energy as well as for example the Navier-Stokes equations that describe the behavior of dynamic fluids. Climate models exist for different parts of the climate system, like the ocean, the atmosphere, the vegetation and the land ice. They can also be applied to different geographical areas: Some models investigate the climate globally, others focus on a specific region. All models have in common that the divide the area of investigation into a (two- or three-dimensional) grid and calculate averaged values on each grid cell. The physical equations used include partial differential equations, which implies that climate models also divide the time frame of investigation into different sections. Starting with an initial configuration, they calculate the values of the climate variables step by step using numerical methods like the Euler scheme. The most comprehensive models are GCMs that

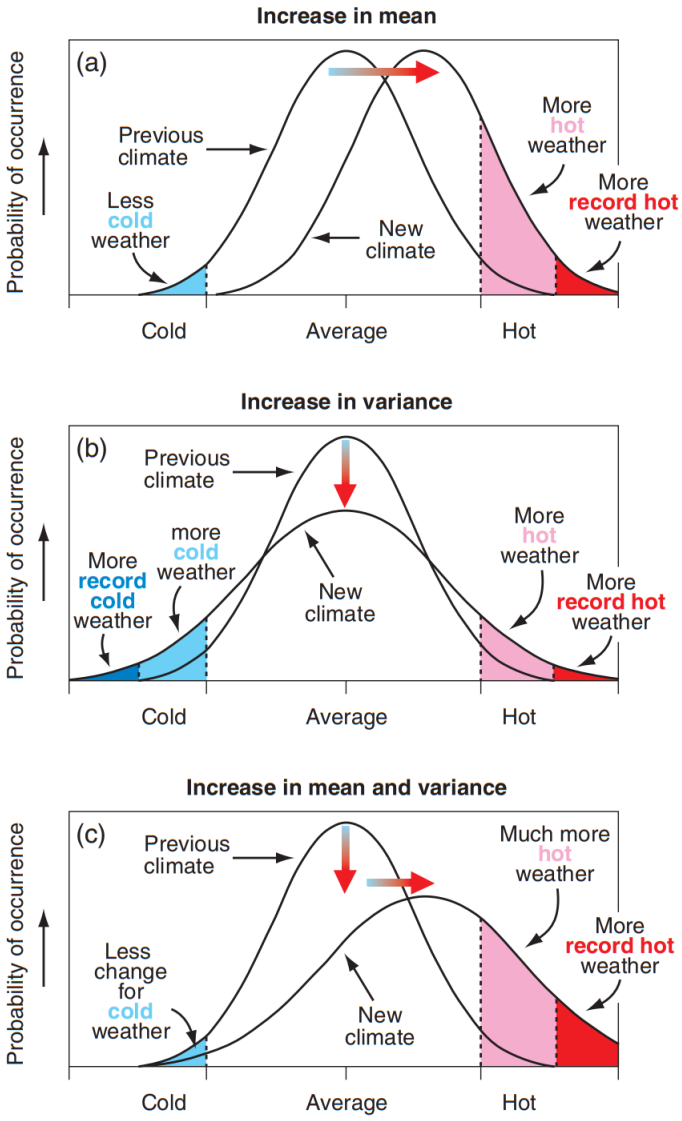

Figure 2.2: Schematic depiction of how changes in mean and in variance can influence the probability of temperature extremes (IPCC, 2001, Figure 2.32).

are applied on a global scale and that consist of different sub-models for the atmosphere, the ocean and for other parts of the climate system. As all the different parts of the climate system are interconnected, the sub-models are intertwined in a process called coupling: A time step is calculated for one of the sub-models, and the resulting values are used as boundary conditions for other ones. This way, the different model parts constantly and successively influence each other.

For the run of a climate model an initial configuration has to be provided, which is based on our knowledge of how the Earth's climate is or how it was at the time for which the model run is done for. In addition to the initial values, the temporal evolution of the external forcings that influence the climate system has to be provided to the model. This includes time series of solar activity and other orbital parameters, data about volcanic activity as well as human-induced greenhouse gas emissions. If the model is used to simulate the climate of the last decades, known data for these forcings can be used, in all other cases the forcing data is based on estimated or predicted values.

Both the spatial division into grid cells and the temporal division into time slices necessarily lead to unavoidable inaccuracies in the calculations and when choosing a grid size, a trade-off has to be made between a higher accuracy and faster computation times. Recent advances in computing have made it possible to steadily increase grid sizes and improve estimation accuracy, but there are still processes in the climate system that happen on scales that are smaller than the grid sizes currently used and that cannot be captured by current climate models. For those processes, simplifications are used that calculate plausible values, sometimes on a stochastic basis. The functions to calculate those values depend on unknown parameters and to determine suitable values for them, a process called "tuning" is applied: The models are run with different sets of parameters and the model results are then evaluated. For models that are used to simulate future climate change, the evaluation is done by using historical runs of the models and by comparing them to data derived from observations. Model runs describing prehistoric conditions like the climate of the Last Glacial Maximum (ca. 21000 years ago) can also be included in the evaluation process, increasing the general applicability of the climate models. The use of tuning necessarily adds uncertainty to the model results and it can not be taken for granted that a tuned parameter value based on historic runs of the models is also valid for a changed climate in the future.

The tuning, the use of averaged values for grid cells and time slices and the fact that parts of the climate system are still insufficiently understood are one large source of uncertainty for climate models (model uncertainty). Another source of uncertainty is the internal variability of the climate system and its random nature. This source of uncertainty cannot be removed, while model uncertainty might decrease as climate models become better in the future. When it comes to future simulations, a third source of uncertainty is that the development of the forcings, especially the anthropogenic ones, is hard to predict. This scenario uncertainty becomes stronger if the predictions of the climate models reach further into the future (Zhang and Chen, 2021).

To assess the accuracy of climate models, it is essential to validate them by comparing the results of historical runs with data from observations. Usual methods for validating climate models include the evaluation of mean values and covariances and the comparison of empirical cumulative distribution functions. These analyses can also be conducted over seasonal and annual averages (climatologies) or along latitudinal/longitudinal transects (Tapiador et al., 2012). When validation is done regarding climate extremes, it is also common to compare the ETCCDI extreme indices (Sillmann et al., 2013; Zhang et al., 2011).

To make the different climate models better comparable, many models participate in the Coupled Model Intercomparison Project (CMIP), coordinated by the Working Group on Coupled Modelling of the World Climate Research Programme. Models participating in CMIP follow common standards regarding the formatting of the model output, and the results are stored in a central location. Additionally, each model runs a set of standardized experimental setups (Diagnostic, Evaluation and Characterization of Klima experiments) as well as a simulation of the historical climate starting from 1850. The current, sixth phase of CMIP has started in 2016. For future simulations, different scenarios of the projected greenhouse gas emissions exist, ranging from a quick reduction of emissions in the next decades to steadily increasing emissions throughout the twenty-first century. These Representative Concentration Pathways (RCP) are used for future runs of the CMIP5 models, in CMIP6 they have been replaced by the Shared Socioeconomic Pathways (SSP), which also take into account predictions for global societal and economic developments (Riahi et al., 2017).

2.4 Observed and predicted changes in extremes

There have been numerous recent examples of devastating climate extreme events, like a heat wave over Siberia in 2020 and one over the Pacific Northwest (USA, Canada) in 2021 causing hundreds of casualties and extensive forest fires (Overland, 2021), or severe floods caused by heavy rainfall in Western Europe in 2021 that killed more than 200 people and caused damage of almost 50 billion Euro (Mohr et al., 2023). These events are not just isolated incidents, as it has been observed that the frequency of high-temperature extremes has increased in many regions of the world since the 1950s, while cold extremes have become less common (Dong et al., 2017). Precipitation extremes show a more mixed picture, with an increase in precipitation extremes in most locations, but to a varying degree, and a decrease in other places (Fischer and Knutti, 2016). Regarding changes in other types of extremes, like the frequency of tropical hurricanes, there is still a high uncertainty, partially due to a lack of data and of consistent measurement methods (Seneviratne et al., 2012). As changes in extremes are currently observable and can be expected to continue in the future, researchers have been increasingly focusing on that topic in recent years (IPCC, 2012; Rahmstorf and Coumou, 2011; Horton et al., 2016).

To investigate how changes in extremes are described by climate models, Kim et al. (2020) have analyzed historical runs of 32 CMIP5 and 30 CMIP6 models globally with respect to the 27 extreme indices by the ETCCDI. They found that the CMIP5 models simulate the extreme temperature indices reasonably well in most regions, with warm biases in South America and mid-latitude Asia and with cold biases in the high latitudes. The warm biases are reduced in the CMIP6 models while the cold ones still remain in them. For precipitation indices, the CMIP6 models show on average results that are better aligned to the observational data sets than the CMIP5 models, but there is a large inter-model spread. This spread, in combination with a tendency to underestimate precipitation extremes, had also been identified for CMIP5 models in a study investigating their performance over the United States (Wuebbles et al., 2014). Zhu et al. (2020) have compared multiple CMIP5 and CMIP6 models with a gridded observational data set, focusing on the simulation of climate extremes in China. They found in general a good agreement of the temperature indices for annual minima, averages and maxima, and a slightly worse agreement with regard to the number of cold nights and warm days.

King et al. (2021) show that a change in extremes caused by anthropogenic influence is detectable already in historical runs of climate models in many locations, especially

for temperature extremes and in the tropics. Fischer et al. (2021) infer that according to climate models, the probability of record-shattering events is expected to increase markedly in the next decades. According to their analyses, heat waves spanning several weeks can be up to seven times more likely in the end of the twenty-first century than they are now. Donat et al. (2020) undertook a detailed evaluation of several historical data sets and climate model runs with respect to extremes, and confirmed that a general warming trend is apparent both in the observational records and in the climate model runs. They point out that climate models fail to reproduce the spatial heterogeneity visible in the observed data, and that it's still unclear what the causes for this mismatch are. For example, a region in the Mid West of the U.S. exhibits a decrease of temperature extremes, in opposition to most other parts of the worlds, but this effect is not reproduced by most climate models. Models also tend to underestimate the decrease of cold extremes. When it comes to precipitation extremes—for which the changes are more spatially heterogeneous than for temperature extremes—climate models and observational data generally agree on the sign of the changes in regions with a good data coverage, but the magnitude of changes is more debated, and the model vary widely in their assessments.

Regarding future changes, CMIP5 models show an increase in extreme temperature on a global level that is approximately in a linear relationship with greenhouse gas emissions (Donat et al., 2020; IPCC, 2012). Pall et al. (2007) explain that the increase in extreme precipitation follows very roughly the increase of water-holding capacity in the atmosphere, which depends exponentially on the air temperature as specified by the Clausius-Clayperon equation and increases by approximately 7% per degree warming. However, they point out that in many regions, changes of the regional wind and ocean current patterns are expected, and those will likely cause changes to be different than the Clausius-Clayperon equation would suggest in some regions. It is also worth nothing that Iles et al. (2020) show that the use of a finer model resolution leads to an improved regional simulation accuracy with respect to extremes.

Chapter 3

Extreme value theory

3.1 Univariate extreme value theory

3.1.1 Introduction to the topic

All definitions of extremes have in common that the events in question are typically rare, which makes statistical investigations difficult. Classical univariate extreme value theory focuses mainly on two approaches to overcome this issue: Block Maxima (BM) and Peaks over Threshold (PoT). The two approaches are based on different definitions of extremes: In the BM approach, maxima are taken over consecutive blocks into which the data have been grouped. In the PoT approach, a fixed threshold is set and only those data values that exceed it are considered. This makes the BM approach applicable for example to investigate ETCCDI extreme indices like Rx1day, the monthly or annual maximum of daily precipitation. The PoT approach can be used for indices like FD, SU, or ID (number of frost/summer/icing days), which are all defined as the number of days for which daily maximum or minimum temperatures are higher or lower than a fixed temperature threshold (see Table 2.1). Note that it is possible to use the PoT approach to investigate extremes *below* a threshold by multiplying the data values with -1. The two methods will now be discussed in more detail.

3.1.2 The Block Maxima approach

The BM approach relies on the fact that under mild regularity conditions, the distribution of the block-wise maxima of the data can be approximated by a Generalized Extreme Value (GEV) distribution. Parametric inference can then be used to estimate the distribution pa-

rameters and draw conclusions about the behavior of the extremes. This approach is often used in climatology and hydrology, examples include Coles et al. (2003); Onwuegbuche et al. (2019); Villarini et al. (2011).

We will now discuss the mathematical theory behind the approach, following McNeil et al. (2015) and de Haan and Ferreira (2006). Let $n \in \mathbb{N}$ and $X_1, \dots, X_n$ be independent and identically distributed (iid) real-valued random variables. Let $M_{(n)} = \max_{i=1}^{n} X_i$. The goal is to investigate the limiting distribution of $M_{(n)}$ as $n \to \infty$. If the X_i have cumulative distribution function (cdf) G, then the cdf of $M_{(n)}$ is G^n, which tends to 0 for all x with $G(x) < 1$. This will make the limiting distribution of $M_{(n)}$ often degenerate, i.e. the random variable is almost surely equal to some fixed value. For example, if $X_i \sim U[0,1]$, then $M_n(n)$ will tend to a random variable which is almost surely equal to 1. To avoid degenerate limits, we introduce normalizing sequences $a_n \in \mathbb{R}$, $b_n > 0$ and assume that

$$\frac{\max(X_1, \dots, X_n) - a_n}{b_n} \xrightarrow[n\to\infty]{\mathrm{d}} Y \tag{3.1}$$

for a non-degenerate random variable Y (the notations $\xrightarrow{d}$ and $\overset{d}{=}$ denote convergence and equality in distribution, respectively). The following convergence of types theorem will give us insights how the choice of the normalizing sequences can affect the limiting distribution:

Theorem 3.1.1. *Let $(Y_n)_{n\in\mathbb{N}}$ be random variables and $a_n \in \mathbb{R}$, $b_n > 0$ and $\tilde{a}_n \in \mathbb{R}$, $\tilde{b}_n > 0$ normalizing sequences. If*

$$\frac{Y_n - a_n}{b_n} \xrightarrow[n\to\infty]{\mathrm{d}} Y \quad , \quad \frac{Y_n - \tilde{a}_n}{\tilde{b}_n} \xrightarrow[n\to\infty]{\mathrm{d}} \tilde{Y} \tag{3.2}$$

and Y, $\tilde{Y}$ not degenerate, then the limits $a = \lim_{n\to\infty} \frac{\tilde{a}_n - a_n}{b_n}$ and $b = \lim_{n\to\infty} \frac{\tilde{b}_n}{b_n}$ exist and

$$\tilde{Y} \overset{\mathrm{d}}{=} \frac{Y - a}{b}. \tag{3.3}$$

This implies that the limiting distribution Y of Eq. (3.1) is unique up to affine transformations. In de Haan and Ferreira (2006) it is further shown that the limiting distribution must be max-stable:

Definition 3.1.1. A random variable Y is *max-stable* if it is non-degenerate and if for all

$n \in \mathbb{N}$ there are $a_n \in \mathbb{R}$, $b_n > 0$:

$$\frac{\max(Y_1, \ldots, Y_n) - a_n}{b_n} \overset{\mathrm{d}}{=} Y \tag{3.4}$$

for independent copies $Y_1, \ldots, Y_n$ of Y.

Theorem 3.1.2. *Let $(X_i)_{i\in\mathbb{N}}$ be iid random variables and $a_n \in \mathbb{R}$, $b_n > 0$, $n \in \mathbb{N}$ normalizing sequences and assume that*

$$\frac{\max(X_1, \ldots, X_n) - a_n}{b_n} \xrightarrow[n\to\infty]{\mathrm{d}} Y \tag{3.5}$$

for some non-degenerate random variable Y. Then, Y must be max-stable.

As shown by Fisher, Tippett and Gnedenko (Fisher and Tippett, 1928; Gnedenko, 1943), max-stable random variables must belong to the parametric family of GEV distributions:

Theorem 3.1.3. *A max-stable random variable must follow a GEV distribution. Its cdf is of the form*

$$F_{\mu,\sigma,\gamma}(x) = \begin{cases} \exp(-\exp(-\frac{x-\mu}{\sigma})) & \gamma = 0 \\ \exp(-(1+\gamma\cdot\frac{x-\mu}{\sigma})^{-\frac{1}{\gamma}}) & \gamma \neq 0, \gamma\frac{x-\mu}{\sigma} > -1 \\ 0 & \gamma > 0, \gamma\frac{x-\mu}{\sigma} \leq -1 \\ 1 & \gamma < 0, \gamma\frac{x-\mu}{\sigma} \leq -1 \end{cases} \tag{3.6}$$

with parameters location (μ), scale ($\sigma > 0$) and shape (γ).

The location and scale parameters of the GEV distribution can roughly be compared to mean and standard deviation of the normal distribution, in the sense that if $Y \sim \mathrm{GEV}(\mu, \sigma, \gamma)$ then $Y + a \sim \mathrm{GEV}(\mu + a, \sigma, \gamma)$ and $bY \sim \mathrm{GEV}(b\mu, b\sigma, \gamma)$. The shape parameter controls how heavy-tailed the distribution is: If $\gamma < 0$, the distribution is a reversed Weibull distribution and has a finite right endpoint at $-1/\gamma$. If $\gamma = 0$, the distribution is a Gumbel distribution and it has exponential tails (like for example also the normal distributions). If $\gamma > 0$, the distribution is a Fréchet distribution with heavy tails. In the latter case, strong positive outliers have a higher chance to occur than for distributions with exponential tails. Note that there is some inconsistency in the literature: In some works, an alternate parametrization of the GEV distribution is used, the shape parameter of that definition is the negative of the shape parameter used here.

If random variables $(X_i)_{i\in\mathbb{N}}$ fulfill Eq. (3.5) for some sequences of normalizing constants, then, according to the convergence of types theorem, the shape parameter γ of the limiting GEV distribution is the same for all sequences of normalizing constants for which convergence occurs. It is said that their distribution is in the *max-domain of attraction* of H_γ, with $H_\gamma = \mathrm{GEV}(0, 1, \gamma)$, and the normalizing sequences can always be chosen such that the normalized maxima converge to a H_γ distributed variable.

When using GEV distributions in practical applications, a fixed, preferably rather large block size n is selected, the data are grouped into blocks of size n and the block-wise maxima are taken. The block-wise maxima are assumed to be GEV distributed. When selecting a block size, a bias-variance tradeoff has to be taken into account: For a low block size, the resulting parameter estimates tend to be biased because the convergence to the GEV distribution holds only asymptotically. A high block size, on the other hand, will lead to a limited amount of block-wise maxima that can be analyzed and therefore to a higher variance in the estimates.

3.1.3 The Peaks over Threshold approach

This work focuses mostly on the BM approach, and that is also the approach that will be used in the following chapters. However, the basic principles of the PoT approach shall be outlined here briefly. The PoT approach investigates the distribution of those parts of the data that exceed a certain fixed threshold. For a random variable X define the conditional excess distribution function F_u as

$$F_u(x) = \Pr(X \leq u + x | X > u) = \frac{F(u + x) - F(u)}{1 - F(u)} \tag{3.7}$$

for a fixed threshold u that is below the right endpoint x_F of X (which can be infinite). The Pickands-Balkema-De Haan theorem states the following:

Theorem 3.1.4. *Let $X_1, \ldots, X_n$ be independent copies of X. If X is in the max-domain of attraction of some H_γ, then it holds*

$$\lim_{u \uparrow x_F} \sup_{0<x<x_F-u} |F_u(x) - G_{\mu,\sigma,\gamma}(x)| = 0. \tag{3.8}$$

$G_{\mu,\sigma,\gamma}(x)$ is the pdf of the generalized Pareto (GP) distribution, which fulfills

$$G_{\mu,\sigma,\gamma}(x) = \begin{cases} 1 - (1 + \gamma\frac{x-\mu}{\sigma})^{-1/\gamma} & \gamma \neq 0 \\ 1 - \exp(-\frac{x-\mu}{\sigma}) & \gamma = 0 \end{cases} \tag{3.9}$$

for parameters μ, $\sigma > 0$, γ.

For thresholds that are high enough, this theorem makes it possible to approximate the distribution of the data that exceed the threshold with a GP distribution.

Like for the BM approach, there exists a bias-variance trade-off for the PoT approach: Choosing a threshold that is too high will lead to few data points exceeding the threshold and to a high variance of the estimators, while a threshold that is too low will result in a bias.

3.1.4 Parameter estimation

To estimate the parameter of a GEV distribution from a data sample, the classical maximum-likelihood estimator (MLE) can be used. We write the parameter tuples as $\theta = (\mu, \sigma, \gamma)$, which are elements of $\Theta = \{(\mu, \sigma, \gamma) \in \mathbb{R}^3 | \sigma > 0\}$, and we denote the density of the corresponding GEV distribution by f_θ. Assuming iid GEV distributed random variables $Y_1, \ldots, Y_n$, the MLE is the estimator that maximizes the product of the densities

$$\hat{\theta} = \text{argmax}_{\theta\in\Theta} \prod_{i=1}^{n} f_\theta(Y_i) \tag{3.10}$$

or equivalently the sum of their logarithms

$$\hat{\theta} = \text{argmax}_{\theta\in\Theta} \sum_{i=1}^{n} \log f_\theta(Y_i). \tag{3.11}$$

Plugging in sample values $y_1, \ldots, y_n$ allows to calculate the maximum-likelihood estimate of the GEV parameters for the sample. It is a well-known result that maximum-likelihood estimators are strongly consistent and asymptotically normal under certain regularity conditions (Harald, 1946). For GEV distributions, these conditions are not fulfilled as they include the requirement that all densities have the same support, whereas for GEV distributions the support depends on the parameter values. It was not until 2017 that Bücher and Segers (2017) showed that strong consistency and asymptotic normality

hold for the MLE for the GEV distribution for the case $\gamma < -\frac{1}{2}$. The other cases are not investigated in detail in Bücher and Segers (2017), but the authors mention earlier works that indicate that asymptotic normality does not hold for $\gamma > \frac{1}{2}$.

An alternative approach is using probability-weighted moment estimators (PWME). For a random variable Y with pdf F, it is possible to define probability-weighted moments $\beta_r := \mathbb{E}[YF(Y)^r]$. In Hosking et al. (1985), it is shown that for the GEV distribution the equality

$$\beta_r = (r+1)^{-1}(\mu - \sigma(1-(r+1)^{\gamma}\Gamma(1-\gamma))/\gamma) \tag{3.12}$$

holds for $\gamma \neq 0, \gamma < 1$. The equations for β_0, β_1 and β_2 can then be solved to obtain formulas for the three GEV parameters, and estimators for the probability-weighted moments can be calculated from a data sample $y_1, \dots, y_n$ by first sorting the sample values in increasing order $y^{(1)} \leq \cdots \leq y^{(n)}$ and then using

$$\hat{\beta}_r = n^{-1} \sum_{i=1}^{n} \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} y^{(i)} \tag{3.13}$$

as an unbiased estimator for β_r. According to Hosking et al. (1985), probability-weighted moment estimators have "low variance and no severe bias, and they compare favorably with estimators obtained by the methods of maximum likelihood".

3.1.5 Seasonality and trend

The convergence of block maxima to a GEV distribution required the underlying data to be iid. For climate data, this assumption is often not plausible, as climate variables can exhibit both seasonal variations and long-term trends. When the blocks are chosen in such a way that the boundaries between blocks are aligned with the seasonal pattern (for example when using annual maxima), it is often still possible to apply GEV distributions. For example, when grouping daily temperature data from the northern hemisphere into annual blocks, the maxima will predominantly be attained in the summer and the winter data could be discarded without changing the results. It can effectively be assumed that the data are iid with a lower block size. An alternative approach to deal with seasonality is to investigate block maxima over different seasons, for example to investigate the seasonal maxima of daily winter (spring, summer, fall) data separately.

To deal with trends in the data, it can be assumed that the trends have no effect on the distribution except for shifting it over time. Filtering approaches can then be used to separate the trend from the data and fit GEV distributions to the detrended data.

If it is desired to explicitly investigate how extremes change over time, it is possible to employ GEV distributions with parameters changing over time. Location, scale and shape of the GEV distribution are then expressed as functions of the time that are parametrized with a parameter vector θ:

$$Y(t) \sim \mathrm{GEV}(\mu_\theta(t), \sigma_\theta(t), \gamma_\theta(t)). \tag{3.14}$$

If a data sample $(t_1, y_1), \ldots, (t_n, y_n)$ is given, maximum-likelihood can be applied to obtain an estimator for θ.

3.2 Multivariate extreme value theory

3.2.1 Max-stable vectors

In the investigation of extremes it is often of interest to also take into account spatial dependency, as extremes that occur at two different locations at the same time can cause particularly high damage. To this end, the considerations of the previous section can be extended to multidimensional random variables. Using component-wise operations, we can define max-stable vectors analogously to the univariate case:

Definition 3.2.1. A random vector Y is *max-stable* if its margins are non-degenerate and there are sequences of vectors $c_n, d_n > 0$ such that:

$$\frac{\max(Y_1, \ldots, Y_n) - c_n}{d_n} \overset{\mathrm{d}}{=} Y \tag{3.15}$$

for independent copies $Y_1, \ldots, Y_n$ of Y.

As in the univariate case, it holds:

Theorem 3.2.1. *Let $(X_i)_{i\in\mathbb{N}}$ be iid random vectors and $c_n, d_n > 0$ sequences of normalizing vectors and assume that*

$$\frac{\max(X_1, \ldots, X_n) - c_n}{d_n} \xrightarrow[n\to\infty]{\mathrm{d}} Y \tag{3.16}$$

for some random vector Y with non-degenerate marginal distributions. Then, Y must be max-stable.

The univariate marginal distributions of a max-stable vector Y must necessarily be GEV distributions. It is common practice to assume that all univariate margins follow the same distribution after a suitable marginal transformation. This follows an idea from copula theory, where the margins of a multivariate distribution are transformed to be uniformly distributed, making it easier to focus on the dependency structure of the distribution. The approach is based on Sklar's theorem (Sklar, 1959):

Theorem 3.2.2. *Let F be a joint d-dimensional probability distribution function with margins $F_1, \ldots, F_n$. Then there exists a copula (i.e. a d-dimensional probability distribution function with $U[0,1]$ distributed margins) C such that*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{3.17}$$

If $F_1, \ldots, F_n$ are absolutely continuous, then C is unique.

In the context of max-stable vectors the margins are usually assumed to have a $\mathrm{GEV}(0, 1, 1)$ distribution (also called a *unit Fréchet distribution*) instead of a uniform distribution. A max-stable vector with unit Fréchet margins is called a *simple max-stable vector*. The distribution of a simple max-stable vector can be described as follows (de Haan and Ferreira, 2006):

Theorem 3.2.3. *Let $Y = (Y^{(1)}, \ldots, Y^{(n)})$ be a simple max-stable vector. Then it holds for all $y_1, \ldots, y_n \in \mathbb{R}$ that*

$$\Pr(Y^{(1)} \leq y_1, \ldots, Y^{(n)} \leq y_n) = \exp(-V(y_1, \ldots, y_n)) \tag{3.18}$$

for a function V that is positive and that is homogeneous of degree -1 (i. e. it fulfills $V(ax_1, \ldots, ax_n) = a^{-1}V(x_1, \ldots, x_n)$ for all $x_1, \ldots, x_n \in \mathbb{R}$ and $a > 0$). V can also be expressed as

$$V(y_1, \ldots, y_n) = \int_{S_n} \max_{i=1,\ldots,n} \frac{w_i}{y_i} \, \mathrm{d}H(w) \tag{3.19}$$

for some finite measure H, with $S_n := \{w \in [0,1]^n | \sum_{i=1}^n w_i = 1\}$.

Because of the absolute continuity of the GEV distribution and Sklar's theorem, V can be uniquely determined also for max-stable vectors that are not simple.

A useful summary measure for the interdependency of a max-stable vector Y is the *extremal coefficient*, which is defined as

$$\theta = V(1, \dots, 1), \tag{3.20}$$

with V as above. The extremal coefficient can take values between 1 and n, with 1 corresponding to comonoticity and n corresponding to stochastic independence of the components of the vector. In the case of a two-dimensional vector $Y = \{Y^{(1)}, Y^{(2)}\}$, the extremal coefficient (which takes values between 1 and 2) can be estimated from a data sample by using the madogram estimator by Ribatet et al. (2015) and Cooley et al. (2006). They rewrite θ as

$$\theta = \frac{1 + 2\nu}{1 - 2\nu} \tag{3.21}$$

with the madogram $\nu = \frac{1}{2}\mathbb{E}[|F_{Y^{(1)}}(Y^{(1)}) - F_{Y^{(2)}}(Y^{(2)})|]$, where $F_{Y^{(1)}}$, $F_{Y^{(2)}}$ are the pdf's of $Y^{(1)}$ and $Y^{(2)}$. For a data sample $(y_1^{(1)}, y_1^{(2)}), \dots, (y_m^{(1)}, y_m^{(2)})$, with $m \in \mathbb{N}$ the sample size, the madogram can be estimated by replacing the pdf's with their empirical counterparts

$$\hat{\nu} = \frac{1}{2m(m+1)} \sum_{i=1}^{m} \Big| \sum_{j=1}^{m} (\mathbf{1}_{y_j^{(1)} \le y_i^{(1)}} - \mathbf{1}_{y_j^{(2)} \le y_i^{(2)}}) \Big|, \tag{3.22}$$

and consequently, the extremal coefficient can be estimated as

$$\hat{\theta} = \frac{1 + 2\hat{\nu}}{1 - 2\hat{\nu}}. \tag{3.23}$$

3.2.2 A clustering algorithm for multivariate extremes

To investigate the dependence structures of a higher-dimensional max-stable vector $(Y^{(1)}, \dots, Y^{(n)})$, $n > 2$, a hierarchical clustering algorithm can be used. Hierarchical clustering is used to group an indexed set of n objects (here, the indices of the components of the random vector) into disjoint subsets. To apply the algorithm, a dissimilarity measure must be provided for all pairs of objects. The dissimilarities can be written in a dissimilarity matrix $D \in \mathbb{R}^{n,n}$, this matrix must fulfill $D_{i,j} = D_{j,i} \ge 0$ and $D_{i,i} = 0$ for all $1 \le i, j \le n$. In addition to that, the desired number of clusters and the linkage criterion (see below) has to be prescribed. The algorithm starts by treating each point as a separate cluster in the first step. In the next step, the pair of points with the lowest dissimilarity is merged into one cluster. In each subsequent step, two of the remaining clusters are merged into one, and this is repeated until the prescribed number of clusters is

reached. The dissimilarity between two clusters that consist of more than one point is determined based on the linkage criterion: If *single linkage* is used, the dissimilarity between two clusters is the minimum of all dissimilarities between points from the clusters. *Complete linkage* uses the maximum of the dissimilarities instead, and *average linkage* the average.

The application of hierarchical clustering to max-stable vectors is due to Bernard et al. (2013). All two-dimensional margins $(Y^{(i)}, Y^{(j)})$, $1 \leq i, j \leq n$ of a max-stable vector are max-stable again and an estimate for their extremal coefficient $\hat{\theta}_{i,j}$ can be calculated. A dissimilarity matrix can then be defined as

$$D_{i,j} := \hat{\theta}_{i,j} - 1. \tag{3.24}$$

Under this measure, two indices i, j are 'similar' and likely to be grouped into the same cluster if the corresponding extremal coefficient is close to 1, so if the variables $Y^{(i)}, Y^{(j)}$ show a high degree of extremal dependence. The clustering algorithm will therefore determine subsets of indices for which the corresponding random variables tend to have concurrent extremes.

3.2.3 Theory of max-stable processes

The extremal coefficients and the resulting clustering algorithms that were presented in the last section only give a summary the dependence structure of extremes at different locations. It is of interest to design mathematical models that can be fitted to given multivariate data and thus allow for inference on the full multivariate distribution. Assuming that the data investigated represent observations at different measurement stations which are located in $\mathbb{R}^2$, it is also desirable to create a model not only for those locations, but also for all other points in the same area. This leads us to considering not random vectors but stochastic processes on a compact subset $\mathcal{S}$ of $\mathbb{R}^2$. The following section on max-stable processes will mostly follow Ribatet (2017) and Ribatet et al. (2015).

A stochastic process $\{Y(s)\}_{s\in\mathcal{S}}$ attributes to each element $s \in \mathcal{S}$ a random variable $Y(s)$. We will often use the short notation Y for $\{Y(s)\}_{s\in\mathcal{S}}$ and, unless stated otherwise, consider from now on only processes with almost surely (a.s.) continuous sample paths. This property implies that the distribution of such a process is completely determined by its finite-dimensional marginal distributions. The definition of max-stability for processes is analogous to the one for random variables and vectors, only using this time continuous

functions for the normalizion:

Definition 3.2.2. A stochastic process Y is *max-stable* if it is non-degenerate and there are sequences of continuous functions $c_n, d_n : \mathcal{S} \to \mathbb{R}$, $d_n(\cdot)$ non-negative such that:

$$\left\{\frac{\max(Y_1(s), \ldots, Y_n(s)) - c_n(s)}{d_n(s)}\right\}_{s\in\mathcal{S}} \overset{\mathrm{d}}{=} \{Y(s)\}_{s\in\mathcal{S}} \tag{3.25}$$

for independent copies $Y_1, \ldots, Y_n$ of Y.

It holds:

Theorem 3.2.4. *Let $(X_i)_{i\in\mathbb{N}}$ be iid stochastic processes and $c_n, d_n : \mathcal{S} \to \mathbb{R}$, $d_n(\cdot)$ non-negative, sequences of continuous functions and assume that*

$$\left\{\frac{\max(X_1(s), \ldots, X_n(s)) - c_n(s)}{d_n(s)}\right\}_{s\in\mathcal{S}} \xrightarrow[n\to\infty]{\mathrm{d}} \{Y(s)\}_{s\in\mathcal{S}} \tag{3.26}$$

for some stochastic process Y with non-degenerate finite-dimensional marginal distributions. Then, Y must be max-stable.

All finite-dimensional marginal distributions of a max-stable process must necessarily be max-stable random vectors, so the results from the previous section are applicable. In particular, we assume as before that (after the application of marginal transformations) all univariate marginal distributions are unit Fréchet and call such a process a *simple max-stable process*. From the previous section, we know that for each max-stable process, the marginal distribution corresponding to some data points $s_1, \ldots, s_n$ must fulfill

$$\Pr(Y(s_1) \leq y_1, \ldots, Y(s_n) \leq y_n) = \exp(-V_{s_1,\ldots,s_n}(y_1, \ldots, y_n)) \tag{3.27}$$

for a function $V_{s_1,\ldots,s_n}$ that is positive and homogeneous of degree -1. The class of functions that are positive and homogeneous of degree -1 is too broad to be written using a finite set of parameters, so in order to perform parametric inference, the set of possible functions has to be restricted in a suitable manner. Assume a parametric subclass of max-stable processes with a parameter space Ψ, which implies in particular that the functions $V_{s_1,\ldots,s_n}$ are parametrized as $V^{\psi}_{s_1,\ldots,s_n}$ with $\psi \in \Psi$ for all $s_1, \ldots, s_n \in \mathcal{S}$. Before discussing possible choices of parametric sub-classes, we first investigate the log-likelihood in more detail. Applying the mixed derivative $\partial s_1 \ldots \partial s_n$, it follows from (3.18) that the likelihood

$f^{\psi}_{s_1,\dots,s_n}$ fulfills

$$f^{\psi}_{s_1,\dots,s_n}(y_1,\dots,y_n) = \exp(-V^{\psi}_{s_1,\dots,s_n}(y_1,\dots,y_n))\cdot \sum_{\tau\in\mathcal{P}_n}(-1)^{|\tau|}\prod_{i=1}^{|\tau|}\frac{\partial^{|\tau_i|}}{\partial y_{\tau_i}}V^{\psi}_{s_1,\dots,s_n}(y_1,\dots,y_n), \quad (3.28)$$

where the elements of $\mathcal{P}_n$ are the partitions τ of $\{s_1,\dots,s_n\}$. Each partition divides $\{s_1,\dots,s_n\}$ into subsets $\tau_1,\dots,\tau_l$, with $|\tau| = l$. In particular, the bivariate densities can be calculated as

$$f^{\psi}_{s_1,s_2}(y_1,y_2) = \exp\Big(-V^{\psi}_{s_1,s_2}(y_1,y_2)\Big)\cdot \Big(\frac{\partial}{\partial y_1}V^{\psi}_{s_1,s_2}(y_1,y_2)\frac{\partial}{\partial y_2}V^{\psi}_{s_1,s_2}(y_1,y_2) - \frac{\partial^2}{\partial y_1\partial y_2}V^{\psi}_{s_1,s_2}(y_1,y_2)\Big). \quad (3.29)$$

In practical applications, a max-stable process is fitted to data that are given on a finite set of locations $\mathcal{T}\subseteq\mathcal{S}$ (for example the locations of weather stations). Denote the (unit Fréchet distributed) data at location $t\in\mathcal{T}$ by $y_t^{(1)},\dots,y_t^{(m)}$ with $m\in\mathbb{N}$ the sample size. The number of partitions of a set with n elements is equal to the Bell number B_n. With increasing n, these numbers soon gets very large: From $n = 10$ on, the Bell numbers are greater than $100,000$, and they exceed the value of one billion for $n \geq 16$. Numerical optimization of the likelihood is computationally not feasible in these cases, and the classical maximum-likelihood approach is therefore usually not applicable to the multivariate density of all data points in $\mathcal{T}$. As a remedy, it is common to use a composite likelihood approach instead: Maximum-likelihood estimation is not carried out by maximizing the full log-likelihood, but instead by maximizing the sum of bivariate marginal log-likelihoods

$$\hat{\psi} = \arg\max_{\psi\in\Psi} L(\psi) \quad (3.30)$$

with

$$L(\psi) = \sum_{i=1}^{m}\sum_{t_1\neq t_2\in\mathcal{T}}\log f^{\psi}_{t_1,t_2}(y^{(i)}_{t_1}, y^{(i)}_{t_2}). \quad (3.31)$$

To reduce numerical complexity, it is also common practice to include only those pairs (t_1,t_2) in Eq. (3.31) for which $\|t_1 - t_2\|$ does not exceed a certain threshold (throughout

this work, $\|\cdot\|$ denotes the Euclidean norm).

The general approach behind composite likelihoods is as follows: Assume that $Y_1, \ldots, Y_n$ are iid random variables with values in $\mathbb{R}^d$, $d \geq 2$ and with pdf $f(y, \theta^\star)$, with θ^*, the true parameter, being an element of the set of possible parameter values $\Theta \subseteq \mathbb{R}^p$. Let further $\{A_i | i \in I\}$ be a countable set of marginal or conditional events on $\mathbb{R}^d$. If asymptotic normality holds for the maximum-likelihood estimator $\hat{\theta}_n$ (remember that under suitable regularity conditions for f and Θ we have that as $n \to \infty$, $\sqrt{n}(\hat{\theta}_n - \theta^*)$ converges in distribution to a normal distribution with zero mean and variance equal to the inverse of the Fisher information), then for the composite likelihood $\mathcal{L}(y, \theta) := \sum_{i \in I} \sum_{j=1}^{n} f(y_j \in A_i, \theta)$ and the corresponding estimator $\hat{\theta}_n^c := \mathrm{argmax}_\theta \mathcal{L}(y, \theta)$ it holds

$$\sqrt{n}(H(\theta^*)J(\theta^*)^{-1}H(\theta^*))^{1/2}(\hat{\theta}_n^c - \theta^*) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0, \mathrm{Id}_p), \tag{3.32}$$

where $M^{1/2}$ denotes the matrix root of M and with $H(\theta^*) = -\mathbb{E}[\nabla^2 \mathcal{L}(Y, \theta^*)]$ and $J(\theta^*) = \mathrm{Var}(\mathcal{L}(Y, \theta^*))$. This property gives the rationale for maximizing the sum of the likelihoods of the bivariate marginal distributions in place of the full likelihood.

In Chapter 6, max-stable models will also be fitted to data on subsets of $\mathcal{T}$. To introduce the notation for this, let $M \subseteq \mathcal{T}$ be a subset of $\mathcal{T}$ with at least two elements, then we write the corresponding composite likelihood as

$$L_M(\psi) = \sum_{i=1}^{m} \sum_{t_1 \neq t_2 \in M} \log f_{t_1,t_2}^{\psi}(y_{t_1}^{(i)}, y_{t_2}^{(i)}) \tag{3.33}$$

and the corresponding estimator as

$$\hat{\psi}_M = \arg\max_{\psi \in \Psi} L_M(\psi). \tag{3.34}$$

3.2.4 Parameteric models for max-stable processes

In order to derive parametric models for max-stable processes, the following stochastic representation for simple max-stable processes (de Haan, 1984; Penrose, 1992) can be used:

Theorem 3.2.5. *Let Y be a simple max-stable process. Then,*

$$\{Y(s)\}_{s \in \mathcal{S}} \overset{\mathcal{D}}{=} \left\{ \max_{i \geq 1} \zeta_i Z^{(i)}(s) \right\}_{s \in \mathcal{S}}, \tag{3.35}$$

where $(\zeta_i)_{i\in\mathbb{N}}$ are the points of a Poisson point process with intensity measure $d\Lambda(\zeta) = \zeta^{-2}d\zeta$ and $\{Z^{(i)}\}_{i\in\mathbb{N}}$ are independent copies of a non-negative stochastic process Z that fulfills $\mathbb{E}[Z(s)] = 1$ for all $s \in \mathcal{S}$.

The process Z is called spectral process of Y. Several representations (using different distributions of Z) can exist for one max-stable process. Using the spectral representation, various parametric models for simple max-stable processes have been constructed by restricting the underlying spectral processes to selected parametric families. One of the first was the *Smith process* (Smith, 1990), which uses as spectral process the process Z given by

$$Z(s) = \phi(s - U_i, \Sigma), \tag{3.36}$$

where ϕ is a two-dimensional Gaussian density with mean 0 and covariance matrix Σ, and the U_i are uniformly distributed on $\mathcal{S}$. The intuition is to give a model for extreme rainfall: The copies of the spectral process correspond to the spatial extents of different storms in the area, and the points of the Poisson process describe the intensity of the storm. Note that the intensity measure of the Poisson process is $d\Lambda(\zeta) = \zeta^{-2}d\zeta$, so with a high probability only few points of the Poisson process will be far away from zero. A realization of a Smith process is depicted in Fig. 3.1a.

The Smith process is quite restrictive, as the only random element in its spectral process is the location of the Gaussian density. A more flexible approach is the *Brown-Resnick process*, derived by Kabluchko et al. (2009) based on Brown and Resnick (1977), which uses as spectral process

$$Z(s) = \exp(\epsilon(s) - \gamma(s)),$$

where ϵ is a Gaussian process with (spatially) stationary increments and γ its semi-variogram. Note that the distribution of Z does not depend on the variance of ϵ, only on the semi-variogram. Following Kabluchko et al. (2009) and Ribatet (2017), a common choice for the semi-variogram is

$$\gamma(h) = \|h/\lambda\|^\alpha,$$

with $\lambda > 0$, $0 < \alpha \leq 2$, making ϵ a fractional Brownian motion. Using Eq. (3.35), it can be calculated that for $s_1, s_2 \in \mathcal{S}$

$$V_{s_1,s_2}(y_1, y_2) = \frac{1}{y_1}\Phi\Big(\frac{\sqrt{2\gamma(h)}}{2} - \frac{\log y_2/y_1}{\sqrt{2\gamma(h)}}\Big) + \frac{1}{y_2}\Phi\Big(\frac{\sqrt{2\gamma(h)}}{2} - \frac{\log y_1/y_2}{\sqrt{2\gamma(h)}}\Big),$$

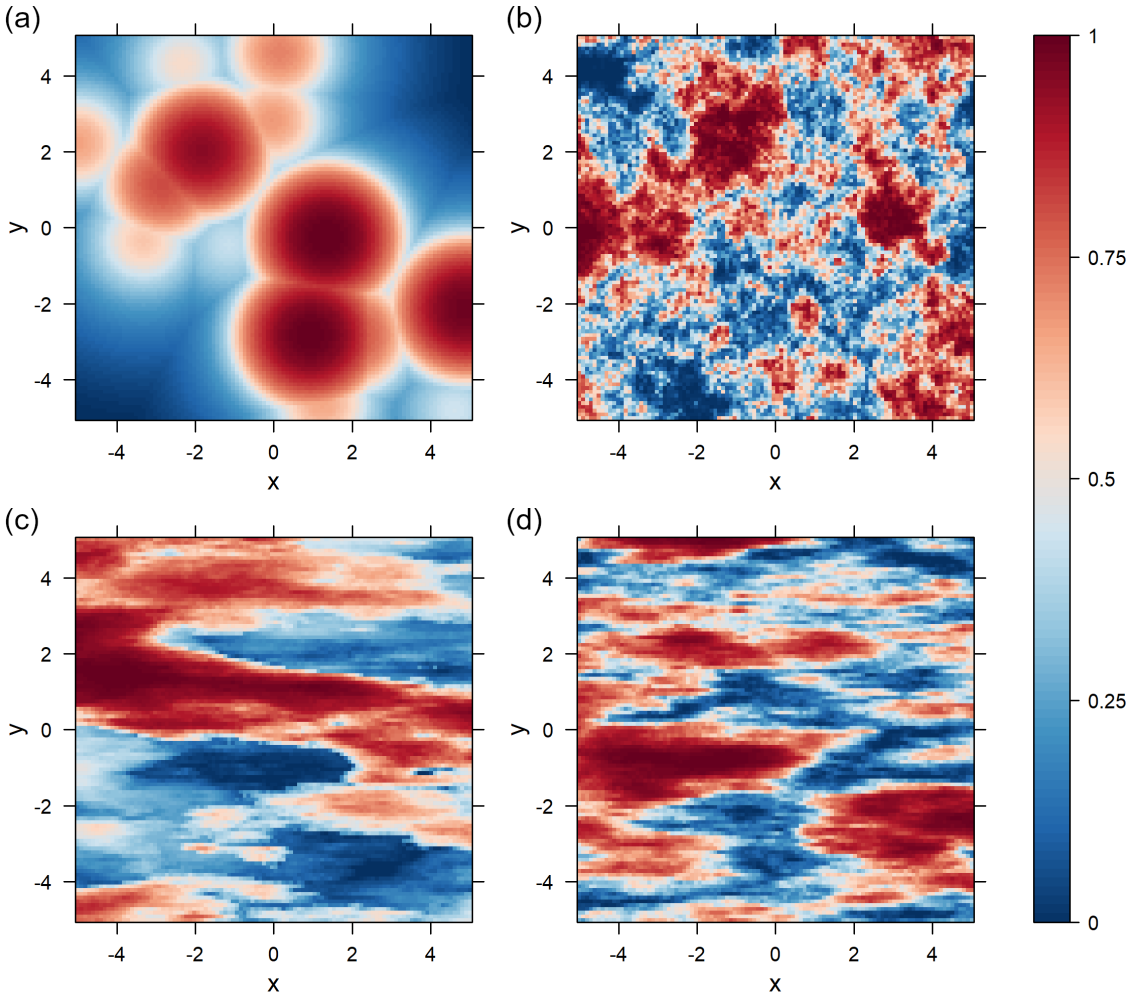

Figure 3.1: Realizations of max-stable processes. (a) Smith process with $\Sigma = \mathrm{Id}_2$, (b) Brown-Resnick process with $\gamma(h) = \|h\|$, (c) Schlather process with $\rho(h) = \exp(-\|h/\|)$, (d) Extremal-t process with ρ as in (c) and $\nu = 10$. For visual reasons all univariate marginal distributions have been transformed to uniform distributions.

where $h = \|s_1 - s_2\|$ and Φ is the standard normal distribution function. A realization of a Brown-Resnick process is depicted in Fig. 3.1b

A different approach is followed by Schlather (2002), who defines the *Schlather process* using as spectral process

$$Z(s) = \sqrt{2\pi}\max(0, G(s)),$$

with G a standard Gaussian process with a covariance function ρ that is spatially stationary (it depends only on the distance $h = \|s_1 - s_2\|$ between the two points $s_1, s_2 \in \mathcal{S}$; we will use the notation $\rho(h)$ instead of $\rho(s_1, s_2)$). A realization of a Schlather process is depicted in Fig. 3.1c.

For the Schlather process, it holds

$$V_{s_1,s_2}(y_1, y_2) = \frac{1}{2}(\frac{1}{y_1} + \frac{1}{y_2})\Big(1 + \sqrt{1 - \frac{2(1+\rho(h))y_1y_2}{(y_1+y_2)^2}}\Big) \tag{3.37}$$

with $h = \|s_1 - s_2\|$. The pairwise extremal coefficients thus fulfill

$$\theta_{s_1,s_2} = 1 + \sqrt{1 - \frac{1+\rho(h)}{2}}. \tag{3.38}$$

Remember that $Y(s_1)$ and $Y(s_2)$ are stochastically independent if and only if the extremal coefficient θ_{s_1,s_2} is equal to 2. For the Schlather process the highest attainable value of the extremal coefficient is $1 + \sqrt{1/2} < 2$, so the bivariate distributions of a Schlather process always exhibit extremal dependence, even if the underlying Gaussian variables are uncorrelated. This is due to the heavy-tailedness of the Poisson process compared to the Gaussian process, causing a tendency for extremes to occur simultaneously as they tend to be caused by high values in the Poisson process.

To avoid this limitation, Opitz (2013) proposed a model using a spectral process that is also heavy-tailed. Schlather's model was extended to the so-called *extremal-t model*, using as spectral process

$$Z(s) = c_\nu \max(0, G(s))^\nu$$

with $\nu \geq 1$ fixed, G as in the Schlather model and c_ν a constant that ensures that $\mathbb{E}[Z(s)] = 1$.

For this process, it holds

$$V_{s_1,s_2}(y_1,y_2) = \frac{1}{y_1} T_{\nu+1}\Big(\frac{(y_2/y_1)^{1/\nu} - \rho(h)}{(\nu+1)^{-1/2}(1-\rho(h)^2)^{1/2}}\Big) + \frac{1}{y_2} T_{\nu+1}\Big(\frac{(y_1/y_2)^{1/\nu} - \rho(h)}{(\nu+1)^{-1/2}(1-\rho(h)^2)^{1/2}}\Big)$$

with h as before and $T_{\nu+1}$ the cdf of the Student-t distribution with $(\nu+1)$ degrees of freedom. A realization of an extremal-t process is depicted in Fig. 3.1d.

One possible choice for the correlation function ρ in the Schlather or extremal-t process is $\rho(h) = \exp(-\|h/\lambda\|^\alpha)$ with $\lambda > 0$ and $0 < \alpha \leq 2$ (Ribatet, 2017). To model anisotropic behavior, Davis et al. (2013) proposed to use

$$\rho(h) = \exp(-\|Ah\|^\alpha) \tag{3.39}$$

instead, with A a 2×2 transformation matrix, meaning that

$$A = \begin{pmatrix} \sin(\gamma)/a & \cos(\gamma)/(a+b) \\ -\cos(\gamma)/(a+b) & \sin(\gamma)/b \end{pmatrix} \tag{3.40}$$

with parameters $a > 0$, $b \geq 0$ and $\gamma \in [0, \pi)$. The level sets of ρ are then not circular, but elliptical with a and $a+b$ proportional to the lengths of the minor and major axes and γ the angle between the x-axis and the major axis of the ellipsis. In climate data, elliptical level sets are preferable over the circular ones of isotropic covariance functions because the presence of ocean currents, a predominant wind direction or topographical boundaries frequently cause the extremal dependence structure of two points to depend not only on their distance, but also on the direction.

All the models discussed up to this point share the property that the finite-dimensional distributions do not change if the data points are shifted in space: they are spatially stationary. For data that cover a heterogeneous area in which dependence structures of extremes are expected to vary depending on the region, these models are therefore of limited applicability. An approach to cover also these cases is the *Huser-Genton process* (Huser and Genton, 2016), which extends the extremal-t processes by using a non-stationary Gaussian process as underlying spectral process. Huser-Genton processes are constructed using Gaussian kernels, making use of an idea by Paciorek and Schervish (2006): For the correlation func-

tion, we do not use a 2×2 transformation matrix A for the whole space like in Eq. (3.39), but instead one matrix A_s for each point $s \in \mathcal{S}$ such that the map $s \mapsto A_s$ is continuous. Let further $\Omega_s = (A_s^T A_s)^{-1}$ and $R(x) = \exp(-x^\alpha)$ for $x \in \mathbb{R}, x \geq 0$. Then, a non-stationary Gaussian process can be constructed such that its covariance structure is given by

$$\rho(s_1, s_2) = |\Omega_{s_1}|^{1/4} |\Omega_{s_2}|^{1/4} \left| \frac{\Omega_{s_1} + \Omega_{s_2}}{2} \right|^{-1/2} R\Big(\sqrt{(s_1 - s_2)^T \Big(\frac{\Omega_{s_1} + \Omega_{s_2}}{2} \Big)^{-1} (s_1 - s_2)} \Big). \quad (3.41)$$

This reduces to the extremal-t process from the previous section if A_s is constant on the whole space. To fit Huser-Genton processes to data, the entries of the matrices A_s are parametrized and then regression is used with suitable covariates like longitude, latitude and altitude.

Another approach to tackle the investigation of spatially heterogeneous data is the use of clustering: The area of investigation is divided into different clusters, and then a stationary max-stable model is fitted to each cluster. The disadvantage of this approach is that the resulting model does not describe the dependency between two points that are in different clusters. We will discuss clustering approaches for max-stable processes in more detail in Chapter 6.

Chapter 4

Validation of climate models using historical data

4.1 Introduction to the chapter

As discussed in Section 2.3, coupled global circulation models are of paramount importance to assess the magnitude of future climate change quantitatively, and it is essential to validate these models by comparing historical runs of the models to observational data. While climate models are able to reproduce many climate phenomena across the globe, their reliability regarding extremes requires additional evaluation. This is the case because of the large impacts of extremes on the natural environment, the economy and the human health (Ciais et al., 2005; Kovats and Kristie, 2006) and also because of the fact that due to the inherent nature of extreme events, their evolution can differ from that of the mean and the variance (Schär et al., 2004; IPCC, 2012). In this chapter, we will investigate and validate historical runs of several CMIP6 models with respect to climate extremes.

We will evaluate the performance of the fully coupled Alfred Wegener Institute-Earth System Model AWI-ESM1.1LR (Shi et al., 2020; Lohmann et al., 2020; Ackermann et al., 2020) in terms of its accuracy regarding variability and extremes of precipitation, putting special focus on spatially concurrent precipitation extremes. The main questions are whether the model is able to accurately reproduce extreme events in different regions and whether spatial dependencies and concurrent extremal events are modeled adequately. We compare model data from a historical run of the AWI-ESM to the global precipitation reanalysis data set CRU TS4.04 (Harris et al., 2020). We start with investigating variability

and extremes locally using empirical statistical parameters and by fitting a GEV distribution to annual precipitation maxima. Then, we use the clustering algorithm by Bernard et al. (2013) (see Chapter 3.2.2) to group the data into different spatial regions based on their similarity in terms of extremal behavior. We extend and generalize the dissimilarity measure by Bernard et al. (2013) to have it take into account also the dissimilarity of the marginal distributions. The resulting clusters are compared for model and observational data and they are used to analyze the ability of the climate model to reproduce spatial dependencies of precipitation extremes.

Our main focus lies on the AWI-ESM and we present our methods using data from this model. We also present a measure for the model accuracy in regard to extremal precipitation, and apply it to a set of different CMIP6 models. Results will be discussed in detail for the AWI-ESM and for the model identified as having the best model accuracy. The methods have also been applied to 26 other models participating in CMIP6, the results for those models can be found in the supplementary material to Contzen et al. (2022).

Model validation in terms of precipitation extremes is already an active research topic. Tabari et al. (2016) investigate the performance of global and regional climate models using the peaks-over-threshold approach. An evaluation of regional and global climate models using extreme precipitation indices is conducted by Bador et al. (2020), revealing a tendency for stronger extremes in regional models. A similar result was obtained by Mahajan et al. (2015) by comparing climate model and observational precipitation data over the United States using GEV distributions. Timmermans et al. (2019) conduct pairwise comparisons of the precipitation extremes of numerous gridded observation-based data sets and find considerable differences between the data sets especially in mountainous regions. Precipitation extremes over India are investigated by Mishra et al. (2014) using GEV distributions and comparisons of indices with a focus on changes over time.

It is also not a new approach to apply clustering algorithms to climate data. Among others, it has been used to define climate zones in the United States (Fovell and Fovell, 1993) and globally (Zscheischler et al., 2012), and to find regions with similar trends in their climatic change over Europe (Carvalho et al., 2016). Those analyses focus on mean values and on their temporal differences, respectively, while we apply clustering specifically to uncover connections regarding climate extremes.

This chapter is structured as follows: After introducing the data sets in Section 4.2, we present the methods used in Section 4.3. The results from applying them to the data are presented in Section 4.4. A section on conclusions and discussions finalizes the chapter.

4.2 Data analyzed in this chapter

The observational data are reanalyzed monthly precipitation data in mm over land (excluding Antarctica) from the CRU TS4.04 data set (Harris et al., 2020; University of East Anglia Climatic Research Unit et al., 2020) with data ranging from 1901 to 2019. We restrict the time frame to the years 1930 to 2014 in order to have a sufficiently large area with non-missing data and to be consistent with the climate model data. The grid size is $0.5° \times 0.5°$, the data have been obtained by interpolating observations from more than 4.000 weather stations using angular distance weighting.

At some locations and time points, no data from nearby weather stations had been available to use for interpolation. In these cases, the creators of the CRU TS4.04 data set used a value from a climatology instead. These climatology values are not very informative in terms of extremes and too many of them would distort the analyses, therefore all grid points with more than 5% climatology values and additionally all grid points with at least twelve consecutive months of climatology values are excluded from our analysis. This results in the exclusions of larger regions in northern and central Africa, in Indonesia, in central Asia and in the polar regions. In the figures showing geographical data in this chapter, those regions are coloured in grey.

The climate model used is the coupled model AWI-ESM1.1LR. It is based on the AWI Earth System Model (AWI-ESM1), which consists of the AWI Climate Model (Sidorenko et al., 2015; Rackow et al., 2018), but with interactive vegetation. The model comprises the atmosphere model ECHAM6 (Stevens et al., 2013), which is run with the T63L47 setup (that is, a horizontal resolution of $1.85° \times 1.85°$ and 47 vertical layers) and the ocean-sea ice model FESOM1.4 (Wang et al., 2014), which employs an unstructured grid, allowing for varying resolutions from 20km around Greenland and in the North Atlantic to around 150km in the open ocean. The land surface processes are computed by the land surface model JSBACH2.11 (Reick et al., 2013). The model considers the surface runoff toward the coasts, deploying a hydrological discharge model that also includes freshwater fluxes by snowmelt (Hagemann and Dümenil, 1997). AWI-ESM1 has been extensively used and

described in the context of paleoclimate changes as well as of changes of the recent and future climate (Shi et al., 2020; Lohmann et al., 2020; Ackermann et al., 2020; Niu et al., 2021). The historical run is documented in Danek et al. (2020) and has been directly used in Ackermann et al. (2020) and Keeble et al. (2021). The model takes furthermore part in CMIP6/PMIP4 activities (Brierley et al., 2020; Brown et al., 2020; Otto-Bliesner et al., 2021; Kageyama et al., 2021a,b).

In our analysis, we restrict the time frame of the model data to the years 1930 to 2014, as in the observational data. We investigate monthly precipitation (sum of convective precipitation and large-scale precipitation) in mm/month. We use bilinear interpolation to scale the reanalysis data to the grid of the atmospheric component of the climate model and take into account only those interpolated grid points that correspond to locations with given observed data, excluding the oceans and the regions with incomplete data mentioned above.

4.3 Methods used in this chapter

We first perform an univariate analysis of extremes, so we investigate the time series of each spatial location (henceforth referred to as grid point) separately, and all operations and analyses described are conducted for each grid point. Since the focus here is not on evaluating the effects of long-time trends, we apply a seasonal-trend decomposition using Loess (Cleveland et al., 1990) on the data and subtract the trend from the data but readd the mean value of the trend, resulting in data for which we assume temporal stationarity. Then, as a first comparison between the data sets, we investigate differences in the empirical mean and empirical standard deviation of the annually maximized precipitation data.

We group the monthly precipitation data from observations and the climate model into one-year block maxima and fit a GEV distribution to the block-wise maxima at each grid point. In our case, we have a relatively small block size of 12 (months per year) and a number of block-wise maxima of 90 (years of investigation). The GEV parameters are estimated using the PWME method. We test the goodness of fit using a one-sided Kolmogorov-Smirnov-test at significance level 5%. The null hypothesis of the test is that the annually maximized data follow the GEV distribution having the probability-weighted

moments estimates as distribution parameters.

We also use the parametric bootstrap method with 2500 resamples to compute 95% confidence intervals for each GEV parameter and for the 95% quantiles of the distributions. Confidence intervals for the GEV parameters based on asymptotic normality also exist for the probability-weighted moments estimators, but, as shown by Hosking et al. (1985), they have a high bias and variance if the shape parameter is far away from zero. In our data, for several time series such a value is estimated for the shape parameter, and comparisons between the confidence intervals based on bootstrap and those based on asymptotic normality also confirmed large differences in these cases. For the sake of methodological consistency and because we also use the bootstrap for the confidence intervals of the 95% quantiles, we calculated the GEV parameter confidence intervals using bootstrap for all time series. Since this method is quite time-consuming, it could also be advocated to choose the method of confidence interval calculation based on the estimated shape parameter value.

To compare the performance of different CMIP6 models, we introduce as a measure for the accuracy of the extremal precipitation an Average Weighted Quantile Difference (AWQD). For this measure, the absolute differences between model and observational 95% GEV quantiles, weighted with the cosine of the latitude, are averaged. The weighting accounts for the fact that the grid cells do not have an equal size for all grid points, and the average is taken because of the different model resolutions. For $\mathcal{G}$ the set of grid points and estimated quantiles $\hat{q}_{0.95,\text{mod}}(g)$ and $\hat{q}_{0.95,\text{obs}}(g)$ for $g \in \mathcal{G}$, we therefore define

$$\text{AWQD} := \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \cos(\text{lat}(g)) \cdot |\hat{q}_{0.95,\text{mod}}(g) - \hat{q}_{0.95,\text{obs}}(g)|. \tag{4.1}$$

In the second part of the analyses, we investigate the spatial distributions of climate extremes. To this end, we use the madogram estimator to calculate extremal coefficients and they apply the clustering algorithm presented in 3.2.2. Remember that this algorithms uses a dissimilarity function based on the estimated extremal coefficients for all pairs of grid points $g_1, g_2 \in \mathcal{G}$:

$$D_0(g_1, g_2) := \hat{\theta}_{X,Y} - 1 \tag{4.2}$$

with X and Y representing the GEV distributions at the grid points g_1 and g_2, respectively.

Note that the extremal coefficient is invariant under rank transformations and especially that it does not depend on the values of the GEV parameters of the marginal distributions. It may be desirable to also include the dissimilarity of the marginal distributions in the clustering. As a further generalized dissimilarity measure we propose

$$D_\lambda(g_1, g_2) := (1-\lambda)D_0(g_1, g_2) + \lambda\Big(\frac{1}{3}d_\mu(g_1, g_2) + \frac{1}{3}d_\sigma(g_1, g_2) + \frac{1}{3}d_\gamma(g_1, g_2)\Big), \tag{4.3}$$

where $\lambda \in [0, 1)$ is a weighting parameter and with

$$d_\mu(g_1, g_2) := \frac{|\hat{\mu}_{g_1} - \hat{\mu}_{g_2}|}{\max_{h_1,h_2\in\mathcal{G}} |\hat{\mu}_{h_1} - \hat{\mu}_{h_2}|} \in [0, 1] \tag{4.4}$$

the normalized distance between the location parameter estimates at the grid points g_1 and g_2 (analogous for d_σ and d_γ). Instead of an equal weighting, it would also be possible to use different weights for d_μ, d_σ and d_γ, but the selection of a set of weights that is clearly better suited to describing GEV distribution dissimilarity is difficult. It could be argued to put more weight on the shape parameter since this parameter describes the heavy-tailedness of the distribution and therefore the strength of its extremes relative to the non-extreme values. On the other hand, we will see in the next section that the uncertainty in the shape parameter estimation is considerably higher than the uncertainty in the estimation of the other two parameters at least for our data, which would speak against weighting shape parameter differences too strongly.

To choose a suitable number of clusters, we consider an approach by Salvador and Chan (2004) called the L-Method. In each step of the hierarchical clustering, the two clusters with minimal dissimilarity are combined, therefore we can plot the number of clusters versus the dissimilarity between them, resulting in a graph called the evaluation graph. The dissimilarity between clusters necessarily grows as the total number of clusters is reduced. The idea of Salvador and Chan (2004) is to find a point from which on the growth rate of the dissimilarity measure increases considerably. It can then be expected that the clusters up to this point combine rather similar data points, while combining them to larger ones would yield artificial results. To determine such a point of change, in the first step, a suitable range of the number of clusters is selected. For our example, we consider different ranges starting with 10 and having no more than 550 clusters. Now, for each possible point of change c in this range, the horizontal axis of the graph is divided into the two parts to the left and the right of that point, and a linear regression line is fitted to each of the two partial

graphs. The root mean squared errors (RMSEs) of the two regression lines are weighted with the number of points involved in the regression analysis and summed up. The point of change with the minimal combined RMSE is chosen as the suitable cluster number. As an alternative method, we set the number of clusters to the highest possible number such that a fixed threshold dissimilarity between clusters is not exceeded (Threshold method). This number can easily be read off of the evaluation graph.

4.4 Results of this chapter

We start with calculating for each grid point the empirical mean and standard deviation of the annually maximized data, as can be seen in Fig. 4.1. In most regions, similar mean values can be observed. A notable overestimation of the annual maxima of monthly precipitation by the climate model takes place in the Himalayas and along the western continent coasts of the Americas. Underestimation occurs most prominently in the Amazon region and parts of Central America, as well as in Bangladesh and East Asia. Looking at the standard deviation, a similar pattern as for the empirical mean can be observed, but with a stronger tendency for underestimation, which occurs also in India and the northern part of Australia. In Fig. 4.2 a) and b), quantile-quantile plots (QQ-plots) of empirical mean and standard deviation are displayed. The quantiles of the empirical mean are in general similar, but the highest quantiles show a strong discrepancy. Regarding the standard deviation, this tendency is much more pronounced, corresponding to the larger areas of underestimation of empirical standard deviation we identified in Fig. 4.1. The difference in empirical mean and the difference in empirical standard deviation are plotted against each other in Fig. 4.2 c). It is visible that in many cases, overestimation (underestimation) of the empirical mean corresponds also to overestimation (underestimation) of the empirical standard deviation. A similar case of heteroscedasticity has also been noted in Lohmann (2018) when investigating Holocene climate.

As pointed out by Katz and Brown (1992), the frequency of extreme events is strongly influenced by changes (or, in this case, misestimation) of the mean as well as of the variance of a distribution. Therefore, an over- and underestimation of extremes can be expected in certain regions based on the results in Figs. 4.1 and 4.2.

Fitting the GEV distributions to the data and applying KS-Tests to check the goodness of fit, the hypothesis of a GEV distribution with the estimated parameters is not rejected

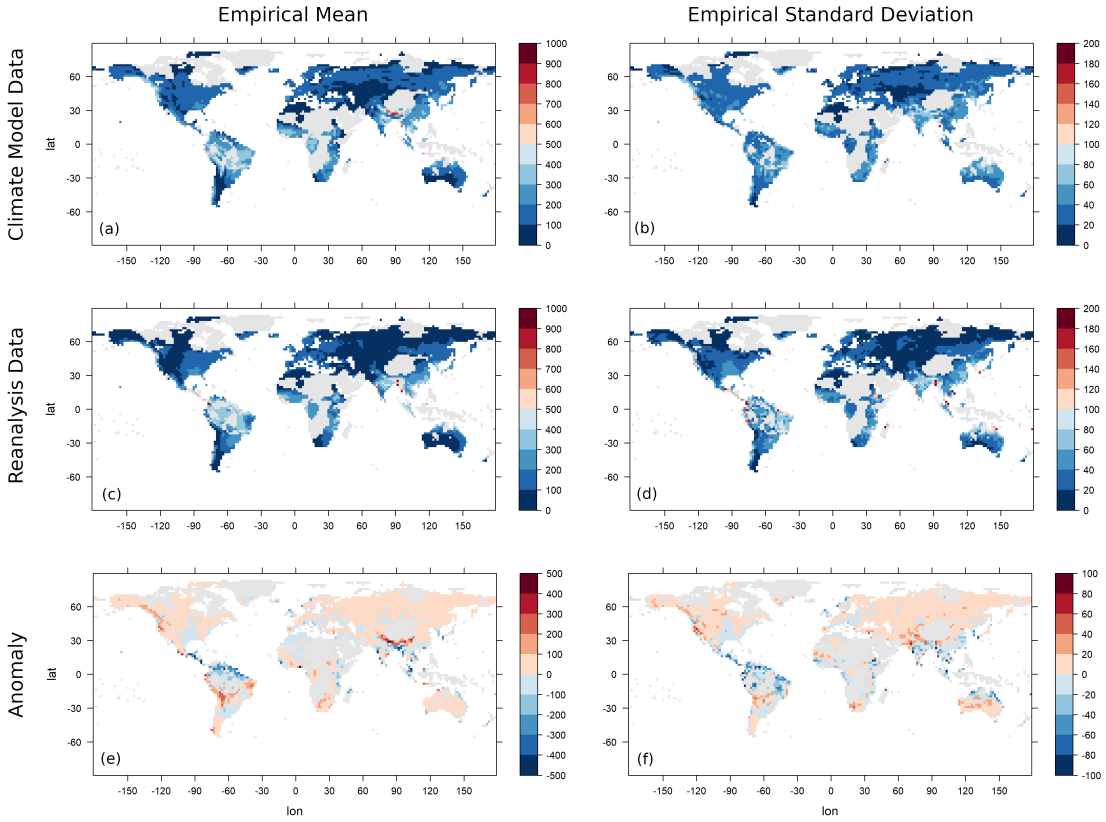

Figure 4.1: The empirical mean (a, c, e) and empirical standard deviation (b, d, f) of the annual maxima of monthly precipitation of the AWI-ESM model data set (a, b) and of the CRU TS4.04 reanalysis data set (c, d) and their difference (model data minus reanalysis data; e, f). Values exceeding the scale limits are truncated. Units are mm/month.

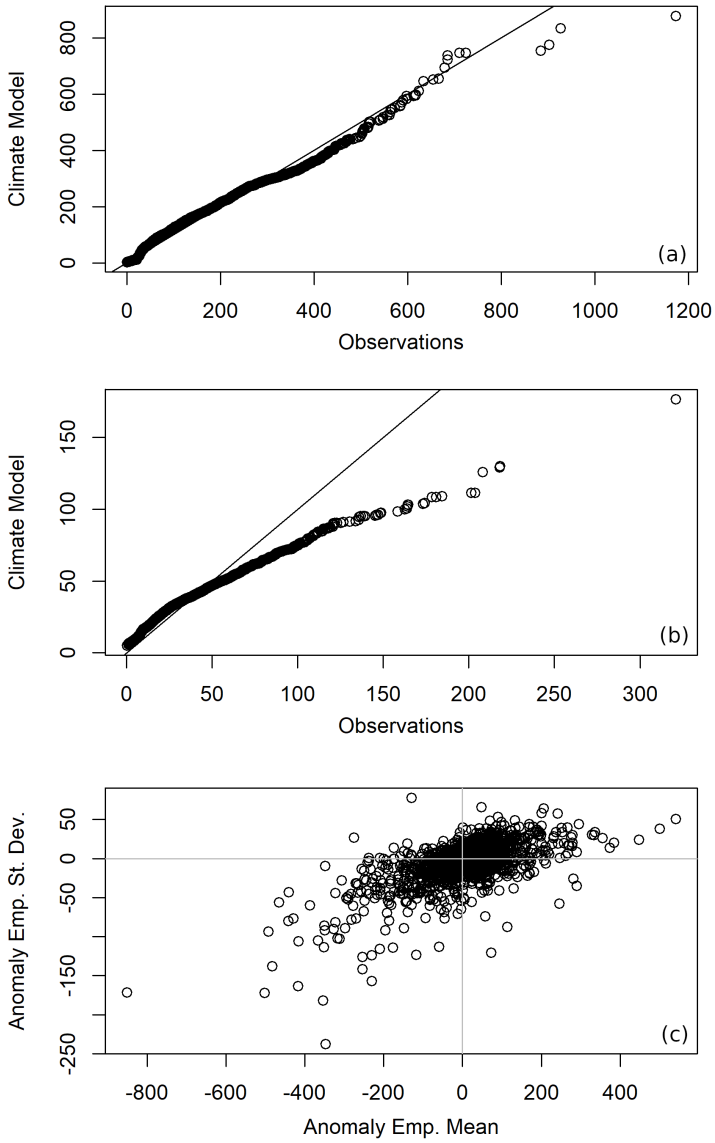

Figure 4.2: QQ-Plots comparing the empirical mean values (a) and the empirical standard deviations (b) of the annually maximized monthly precipitation of the CRU TS4.04 reanalysis data set and of the AWI-ESM model data set. Deviance of empirical mean and standard deviation plotted against each other (c). Units are mm/month.

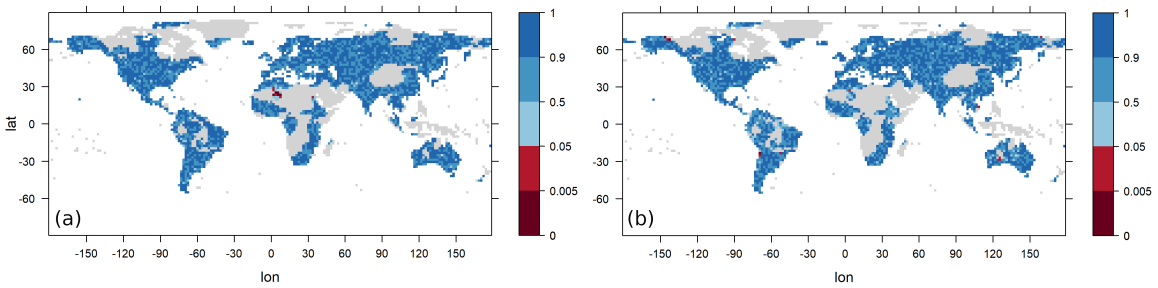

Figure 4.3: P-values of Kolmogorov-Smirnov tests for the hypothesis that the data follow a GEV distribution with parameters estimated using probability-weighted moments. Test results for the AWI-ESM climate model (a) and for the CRU TS4.04 reanalysis data (b).

for nearly all grid points in both observational and climate model data, except for parts of the Sahara and some isolated points.

The estimated parameter values are depicted in Fig. 4.4. In Fig. 4.5, the differences between model and observational parameters are depicted. Shaded areas are areas in which the model parameter falls into the 95% confidence interval of the corresponding observation parameter and vice versa. We can observe a similarity between the anomaly of the location parameters and the anomaly of the empirical means discussed above, and likewise a similarity between the anomalies of scale parameters and empirical standard deviations. For the location parameter, we observe high differences quite often, and the parameters estimated for one data set seldom fall into the confidence interval derived from the other data set. The estimated scale parameters are covered more often by the confidence intervals derived from the other data set, although there are also large regions with a high difference in the two estimates. The estimated shape parameters are covered by the confidence intervals at many locations, but it needs to be noted that the estimator of the shape parameter is known to be sensitive to small variations in the data. Therefore, the confidence intervals calculated using the parametric bootstrap tend to be large and not particularly informative. In Fig. 4.6, the anomalies of the 95% upper quantiles of the estimated GEV distributions are depicted, again with shaded areas indicating quantiles lying within the confidence levels determined using parametric bootstrap. Climate extremes are most strongly overestimated by the model in the mountainous regions of the Himalaya, the Andes and the Rocky Mountains. An underestimation of climate extremes takes place most notably in the Amazon region and parts of eastern Asia. This corresponds well to the regions of over- and underestimation of the empirical means and standard deviations and the implications of

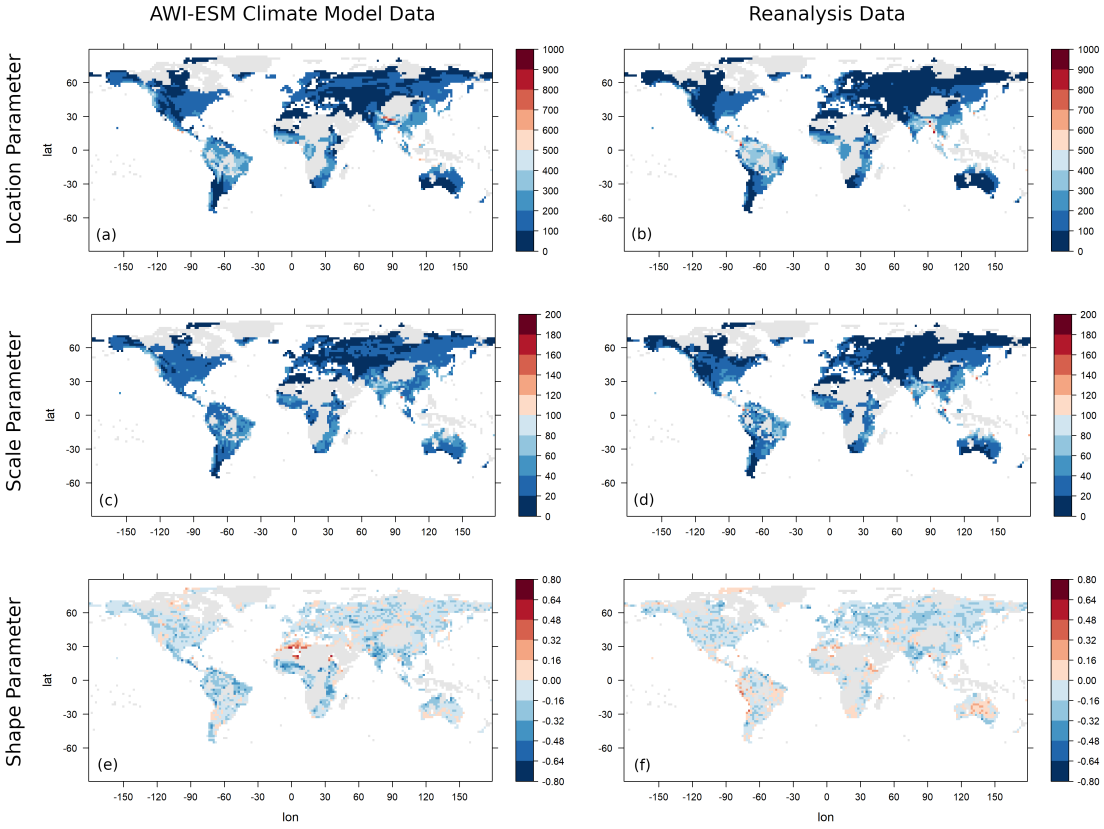

Figure 4.4: The estimated GEV parameters location (a, b), scale (c, d) and shape (e, f) for AWI-ESM climate model data (a, c, e) and for reanalysis data (b, d, f). Values exceeding the scale limits are truncated. Units are mm/month.

such misestimations discussed above.

We apply the hierarchical clustering algorithms using the two dissimilarity measures D_0 and $D_{0.25}$ as introduced in the previous section. The numbers of clusters determined using the L-Method with selected cluster ranges (from 10 to a maximal number of clusters m) and using the threshold method with selected threshold dissimilarities h is documented in Table 4.1.

The results of the L-Method seem to depend rather strongly on the data set investigated and the value of m (compare for example the results for $m = 250$ and $m = 300$ for measure D_0), making this method less suitable for the comparison of two data sets.

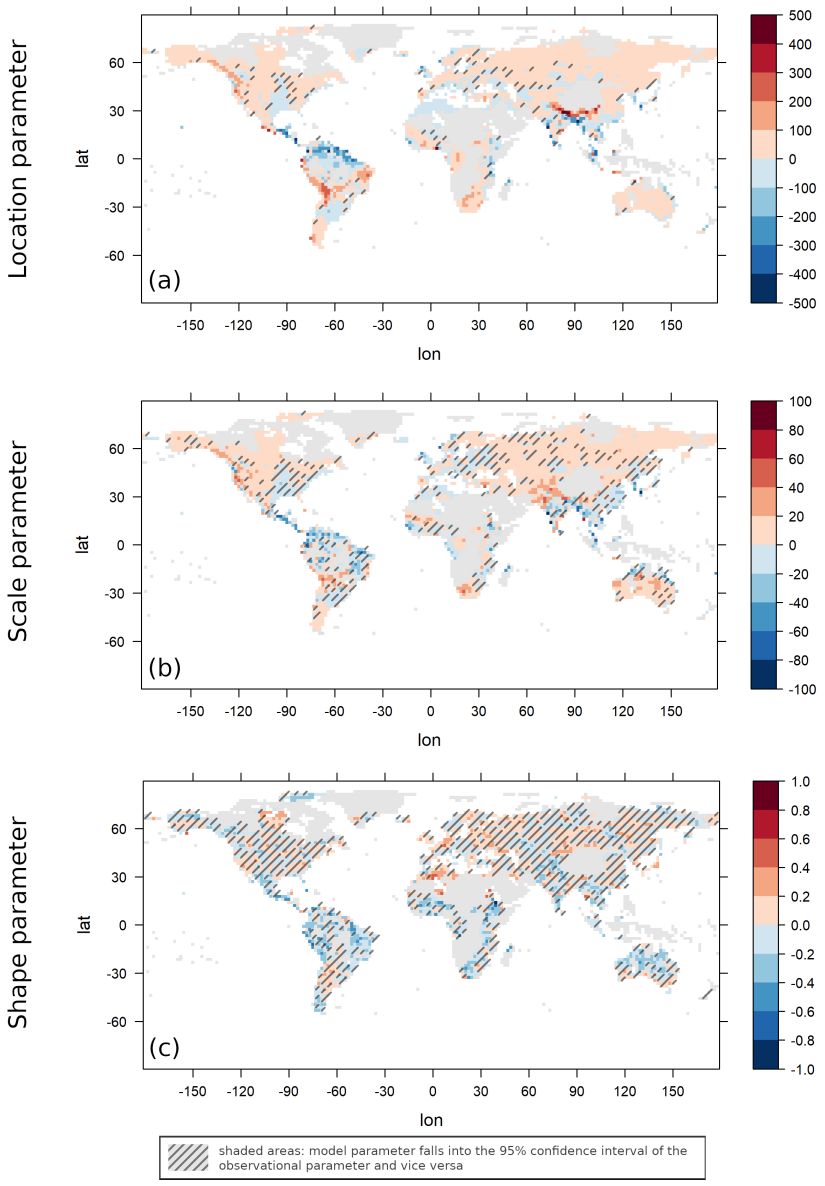

Figure 4.5: Difference between AWI-ESM model and observational GEV parameter estimates: Location parameter (a), scale parameter (b) and shape parameter (c). Values exceeding the scale limits are truncated. Units are mm/month.

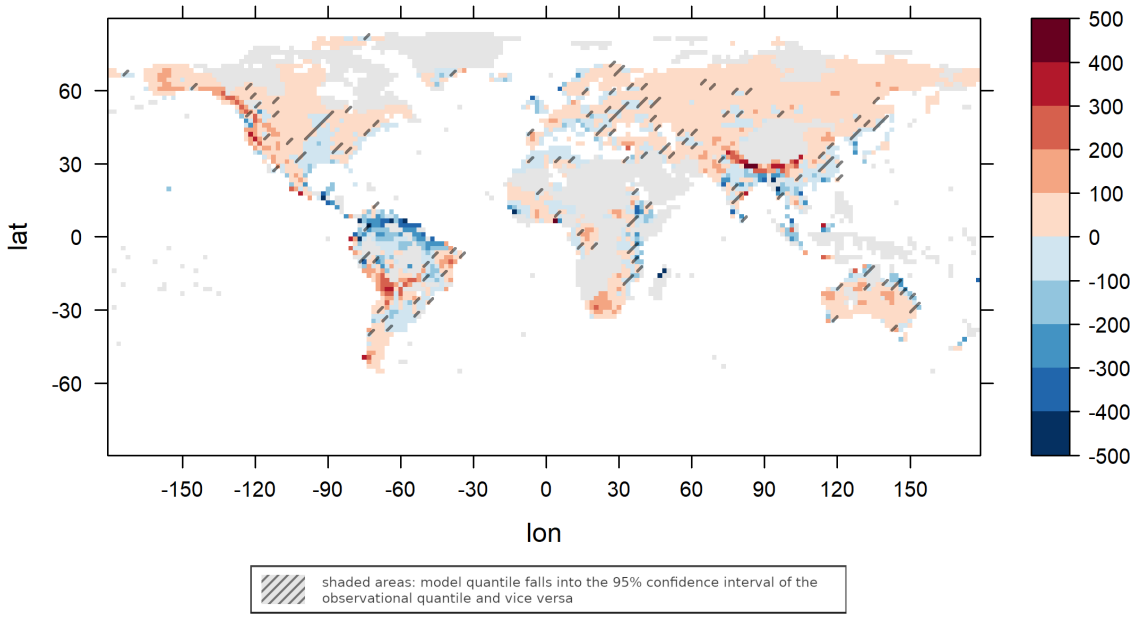

Figure 4.6: Difference of the 0.95-quantiles of the estimated GEV distribution for AWI-ESM model and observational data. Values exceeding the scale limits are truncated. Units are mm/month.

Table 4.1: The number of clusters for AWI-ESM climate model and observational data determined with the L-Method (above the middle line) and the threshold method (below the middle line) for different ranges/thresholds and for dissimilarity measure D_0 (top) and $D_{0.25}$ (bottom).

$\mathbf{D_0}$	AWI-ESM	Observations
$m = 250$	64	146
$m = 300$	148	148
$m = 400$	200	296
$m = 500$	234	291
$h = 0.85$	143	127
$h = 0.825$	188	177
$h = 0.8$	232	221
$h = 0.775$	280	254

$\mathbf{D_{0.25}}$	AWI-ESM	Observations
$m = 250$	187	102
$m = 300$	165	142
$m = 400$	223	140
$m = 500$	232	265
$h = 0.675$	118	109
$h = 0.65$	165	167
$h = 0.625$	219	220
$h = 0.6$	281	265

The threshold method generally predicts a similar, but in most cases slightly lower cluster number for observational data than for climate model data. In Fig. 4.7, the clusters for both data sets are depicted using the threshold method for dissimilarity measure D_0 with threshold $h = 0.825$ as well as for dissimilarity measure $D_{0.25}$ with threshold $h = 0.65$.

To exemplify the differences and similarities in the clusterings, we have a closer look at Europe in the D_0-clusterings. In the model data, there is one cluster covering western Spain and Portugal, one cluster covering eastern Spain, and one cluster consisting of southern France and Italy. Great Britain and Denmark are in the same cluster, the northern parts of France together with Belgium and the Netherlands in another one. One cluster covers Germany and Switzerland, and in Eastern Europe we see several clusters covering larger areas in the longitudinal direction, for example one cluster over Poland, one over Ukraine, and one over Turkey and Greece. The clusters in the observational dataset show a slightly different picture: Here, the whole Iberian Peninsula is in one cluster, and one large cluster extends over northern France, Belgium, the Netherlands and Germany to the

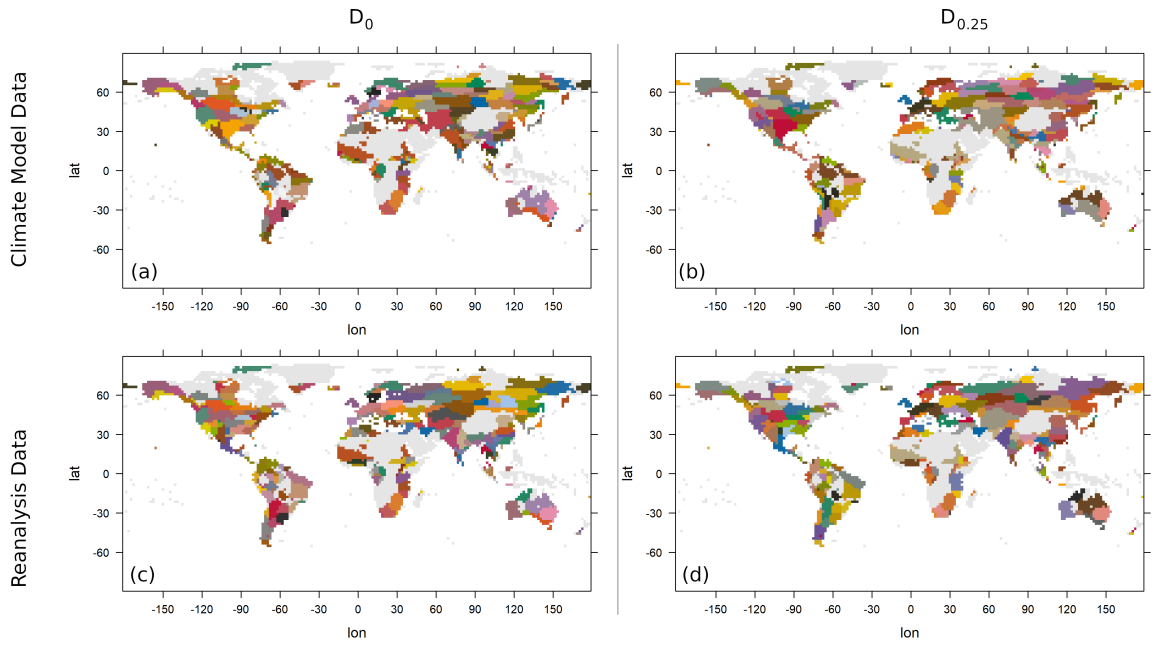

Figure 4.7: Clustering of AWI-ESM model data (a, b) and observational data (c, d) with the dissimilarity measure D_0 and threshold $h = 0.825$ (a, c) and with dissimilarity measure $D_{0.25}$ and threshold $h = 0.65$ (b, d).

western parts of Poland. On the other hand, Great Britain and Denmark are now in two separate clusters. Regarding other parts of the world, it is worth noting that in all four clusterings a large cluster cluster covering the Sahara (or at least all parts of it for which there are observations available) can be identified. There are no clusters extending over two regions that are very far apart from each other, and in general clusters tend to cover more area in the longitudinal direction than in the latitudinal one.

For the AWI-ESM, we calculated an AWQD of 52.98, making it the third-best of all 27 CMIP6 models analyzed. The AWQDs of the other 26 CMIP6 models analyzed are given in Table 4.4. In Fig. 4.8, the AWQDs are plotted against the model resolution (the total number of model grid points in units of 10^4). A linear regression (red line; intercept: 73.310, slope: -2.368) indicates that models with a higher resolution have a tendency to describe extremal precipitation better. A test on the significance of the slope parameter (null hypothesis of the slope parameter being equal to zero) was significant at the 5% level with a p-value of 0.0357.

Table 4.2 . Analyzed CMIP6 models and their spatial resolutions and AWQDs.

Model name	**Reference to data set**	**Resolution**	**AWQD**
ACCESS-CM2	Dix et al. (2019)	192×144	**76.73**
ACCESS-ESM1.5	Ziehn et al. (2019)	192×143	**85.86**
AWI-CM1.1MR	Semmler et al. (2018)	384×192	**54.70**
AWI-ESM1.1LR	Danek et al. (2020)	192×96	**52.98**
BCC-CSM2-MR	Wu et al. (2018)	320×160	**83.00**
CAMS-CSM1.0	Rong (2019)	320×160	**54.35**
CanESM5	Swart et al. (2019)	128×64	**88.72**
CESM2	Danabasoglu (2019a)	288×90	**63.27**
CESM2-FV2	Danabasoglu (2019b)	144×96	**62.99**
CESM2-WACCM-FV2	Danabasoglu (2019c)	144×96	**61.51**
CNRM-CM6-1	Voldoire (2018)	256×128	**70.57**
EC-Earth3-Veg-LR	EC-Earth Consortium (2020)	320×160	**44.71**
FGOALS-f3-L	Yu (2019)	288×180	**60.86**
FGOALS-g3	Li (2019)	180×80	**83.08**
GISS-E2.1G	NASA/GISS (2018)	144×90	**69.92**
HadGEM3-GC31-LL	Ridley et al. (2019)	192×144	**75.83**
INM-CM4.8	Volodin et al. (2019b)	180×120	**67.25**
INM-CM5.0	Volodin et al. (2019a)	180×120	**74.32**
IPSL-CM6A-LR	Boucher et al. (2018)	144×143	**78.26**
MIROC-ES2L	Hajima et al. (2019)	128×64	**59.95**
MPI-ESM1.2-HR	Jungclaus et al. (2019)	384×192	**54.16**

Continued on next page

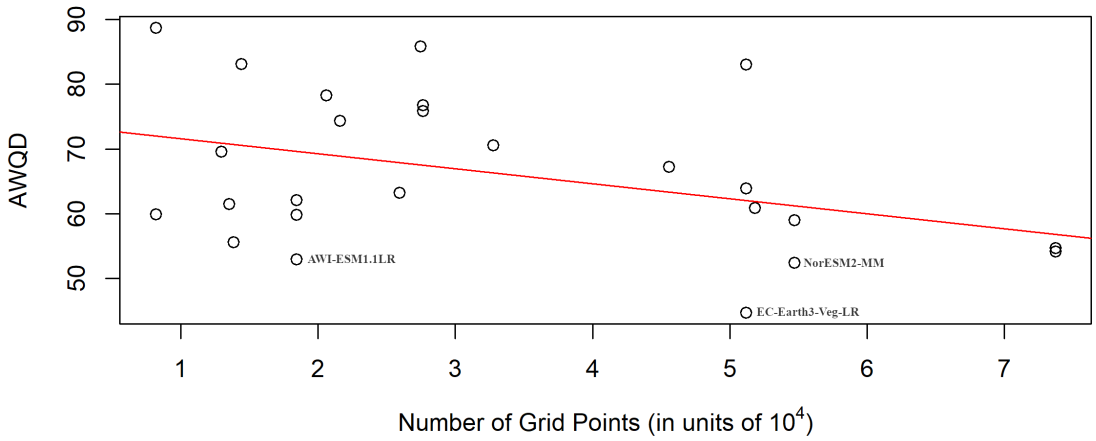

Figure 4.8: The Average Weighted Quantile Difference (AWQD) of the 27 CMIP6 models considered plotted against the model resolution (number of model grid points in units of 10^4). In red: Linear regression line (intercept 73.310, slope -2.368).

Table 4.2 – *Continued*

Model name	**Reference to data set**	**Resolution**	**AWQD**
MPI-ESM1.2-HAM	Neubauer et al. (2019)	192×96	**59.87**
MRI-ESM2.0	Yukimoto et al. (2019)	320×160	**63.92**
NESM3	Cao and Wang (2019)	192×96	**62.12**
NorESM2-LM	Seland et al. (2019b)	144×96	**55.60**
NorESM2-MM	Seland et al. (2019a)	288×192	**52.43**
TaiESM1.0	Lee and Liang (2020)	288×192	**58.99**

The best model in terms of the AWQD is the high-resolution model EC-Earth3-Veg-LR (EC-Earth Consortium, 2020) with a value of 44.71. We will now discuss results for this model in more detail, while results for the other models can be found in the supplement. For the EC-Earth3-Veg-LR, the estimated GEV parameters and anomalies are depicted in Fig. 4.9. The differences of the 95% quantiles are depicted in Fig. 4.10. The numbers of clusters determined using the L-Method and the threshold method are found in Table 4.3 and images of clusterings are depicted in Fig. 4.11. QQ-Plots and plots of KS-Tests

are similar to the corresponding plots for the AWI-ESM. The EC-Earth3-Veg-LR model predicts climate extremes better than the AWI-ESM in the Himalayas and in the Amazon region (compare Fig. 4.6 to Fig. 4.10), while it overestimates precipitation extremes more strongly than the AWI-ESM at the western coast of South America. The number of clusters is in general higher than for the AWI-ESM, in part probably due to the higher model resolution (320×160 compared to 192×96). Note that this increased resolution is also the reason for the different values for the cluster numbers of the reanalysis data in Tables 4.1 and 4.3, because reanalysis data were in each case interpolated to the climate model resolution. When comparing again the clusters over Europe using the D_0 dissimilarity measure, it can be observed that in the western part of Europe, model and observational clusters are in general similar, with only slight differences over the Iberian Peninsula and with an area covering southern France and northern Italy that is in one cluster in the model data and in two different clusters in the observational data. In Eastern Europe and Scandinavia, the differences between the clusterings are larger and it is more difficult to see correspondences. The general remarks that have been made about the clusterings while discussing the AWI-ESM data also apply here.

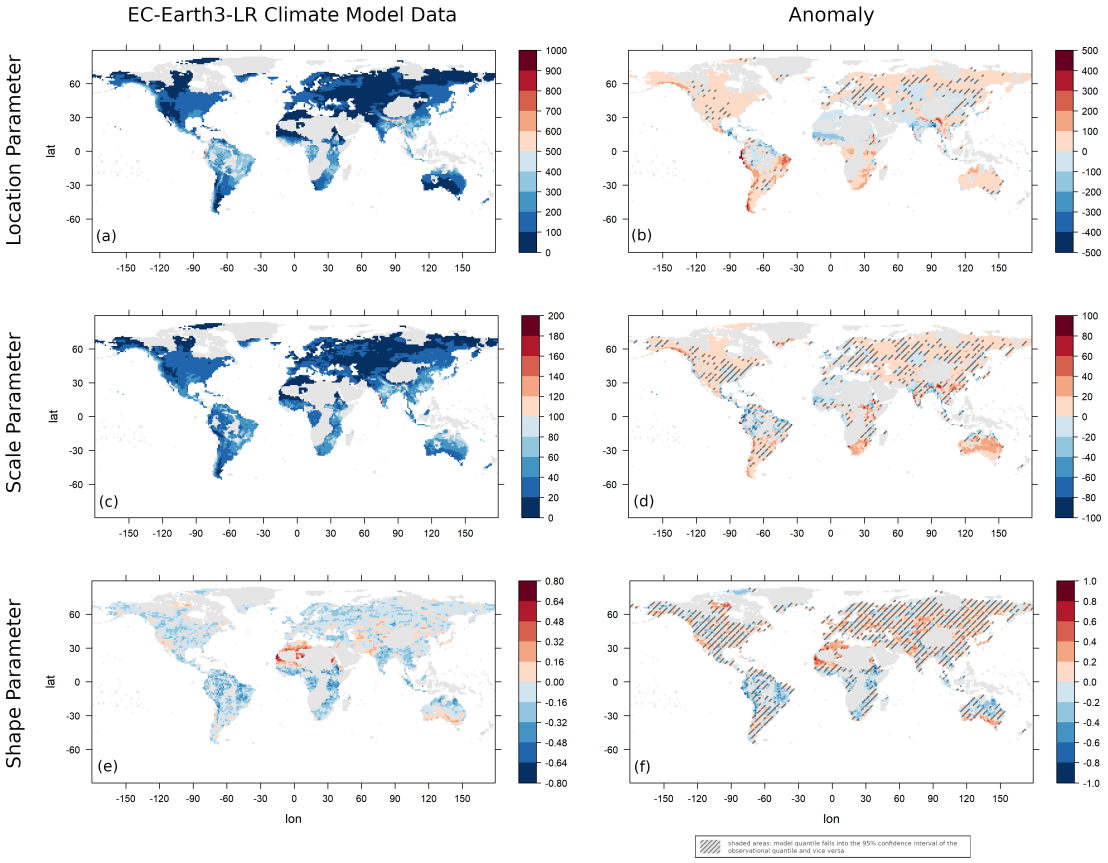

Figure 4.9: EC-Earth3-Veg-LR climate model estimated GEV parameters (a, c, e) and their anomaly compared to the reanalysis GEV parameters (b, d, f). The GEV parameters are location (a, b), scale (c, d) and shape (e, f). Values exceeding the scale limits are truncated. Units are mm/month.

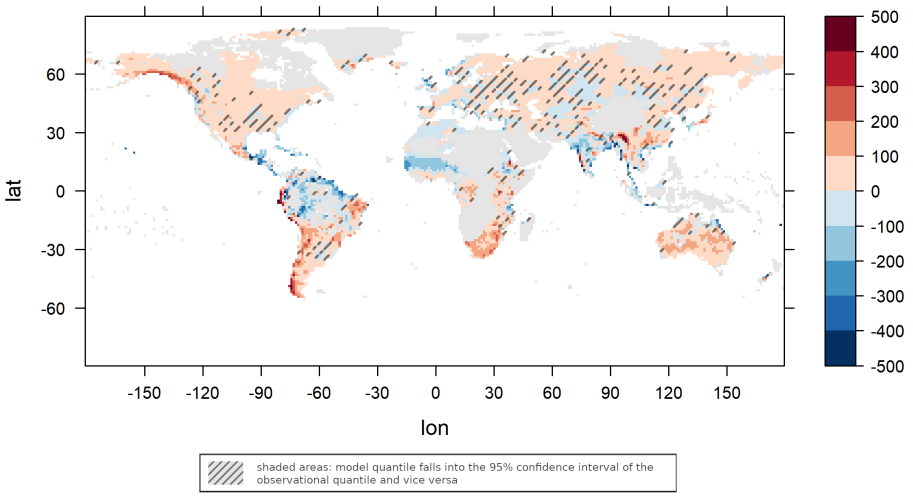

Figure 4.10: Difference of the 0.95-quantiles of the estimated GEV distribution for EC-Earth3-Veg-LR model and observational data. Values exceeding the scale limits are truncated. Units are mm/month.

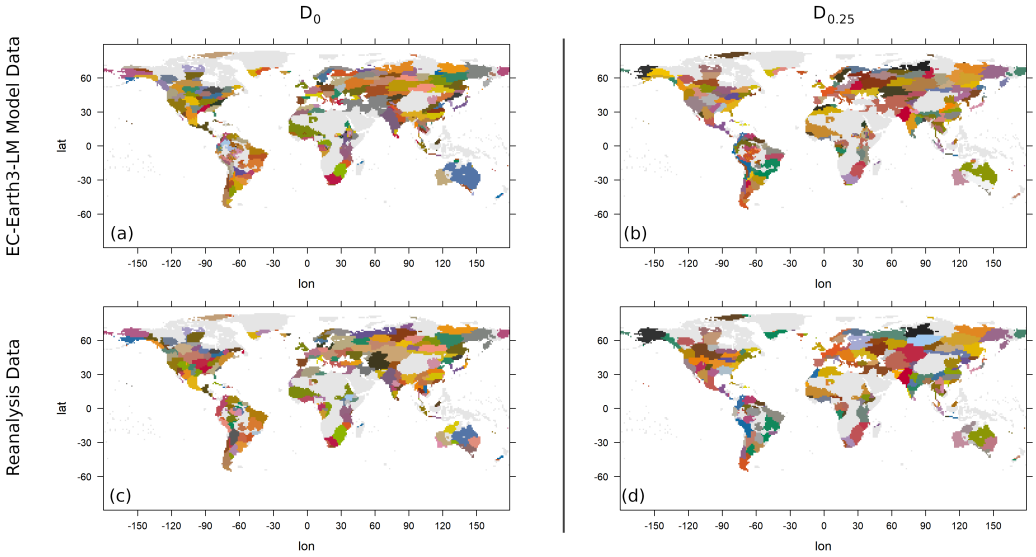

Figure 4.11: Clustering of EC-Earth3-Veg-LR model data (a, b) and observational data (c, d) with the dissimilarity measure D_0 and threshold $h = 0.825$ (a, c) and with dissimilarity measure $D_{0.25}$ and threshold $h = 0.65$ (b, d).

Table 4.3: The number of clusters for EC-Earth3-Veg-LR climate model and observational data determined with the L-Method (above the middle line) and the threshold method (below the middle line) for different ranges/thresholds and for dissimilarity measure D_0 (top) and $D_{0.25}$ (bottom).

$\mathbf{D_0}$	EC-Earth3-Veg-LR	Observations
$m = 250$	76	89
$m = 300$	141	90
$m = 400$	181	94
$m = 500$	184	272
$h = 0.85$	173	145
$h = 0.825$	224	186
$h = 0.8$	299	240
$h = 0.775$	366	272

$\mathbf{D_{0.25}}$	EC-Earth3-Veg-LR	Observations
$m = 250$	113	67
$m = 300$	117	67
$m = 400$	129	154
$m = 500$	146	282
$h = 0.675$	131	116
$h = 0.65$	203	166
$h = 0.625$	276	225
$h = 0.6$	358	279

4.5 Conclusions of this chapter and discussion

We presented approaches and methods to validate climate model outputs by comparing their extremal behavior to the extremal behavior of observational data. To illustrate these methods, we compared precipitation extremes between the AWI-ESM and the CRU TS4.04 data set of reanalyzed observations. After an analysis of empirical statistical parameters, we fitted GEV distributions to the data and analyzed the differences in estimated parameters. Then we continued with an analysis of spatial concurrence of extremes based on a hierarchical clustering approach and a dissimilarity measure derived from bivariate copula theory. While the empirical statistics are similar for many parts of the world, we can also identify larger regions of over- and underestimation of empirical means and standard deviations by the climate model. These misestimations often go hand in hand with a similar misestimation of the standard deviation (heteroscedasticity), although for the standard deviation a stronger tendency for underestimation can be observed. Misestimations of mean and standard deviations translate into a misestimation of extreme values, and this

can be confirmed by the comparison of the fitted GEV distribution parameters and the 0.95-quantiles derived from them. The shape parameter, indicative of the heavy-tailedness of the distribution, is in general similar between model and observational data, but because of the difficulties in reliably estimating this parameter from data (that are in turn a result of the rareness of extreme events in the data) these results have to be taken with caution.

The cluster analysis based on spatial dependencies and the occurrence of concurrent extremes shows that there is generally a good agreement between identified clusters. Also the number of clusters is in general similar, with a slight tendency for a higher cluster number in the model data. Since it is mostly large-scale weather events and teleconnections contributing to concurrent climate extremes, this may indicate that the basic physical behavior underlying them is in general well captured by the AWI-ESM. Further analyses can be conducted to investigate in detail the reasons for different clusterings over selected regions.

In addition to the AWI-ESM, several other CMIP6 models are also analyzed. A comparison of the model accuracy, measured using an averaged quantile difference, shows a tendency for higher-dimensional models to capture extremal behavior better.

In addition to model validation, the definition of regions with concurrent extremes may turn out useful for assessments of risks in an economical context and for insurance. It needs to be noted, though, that extremes in climate models and in gridded reanalysis data sets tend to be damped because of the spatial averaging performed during the creation of the data (Bador et al., 2020). Another possible field of application is palaeoclimatology. The spatial distribution of precipitation extremes is known to have changed markedly in the past (Lohmann et al., 2020; Ionita et al., 2021), and clustering based on climate models could be used to generalize the sparse existing palaeoclimatic data to larger regions.

Chapter 5

Long-term evolution of temperature extremes

5.1 Introduction to the chapter

In this chapter, we use climate model data to investigate future changes in temperature extremes on a global level. While a rising trend in frequency and magnitude of temperature extremes is currently observed in many regions of the world (Rahmstorf and Coumou, 2011; Lorenz et al., 2019; Seneviratne et al., 2021), these are not taking place uniformly around the globe, but instead they show a strong dependency on the geographic location and its climatic conditions (McBride et al., 2022; Twardosz et al., 2021; Trenberth et al., 2007). However, the focus of most studies is on regional investigations of the development of temperature extremes. In addition, the evolution of the extremes in future decades and centuries is less well understood, as mostly only the near future is investigated (Rummukainen, 2013).

Changes in the expected frequency of extreme events can be caused by changes in various statistical parameters, like the mean and the variance (Parey et al., 2013). In addition to that, starting time and duration of changes can also vary in different regions. In this chapter, we develop statistical models to investigate changes in temperature extremes in a warming climate on a global scale and for a period of investigation spanning several centuries. In order to gain insights into future changes of extremes, four Earth system models will be analyzed with respect to daily temperatures from historical and future simulations ranging from 1850 to 2300.

It is expected that the rate of change of extremes will increase in the near future (Smith, 1990). Under the premise that mankind will be able to slow and ultimately end the increase of atmospheric CO_2 emissions someday, it can be expected that in consequence, changes in extreme temperatures will gradually slow down as the climate system will be tending toward a new equilibrium state (King et al., 2021), although it may still take centuries for a new stationary state to be completely reached due to slow-changing components of the climate system (Hansen et al., 2005). Taking these considerations together, we can expect changes in extreme temperature to follow in general a slow–fast–slow pattern over time. To describe a transition from an initial value to a final one that starts slowly, then speeds up and finally decelerates again when approaching the new value, it is common practice to use a logistic function, which exhibits a characteristic S-shaped form. The first application of logistic functions in modeling is due to Verhulst, who designed a logistic growth model to describe the development of biological populations in 1845 (Verhulst, 1845). The motivation in the ecological context is that the population growth is slow at the beginning (limited by the small population size) as well as at the end (limited by the lack of natural resources). The logistic growth model has been successfully applied in biology and epidemiology—a recent example being its application to the spreading of the coronavirus disease 2019 (Shen, 2020)—and this has motivated its use as a general model to describe changes from one state to another in fields as varied as linguistics (Altmann, 1983), medicine (Yano et al., 1998) or economics (Kwasnicki, 2013).

To model extremes in a changing climate, we will use non-stationary GEV distributions with time-dependent distribution parameters. The changes in the distribution parameters will be described using logistic functions. After fitting the statistical models to the data, we will analyze the estimated distribution parameters in detail, and we will use the estimates also to investigate future changes in the distribution quantiles.

Changes in the expected frequency of extreme events can be caused by changes in the mean values of the GEV distributions, changes in their variability, changes in their heavy-tailedness or by a combination of these factors (Katz and Brown, 1992; Cooney, 2012; Lewis and King, 2017). The application of non-stationary GEV distributions enables us to investigate which factors contribute to what extent at different geographic locations. In addition, we will investigate whether changes in the different distribution parameters occur simultaneously or if changes in some parameters precede changes in others.

Several non-stationary models based on GEV distributions have been proposed to describe the influence of climate change on climate extremes: In Panagoulia et al. (2014), a GEV distribution with the parameters polynomially depending on time was proposed and its application was showcased using precipitation data from Greece. In a similar way, in Sarhadi and Soulis (2017), non-stationary models with different degrees of freedom were constructed and evaluated using Bayesian inference and Markov chain Monte Carlo techniques. In Tian et al. (2020), an idea first proposed in Cannon (2010) was extended and neural networks were used to choose between a variety of non-stationary models with different covariates that can interact with each other. The approach of combining GEV distributions with logistic functions gives us the possibility to investigate developments in extreme temperature over a time span of several centuries and on a global level and to research how changes in extreme temperatures will unfold in different regions.

The rest of this chapter is organized as follows: In Section 5.2, the temperature data sets are presented. We discuss the logistic models and the model-fitting algorithm we use in Section 5.3. Before presenting the results of applying the logistic models to the data in Section 5.5, a simulation study is conducted to investigate the accuracy of the model fitting algorithm in Section 5.4. A discussion (Section 5.6) follows, and a section on conclusions and an outlook finalize the chapter.

5.2 Climate data analyzed in this chapter

We investigate daily temperature data at two meters above surface from four global earth system models. For each earth system model, the data consist of a simulation of the historical climate from 1850 to 2005 and a future simulation from 2005 to 2300 that follows the representative concentration pathway RCP8.5 of the Intergovernmental Panel on Climate Change IPCC (Riahi et al., 2007). The RCP8.5 scenario provides atmospheric CO_2 values until the year 2100. For the years after 2100, the climate model runs with prescribed CO_2 values that are set to the value of the year 2100, see Fig 5.1a. The four Earth system models used are the model bcc-csm1-1 from the Beijing Climate Center (Wu et al., 2014, in the following: "BCC"), the model CCSM4 from the National Center for Atmospheric Research NCAR (Gent et al., 2011, "CCSM4"), the CSIRO-Mk3-6-0 (Jeffrey et al., 2013, "CSIRO"), and the MPI-ESM-LR from the Max Planck Institute for Meteorology in Hamburg, Germany (Giorgetta et al., 2013, "MPI-ESM"). All models

take part in the Climate Model Intercomparison Project CMIP5 (Taylor et al., 2012). In the plots, the coastline boundaries have been obtained from Natural Earth Version 4.2.0 (http://www.naturalearthdata.com/).

In Fig 5.1b, the evolutions of the annual global mean temperature that are predicted by the four Earth system models are displayed. They roughly follow an S-shaped form for each model, but differ strongly among the different climate models in terms of timing and magnitude of the changes.

5.3 Statistical models for changes in extremes over time

To model the effects of changes in the climate, the GEV distributions we use need to have time-dependent distribution parameters. Due to the reasoning laid out in the introduction, we choose logistic functions to describe the change of the GEV parameters over time. The logistic function we use as the basis for our models is given by

$$f(x) = \frac{1}{1 + \exp(-2 \cdot \log(19) \cdot x)}. \tag{5.1}$$

It describes a growth limited by 0 for $x \to -\infty$ and by 1 for $x \to \infty$ with the highest growth rate at $x = 0$. The constant $2 \cdot \log(19)$ in the exponential function is used for better interpretability of the parameters of the models we will present below, it ensures that 90% of the change from 0 to 1 takes place in the interval $[-\frac{1}{2}, \frac{1}{2}]$. We use the function in our models in the following way: For each of the three GEV parameters $p \in \{\mu, \sigma, \gamma\}$ we describe its temporal development as

$$\hat{p}(t) = p_s + p_c \cdot f\Big(\frac{t-a}{b}\Big). \tag{5.2}$$

The model parameter p_s describes the "initial state" and p_c describes the total magnitude of the change. The model parameters a and b control the timing of the change. Parameter a indicates the time point at which the growth rate is highest (which is also the time point at which exactly half of the change from p_s to $p_s + p_c$ is completed) and parameter b indicates the approximate duration of the change (in the sense that 90% of the total change takes place in the time span $[a-\frac{b}{2}, a+\frac{b}{2}]$). See also Fig 5.2 for a visualization.

This leads to the following model:

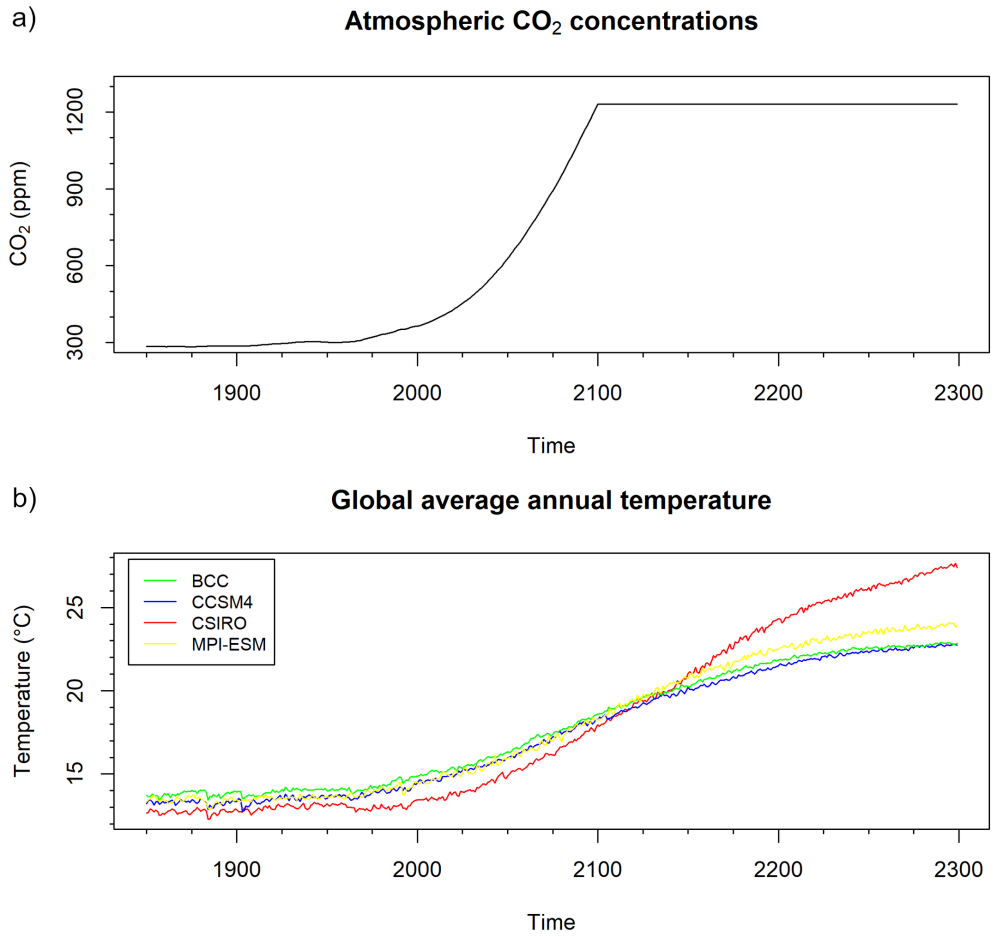

Figure 5.1: **Atmospheric CO_2 concentration and global annual mean temperature.** Panel a: The atmospheric CO_2 concentration (in ppm) that was used for the model runs. The CO_2 concentration follows the RCP8.5 scenario (Riahi et al. (2007)) until 2100 and is kept constant afterwards. Panel b: The annual global mean temperature (in °C) according to the climate model runs.

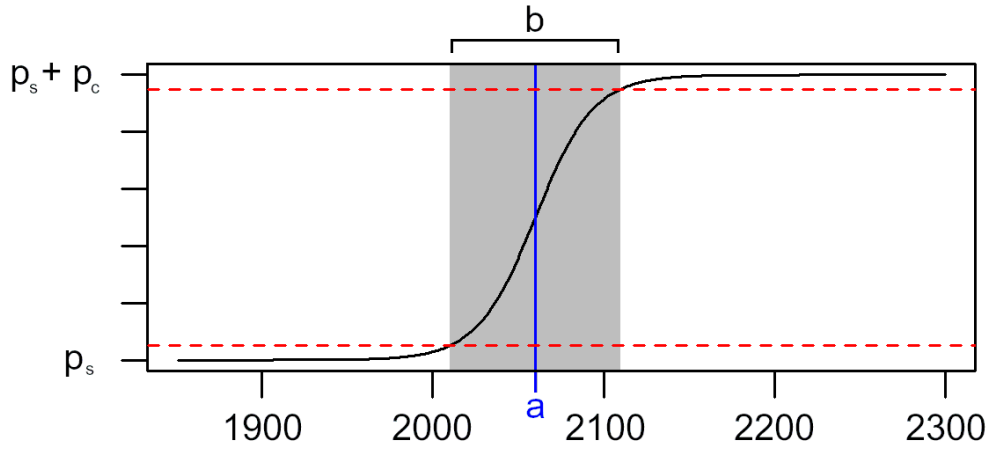

Figure 5.2: **Visualization of the parameter values of the logistic models.** A sigmoidal curve following Eq (5.2) with parameters $a = 2060$ and $b = 100$ is displayed. Parameter a corresponds to the time point at which half of the transition from p_s to $p_s + p_c$ is completed. Ninety percent of this transition take place within the interval $[a - \frac{b}{2}, a + \frac{b}{2}]$, so parameter b describes the approximate time span of the transition.

Model 1a. The three GEV parameters location, scale and shape are described using a logistic curve, using for each parameter a different initial value and amount of change. The parameters a and b are the same for location, scale and shape.

$$\begin{aligned}\hat{\mu}(t) &= \mu_s + \mu_c \cdot f\left(\tfrac{t-a}{b}\right)\\ \hat{\sigma}(t) &= \sigma_s + \sigma_c \cdot f\left(\tfrac{t-a}{b}\right)\\ \hat{\gamma}(t) &= \gamma_s + \gamma_c \cdot f\left(\tfrac{t-a}{b}\right)\end{aligned}$$

As pointed out in Schär et al. (2004), the evolution of extreme events may be different from that of mean and variance (which may show different behaviors among themselves). It may therefore be necessary to allow for changes in location, scale and shape to take place at different times and over different durations. This leads to the following more complex model:

Model 1b. This model is the same as Model 1a, but with individual parameters a_μ, b_μ,

a_σ, b_σ and a_γ, b_γ being used for location, shape and scale of the GEV distribution.

$$\begin{aligned}\hat{\mu}(t) &= \mu_s + \mu_c \cdot f\Big(\tfrac{t-a_\mu}{b_\mu}\Big)\\ \hat{\sigma}(t) &= \sigma_s + \sigma_c \cdot f\Big(\tfrac{t-a_\sigma}{b_\sigma}\Big)\\ \hat{\gamma}(t) &= \gamma_s + \gamma_c \cdot f\Big(\tfrac{t-a_\gamma}{b_\gamma}\Big)\end{aligned}$$

When applying non-stationary GEV distributions, it is often assumed that the only time-dependent parameters are location and scale, while the shape parameters is kept constant (Nogaj et al., 2007; Panagoulia et al., 2014). This approach leads us to a second type of model:

Model 2a. The GEV parameters location and scale are described using a logistic curve, using for each parameter a different initial value and amount of change. The parameters a and b are the same for location and scale. The shape parameter is kept constant over the whole time interval.

$$\begin{aligned}\hat{\mu}(t) &= \mu_s + \mu_c \cdot f\Big(\tfrac{t-a}{b}\Big)\\ \hat{\sigma}(t) &= \sigma_s + \sigma_c \cdot f\Big(\tfrac{t-a}{b}\Big)\\ \hat{\gamma}(t) &= \gamma_{const}\end{aligned}$$

Model 2b. This model is the same as Model 2a, but with individual parameters a_μ, b_μ, a_σ, b_σ being used for location and scale of the GEV distribution.

$$\begin{aligned}\hat{\mu}(t) &= \mu_s + \mu_c \cdot f\Big(\tfrac{t-a_\mu}{b_\mu}\Big)\\ \hat{\sigma}(t) &= \sigma_s + \sigma_c \cdot f\Big(\tfrac{t-a_\sigma}{b_\sigma}\Big)\\ \hat{\gamma}(t) &= \gamma_{const}\end{aligned}$$

The logistic function, as used in the models above, has the limitation that the inflection point (the point of the strongest growth) is exactly in the middle of the curve, having always a value of $p_s + \frac{1}{2}p_c$. To allow for more flexibility, a generalized function that was proposed in Richards (1959) can be used. For $\beta > 0$, the Richards function is defined as

$$g_\beta(x) = \Big(1 + (2^\beta - 1) \cdot \exp\Big(-\log\Big(\frac{0.95^{-\beta} - 1}{0.05^{-\beta} - 1}\Big) \cdot x\Big)\Big)^{-\frac{1}{\beta}}. \tag{5.3}$$

We use it to describe the time-changing GEV parameters $p \in \{\mu, \sigma, \gamma\}$:

$$\hat{p}(t) = p_s + p_c \cdot g_\beta\Big(\frac{t-a}{b}\Big). \tag{5.4}$$

The interpretation of the parameters p_s and p_c remains unchanged. The parameter a describes, as before, the time point at which the model attains the midpoint of the change (the value $p_s + \frac{1}{2}p_c$). In the previous models, this was also the point of the highest growth rate, while here, the inflection point depends on the value of the parameter β. For $\beta = 1$, the model reduces to the previous model (g_1 is equal to f), while the inflection occurs at a later time point than a for $\beta > 1$ and at an earlier time point for $\beta < 1$. The parameter $b > 0$ controls the velocity of the change in such a way that the change from $p_s + \frac{1}{20}p_c$ to $p_s + \frac{19}{20}p_c$ (90% of the total amount of change) takes place in an interval of length b. Because of the asymmetry of the function g_β for $\beta \neq 1$, this interval is no longer $[a - \frac{b}{2}, a + \frac{b}{2}]$, but shifted to the left for $\beta > 1$ and to the right for $\beta < 1$. In Fig 5.3, plots of the model function for different parameter values are depicted.

Using the Richards function g_β instead of f in the previous models gives us four models that we denote by adding the letter R to the model name. Compared to the models using the function f, Model 1aR and 2aR have one additional model parameter β, while Model 1bR and 2bR feature additional model parameters for the non-constant GEV parameters β_μ, β_σ and (only Model 1bR) β_γ.

Non-stationary GEV distributions can be fitted to data using Maximum Likelihood Estimators, see Mudelsee (2014), Chapter 6.3 and El Adlouni et al. (2007). In the numerical optimization, the fitting algorithm L-BGFS-B is used. For this purpose, the models are reparametrized to no longer use the parameters p_c for $p \in \{\mu, \sigma, \gamma\}$ describing the magnitude of change, but parameters $p_e := p_s + p_c$ describing the values after the change instead. This makes it possible to ensure in an easy way that all values of $\sigma(t)$ are positive (using the condition $\sigma_e > 0$ instead of the equivalent $-\sigma_c < -\sigma_s$). To determine suitable starting values for the parameters μ_s and σ_s, a stationary GEV distribution is fitted to the first quarter of the data of the time series investigated, yielding estimates $\hat{\mu}$ and $\hat{\sigma}$, and starting values are selected randomly from the intervals $[\hat{\mu} - 5, \hat{\mu} + 5]$ and $[\max(0, \hat{\sigma} - 5), \hat{\sigma} + 5]$. The same is done for the parameters μ_e and σ_e using the last quarter of the time series data. Since the estimation of the shape parameter is not very reliable for small samples, starting values for γ_s, γ_e or γ_{const} are not determined that way, but chosen randomly from the interval $[-1, 1]$. Random selection from an interval is also done for all other

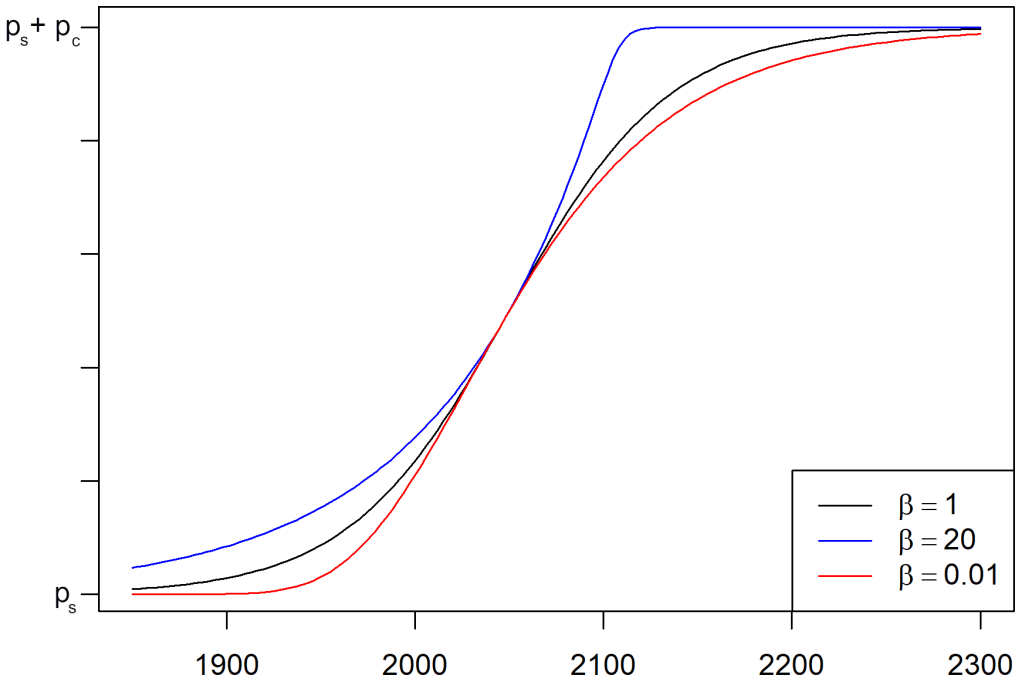

Figure 5.3: **Visualization of the parameter β of the Richards function.** The plot shows Richards functions g_β for different values of β. The Richards function for $\beta = 1$ is identical to the logistic function f. The point of the highest growth rate is shifted to the right for $\beta > 1$ and to the left for $\beta < 1$. For all lines depicted, the other parameters used are $a = 2050$, $b = 100$.

model parameters using suitable, large intervals to select values from. The stationary GEV distributions are fitted using the R package "EnvStats" (Millard, 2013). Part of our R-code is based on work by Takahito Mitsui in the context of Mitsui and Crucifix (2017). The optimization algorithm is run several times with different starting values in order to find a global maximum of the likelihood function.

To choose the best model out of the different models presented here, we apply the Bayesian Information Criterion (BIC; Schwarz, 1978). To test the goodness-of-fit of the models, note that a $\mathrm{GEV}(\mu, \sigma, \gamma)$-distributed random variable can be transformed to a $\mathrm{GEV}(1, 1, 1)$ distribution (a so-called unit Fréchet distribution) by applying the transformation

$$G_{\mu,\sigma,\gamma}(z) = \max\left(0, 1 - \gamma \cdot \left(\frac{z-\mu}{\sigma}\right)\right)^{-\frac{1}{\gamma}}. \tag{5.5}$$

By applying $G_{\hat{\mu},\hat{\sigma},\hat{\gamma}}(z)$ with the (time-dependent) estimated model parameters to the data, we obtain for each grid point a time series that is unit Fréchet distributed if the model assumptions are true. We test the hypothesis of the transformed data being unit Fréchet distributed using a one-sample Kolmogorov-Smirnov test (Stephens, 1970).

5.4 A simulation study to investigate the statistical models

Before applying the statistical models to climate data, we first test how accurately model parameters can be estimated under ideal conditions. For each of the eight models presented above, we prescribe values for the model parameters, simulate data following the corresponding non-stationary GEV distribution, and fit the model to the data. We then compare the estimated model parameters with the true ones.

In addition to that, we test how susceptible the models that use the logistic function are to model misspecification. To this end, we simulate data from the models as above, but replacing the function f from Eq (5.1) with the following three functions of a similar sigmoidal shape that are known for example as activation functions for neural networks (Menon et al., 1996; Bagul and Chesneau, 2021):

$$g_1(x) = \tfrac{1}{2} + \tfrac{1}{\pi}\arctan\left(\tfrac{\pi}{4}x\right) \tag{5.6}$$

$$g_2(x) = \tfrac{1}{2} + \frac{x}{4\sqrt{1+\frac{x^2}{4}}} \tag{5.7}$$

$$g_3(x) = \tfrac{1}{2} + \tfrac{1}{2}\mathrm{erf}\left(\tfrac{x\sqrt{\pi}}{4}\right). \tag{5.8}$$

The simulated data follow a non-stationary GEV distribution with time-dependent distribution parameters μ_t, σ_t, γ_t. We then calculate estimates $\hat{\mu}_t$, $\hat{\sigma}_t$, $\hat{\gamma}_t$ by fitting the original statistical model (using function f) to the data, and we calculate the time-integrated squared difference of given and estimated GEV parameters

$$\int_{t\in T}(p_t - \hat{p}_t)^2 \mathrm{d}t \tag{5.9}$$

for the three GEV parameters $p \in \{\mu, \sigma, \gamma\}$.

We simulate 1000 time series of length 150 for each logistic model using the parameters $\mu_s = 20$, $\mu_c = 10$, $\sigma_s = 2, \sigma_c = 1$. The parameters for the shape parameter are $\gamma_s = 0.1$ and

$\gamma_c = 0.1$ (models with a varying shape parameter) or $\gamma_{const} = 0.1$ (models with a constant shape parameter). For the models describing a simultaneous change in all parameters we use the parameters $a = 2075$, $b = 30$, otherwise we use $a_\mu = 2050$, $b_\mu = 30$, $a_\sigma = 2075$, $b_\sigma = 30$ and, if applicable, $a_\gamma = 2100$ and $b_\gamma = 30$. The models based on the Richards function instead of the logistic function additionally have a parameter β (or parameters β_μ, β_σ, β_γ, respectively) equal to 5.

It turns out that for each parameter, the estimation quality is similar for all models in which the parameter occurs. In particular, the estimation is not more inaccurate for the more complex models with a higher number of parameters. For each parameter, boxplots of the estimates are depicted in Fig 5.4. Since the estimates are similar for each model, only the boxplot for one model per parameter is depicted. The boxplots indicate that the start and change values of the GEV parameters are in general well estimated, and the same is true for the parameter a and a_μ if they exist in the model. The estimates for parameters b and b_μ are in most cases close to the true value, but there are also some cases of a considerable misestimation (with a true parameter value of 30, the estimates take values of up to 120). The parameters describing a separate change in scale, a_σ and b_σ, are estimated much worse than the other ones, estimates that are far away from the true value occur regularly. In addition, parameter b_σ is in most cases underestimated, with the median of the estimates being far lower than the true value, while cases of a strong overestimation of this parameters also occur. The same can be said for the parameters a_γ and b_γ, but their estimation accuracy is even lower.

The estimation of the additional β parameters that appear in the models using the Richards function turned out to be very problematic for all models. The estimated values are usually far away from the true ones, and even the medians of the estimates are between 50 and 75 and not even close to the true parameter values of 5. A reliable estimation of the β parameter of the Richards function seems to be impossible in general using the method we employed here. Because of that, the models using the Richards function will not be considered further and only the models using the logistic function will be applied to the data. It was considered also to exclude Model 1b because of the high estimation inaccuracy of the parameters a_γ and b_γ, but for the sake of completeness, the model was kept. In the next section it will be seen that this model is rarely favored by the BIC anyway.

The results of the simulation study investigating model misspecification due to other

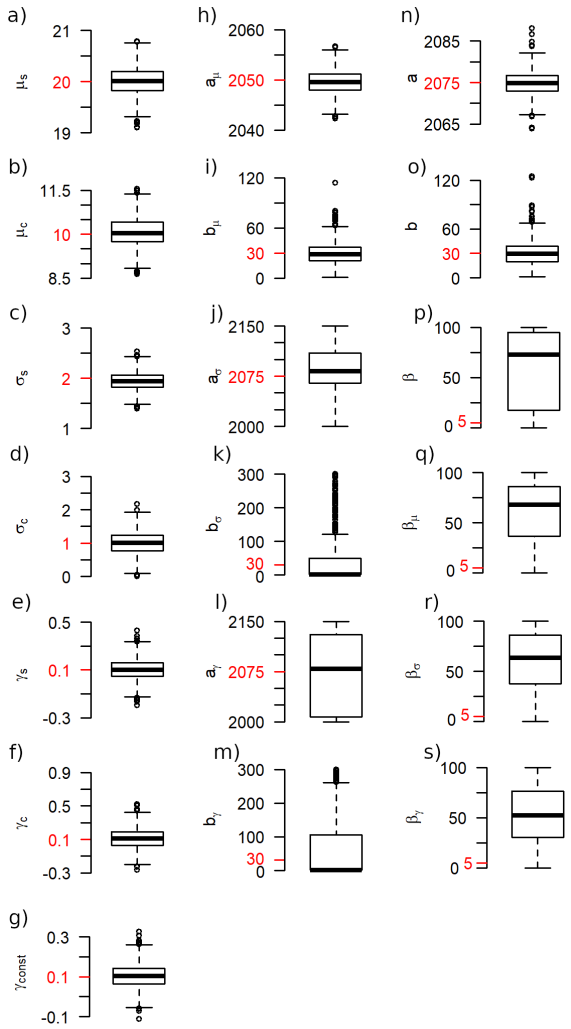

Figure 5.4: **The accuracy of the parameter estimation for simulated data.** The accuracy of the maximum likelihood estimators is investigated by applying the models to data that were generated following the respective model. For each parameter, a boxplot of the estimates is depicted, with the true parameter value indicated in red. Since the results for each parameter are very similar across the models, only one boxplot is presented per parameter. The model depicted is 1a (a-f, n, o), 1b (h-m), 2a (g), 1aR (p) and 1bR (q-s).

logistic functions than f are given in Table 5.1. Results are printed only for Model 1a, but are similar for the other logistic models. For data that were created using one of the functions g_i, the errors are similar to those using function f. Therefore, model misspecification caused by the usage of different sigmoidal functions does not have a strong negative impact on the

estimation accuracy and there is no need for using different functions than f when applying the models to the data.

Function used	Location	Scale	Shape
$f(t)$	0.232	0.057	0.008
$g_1(t)$	0.230	0.060	0.008
$g_2(t)$	0.278	0.060	0.008
$g_3(t)$	0.235	0.060	0.008

Table 5.1: **The influence of model misestimation on the estimation accuracy.** The squared difference of constructed and estimated GEV parameters is depicted for the GEV parameters location, scale and shape. The data were simulated using the sigmoidal function in the left-most column of the table while the model that was fitted to the data always uses the function $f(t)$. The model used is Model 1a, results for the other models are similar. The errors are averaged over 5000 iterations.

5.5 Application of the models to the data

As mentioned before, the only statistical models that are applied to the data are the four models based on the logistic function. We apply the models to four different climate simulations. In Fig 5.5, the best model according to the BIC is depicted at each grid point for the four data sets. It can be noted that the statistical models with a constant shape parameter (Model 2a and 2b) are often preferred over those with a varying shape parameter; one of these models is selected for at least 80% of the grid points for all data sets. There are many smaller regions in which a model with a varying shape parameter is preferred, a clear interconnection between those regions could not be identified. On the other hand, a pattern is visible regarding the question whether a model with simultaneous changes in location and scale (and, if applicable, shape) parameter is selected or not: Models with individual change parameters for the different GEV parameters are preferred almost exclusively in high-latitude regions. In particular, they are preferred throughout the whole region around the North Pole from ca. $80°$N onward for all four data sets, and for some data sets in a varying degree also in the high southern latitudes. In the other regions statistical models with a simultaneous change in the GEV parameters are predominant.

To investigate the magnitude of changes in extremes, in Fig 5.6, the difference of the 95% quantile of the fitted GEV distribution in the year 2300 and the quantile of the

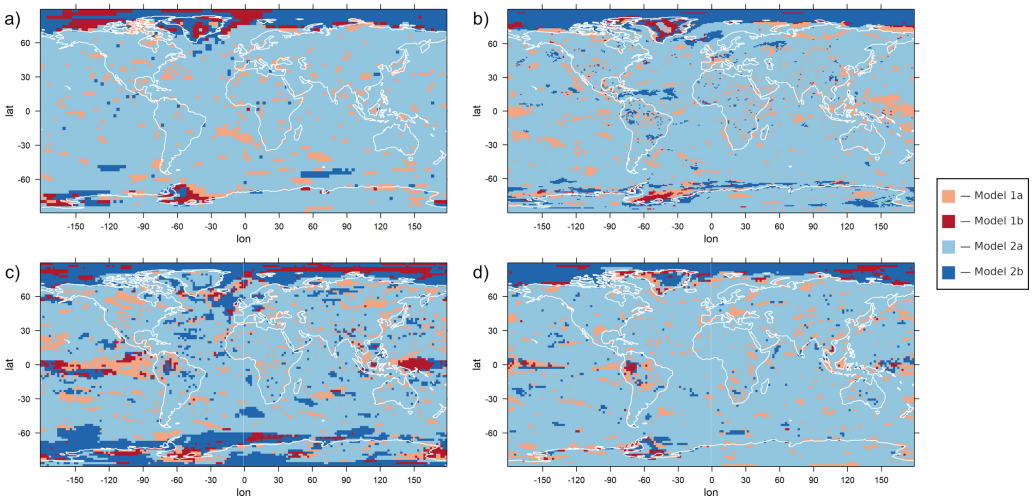

Figure 5.5: **The preferred model according to the Bayesian Information Criterion for each grid point.** The logistic models 1a, 1b, 2a and 2b are applied to the yearly maxima of daily temperature data and the BIC is used to determine the optimal one out of these for each grid point. Data set used: BCC (a), CCSM4 (b), CSIRO (c), MPI-ESM (d).

distribution in the year 1850 is depicted for each grid point and each data set. The statistical model used to calculate these values is the one that is preferred by the BIC at that grid point. While the magnitude of changes varies considerably depending on the data set, some general tendencies can be identified for all climate model outputs: The quantiles show in general an increasing trend, regions where the quantiles stay the same or decrease are an exception for all data sets. The quantile changes are higher over land than over the ocean, and in most data sets, particularly high changes can be detected in Europe, North America and parts of Siberia. Compared to the changes in other land regions of the world, Greenland shows an unusually small increase, in some models even partially a decrease.

For the now following investigation of the individual parameters of the statistical models, we will focus on model CCSM4 (depicted in panel b of Figs 5.5 and 5.6), which has the highest resolution of the investigated data sets. The results for the other models are in general similar, significant deviations will be briefly discussed later.

All four statistical models we use share the parameters μ_s, μ_c, σ_s and σ_c describing the starting value and the magnitude of change of the location and scale parameters. Parameters γ_s and γ_c are estimated only for Models 1a and 2a, but for the other models we can define γ_s as the constant estimate of the shape parameter and γ_c as equal to zero.

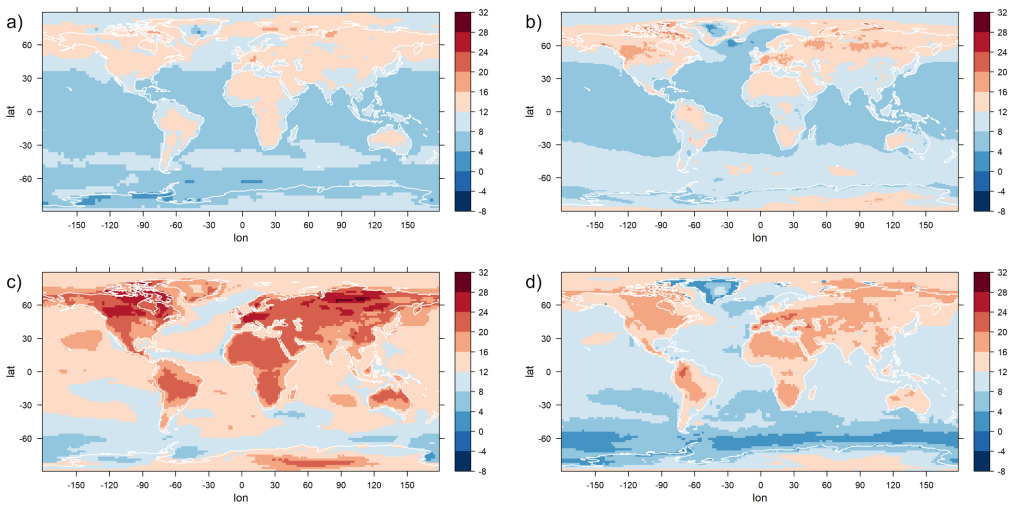

Figure 5.6: **Changes in the 95%-quantiles.** For each grid point, the change in the 95%-quantile of the fitted GEV distribution over the time interval from 1850 to 2300 is depicted. The GEV distributions are estimated by fitting the logistic models to yearly maxima of daily temperature data. For each grid point, the statistical model that is preferred by the Bayesian Information Criterion is used. Units are $^{\circ}$C. Data set used: BCC (a), CCSM4 (b), CSIRO (c), MPI-ESM (d).

Using this definition, the values of the six parameters are depicted in Fig 5.7. For each grid point, we show the estimates of the model that is preferred by the BIC at that grid point.

As expected, the starting value of the location parameter depends highly on the latitude and the climate zone of the grid point investigated. The starting values of the scale parameter show a dependency on the continentality of the climate: the scale parameter is lowest over the oceans and highest in the very continental regions of Siberia, Alaska and northern Canada. It is also relatively high in Antarctica. The starting values of the shape parameter are quite homogeneous, attaining mostly slightly negative values that indicate that no strong positive extremes are present. The only exception to this are some regions in the Arctic Ocean, north of the regions with the high scale parameter discussed above. In these regions, the high values of the shape parameter together with low values of the scale parameter indicate a climate that is in general fairly homogeneous, but with occasionally strong outliers.

Investigating now the parameters describing the magnitude of change in the GEV

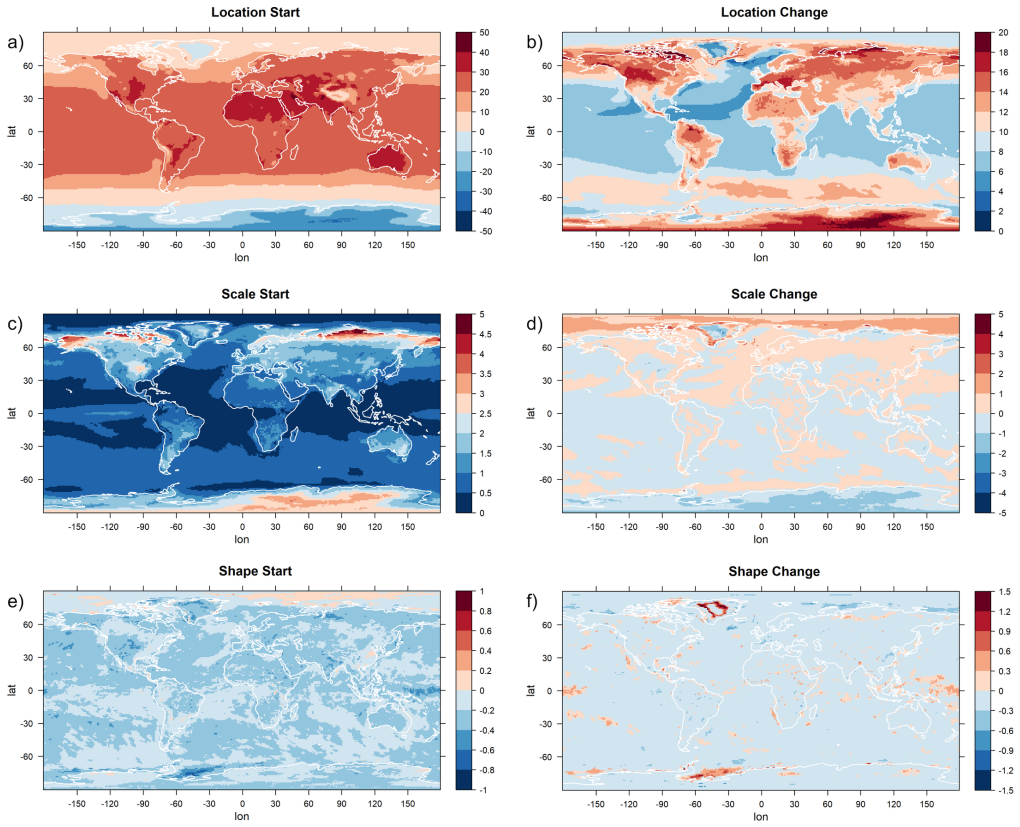

Figure 5.7: **The estimates for the starting values and the amounts of change of the GEV parameters.** The estimates for the parameters μ_s (a), σ_s (c), γ_s (e) and μ_c (b), σ_c (d), γ_c (f), describing starting value and total amount of change over time of the GEV parameters location, scale and shape. The models are applied to yearly maxima of daily temperature data of the climate model CCSM4. For each grid point, the estimates of the model that was preferred by the Bayesian Information Criterion are depicted. Units are °C.

parameters, we detect strong changes in the location parameter especially over land masses, with an increase of up to 20 °C occurring in Europe and the central parts of North America. The highest changes, however, occur in the high-latitude regions that also feature a high initial shape parameter. Over the ocean, the changes in the location parameter are in general much smaller, especially in the Northern Hemisphere.

The scale parameter remains mostly the same in most regions, with a tendency to a slight increase. The most notable changes occur in the Arctic, where the scale parameter increases considerably, and in Antarctica, where it decreases. In most regions, the shape

parameter is predicted to not undergo a change, as statistical models with a constant shape parameter are preferred by the BIC. Regions exhibiting a change in the shape parameter are in the Pacific near the equator and south of South America near Antarctica. A notable exception is a part of Greenland which shows not only a very strong increase in the shape parameter, but also unusually small increases in the location and the scale parameters.

Since a shift in the location parameter of a GEV distribution directly implies an equal shift in the quantiles, it is not surprising that the changes in 95% quantiles (Fig 5.6b) show a similar structure than the changes in the location parameter. The quantile changes are also affected by the changes in the scale parameter, therefore in Antarctica they are lower than the change in the location parameter would suggest (due to a decrease in the scale parameter), and in the Arctic they are higher (due to an increase in the scale parameter that is stronger than elsewhere).

We now turn our attention to the parameters describing at which time the changes take place. Models 1a and 2a have one parameter describing the time of change and one parameter describing its duration that are used for all three GEV parameters simultaneously. In Fig 5.8, these parameters are depicted. As before, for each grid point, the estimates of the statistical model that was favored by the BIC are depicted. If the selected model at a grid point is not one of Model 1a or 2a, the grid point is grayed out. For most grid points, the time around which the change takes place is between 2075 and 2125 and the duration of the change is between 240 and 360 years. The most notable exception to this is the northern Atlantic Ocean, a region in which the duration of the changes tends to be much longer and highest change rate tends to occur much later. Changes that start unusually late occur also off the coast of Antarctica. Both regions are characterized by a ventilation of the deeper layers of the ocean providing an enhanced effective heat capacity dampening the warming signal (Lohmann, 2020, see e.g.).

The other two statistical models, models 1b and 2b, have individual change parameters for the location and the scale (and, in the case of Model 1b, for the shape) parameter. These values are depicted in Figure 5.9 for the grid points at which one of those models is chosen. At all grid points, strong differences between the parameters corresponding to location and those corresponding to scale can be seen, explaining why models that allow for individual changes in the different parameters perform better there. We focus on the largest contiguous region for which one of the models is selected, which is the area

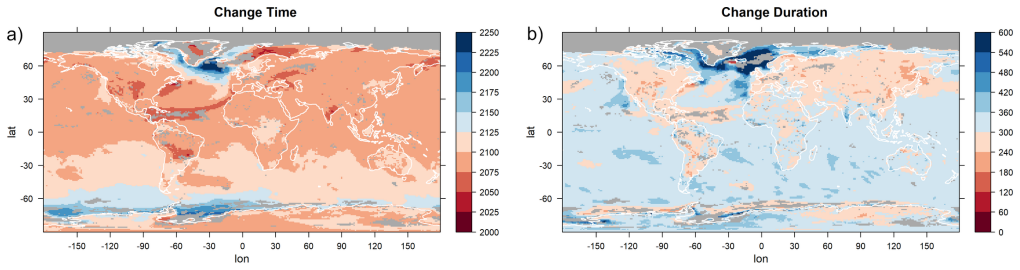

Figure 5.8: **The estimates for timing (a) and duration (b) of change, depicted for statistical models with a simultaneous change in all GEV parameters.** The models are applied to yearly maxima of daily temperature data of the climate model CCSM4. For each grid point, the estimates of the model that was preferred by the Bayesian Information Criterion are depicted. If the preferred model at a certain grid point does not feature parameters for simultaneous changes in the GEV parameters, the grid point is grayed out. Units are years.

around the North Pole. In this region, changes in the scale parameter take place much earlier than those in the location parameter (2000-2100 compared to 2150-2200), and the scale parameter also changes considerably more rapidly than the location parameter (a duration of change of 0-120 years compared to 180-300 years).

To illustrate the four statistical models further, for each of them one grid point where the model is preferred by the BIC is selected. In Fig 5.10, Fig 5.11, Fig 5.12 and Fig 5.13, the time series for those grid points are depicted, together with the modeled time-dependent GEV parameters and the median and the upper and lower 95% quantiles of the modeled GEV distribution.

A first visual inspection indicates that the models seem to fit the data reasonably well. The most common model is Model 2a, showcased in Fig 5.12 for the grid point 0° N, 0° E. A clear logistic shape is visible in the time series of that grid point, which is reflected by a corresponding change over time of the location parameter. The scale parameter slightly decreases over time, while the shape parameter stays constant in this model. As already mentioned, Model 2a is preferred at most grid points, and the corresponding time series are usually similar to the one presented here.

While the shape parameter is constant in Model 2a, it undergoes a change over time in Model 1a, for which an example is depicted in Fig 5.10 (grid point 0° N, 180° E in the Pacific Ocean). The shape parameter shows an increase over time here, while the

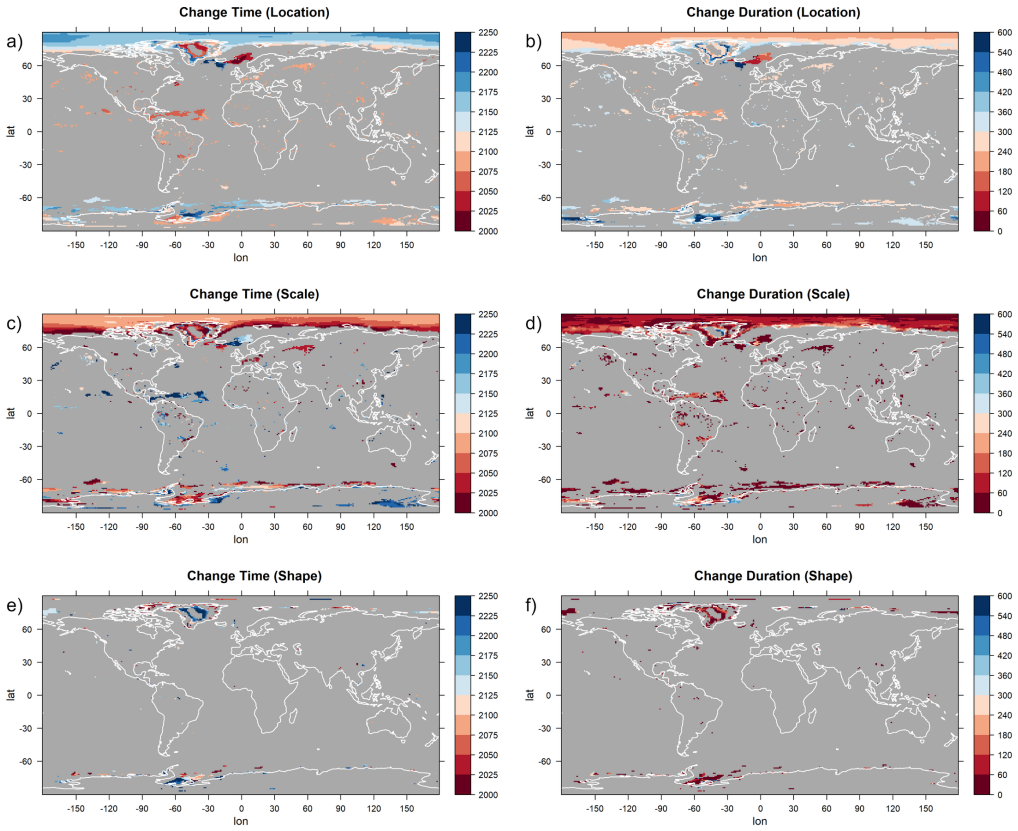

Figure 5.9: **The estimates for timing (a, c, e) and duration (b, d, f) of change for the location (a, b), the scale (c, d) and the shape (e, f) parameter, depicted for statistical models with separate changes in the GEV parameters.** The models are applied to yearly maxima of daily temperature data of the climate model CCSM4. For each grid point, the estimates of the model that was preferred by the Bayesian Information Criterion are depicted. If the preferred model at a certain grid point does not feature parameters for separated changes in the different GEV parameters, the grid point is grayed out. Units are years.

scale parameter decreases at the same time. This indicates a shift to a climate with less variability in general, but more outliers than before. Model 1a is common in parts of the Pacific Ocean and it also appears in several small regions around the world.

Model 2b is predominant in the region around the North Pole, an example is depicted in Fig 5.13 for the grid point $85°$ N, $0°$ E. This model keeps the shape parameter constant and allows for sigmoidal changes in the location and scale parameters with different velocities and at different points in time. For the grid points near the North Pole, changes in the scale parameter are quicker and occur earlier than changes in the location parameter,

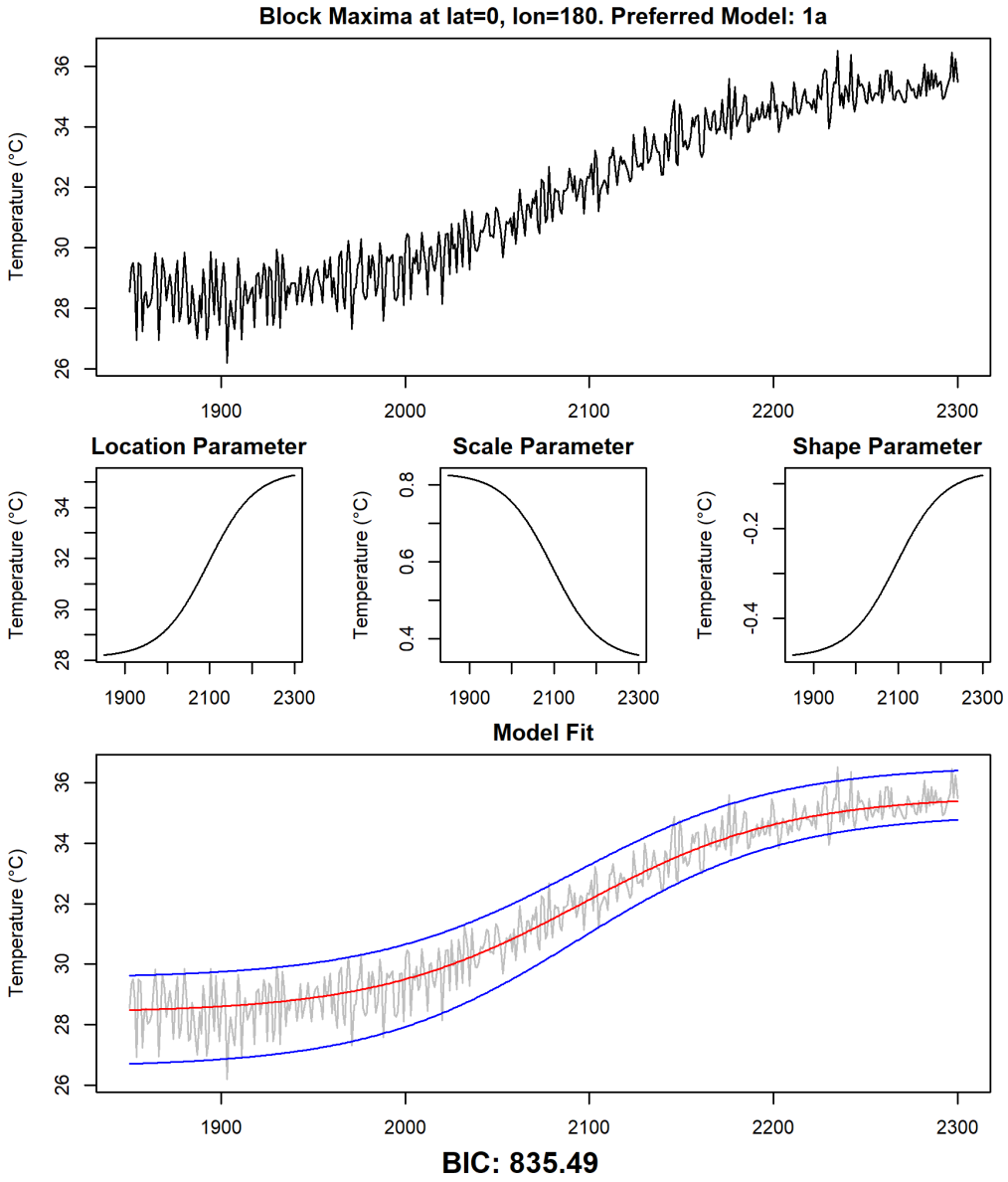

Figure 5.10: **Detailed examination of data and fitted models at grid point** 0° **N,** 180°**E.** The yearly maxima of daily temperature data of the climate model CCSM4 at grid point 0° N, 180°E, together with the non-stationary GEV parameter estimates of the preferred model at this grid point (Model 1a) and the median of the estimated distribution (red line) as well as the 95% confidence interval (blue lines).

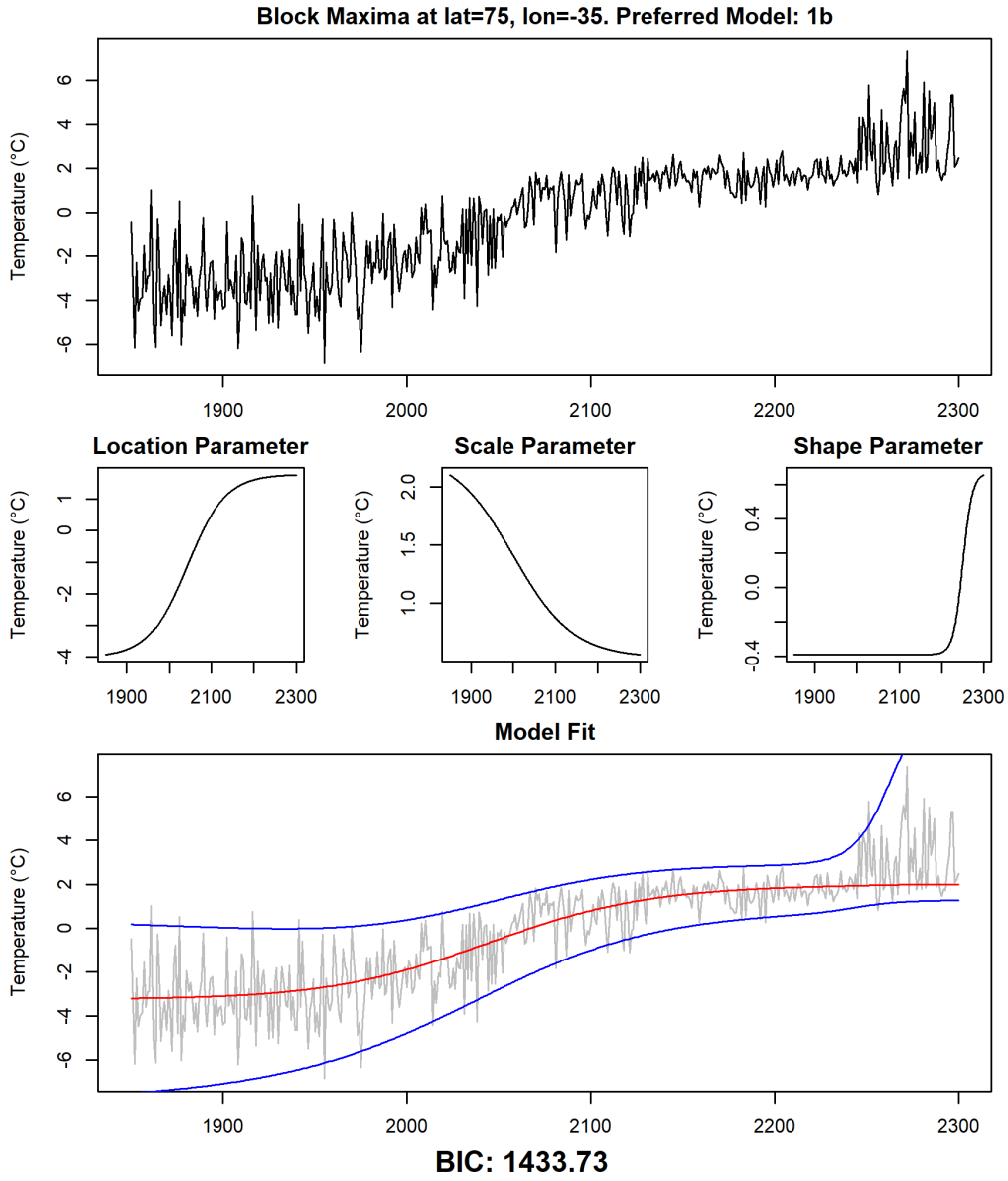

Figure 5.11: **Detailed examination of data and fitted models at grid point** $75°$ **N,** $35°$ **W.** The same analysis as in Fig 5.10 for grid point $75°$ N, $35°$ W (Model 1b).

as is also seen in Fig 5.13: In the first 200 years of the investigation period, the variability of the data is very low, and then it increases rather quickly in the years 2050 through 2100, while gradual changes in the location parameter follows later.

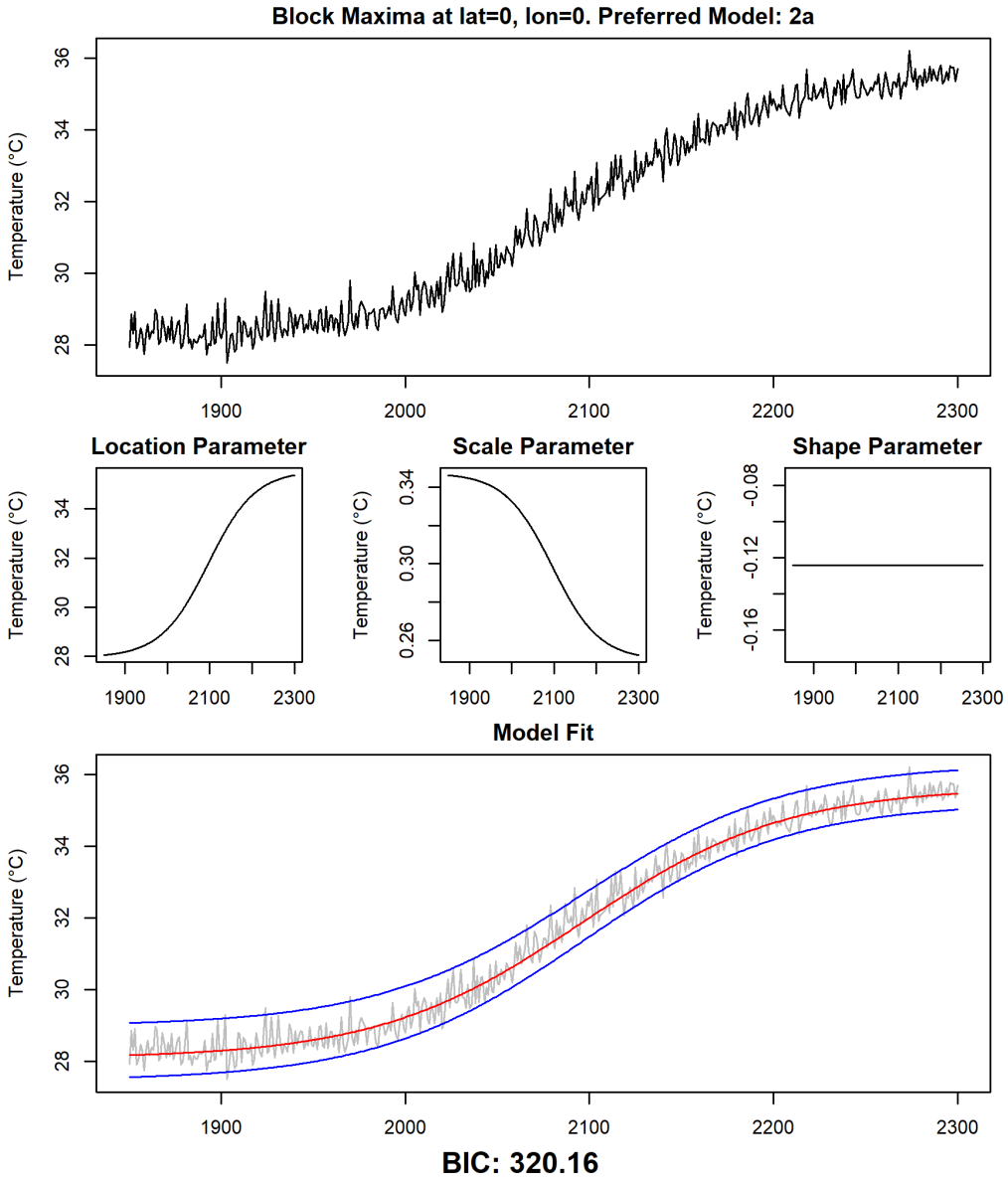

Figure 5.12: **Detailed examination of data and fitted models at grid point** 0° **N,** 0° **E.** The same analysis as in Fig 5.10 for grid point 0° N, 0° E (Model 2a).

The fourth statistical model is Model 1b, the most complex of the statistical models we use and the only one that allows for changes in all three GEV parameters at different

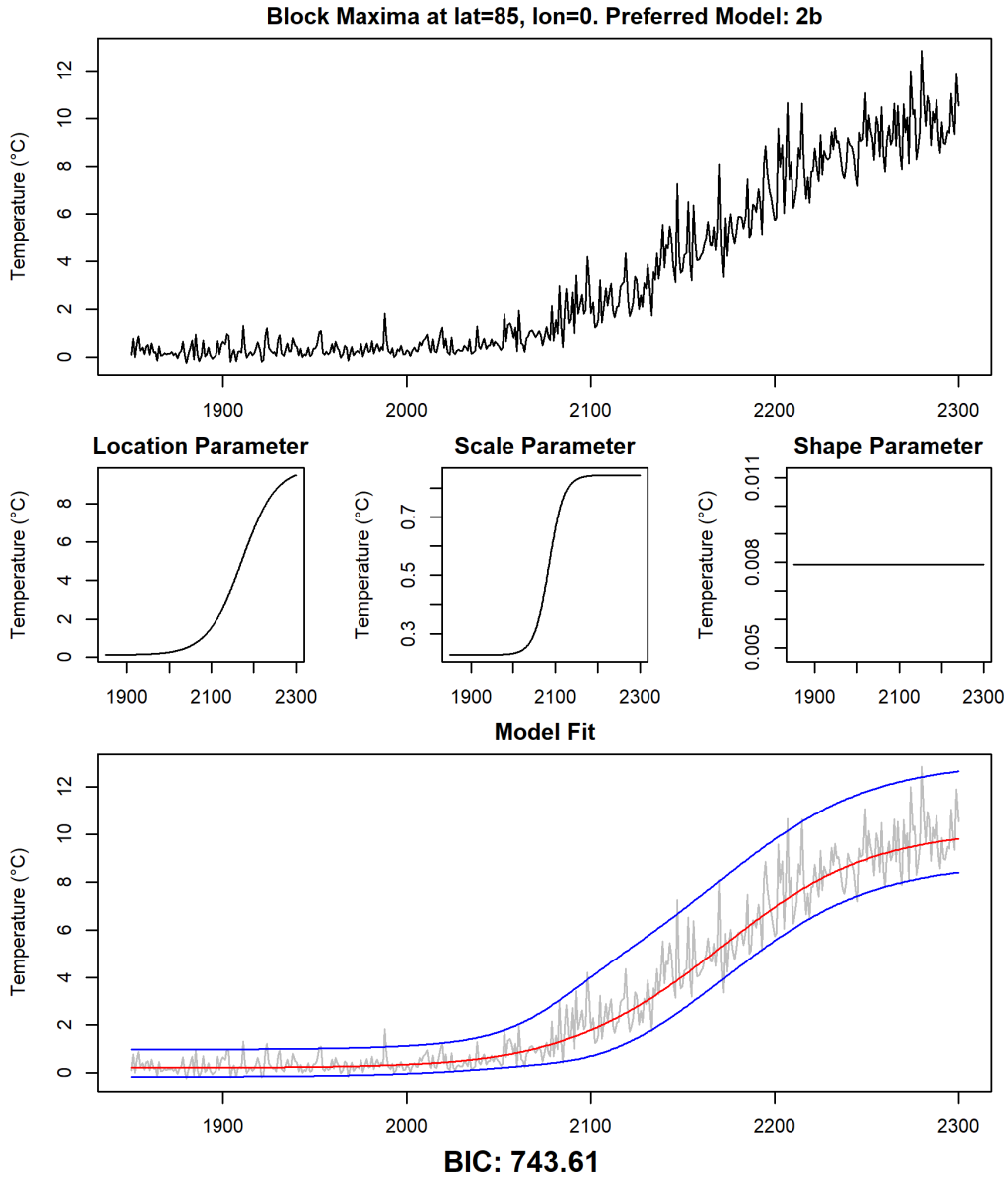

Figure 5.13: **Detailed examination of data and fitted models at selected grid point** $85°$ **N,** $0°$ **E.** The same analysis as in Fig 5.10 for grid point $85°$ N, $0°$ E (Model 2b).

speeds and points in time. There are only few regions where this model is preferred, one of them is a part of Greenland including the grid point $75°$ N, $35°$ W, which is analyzed in Fig 5.11. Besides the increase in the location parameter, we can detect here a strong

decrease of the scale parameter that takes place mostly between 2000 and 2100. The shape parameter shows a pronounced increase after the year 2100. These model parameters indicate a complex behavior of the underlying time series that involves different kinds of changes at different points in time.

The goodness of fit of the statistical models is tested using a Kolmogorov-Smirnov test at significance level 5%, which is applied for each grid point to the results of the model that is preferred at that grid point by the BIC. There are only few grid point for which the hypothesis of the data following the modeled non-stationary GEV distribution is rejected, see Fig 5.14. It is important to keep in mind that the non-rejection of the hypothesis does not mean its confirmation, but still, this result is a promising indicator for the general applicability of the statistical models.

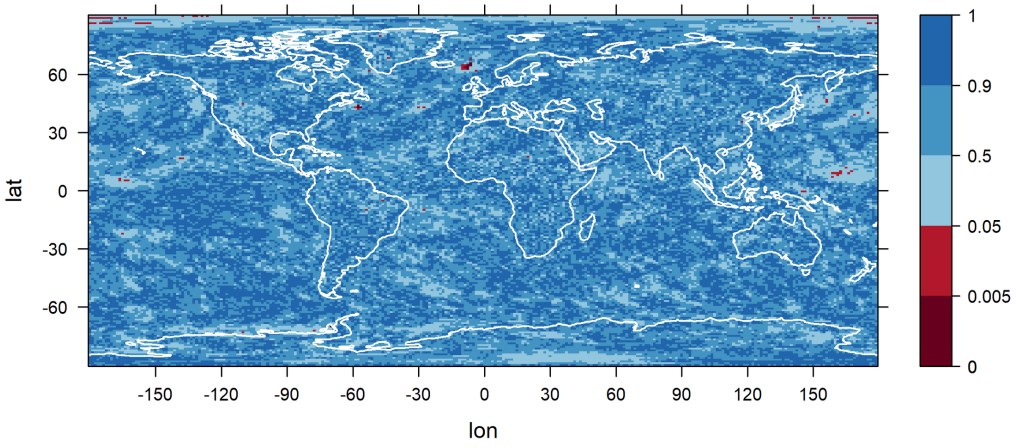

Figure 5.14: **P-values of the Kolmogorov-Smirnov test to investigate the goodness of fit.** A Kolmogorov-Smirnov test is applied for each grid point to the results of the statistical model that is preferred at that grid point by the BIC. The data used are yearly maxima of daily temperatures of the climate model CCSM4. Grid points at which the hypothesis of the data being GEV distributed with the modeled time-dependent GEV parameters is rejected at significance level 5% are colored in red.

A detailed analysis of the different parameters of the statistical models was presented here for the earth system model CCSM4. For the other models, we briefly mention some key differences to the CCSM4 model. The estimated parameter values that are depicted

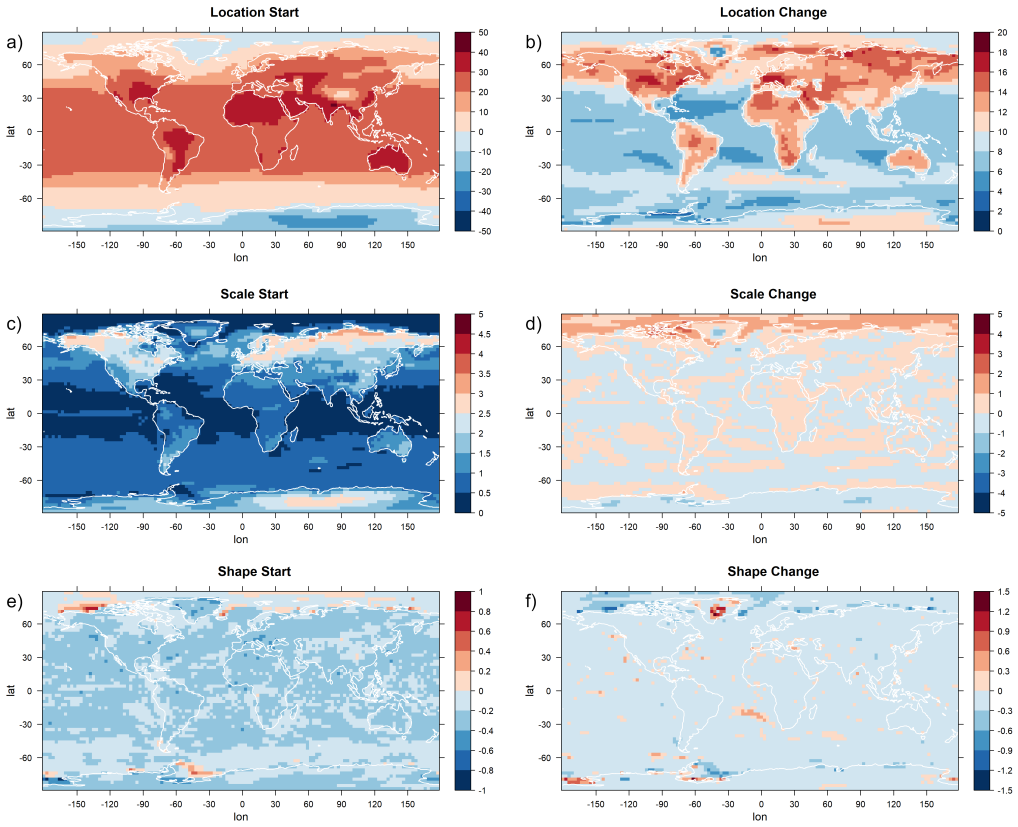

Figure 5.15: **As Fig. 5.7, but for model BCC.**

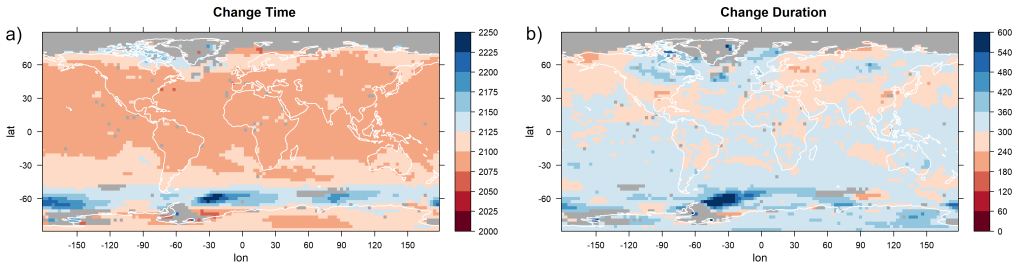

Figure 5.16: **As Fig. 5.8, but for model BCC.**

in Fig. 5.7 through 5.9 for CCSM4 are depicted for the other climate models in Fig. 5.15 through 5.23. The starting values of the three estimated GEV parameters are very similar for all four earth system models (compare Fig 5.7 and 5.15, 5.18, 5.21, panels a, c, e). CCSM4 tends to lead to higher values of the scale starting parameter in the high northern latitudes than the other models. This parameter also shows different values for Antarctica

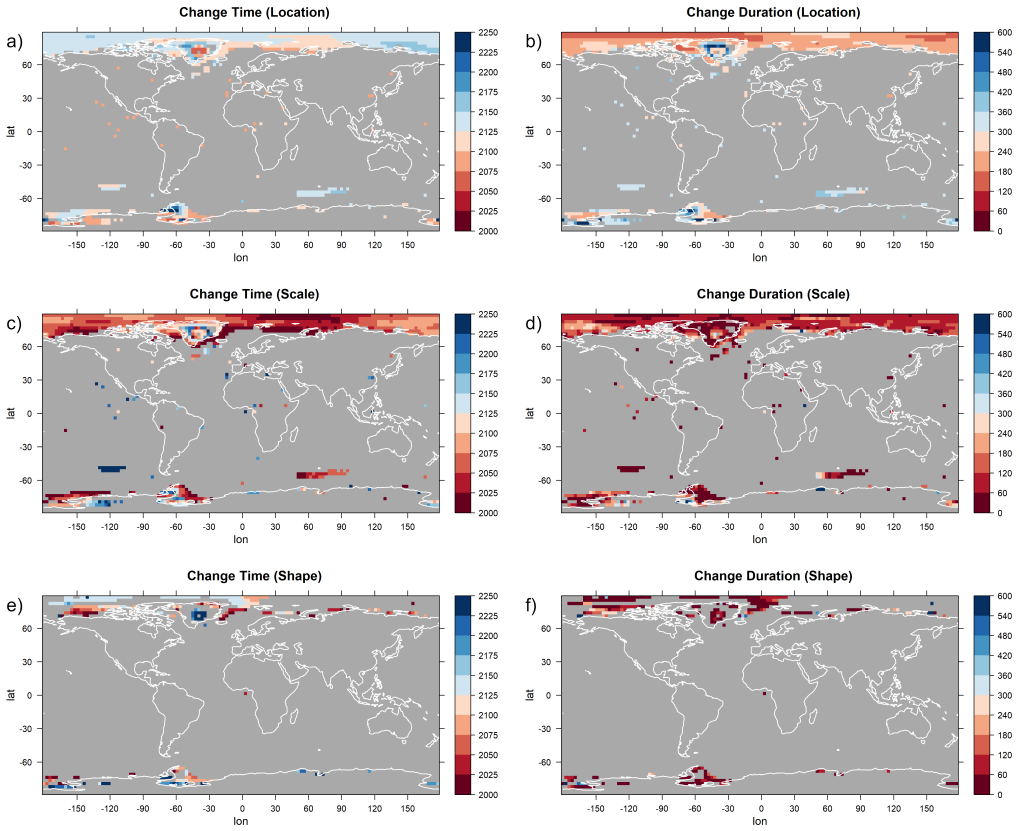

Figure 5.17: **As Fig. 5.9, but for model BCC.**

among the different earth system models. The changes in the location parameter (panel b of the figures) show a clear land-sea distinction in all four earth system models. In CSIRO, its values are considerably higher than in the other models, resulting also in the large difference in 95% quantiles compared to the other three models (Fig 5.6). This climate model also shows a region south of Africa near Antarctica with an unusually high scale change (panel d) and a high negative shape change (panel f) that is not identified in the other climate models. Other than that, all models agree that changes in scale are in general not high, with the exception of the high latitudes that show a strong increase in scale in the north and a decrease in scale in the south.

The timing of the simultaneous changes in all parameters also indicates a marked difference between CSIRO and the other climate models (compare Fig 5.8 and 5.16, 5.19, 5.22). In CSIRO, the time of the highest change rate (panel a of the figures) is in general approximately 50 years later than in the other models. Besides that, disagreements regard-

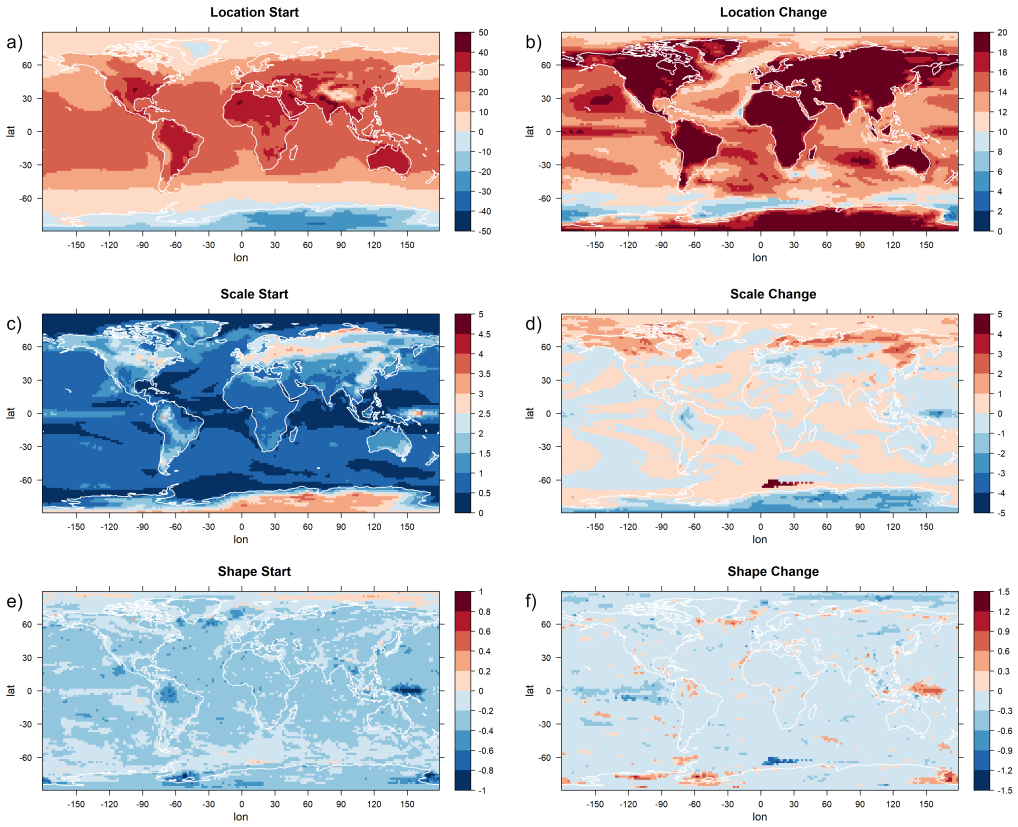

Figure 5.18: **As Fig. 5.7, but for model CSIRO.**

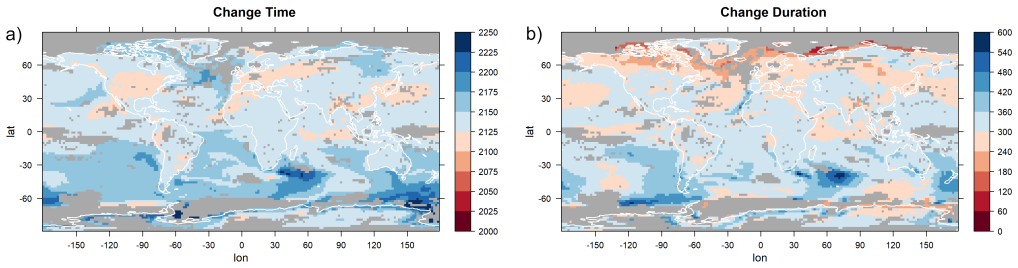

Figure 5.19: **As Fig. 5.8, but for model CSIRO.**

ing the timing of changes exist also for the Indian and Pacific Ocean in the high southern latitudes, for which some models predict a change that starts later and takes longer than in the CCSM4 model. The prolonged changes in the North Atlantic Ocean that can be seen in panel b of Fig 5.8 for CCSM4 are not detected for the other earth system models (panel b of 5.16, 5.19, 5.22) and in general, it can be said that the four earth system models show large

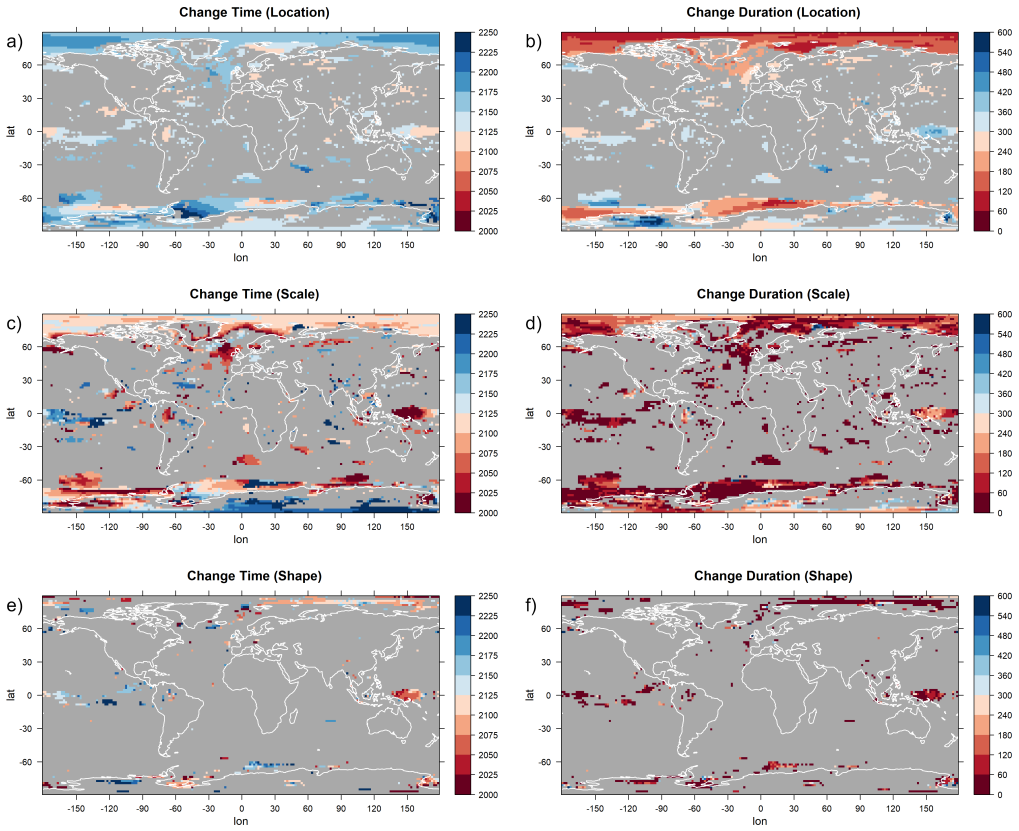

Figure 5.20: **As Fig. 5.9, but for model CSIRO.**

differences in the higher latitudes, both as to which statistical model is selected (see again Fig 5.5) and what parameter values are estimated. A common feature of all models is that near the North Pole, statistical models with separate changes in the location and the scale parameters are preferred and that the changes in scale precede the changes in location and also happen more quickly. The models disagree with regard to which regions use a constant and which ones use a variable shape parameter. For the CSIRO earth system model, separate changes in the location and the scale parameter are predicted in more regions than for the other three models, including large parts of Antarctica and the Pacific Ocean near the Equator (compare Fig 5.9 and 5.17, 5.20, 5.23).

5.6 Discussion of the results of this chapter

We present statistical models for extreme temperature that are applied to global climate data that span several hundred years and are influenced by climate change. While it is

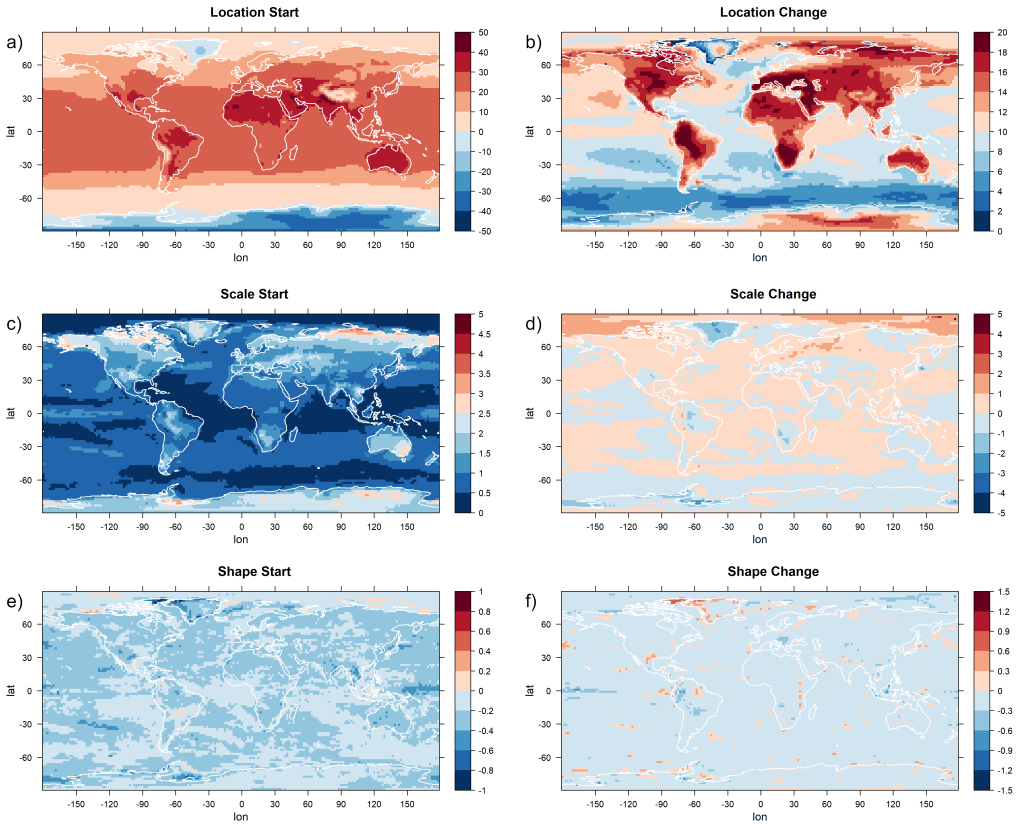

Figure 5.21: **As Fig. 5.7, but for model MPI-ESM.**

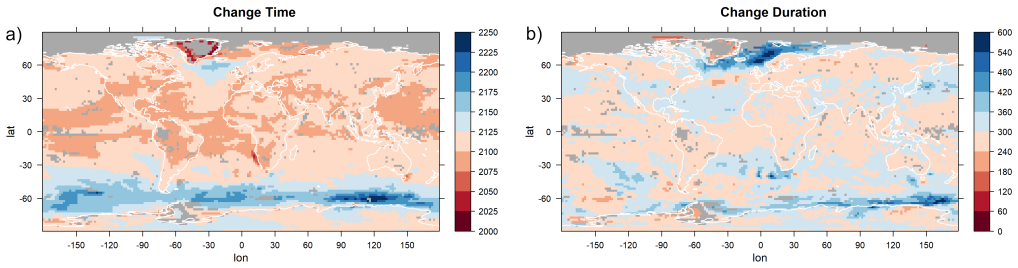

Figure 5.22: **As Fig. 5.8, but for model MPI-ESM.**

not a new approach to use non-stationary GEV distributions to investigate the development of climate extremes, most studies assume a dependency of the GEV distribution parameters on time that is either linear/polynomial (Casati and de Elía, 2014; Kharin and Zwiers, 2005; Sarhadi and Soulis, 2017; Mahajan et al., 2015) or exponential (Hanel and Buishand, 2010). Consequently, the models are usually applied to data covering

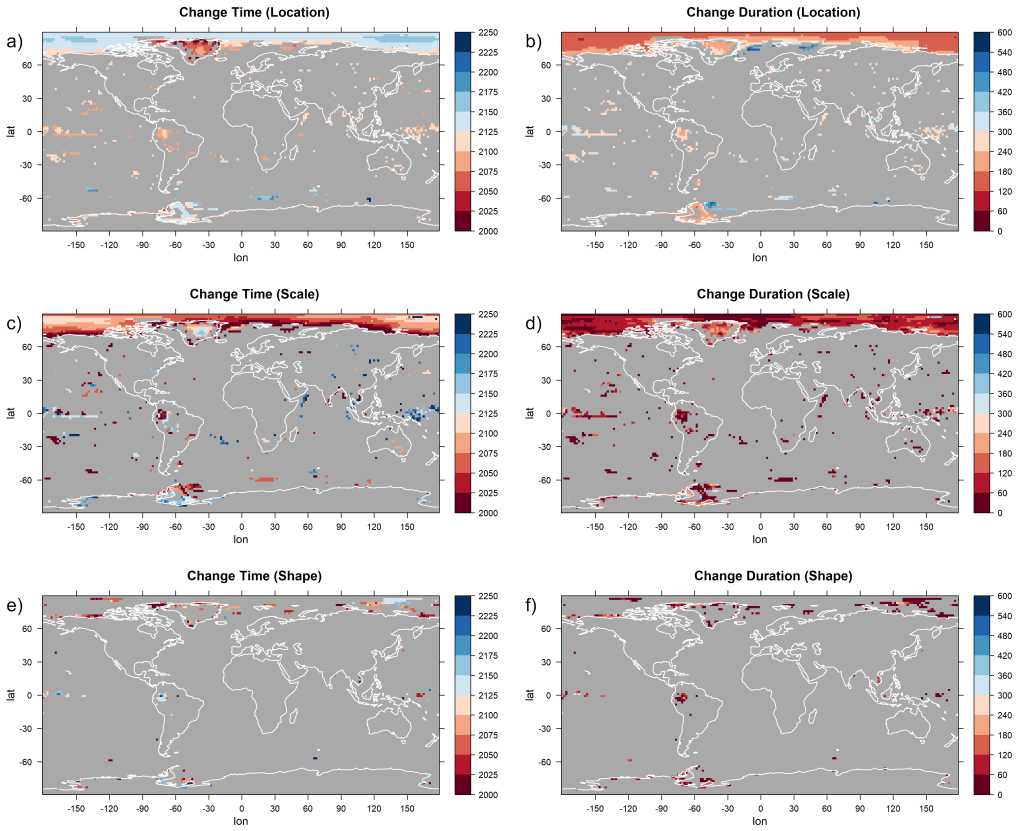

Figure 5.23: **As Fig. 5.9, but for model MPI-ESM.**

not more than 100 years. If the goal is to investigate changes in extremes on a longer time scale, the time frame is usually split up into several intervals of short length and stationary GEV distributions are fitted to each one. Then, their parameters are compared. This approach was used for global precipitation in Russo and Sterl (2012), for precipitation and temperature in Australia in Perkins et al. (2014) and for summer temperature in the United States in Hogan et al. (2019). In Slater et al. (2021), annual maxima of daily temperature data from several CMIP6 models were investigated and stationary GEV distributions were fitted to the data at different time intervals. When stationary distributions are applied, it needs to be assumed that the changes in the investigated time intervals are not large. It it also more difficult to make statements regarding the temporal aspects of the changes. Non-stationary GEV distribution are advantageous in this regard, although it can be difficult to find suitable parametrizations for the parameters.

Our approach of combining logistic functions with GEV distributions to describe

climate extremes has not been used before to our knowledge. Logistic functions have, however, been used to describe historical CO_2 emissions in many countries (Meng and Niu, 2011; Çiğdem Köne and Büke, 2010) and have also been applied to future projections of greenhouse gas concentrations (Pérez-Suárez and López-Menéndez, 2015). Climate change is closely connected to CO_2 concentrations, and the mean global temperature has been shown to be in an approximately linear relation to them (Matthews et al., 2009). This further supports the idea of using logistic functions to describe extreme events under climate change as well.

The results in Slater et al. (2021) that were obtained by fitting stationary GEV distributions to CMIP6 model results are mostly in line with the results of this chapter, even though we used a different greenhouse gas emission scenario (RCP8.5 vs. SSP370). In both works, it is noted that the location parameter changes strongly over land and that this contributes to a large extent to the changes in extremes. A large increase in the shape parameter over the Arctic was detected in Slater et al. (2021), and for the scale parameter, they identify a tendency for an increase over time in the tropics and a decrease over time in high-latitudes. Our study also identifies an increase of scale in the low latitudes and a decrease of scale in Antarctica, but results for the Arctic are different. Our statistical models do not predict an increase in the shape parameter in the Arctic, but instead an incrase of the scale parameter while the shape parameter stays constant (Fig 5.7).

As discussed in the previous section, the Arctic region is unusual with regard to model selection: It is one of the few regions in which statistical models featuring non-simultaneous changes in location and scale parameter are preferred. In addition to that, the changes in the scale parameter are higher than in all other regions. Both results can be explained with the permanent presence of ice in the Arctic: the temperature of melting ice does not exceed $0\,^{\circ}\mathrm{C}$, therefore the annual maximum of daily temperature is close to $0\,^{\circ}\mathrm{C}$ as long as ice is present all year long, resulting in a very low variability of the annual maxima. It is indeed shown (Taylor et al., 2012) that in the RCP8.5 run of the CCSM4 earth system model, the Arctic becomes ice-free in the 2060s, which is also the period of time at which the variability of the time series starts to increase (Fig 5.13). After that, the value of the scale parameter is comparable to other land regions. This process also explains the complex behavior of the time series in Greenland in (Fig 5.11). Annual maxima are below $0\,^{\circ}\mathrm{C}$ at the beginning of the investigation period. Due to increasing temperatures, ice begins to melt and the variability decreases, as the annual

maxima are permanently close to $0\,^{\circ}\mathrm{C}$ in the years 2100 through 2200. After that, ice has completely melted in summer and the variability increases again from 2200 onwards. This example also shows the limitations of the logistic models we present here: The increased variability in the years 2250 through 2300 is most likely better modeled by a high scale parameter than by a high shape parameter. This is also indicated by the results of fitting a stationary GEV distribution to the values of the time series in Fig 5.11 in the years 2250 through 2300. The fitted values are 1.44 for the scale parameter and 0.12 for the shape parameter. It is, however, not possible to model a change of the scale parameter going from a high value to a low value and then back to a high value again using a sigmoid function, so the BIC favors a model with a high shape parameter in later years instead.

It also needs to be emphasized that logistic functions are suitable for the modeling of future climate only under the condition of a cessation of greenhouse gas emissions in the future. For model data that are based on other scenarios, different functions have to be used, although logistic functions might also be useful for describing data that show a continuously rising trend in the extremes. In particular, the extraction and storage of atmospheric CO_2 in order to revert some consequences of climatic changes (and to prevent others) are more and more discussed. This is reflected by the SSP scenarios (replacing the RCP scenarios) used in the newer IPCC reports (IPCC, 2021), of which some predict a reduction of the atmospheric CO_2 levels starting in the second half of the century. A possible extension of the logistic models for such a scenario is based on the double logistic function

$$p_s + p_{c,1} \cdot f\left(2 \cdot \log(19) \cdot \frac{t - a_{p,1}}{b_{p,1}}\right) + p_{c,2} \cdot f\left(2 \cdot \log(19) \cdot \frac{t - a_{p,2}}{b_{p,2}}\right). \tag{5.10}$$

In this formula, two logistic function are combined, allowing for the description of a change from one state to another that is not completed, but instead reverted mid-way and that finally settles on an intermediate value. Such a model could also be useful with the RCP data sets used here to model the behavior in Greenland region for which sigmoidal functions are of limited suitability.

Besides that, the methodology presented in this chapter is not restricted to a specific application. The models or variations of them can also be applied to other data sets and other scientific questions regarding changes of extremes over time. Thus, the development of this methodology is a scientific contribution of its own right.

Our focus is on the univariate analysis of temperature extremes: for each grid point, the time series of temperature data is investigated separately from all others. For a better understanding of climate extremes it is important to also investigate multivariate distributions. For example, climate extremes that take place simultaneously over a large region are especially problematic because of high damages for economies and possible difficulties in providing necessary medical or humanitarian aid. It would therefore be interesting to use spatio-temporal models to describe climate extremes. We will discuss this also in the concluding remarks at the end of this work.

5.7 The main conclusions of this chapter

We have designed and applied statistical models for the development of temperature extremes over several centuries that allow us to investigate the magnitude and the timing of the changes in temperature extremes. In addition, the models differentiate between changes in the mean, the variability and the distributional shape of the estimated non-stationary GEV distributions. We summarize the conclusions in the following main points:

1. A strong increase in the 95% quantiles of the annual temperature maxima could be detected in most regions of the world. In these regions, extremes will continue to rise and reach unprecedented strengths in the future. This is true especially over continents and corresponds to the well-known fact that global warming is stronger over land than over the oceans or in coastal regions (Byrne and O'Gorman, 2018; McBride et al., 2022). However, we find a disagreement between the earth system models in terms of the total magnitude of the changes.

2. The development of extremes depends highly on the geographic region. Geographically varying developments can be detected not only with regard to the magnitude of changes, but also with regard to their timing, and to the extent to which the changes in extreme events are caused by changes in the location, the scale or the shape of the distribution of the annual maxima. For example, changes in the North Atlantic Ocean are slower than elsewhere.

3. Changes in location and scale of the distributions are predicted to take place simultaneously in most regions. Most earth system models agree that the highest rate of change is reached in the time between 2050 and 2100 over land and most parts of the

oceans. In some high-latitude areas, changes over oceans start much later and last longer. The velocity of the change tends to be higher over land than over the oceans. Taking this together with conclusion 2, we can expect large and rather rapid changes in temperature over land masses over the course of about 100 years.

4. Non-simultaneous changes in the parameters are predicted is the region around the North Pole, in which an abrupt increase in variability is followed by a gradual increase of mean values. This is probably caused by the effects of the melting of sea ice. The earth system models disagree about the nature of the changes in Antarctica and in Greenland, which could hint to insufficient representations of polar processes in climate models such as feedbacks with the cryosphere (Ackermann et al., 2020, e.g.). In addition, the statistical models presented here might not be suitable to describe the complex changes that are predicted for those regions.

Chapter 6

Clustering algorithms for max-stable processes

6.1 Introduction to the chapter

In Chapter 3.2.1, we have laid out how stochastic processes on a compact subset of $\mathbb{R}^2$ can be used to investigate spatial dependencies in a certain region of investigation. We have presented different parametric sub-classes of max-stable processes that can be used for parametric inference, like the Brown-Resnick process or the extremal-t process and have pointed out that because of their spatial stationarity, their application is limited to areas that are small and rather homogeneous.

One of the possible approaches to apply max-stable processes also to data on larger regions is the application of a clustering algorithm. The algorithm divides the area of investigation into smaller sets, onto which stationary max-stable processes can be fitted. This idea has recently been proposed by Saunders et al. (2021), using as clustering algorithm the algorithm by Bernard et al. (2013) that we have used in a different context in Chapter 4. In this chapter, we will build up on this approach. While clustering is a promising idea to make stationary max-stable processes applicable to data that cover a large and inhomogeneous area, we will see that the algorithm by Bernard et al. (2013) does not necessarily produce regions in which spatial stationarity is a reasonable assumption. We will illustrate this using a simple example of a non-stationary max-stable processes, and we will propose an alternative clustering algorithm and compare the two algorithms in a simulation study.

This chapter is structured as follows: In the next section, we will discuss the clustering algorithm by Saunders et al. (2021), the limitations we see with it, and our suggested alternative. Using a simulation study, we investigate the performance of the clustering algorithms in Section 6.3. A section on conclusions and a discussion finalize the chapter.

6.2 Discussion of different clustering algorithms

6.2.1 Clustering based on extremal coefficients

Saunders et al. (2021) suggest to use the hierarchical clustering algorithm by Bernard et al. (2013) to split the area of investigation into smaller subsets. Recall from Section 3.2.1 that the dissimilarity measure used in this algorithm is

$$D_1(s_1, s_2) := \hat{\theta}_{s_1,s_2} - 1, \tag{6.1}$$

with $\hat{\theta}_{s_1,s_2}$ the madogram estimator for the pairwise extremal coefficent θ_{s_1,s_2} for grid points $s_1, s_2 \in \mathcal{S}$. In the following, we will use the term Extremal Dependence Clustering (EDC) for the clustering based on this dissimilarity function.

The EDC algorithm is performed with the goal of defining regions to which spatially stationary max-stable processes are fitted. It should therefore group points together in such a way that within the clusters stationarity can be assumed. The dissimilarity measure by Bernard et al. (2013) is based on the comparison of extremal coefficients and therefore groups together points with a tendency for concurrent extremes. Within such a cluster, pairwise extremal dependencies tend to be high in general, which might reduce the possible extent of spatial non-stationarity. Nevertheless, spatial stationarity is not a justified assumption within the clusters defined that way, and the dissimilarity measure by Bernard et al. (2013) was not designed with the intention of finding such regions.

We illustrate this using a concrete example of a non-stationary max-stable process. We can construct such a process using the approach by Huser and Genton (2016) that we discussed in Chapter 3.2.4. As a simple example we use a Huser-Genton process on the set $\mathcal{S} = [-5, 5] \times [-5, 5]$ with matrix parameters $a_s = 2$ constant, $b_s = (x + 5)/2$, $s = (x, y) \in \mathcal{S}$ and $\gamma_s = 0$ constant. The global model parameters are $\nu = 5$ and $\alpha = 1$. This process is obviously stationary on the sets $\{x\} \times [-5, 5]$ for all $x \in [-5, 5]$, and if we investigate a

vertical stripe of the form $[x-\epsilon, x+\epsilon] \times [-5,5]$, $\epsilon > 0$ small, the values of a_s, b_s and γ_s in that region are very similar and stationarity is a reasonable approximation. The clustering based on extremal coefficients, however, does not result in clusters of such a form. In Fig. 6.1 we show the pairwise extremal coefficients $\theta_{s,t}$, $t \in \mathcal{S}$ for four selected values of s: $s_1 = (-3,2)$, $s_2 = (3,2)$, $s_3 = (-3,-2)$ and $s_4 = (3,-2)$. It can be observed also from Fig. 6.1 that the dependence structures for the points with the same value of x are identical and that it would therefore be reasonable to group them into the same cluster. However, the extremal coefficient θ_{s_1,s_2} is close to 2, so the points s_1 and s_2 will likely not be grouped into the same cluster by the EDC clustering. The same holds for the points s_3 and s_4. Instead, pairs of points with a low extremal coefficient, like for example $(2,2)$ and $(3,2)$ will be grouped together even though the dependency structures around these points differ. Indeed, if we apply the EDC clustering to the true values of the extremal coefficients, we obtain the clusters depicted in Fig. 6.2, confirming the theoretical considerations we just made.

6.2.2 Clustering based on local estimates

An approach that is expected to be more suitable to finding spatially stationary regions is based on the direct comparison of the structures of extremal dependence around two points. We assume that for each point $s \in \mathcal{S}$, the extremal coefficients form a locally elliptic structure, that is, for all points $t \in \mathcal{S}$ in the vicinity of s the extremal coefficient $\theta_{s,t}$ can be approximated by $R_\alpha(\|A_s(s-t)\|)$ for some transformation matrix A_s and some $\alpha \in (0,2]$ (fixed on the whole process), with $R_\alpha(x) = \exp(-x^\alpha)$ as in Section 6.2.1. This is obviously true for stationary extremal-t processes and Huser and Genton (2016) show that it is also true for the non-stationary processes they designed, so we do not consider this to be a too severe restriction. Fixing a small $\epsilon > 0$ and choosing values for the global parameters α and ν, we obtain for each of the locations $t \in \mathcal{T}$ at which data are given an estimate for A_t by using composite maximum likelihood on the sum of the pairwise log-likelihoods for all pairs (t,u), $u \in U(t) := \{u \in \mathcal{T} \,\big|\, \|t-u\| < \epsilon\}$. Depending on the spatial structure of $\mathcal{T}$, some isolated points may have to be excluded beforehand to ensure that $U(t)$ is always of a sufficient size.

In order to identify structures in the estimates better and to reduce the influence of outliers, we apply spatial smoothing to the estimated parameters a, b, and γ at every point. In our application, to smooth the values at $t \in \mathcal{T}$ we used local averages on a vicinity of t. Other spatial smoothing algorithms, for example kernel smoothing, are also possible

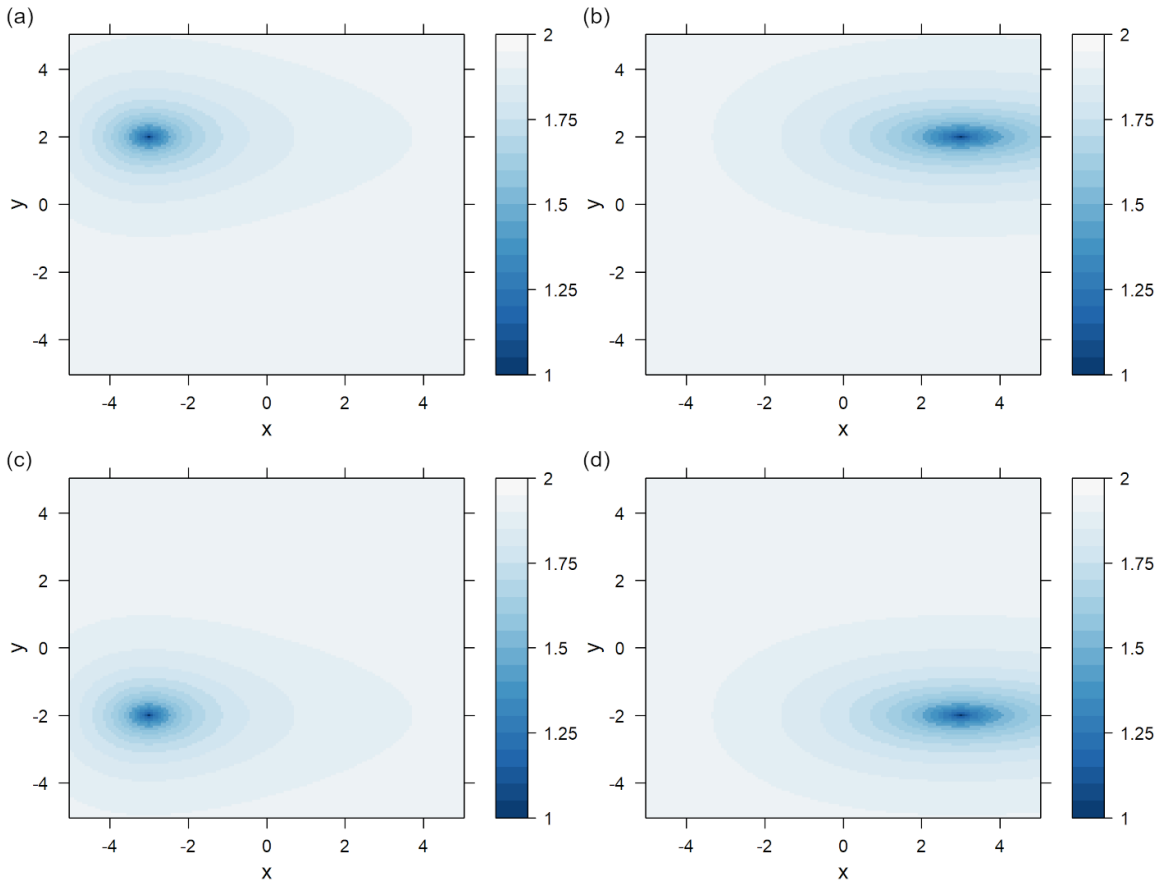

Figure 6.1: Illustration of the non-stationary max-stable process on $\mathcal{S} = [-5, 5] \times [-5, 5]$ given by the matrix parameters $a_s = 2$, $b_s = (x + 5)/2$, $s = (x, y) \in \mathcal{S}$ and $\gamma_s = 0$ and the global parameters $\nu = 5, \alpha = 1$. Displayed are the pairwise extremal coefficients for the points in $\mathcal{S}$ relative to **(a)** $s_1 = (-3, 2)$, **(b)** $s_2 = (3, 2)$, **(c)** $s_3 = (-3, -2)$ and **(d)** $s_4 = (3, -2)$.

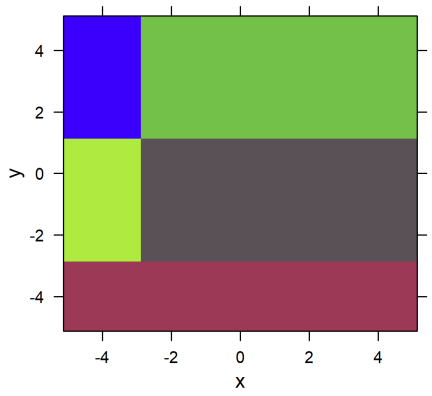

Figure 6.2: The results of the EDC clustering algorithm, applied to the true extremal coeffients of the non-stationary max-stable process from Fig. 6.1 using a number of clusters of 5.

(see for example Wand and Jones, 1994). This results in an estimated matrix $\hat{A}_t$ for each location $t \in \mathcal{T}$. If we group points with similar values of $\hat{A}_t$ into one cluster, it is reasonable to assume that within this cluster spatial stationarity holds approximately.

To determine which of the matrices are considered 'similar', we compare the shapes of the elliptical level sets corresponding to them using the Jaccard index. Let $L_s := \{h \in \mathbb{R}^2 \big| \, ||\hat{A}_s h|| \geq 0.5\}$. We define a dissimilarity measure as

$$D_2(t_1, t_2) = 1 - \frac{|L_{t_1} \cap L_{t_2}|}{|L_{t_1} \cup L_{t_2}|} \tag{6.2}$$

with $|\cdot|$ denoting the area. The value of 0.5 in the definition of L_t is arbitrary, any other value in $(0, 1)$ would yield the same result. We will use the term Local Estimates Clustering (LEC) for this clustering from now on.

6.2.3 A method to compare different clusterings

To compare the two algorithms, we investigate which of the resulting clusterings is better suited to fitting stationary max-stable processes to it. Assume that we have fitted both clustering algorithms to some data and that we have fitted a stationary max-stable process to each of the resulting clusters. Denote the clusters of the EDC algorithm by $\{\sigma_1, \ldots, \sigma_p\}$ and the clusters of the LEC algorithm by $\{\tau_1, \ldots, \tau_q\}$. Each of these elements is a subset

of $\mathcal{T}$, the different clusters in one clustering are disjoint and their union is the whole set $\mathcal{T}$. Define $\mathcal{M} = \{\sigma_1, \ldots, \sigma_p, \tau_1, \ldots, \tau_q\}$ as the set of all clusters. Remember that for each cluster $M \in \mathcal{M}$, the estimated parameters of the corresponding max-stable process $\hat{\psi}_M$ have been calculated by maximizing the composite likelihood L_M from Eq. (3.33).

Note that for either clustering algorithm, max-stable processes are fitted only to the data within the same cluster, so if two points fall into two different clusters, a statistical model for their dependency is not provided. For this reason, it is not possible to compute or compare the composite likelihoods on the whole set $\mathcal{T}$ from Eq. (3.31), which would otherwise be the standard approach for evaluating the goodness of fit. The likelihoods of the fitted processes of the different clusters are not comparable to each other because each one of them is based on different underlying data. However, it is possible to calculate composite likelihoods on the intersections of clusters of the two clusterings, that is, on the sets $\upsilon_{ij} = \sigma_i \cap \tau_j$, provided they contain two or more elements. On each of these intersections a stationary max-stable process has been fitted for both algorithms, and by comparing the likelihoods the goodness of fit of the processes on this area can be compared. This leads to a measure for the goodness of fit on υ_{ij} for both algorithms:

$$\hat{L}_{i,j}^{EDC} = L_{\upsilon_{ij}}(\hat{\psi}_{\sigma_i}), \qquad \hat{L}_{i,j}^{LEC} = L_{\upsilon_{ij}}(\hat{\psi}_{\tau_j}). \tag{6.3}$$

Note that both models have the same number of parameters, so we can compare the likelihoods directly and do not need a penalty term as in the Akaike or Bayesian Information Criterion.

6.3 A simulation study to investigate the clustering algorithms

In this section, we compare the two clustering algorithms by means of a simulation study. To do this, we simulate data from the Huser-Genton model we already investigated in Section 6.2.1. Remember that we use for this model as global parameters $\alpha = 1$ and $\nu = 5$ and as parameters for the local dependencies $a_s = 2$, $b_s = (x + 5)/2$ and $\gamma_s = 0$ for $s = (x, y) \in \mathcal{S}$ (see Fig. 6.1). We choose a horizontal and vertical resolution of the space $\mathcal{S} = [-5, 5] \times [-5, 5]$ equal to 0.2 and simulate data from processes with 250 independent observations.

For the clustering algorithms, we choose a number of clusters equal to five in both

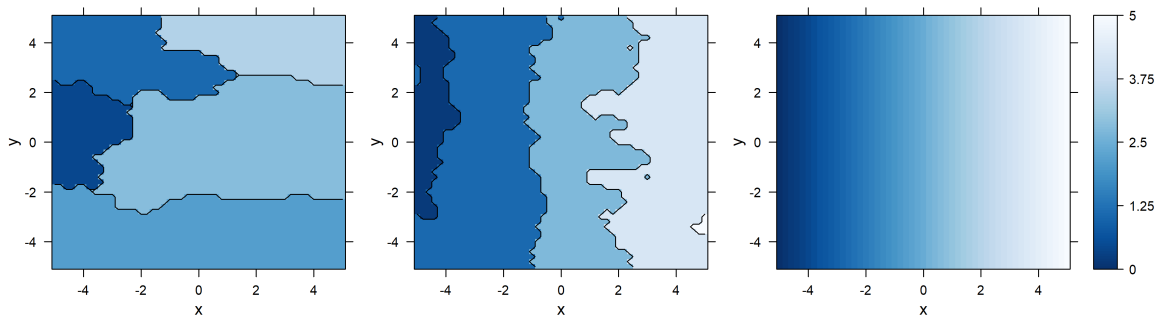

Figure 6.3: Results of the EDC algorithm (**a**) and the LEC algorithm (**b**) applied to simulated data of a Huser-Genton process as in Fig. 6.1 with 250 observations. The number of clusters used is five. The colors within the clusters show the estimated value for parameter b on the stationary-max-stable process that has been fitted to the data on the cluster. For reference, the true values of the parameter b_s used to simulate the data are depicted in (**c**).

algorithms. In a first investigation, we apply the algorithms using as global parameters the true values of ν and α. In Fig. 6.3a and Fig. 6.3b, the clusters produced by the two algorithms are displayed. Stationary max-stable processes are fitted to the data in the clusters, and the color inside each cluster in Fig. 6.3a and Fig. 6.3b depicts the value of the corresponding estimate for the parameter b. For reference, the true values of the parameter b_s are depicted in Fig. 6.3c. The true values of the other two parameters a_s and γ_s are constant over the whole space; their estimates are also similar for all clusters and are not depicted. It can be observed that the clusters of the EDC algorithm are similar to those derived when applying the EDC algorithm using the true values (Fig. 6.2). In particular, as in the theoretical case, there is considerable variation in the true values of b_s within some of the clusters. A fitted stationary process cannot account for that variation. The LEC algorithm results in clusters that form vertical stripes, and on these clusters there is less variation in the true values of b_s. The fitted values on the clusters are therefore often closer to the true values than for the EDC algorithm (compare Fig. 6.3b and Fig. 6.3c).

The two clustering algorithms are compared using the method described in Section 6.2.3. In Fig. 6.4, we depict the intersections of the clusters of the two algorithms. The color of each region indicates which algorithm has the better goodness of fit there (darker color — EDC, lighter color — LEC). It can be observed that the LEC algorithm results in a better goodness of fit on most regions.

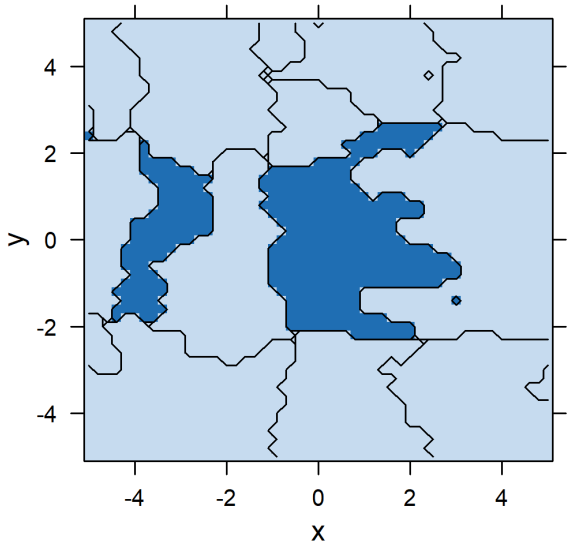

Figure 6.4: Comparison of the goodness of fit of the two clustering algorithms. For each intersection of the clusters of the LEC and the EDC algorithm from Fig. 6.3, the goodness of fit of the two algorithms is compared. Dark blue color indicates that the EDC algorithm has a better goodness of fit, light blue colors indicates a better goodness of fit of the LEC algorithm.

The above analyses present the results for just one simulation and using the true values for ν and α. For a more general investigation, the simulation of a non-stationary process and its investigation is repeated another 24 times. As the true values of ν and α are not known in practical applications, we conduct the analyses also for other parameter values, using for ν the values $3, 5,$ and 7 and for α the values $0.7, 1.0, 1.3$. The resulting clusters are of course slightly different each time, but the general structures that can be identified in Fig. 6.3 stay the same (not depicted). In Fig. 6.5, for each combination of the values for ν and α, we depict for each point $s \in \mathcal{S}$ the percentage of the 24 simulations for which the LEC algorithms has a better goodness of fit on the cluster the point is in. It can be observed that for each choice of the global parameter values and throughout the area of investigation, the LEC algorithm exhibits at each point a better goodness of fit for more than 75% of the realizations.

As a second example for a non-stationary process we use a process with parameters $a_s = 1$ constant, $b_s = 3$ constant and $g_s = (-x + 5) \cdot \pi/2$. The dependence structure around each point is locally an ellipse with the angle of the major axis rotating clockwise with increasing x (vertical for $x = -5$, horizontal for $x = 0$, vertical again for $x = 5$). In Fig. 6.6a and Fig. 6.6b we depict for one simulation the clusters produced by the two algorithms, with the colors of the clusters indicating the value of the estimate of parameter g of the fitted processes. We depict the true values of parameter g_s in Fig. 6.6c. Again, the clusters of the EDC algorithm group points with a high interdependency together, so their shape follows the rotation of the ellipses (this is visible especially well in the cluster in the middle of Fig. 6.6a and the two clusters above it), while the LEC algorithm results in clusters that form vertical stripes and reconstruct the spatial structure in the parameter values g_s, $s \in \mathcal{S}$ better. The analysis of the goodness of fit yields similar results as for the first algorithm, with the LEC algorithm being preferred in at least 75% of the cases for every point.

As a third example, we use values of $a_s = (7.5 - \|s\|)/2 + 1$, $b_s = 0$, $g_s = 0$. This time, the true spatial structure is a bit different, it does not feature vertical stripes, but instead a circular structure with the values of g_s depending on the distance of s to the center. This circular structure is visible in the clusters of the LEC algorithm (Fig. 6.7b). It is not reproduced by the EDC algorithm, which results in clusters that are quite uninformative this time, as the fitted values for a are very similar for each cluster (Fig. 6.7a). The results for the goodness of fit are similar to those for the two examples before.

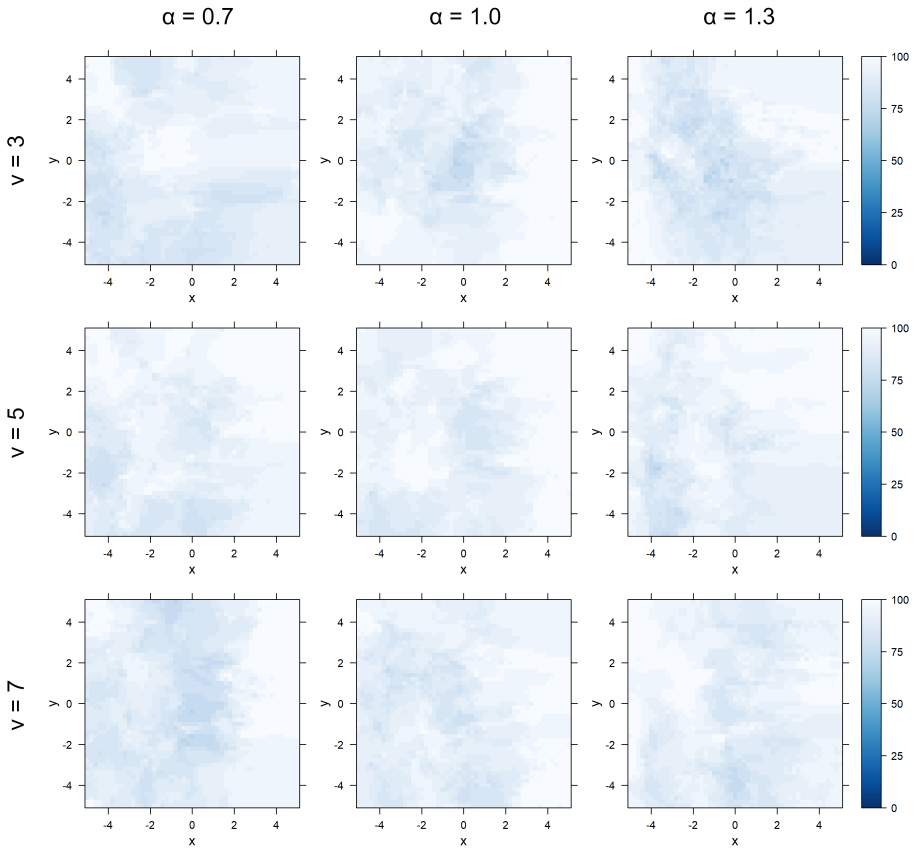

Figure 6.5: The analysis depicted in Fig. 6.4 is repeated for 24 independent simulations and the clustering algorithms are applied using different values for the global parameters ν and α (true values: 5 and 1.0, respectively). For each point $s \in \S$, the percentage of results for which the LEC algorithm has a better goodness of fit than EDC algorithm on the region the point is in is depicted.

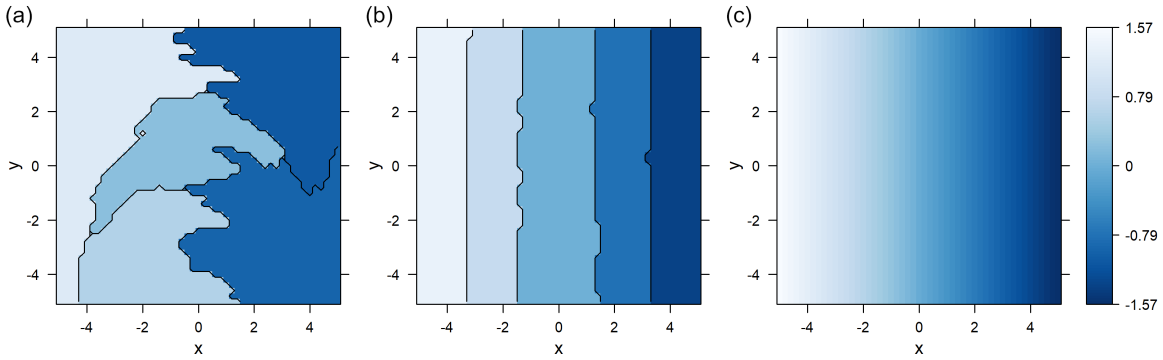

Figure 6.6: Results of the EDC algorithm (**a**) and the LEC algorithm (**b**) applied to simulated data of a Huser-Genton process with parameters $a_s = 1$ constant, $b_s = 3$ constant and $g_s = (-x + 5) \cdot \pi/2$ with 250 observations. The number of clusters used is five. The colors within the clusters show the estimated value for parameter g on the stationary-max-stable process that has been fitted to the data on the cluster. For reference, the true values of the parameter g_s are depicted in (**c**).

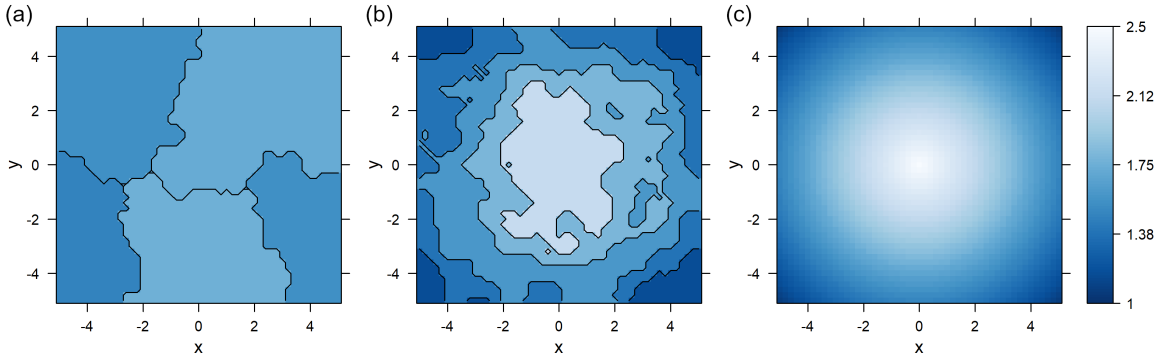

Figure 6.7: Results of the EDC algorithm (**a**) and the LEC algorithm (**b**) applied to simulated data of a Huser-Genton process with parameters $a_s = (7.5 - \|s\|)/2 + 1$, $b_s = 0$, $g_s = 0$ with 250 observations. The number of clusters used is five. The colors within the clusters show the estimated value for parameter a on the stationary-max-stable process that has been fitted to the data on the cluster. For reference, the true values of the parameter a_s are depicted in (**c**).

6.4 Conclusions of this chapter and discussion

We have discussed a clustering algorithm by Saunders et al. (2021) that is used in multivariate extreme value theory to group an area of investigation into smaller regions. The obtained regions are then used to fit parametric stationary max-stable processes to the data. This allows the application of such processes also to data for which stationarity on the whole area cannot be assumed. While clustering is in general a sensible and valid approach to the problem, the dissimilarity measure used previously is not necessarily suitable to find regions in which stationarity can be assumed, as we have discussed for one concrete example of a non-stationary max-stable process. We propose a different dissimilarity measure based on local estimates and demonstrate in a simulation study that for three different examples of non-stationary data we used, it indeed reconstructs the dependency structures of the data more accurately and that the processes that were fitted to the data on the clusters mostly have a better goodness of fit. For more general results regarding the performance of the two methods, a more formal and mathematically more strict analysis is required.

The clusters derived using the method by Saunders et al. (2021) do have a useful and meaningful interpretation in another context: They show regions in which there is a high extremal dependency between pairs of points. Therefore, a large cluster in a certain area indicates a tendency for more large-scale extreme events in that region. Finding such clusters is of relevance for example in the context of insurances or risk management.

The clustering approach in this chapter and the subsequent fitting of max-stable processes requires choosing two global parameters, and so far, we have not found a systematic method to do so. For exactly two different choices of the values, the clusters could be calculated and the goodness of fit of the processes could be compared using the measure from Section 6.2.3. Unfortunately, this measure does not allow for the comparison of more than two clusterings at the same time. Carrying out a lot of pairwise comparisons is time-consuming and does not seem to be a very convenient approach. Besides, there is no guarantee that the results of these pairwise comparisons do not contradict each other. A more sophisticated approach is definitely desirable here.

A more general limitation with the approach of using cluster algorithms and then fitting regional max-stable models is that while those models can be used to describe the data within one cluster and also give meaningful information about how dependence

structures vary spatially, they do not enable us to model directly the dependency between two points that are in different clusters. In this regard, the clustering approach is inferior to non-stationary max-stable processes like the ones presented in Huser and Genton (2016) (which are on the other hand more difficult to apply and require the availability of suitable covariates). It is an interesting future research direction to use the max-stable processes that have been fitted on the clusters and try to find a way to combine and extend them to a parsimonious process covering the whole area of investigation.

Another possible application of the clustering algorithms is the spatio-temporal investigation of extremes. When modeling changes over time using GEV distributions with time-dependent parameters like it was done in the last chapter, the data can be transformed to have a unit Fréchet distribution over the whole time frame of investigation. By applying clustering algorithms to the transformed data at different time windows, it is possible to combine the temporal investigation of changes in extremes with an investigation of changes in the spatial dependence structure.

Chapter 7

Final considerations

7.1 Conclusions

In this chapter, we will recapitulate the results from the previous chapters and outline how they contributed to answering the research questions we laid out in the introduction. Afterwards, we will give a brief outlook and discuss potential areas for further research based on the methods and results used here.

1. *Using historical runs of climate models, how well do the statistical properties of climate extremes in the simulated data match with those of observed climate extremes?*

To address this question, we compared annual precipitation maxima between climate models, in particular the AWI-ESM, and an observational data set. We compared empirical statistical parameters and the estimated parameters of GEV distributions which we fitted to the data. We also used a clustering approach to perform an analysis of the spatial concurrence of extremes in both data sets. As a result, we could observe that the distributions of historical extremes are similar for both data sets in many regions of the world. However, we also identified larger regions of over- and underestimation of extremes by the climate model. These misestimations often consist of a misestimation of both the location and the scale parameter of the GEV distribution, while the shape parameter of the GEV distribution (which indicates the heavy-tailedness) is in general similar between model and observational data. Because of the difficulties in reliably estimating the shape parameter this last result has to be taken with caution.

The areas in which misestimation takes place are predominantly mountainous areas like the Andes or the Himalayas. This might be the case due to the fact that the territory in those areas is very fine-grained, and the climate is very much influenced by local small-scale topography. As climate models perform their calculations on spatially averaged grid cells, a complex and variable topography inside a grid cell cannot be adequately modeled and the resulting extremes are not accurately described. In addition to the AWI-ESM, several other CMIP6 models are also analyzed, and a comparison of the model accuracy between the models shows a tendency for the higher-dimensional models to capture the extremal behavior better, supporting this explanation further.

The cluster analysis we performed identifies regions with a tendency for concurrent extremes. There is generally a good agreement between the clusters for model and observational data. Also the estimated number of clusters is in general similar for both data sets, with a slight tendency for a higher cluster number for the model data. Since it is mostly large-scale weather events and teleconnections contributing to concurrent climate extremes, this may indicate that the basic physical behavior underlying them is in general well captured by the climate models. The lower number of clusters when applying the clustering algorithms to climate model data might indicate that the complexity and regional variability of the climate system is not completely reproduced by the models.

To conclude, the statistical distributions of extremes are in general well met by the climate models, which motivates the use of climate models also to predict future changes in extremes. In mountainous regions, however, the accuracy of the models is reduced, indicating that the spatial resolution of the models is not sufficient for an accurate analysis of those regions. The future predictions of extremes in those areas should therefore be taken with caution. The accuracy of the modeling of extremes in mountainous areas may be improved in future generations of climate models that have a higher spatial resolution.

2. *How do climate models predict climate extremes to change in the future?*

To address this question, we analyzed future simulations of temperature extremes for several climate models. We fitted non-stationary GEV distributions to the data, with the parameters of the distributions changing in a slow–fast–slow pattern. The changes in the different distribution parameters are allowed to take place independently and at varying time periods, and we tested and compared several different statistical models.

We can conclude that according to the climate models, the total magnitude of extremes, measured by analyzing the 95% quantiles of the annual maxima, will increase strongly over time in most regions of the world. A particularly large change is predicted over land, and this corresponds to the well-known fact that global warming is stronger over land than over the ocean. The different models in general agree on this pattern, however, they differ widely in terms of the predicted amounts of the change.

A geographically varying development can be detected not only with regard to the magnitude of changes, but also with regard to their timing, and to the extent to which the changes in extreme events are caused by changes in the location, the scale or the shape of the distribution of the annual maxima. Changes over the ocean take place more gradually while changes over land are faster. In most regions, changes in mean and variance take place simultaneously while the shape parameter of the distribution is predicted to stay constant. In the Arctic region, however, a different picture emerges: There, climate variability is predicted to increase rather quickly in the second half of the twenty-first century, probably due to the melting of ice, whereas changes in the mean values take longer and come into effect later.

3. *How can statistical methods to describe the spatial structure of extremes be improved?*

To address this question, we investigated regionalization approaches for max-stable processes. Max-stable processes can be used for the investigation of climate extremes and their spatial dependencies on a continuous area. Most existing parametric models of max-stable processes assume spatial stationarity and are therefore not suitable for the application to data that cover a large and heterogeneous area. For this reason, it has recently been proposed to use a clustering algorithm to divide the area into smaller regions and fit parametric max-stable processes to the data within those regions.

This clustering algorithm groups points together if their pairwise extremal coefficient is low, which implies a strong extremal dependency of the points. As pointed out in the chapter, this is not necessarily a good criterion for the clustering, as there are cases in which it results in regions on which spatial stationarity is not a reasonable assumption. We propose an alternative clustering algorithm which is based on the comparison of local estimates of the dependence structures around the different points. In a simulation study

we demonstrate that this new approach can lead to improved results.

7.2 Outlook

The validation of climate models demonstrated that current global models still have difficulties in mountainous areas with their fine-grained topography. As mentioned, future climate models with a higher spatial resolution could provide an improvement there. Another approach that is currently already applicable is the investigation of regional climate models. Those models can be run with higher resolutions than the global models, and a comparison of global and regional climate models in terms of extremes would provide valuable insights.

This investigation of future changes in climate extremes demonstrated that extremes are predicted to change strongly in the future, highlighting the dangers that are associated with that. The currently observable changes in magnitude and frequency of extreme events already cause "widespread adverse impacts on food and water security, human health and on economies and society and related losses and damages to nature and people" (IPCC, 2023). With even stronger extremes in the future, these problems will become more severe. In our analysis it could be observed that changes in extremes depend strongly on the region, and this regionally varying behavior needs to be taken into account when assessing the impacts of climate change. However, our study also pointed out that there are still large discrepancies between the predictions of the different models, even though they have been run using the same RCP scenario, indicating that the extent of the changes is still controversial. Further investigations on the causes of those discrepancies would be instructive.

The study of the different clustering algorithms to model spatial dependencies of extremes suggested that our newly proposed method can address some limitations of the previous one. As of now, these results rely mostly on simulation studies. A mathematically more sound treatment of the topic, maybe including proofs of some of the properties of the different cluster algorithm, would be desirable. Afterwards, an application of the methods to climate model data would be a logical next step to investigate spatial dependencies of extremes in climate model data in more detail.

The clustering algorithms can also be applied to data that is not stationary over

time and to which time-dependent GEV distributions have been fitted (like in Chapter 5). When working with such data, clusters can be calculated for subsets of the data on different time intervals. The comparison of the clusters would shed light on how spatial dependencies change over time. Such an investigation could be a first step towards a complete spatio-temporal model for extremes on a global scale.

Another possible field of application of the results in this work is the investigation of risks associated with extremes. The analysis of such risks is crucial in particular for insurance and reinsurance companies. While univariate risk analysis is well-established, multivariate spatial analysis of risks is still an active research topic. Recently, one approach to investigate spatial risks using max-stable processes has been developed by Koch (2017). It would be interesting to combine this approach with non-stationary max-stable models or with clustering approaches, and on the other hand also to investigate possible temporal changes of risks. According to the IPCC, "[d]epending on the level of global warming, the assessed long-term impacts [of climate change] will be up to multiple times higher than currently observed for 127 identified key risks, e.g., in terms of the number of affected people and species. Risks [...] are projected to become increasingly severe with every increment of global warming" (IPCC, 2023). The methods and approaches presented in this work and the results that were derived from them might offer possibilities to quantify future risks more accurately and, in doing so, to lay ground for the evaluation of suitable adaptive measures to mitigate their effects.

Bibliography

L. Ackermann, C. Danek, P. Gierz, and G. Lohmann. AMOC Recovery in a Multicentennial Scenario Using a Coupled Atmosphere-Ocean-Ice Sheet Model. *Geophysical Research Letters*, 47(16), 2020. doi: 10.1029/2019GL086810.

G. Altmann. The Piotrowski law and its generalizations. In *Exact language change research. Theoretical contributions, statistical analyzes and work reports*, volume 49, pages 45–90. edition herodot, Göttingen, 1983.

M. Bador, J. Boé, L. Terray, L. V. Alexander, A. Baker, A. Bellucci, R. Haarsma, T. Koenigk, M.-P. Moine, K. Lohmann, D. A. Putrasahan, C. Roberts, M. Roberts, E. Scoccimarro, R. Schiemann, J. Seddon, R. Senan, S. Valcke, and B. Vanniere. Impact of Higher Spatial Atmospheric Resolution on Precipitation Extremes Over Land in Global Climate Models. *Journal of Geophysical Research: Atmospheres*, 125(13), 2020. doi: 10.1029/2019JD032184.

Y. J. Bagul and C. Chesneau. Sigmoid functions for the smooth approximation to the absolute value function. *Moroccan Journal of Pure and Applied Analysis*, 7(1):12–19, 2021. doi: doi:10.2478/mjpaa-2021-0002.

E. Bernard, P. Naveau, M. Vrac, and O. Mestre. Clustering of Maxima: Spatial Dependencies among Heavy Rainfall in France. *Journal of Climate*, 26(20):7929–7937, 2013. doi: 10.1175/JCLI-D-12-00836.1.

O. Boucher, S. Denvil, G. Levavasseur, A. Cozic, A. Caubel, M.-A. Foujols, Y. Meurdesoif, P. Cadule, M. Devilliers, J. Ghattas, N. Lebas, T. Lurton, L. Mellul, I. Musat, J. Mignot, and F. Cheruy. IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP historical. Version 20180803. Earth System Grid Federation [data set], 2018. URL `https://doi.org/10.22033/ESGF/CMIP6.5195`.

C. M. Brierley, A. Zhao, S. P. Harrison, P. Braconnot, C. J. R. Williams, D. J. R. Thornalley, X. Shi, J.-Y. Peterschmitt, R. Ohgaito, D. S. Kaufman, M. Kageyama, J. C. Hargreaves,

M. P. Erb, J. Emile-Geay, R. D'Agostino, D. Chandan, M. Carré, P. J. Bartlein, W. Zheng, Z. Zhang, Q. Zhang, H. Yang, E. M. Volodin, R. A. Tomas, C. Routson, W. R. Peltier, B. Otto-Bliesner, P. A. Morozova, N. P. McKay, G. Lohmann, A. N. Legrande, C. Guo, J. Cao, E. Brady, J. D. Annan, and A. Abe-Ouchi. Large-scale features and evaluation of the PMIP4-CMIP6 midHolocene simulations. *Climate of the Past*, 16(5):1847–1872, 2020. doi: 10.5194/cp-16-1847-2020.

B. M. Brown and S. I. Resnick. Extreme Values of Independent Stochastic Processes. *Journal of Applied Probability*, 14(4):732–739, 1977. doi: 10.2307/3213346.

J. R. Brown, C. M. Brierley, S.-I. An, M.-V. Guarino, S. Stevenson, C. J. R. Williams, Q. Zhang, A. Zhao, A. Abe-Ouchi, P. Braconnot, E. C. Brady, D. Chandan, R. D'Agostino, C. Guo, A. N. LeGrande, G. Lohmann, P. A. Morozova, R. Ohgaito, R. O'ishi, B. L. Otto-Bliesner, W. R. Peltier, X. Shi, L. Sime, E. M. Volodin, Z. Zhang, and W. Zheng. Comparison of past and future simulations of ENSO in CMIP5/PMIP3 and CMIP6/PMIP4 models. *Climate of the Past*, 16(5):1777–1805, 2020. doi: 10.5194/cp-16-1777-2020.

A. Bücher and J. Segers. On the maximum likelihood estimator for the Generalized Extreme-Value distribution. *Extremes*, 20:839–872, 2017. doi: 10.1007/s10687-017-0292-6.

U. Büntgen, V. S. Myglan, F. C. Ljungqvist, M. McCormick, N. di Cosmo, M. Sigl, J. Jungclaus, S. Wagner, P. J. Krusic, J. Esper, J. O. Kaplan, M. A. C. de Vaan, J. Luterbacher, L. Wacker, W. Tegel, and A. V. Kirdyanov. Cooling and societal change during the Late Antique Little Ice Age from 536 to around 660 AD. *Nature Geoscience*, 9(3):231–236, 2016. doi: 10.1038/ngeo2652.

M. P. Byrne and P. A. O'Gorman. Trends in continental temperature and humidity directly linked to ocean warming. *Proceedings of the National Academy of Sciences*, 115(19):4863–4868, 2018. doi: 10.1073/pnas.1722312115.

A. J. Cannon. A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology. *Hydrological Processes*, 24(6):673–685, 2010. doi: 10.1002/hyp.7506.

J. Cao and B. Wang. NUIST NESMv3 model output prepared for CMIP6 CMIP historical. Version 20190701. Earth System Grid Federation [data set], 2019. URL `https://doi.org/10.22033/ESGF/CMIP6.8769`.

O. Cardona, M. van Aalst, J. Birkmann, M. Fordham, G. McGregor, R. Perez, R. Pulwarty, E. Schipper, , and B. Sinh. Determinants of risk: exposure and vulnerability. In C. Field,

V. Barros, T. Stocker, D. Qin, D. Dokken, K. Ebi, K. M. M.D. Mastrandrea, G.-K. Plattner, S. Allen, M. Tignor, , and P. Midgley, editors, *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*, pages 65–108. Cambridge University Press, 2012.

T. A. Carleton and S. M. Hsiang. Social and economic impacts of climate. *Science*, 353(6304), 2016. doi: 10.1126/science.aad9837.

M. Carvalho, P. Melo-Gonçalves, J. Teixeira, and A. Rocha. Regionalization of Europe based on a K-Means Cluster Analysis of the climate change of temperatures and precipitation. *Physics and Chemistry of the Earth Parts A/B/C*, 94:22–28, 2016. doi: 10.1016/j.pce.2016.05.001.

B. Casati and R. de Elía. Temperature Extremes from Canadian Regional Climate Model (CRCM) Climate Change Projections. *Atmosphere-Ocean*, 52(3):191–210, 2014. doi: 10.1080/07055900.2014.886179.

P. Ciais, M. Reichstein, N. Viovy, A. Granier, J. Ogée, V. Allard, M. Aubinet, N. Buchmann, C. Bernhofer, A. Carrara, F. Chevallier, N. D. Noblet, A. D. Friend, P. Friedlingstein, T. Grünwald, B. Heinesch, P. Keronen, A. Knohl, G. Krinner, D. Loustau, G. Manca, G. Matteucci, F. Miglietta, J. M. Ourcival, D. Papale, K. Pilegaard, S. Rambal, G. Seufert, J. F. Soussana, M. J. Sanz, E. D. Schulze, T. Vesala, and R. Valentini. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, 437:529–533, 2005. doi: 10.1038/nature03972.

R. Cleveland, W. Cleveland, J. McRae, and I. Terpenning. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–33, 1990.

S. Coles, L. R. Pericchi, and S. Sisson. A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology*, 273(1):35–50, 2003. doi: 10.1016/S0022-1694(02)00353-0.

J. Contzen, T. Dickhaus, and G. Lohmann. Variability and extremes: statistical validation of the Alfred Wegener Institute Earth System Model (AWI-ESM). *Geoscientific Model Development*, 15(4):1803–1820, 2022. doi: 10.5194/gmd-15-1803-2022.

J. Contzen, T. Dickhaus, and G. Lohmann. Long-term temporal evolution of extreme temperature in a warming Earth. *PLOS ONE*, 18(2):1–30, 2023. doi: 10.1371/journal.pone.0280503.

J. Contzen, T. Dickhaus, and G. Lohmann. Regionalization approaches for the spatial analysis of extremal dependence. *Extremes*, in press.

D. Cooley, P. Naveau, and P. Poncet. Variograms for spatial max-stable random fields. In P. Bertail, P. Soulier, and P. Doukhan, editors, *Dependence in Probability and Statistics*. Springer, New York, 2006.

C. M. Cooney. Managing the Risks of Extreme Weather: IPCC Special Report. *Environmental Health Perspectives*, 120(2):a58–a58, 2012. doi: 10.1289/ehp.120-a58.

G. Danabasoglu. NCAR CESM2 model output prepared for CMIP6 CMIP historical. Version 20190514. Earth System Grid Federation [data set], 2019a. URL `https://doi.org/10.22033/ESGF/CMIP6.7627`.

G. Danabasoglu. NCAR CESM2-FV2 model output prepared for CMIP6 CMIP historical. Version 20200226. Earth System Grid Federation [data set], 2019b. URL `https://doi.org/10.22033/ESGF/CMIP6.11297`.

G. Danabasoglu. NCAR CESM2-WACCM-FV2 model output prepared for CMIP6 CMIP historical. Version 20200226. Earth System Grid Federation [data set], 2019c. URL `https://doi.org/10.22033/ESGF/CMIP6.11298`.

C. Danek, X. Shi, C. Stepanek, H. Yang, D. Barbi, J. Hegewald, and G. Lohmann. AWI AWI-ESM1.1LR model output prepared for CMIP6 CMIP historical. Version 20200212. Earth System Grid Federation [data set], 2020. URL `https://doi.org/10.22033/ESGF/CMIP6.9328`.

R. A. Davis, C. Klüppelberg, and C. Steinkohl. Max-stable processes for modeling extremes observed in space and time. *Journal of the Korean Statistical Society*, 42(3):399–414, 2013. doi: 10.1016/j.jkss.2013.01.002.

L. de Haan. A spectral representation for max-stable processes. *The Annals of Probability*, 12(4):1194–1204, 1984. doi: 10.1214/aop/1176993148.

L. de Haan and A. Ferreira. *Extreme value theory: An introduction*. Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006.

R. J. Delmas. Environmental information from ice cores. *Reviews of Geophysics*, 30(1):1–21, 1992. doi: 10.1029/91RG02725.

T. Dinku. Challenges with availability and quality of climate data in africa. In A. M. Melesse, W. Abtew, and G. Senay, editors, *Extreme Hydrology and Climate Variability*, pages 71–80. Elsevier, 2019. ISBN 978-0-12-815998-9. doi: 10.1016/B978-0-12-815998-9.00007-5.

M. Dix, D. Bi, P. Dobrohotoff, R. Fiedler, I. Harman, R. Law, C. Mackallah, S. Marsland, S. O'Farrell, H. Rashid, J. Srbinovsky, A. Sullivan, C. Trenham, P. Vohralik, I. Watterson, G. Williams, M. Woodhouse, R. Bodman, F. B. Dias, C. M. Domingues, N. Hannah, A. Heerdegen, A. Savita, S. Wales, C. Allen, K. Druken, B. Evans, C. Richards, S. M. Ridzwan, D. Roberts, J. Smillie, K. Snow, M. Ward, and R. Yang. CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 CMIP historical. Version 20210607. Earth System Grid Federation [data set], 2019. URL `https://doi.org/10.22033/ESGF/CMIP6.4271`.

M. G. Donat, J. Sillmann, and E. M. Fischer. Chapter 3 - Changes in climate extremes in observations and climate model simulations. From the past to the future. In J. Sillmann, S. Sippel, and S. Russo, editors, *Climate Extremes and Their Implications for Impact and Risk Assessment*, pages 31–57. Elsevier, 2020. ISBN 978-0-12-814895-2. doi: 10.1016/B978-0-12-814895-2.00003-3.

B. Dong, R. T. Sutton, and L. Shaffrey. Understanding the rapid summer warming and changes in temperature extremes since the mid-1990s over Western Europe. *Climate Dynamics*, 48:1537–1554, 2017. doi: 10.1007/s00382-016-3158-8.

P. M. J. Douglas, M. Pagani, M. A. Canuto, M. Brenner, D. A. Hodell, T. I. Eglinton, and J. H. Curtis. Drought, agricultural adaptation, and sociopolitical collapse in the Maya Lowlands. *Proceedings of the National Academy of Sciences*, 112(18):5607–5612, 2015. doi: 10.1073/pnas.1419133112.

EC-Earth Consortium. EC-Earth-Consortium EC-Earth3-Veg-LR model output prepared for CMIP6 CMIP historical. Version 20200217. Earth System Grid Federation [data set], 2020. URL `https://doi.org/10.22033/ESGF/CMIP6.4707`.

S. El Adlouni, T. B. M. J. Ouarda, X. Zhang, R. Roy, and B. Bobée. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, 43(3), 2007. doi: 10.1029/2005WR004545.

ETCCDI. The ETCCDI Climate Change Indices, 2009. URL `etccdi.pacificclimate.org/list{_}27{_}indices.shtml`. Last Access: 04/20/2023.

E. Fischer and R. Knutti. Observed heavy precipitation increase confirms theory and early models. *Nature Climate Change*, 6:986–991, 2016. doi: 10.1038/nclimate3110.

E. M. Fischer, S. Sippel, and R. Knutti. Increasing probability of record-shattering climate extremes. *Nature Climate Change*, 11(8):689–695, 2021. doi: 10.1038/s41558-021-01092-9.

R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190, 1928. doi: 10.1017/S0305004100015681.

R. G. Fovell and M.-Y. C. Fovell. Climate zones of the Conterminous United States defined using cluster analysis. *Journal of Climate*, 6(11):2103–2135, 1993. doi: 10.1175/1520-0442(1993)006\textless2103:CZOTCU\textgreater2.0.CO;2.

P. R. Gent, G. Danabasoglu, L. J. Donner, M. M. Holland, E. C. Hunke, S. R. Jayne, D. M. Lawrence, R. B. Neale, P. J. Rasch, M. Vertenstein, P. H. Worley, Z.-L. Yang, and M. Zhang. The Community Climate System Model Version 4. *Journal of Climate*, 24(19):4973–4991, 2011. doi: 10.1175/2011JCLI4083.1.

M. A. Giorgetta, J. Jungclaus, C. H. Reick, S. Legutke, J. Bader, M. Böttinger, V. Brovkin, T. Crueger, M. Esch, K. Fieg, K. Glushak, V. Gayler, H. Haak, H.-D. Hollweg, T. Ilyina, S. Kinne, L. Kornblueh, D. Matei, T. Mauritsen, U. Mikolajewicz, W. Mueller, D. Notz, F. Pithan, T. Raddatz, S. Rast, R. Redler, E. Roeckner, H. Schmidt, R. Schnur, J. Segschneider, K. D. Six, M. Stockhause, C. Timmreck, J. Wegner, H. Widmann, K.-H. Wieners, M. Claussen, J. Marotzke, and B. Stevens. Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, 5(3):572–597, 2013. doi: 10.1002/jame.20038.

B. Gnedenko. Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44(3):423–453, 1943. doi: 10.2307/1968974.

S. Hagemann and L. Dümenil. A parametrization of the lateral waterflow for the global scale. *Climate Dynamics*, 14:17–31, 1997. doi: 10.1007/s003820050205.

T. Hajima, M. Abe, O. Arakawa, T. Suzuki, Y. Komuro, T. Ogura, K. Ogochi, M. Watanabe, A. Yamamoto, H. Tatebe, M. A. Noguchi, R. Ohgaito, A. Ito, D. Yamazaki, A. Ito, K. Takata, S. Watanabe, M. Kawamiya, and K. Tachiiri. MIROC MIROC-ES2L model output prepared for CMIP6 CMIP historical. Version 20191118. Earth System Grid Federation [data set], 2019. URL `https://doi.org/10.22033/ESGF/CMIP6.5602`.

M. Hanel and T. A. Buishand. On the value of hourly precipitation extremes in regional climate model simulations. *Journal of Hydrology*, 393(3):265–273, 2010. doi: 10.1016/j.jhydrol.2010.08.024.

J. Hansen, L. Nazarenko, R. Ruedy, M. Sato, J. Willis, A. D. Genio, D. Koch, A. Lacis, K. Lo, S. Menon, T. Novakov, J. Perlwitz, G. Russell, G. A. Schmidt, and N. Tausnev. Earth's Energy Imbalance: Confirmation and Implications. *Science*, 308(5727):1431–1435, 2005. doi: 10.1126/science.1110252.

C. Harald. *Mathematical methods of statistics*. Princeton University Press, Princeton, 1946.

I. Harris, T. Osborn, P. Jones, and D. Lister. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7, 2020. doi: 10.1038/s41597-020-0453-3.

S. He, Y. Peng, Y. Jin, B. Wan, and G. Liu. Review and Analysis of Key Techniques in Marine Sediment Sampling. *Chinese Journal of Mechanical Engineering*, 33(66), 2020. doi: 10.1186/s10033-020-00480-0.

G. Helle and G. H. Schleser. Interpreting climate proxies from tree-rings. In H. Fischer, T. Kumke, G. Lohmann, G. Flöser, H. Miller, H. von Storch, and J. F. W. Negendank, editors, *The Climate in Historical Times: Towards a Synthesis of Holocene Proxy Data and Climate Models*, pages 129–148. Springer, Berlin, Heidelberg, 2004. ISBN 978-3-662-10313-5. doi: 10.1007/978-3-662-10313-5{_}8.

E. Hogan, R. E. Nicholas, K. Keller, S. Eilts, and R. L. Sriver. Representation of U.S. Warm Temperature Extremes in Global Climate Model Ensembles. *Journal of Climate*, 32(9):2591–2603, 2019. doi: 10.1175/JCLI-D-18-0075.1.

J. R. Holton and G. J. Hakim. *An Introduction to Dynamic Meteorology (Fifth Edition)*. Academic Press, Boston, 2013. doi: 10.1016/B978-0-12-384866-6.00030-1.

R. Horton, J. Mankin, C. Lesk, E. Coffel, and C. Raymond. A review of recent advances in research on extreme heat events. *Current Climate Change Reports*, 2(4):242–259, 2016. doi: 10.1007/s40641-016-0042-x.

J. Hosking, J. Wallis, and E. Wood. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261, 1985. doi: 10.1080/00401706.1985.10488049.

R. Huser and M. G. Genton. Non-Stationary Dependence Structures for Spatial Extremes. *Journal of Agricultural, Biological, and Environmental Statistics*, 21:470–491, 2016. doi: 10.1007/s13253-016-0247-4.

C. E. Iles, R. Vautard, J. Strachan, S. Joussaume, B. R. Eggen, and C. D. Hewitt. The benefits of increasing resolution in global and regional climate simulations for European climate extremes. *Geoscientific Model Development*, 13(11):5583–5607, 2020. doi: 10.5194/gmd-13-5583-2020.

M. Ionita, M. Dima, V. Nagavciuc, and G. Lohmann. Past megadroughts in central Europe were longer, more severe and less warm than modern droughts. *Communications Earth and Environment*, 2(61), 2021. doi: 10.1038/s43247-021-00130-w.

IPCC. *Climate Change 2001: The Scientific Basis. Contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 2001.

IPCC. *Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 2012. doi: 10.13140/2.1.3117.9529.

IPCC. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2021.

IPCC. Summary for Policymakers. In Core Writing Team, H. Lee, and J. Romero, editors, *Climate Change 2023: Synthesis Report. A Report of the Intergovernmental Panel on Climate Change. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva, Switzerland, (in press), 2023.

S. Jeffrey, L. Rotstayn, M. Collier, S. Dravitzki, C. Hamalainen, C. Moeseneder, K. Wong, and J. Syktus. Australia's CMIP5 submission using the CSIRO-Mk3.6 model. *Australian Meteorological and Oceanographic Journal*, 63:1–13, 2013. doi: 10.22499/2.6301.001.

P. D. Jones. Early european instrumental records. In P. D. Jones, A. E. J. Ogilvie, T. D. Davies, and K. R. Briffa, editors, *History and Climate: Memories of the Future?*, pages 55–77. Springer US, Boston, MA, 2001. ISBN 978-1-4757-3365-5. doi: 10.1007/978-1-4757-3365-5{_}4.

B. Jongman, S. Hochrainer-Stigler, L. Feyen, J. Aerts, R. Mechler, W. Botzen, L. Bouwer, G. Pflug, R. Rojas, and P. Ward. Increasing stress on disaster-risk finance due to large floods. *Nature Climate Change*, 4:264–268, 2014. doi: 10.1038/NCLIMATE2124.

J. Jungclaus, M. Bittner, K.-H. Wieners, F. Wachsmann, M. Schupfner, S. Legutke, M. Giorgetta, C. Reick, V. Gayler, H. Haak, P. de Vrese, T. Raddatz, M. Esch, T. Mauritsen, J.-S. von Storch, J. Behrens, V. Brovkin, M. Claussen, T. Crueger, I. Fast, S. Fiedler, S. Hagemann, C. Hohenegger, T. Jahns, S. Kloster, S. Kinne, G. Lasslop, L. Kornblueh, J. Marotzke, D. Matei, K. Meraner, U. Mikolajewicz, K. Modali, W. Müller, J. Nabel, D. Notz, K. Peters-von Gehlen, R. Pincus, H. Pohlmann, J. Pongratz, S. Rast, H. Schmidt, R. Schnur, U. Schulzweida, K. Six, B. Stevens, A. Voigt, and E. Roeckner. MPI-M MPI-ESM1.2-HR model output prepared for CMIP6 CMIP historical. Version 20190710. Earth System Grid Federation [data set], 2019. URL `https://doi.org/10.22033/ESGF/CMIP6.6594`.

Z. Kabluchko, M. Schlather, and L. de Haan. Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 37(5):2042–2065, 2009. doi: 10.1214/09-AOP455.

M. Kageyama, S. P. Harrison, M.-L. Kapsch, M. Lofverstrom, J. M. Lora, U. Mikolajewicz, S. Sherriff-Tadano, T. Vadsaria, A. Abe-Ouchi, N. Bouttes, D. Chandan, L. J. Gregoire, R. F. Ivanovic, K. Izumi, A. N. LeGrande, F. Lhardy, G. Lohmann, P. A. Morozova, R. Ohgaito, A. Paul, W. R. Peltier, C. J. Poulsen, A. Quiquet, D. M. Roche, X. Shi, J. E. Tierney, P. J. Valdes, E. Volodin, and J. Zhu. The PMIP4 Last Glacial Maximum experiments: preliminary results and comparison with the PMIP3 simulations. *Climate of the Past*, 17(3):1065–1089, 2021a. doi: 10.5194/cp-17-1065-2021.

M. Kageyama, L. C. Sime, M. Sicard, M.-V. Guarino, A. de Vernal, R. Stein, D. Schroeder, I. Malmierca-Vallet, A. Abe-Ouchi, C. Bitz, P. Braconnot, E. C. Brady, J. Cao, M. A. Chamberlain, D. Feltham, C. Guo, A. N. LeGrande, G. Lohmann, K. J. Meissner, L. Menviel, P. Morozova, K. H. Nisancioglu, B. L. Otto-Bliesner, R. O'ishi, S. Ramos Buarque, D. Salas y Melia, S. Sherriff-Tadano, J. Stroeve, X. Shi, B. Sun, R. A. Tomas, E. Volodin, N. K. H. Yeung, Q. Zhang, Z. Zhang, W. Zheng, and T. Ziehn. A multi-model CMIP6-PMIP4 study of Arctic sea ice at 127 ka: sea ice data compilation and model differences. *Climate of the Past*, 17(1):37–62, 2021b. doi: 10.5194/cp-17-37-2021.

T. R. Karl, N. Nicholls, and A. Ghazi. CLIVAR/GCOS/WMO Workshop on Indices and Indicators for Climate Extremes Workshop Summary. In T. R. Karl, N. Nicholls, and A. Ghazi, editors, *Weather and Climate Extremes: Changes, Variations and a Perspective from*

the Insurance Industry, pages 3–7. Springer Netherlands, Dordrecht, 1999. doi: 10.1007/978-94-015-9265-9{_}2.

R. Katz and B. Brown. Extreme events in a changing climate: Variability is more important than averages. *Climatic Change*, 21:289–302, 1992. doi: 10.1007/BF00139728.

J. Keeble, B. Hassler, A. Banerjee, R. Checa-Garcia, G. Chiodo, S. Davis, V. Eyring, P. T. Griffiths, O. Morgenstern, P. Nowack, G. Zeng, J. Zhang, G. Bodeker, S. Burrows, P. Cameron-Smith, D. Cugnet, C. Danek, M. Deushi, L. W. Horowitz, A. Kubin, L. Li, G. Lohmann, M. Michou, M. J. Mills, P. Nabat, D. Olivié, S. Park, Ø. Seland, J. Stoll, K.-H. Wieners, and T. Wu. Evaluating stratospheric ozone and water vapour changes in CMIP6 models from 1850 to 2100. *Atmospheric Chemistry and Physics*, 21(6):5015–5061, 2021. doi: 10.5194/acp-21-5015-2021.

V. V. Kharin and F. W. Zwiers. Estimating Extremes in Transient Climate Change Simulations. *Journal of Climate*, 18(8):1156–1173, 2005. doi: 10.1175/JCLI3320.1.

Y.-H. Kim, S.-K. Min, X. Zhang, J. Sillmann, and M. Sandstad. Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. *Weather and Climate Extremes*, 29: 100269, 2020. doi: 10.1016/j.wace.2020.100269.

A. D. King, A. R. Borowiak, J. R. Brown, D. J. Frame, L. J. Harrington, S.-K. Min, A. Pendergrass, M. Rugenstein, J. M. K. Sniderman, and D. A. Stone. Transient and Quasi-Equilibrium Climate States at 1.5°C and 2°C Global Warming. *Earth's Future*, 9(11), 2021. doi: 10.1029/2021EF002274.

E. Koch. Spatial risk measures and applications to max-stable processes. *Extremes*, 20(3): 635–670, 2017. doi: 10.1007/s10687-016-0274-0.

U. Koedel, C. Schuetze, P. Fischer, I. Bussmann, P. K. Sauer, E. Nixdorf, T. Kalbacher, V. Wichert, D. Rechid, L. M. Bouwer, and P. Dietrich. Challenges in the evaluation of observational data trustworthiness from a data producers viewpoint. *Frontiers in Environmental Science*, 9, 2022. doi: 10.3389/fenvs.2021.772666.

R. S. Kovats and L. E. Kristie. Heatwaves and public health in Europe. *European Journal of Public Health*, 16(6):592–599, 2006. doi: 10.1093/eurpub/ckl049.

W. Kwasnicki. Logistic growth of the global economy and competitiveness of nations. *Technological Forecasting and Social Change*, 80(1):50–76, 2013. doi: 10.1016/j.techfore.2012.07.007.

W.-L. Lee and H.-C. Liang. AS-RCEC TaiESM1.0 model output prepared for CMIP6 CMIP historical. Version 20210416. Earth System Grid Federation [data set], 2020. URL `https://doi.org/10.22033/ESGF/CMIP6.9755`.

S. C. Lewis and A. D. King. Evolution of mean, variance and extremes in 21st century temperatures. *Weather and Climate Extremes*, 15:1–10, 2017.

L. Li. CAS FGOALS-g3 model output prepared for CMIP6 CMIP historical. Version 20190818. Earth System Grid Federation [data set], 2019. URL `https://doi.org/10.22033/ESGF/CMIP6.3356`.

G. Lohmann. ESD Ideas: The stochastic climate model shows that underestimated Holocene trends and variability represent two sides of the same coin. *Earth System Dynamics*, 9(4):1279–1281, 2018. doi: 10.5194/esd-9-1279-2018.

G. Lohmann. Temperatures from energy balance models: the effective heat capacity matters. *Earth System Dynamics*, 11(4):1195–1208, 2020. doi: 10.5194/esd-11-1195-2020.

G. Lohmann, M. Butzin, N. Eissner, X. Shi, and C. Stepanek. Abrupt Climate and Weather Changes Across Time Scales. *Paleoceanography and Paleoclimatology*, 35(9), 2020. doi: 10.1029/2019PA003782.

R. Lorenz, Z. Stalhandske, and E. M. Fischer. Detection of a Climate Change Signal in Extreme Heat, Heat Stress, and Cold in Europe From Observations. *Geophysical Research Letters*, 46(14):8363–8374, 2019. doi: 10.1029/2019GL082062.

S. Mahajan, K. Evans, M. Branstetter, V. Anantharaj, and J. Leifeld. Fidelity of Precipitation Extremes in High Resolution Global Climate Simulations. *Procedia Computer Science*, 51:2178–2187, 2015. doi: 10.1016/j.procs.2015.05.492.

H. D. Matthews, N. P. Gillett, P. A. Stott, and K. Zickfeld. The proportionality of global warming to cumulative carbon emissions. *Nature*, 459:829–832, 2009. doi: 10.1038/nature08047.

C. M. McBride, A. C. Kruger, and L. Dyson. Trends in probabilities of temperature records in the non-stationary climate of South Africa. *International Journal of Climatology*, 42(3):1692–1705, 2022. doi: 10.1002/joc.7329.

A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: Concepts, techniques and tools. Revised edition*. Economics Books. Princeton University Press, 2015.

M. Meng and D. Niu. Modeling CO2 emissions from fossil fuel combustion using the logistic equation. *Energy*, 36(5):3355–3359, 2011. doi: 10.1016/j.energy.2011.03.032.

A. Menon, M. Kishan, C. K. Mohan, and S. Ranka. Characterization of a Class of Sigmoid Functions with Applications to Neural Networks. *Neural Networks*, 9(5):819–835, 1996. doi: 10.1016/0893-6080(95)00107-7.

S. Millard. EnvStats, an R Package for Environmental Statistics., 2013. URL `https://cran.r-project.org/web/packages/EnvStats/index.html`. Last Access: 02/12/2021.

V. Mishra, D. Kumar, A. R. Ganguly, J. Sanjay, M. Mujumdar, R. Krishnan, and R. D. Shah. Reliability of regional and global climate models to simulate precipitation extremes over India. *Journal of Geophysical Research: Atmospheres*, 119(15):9301–9323, 2014. doi: 10.1002/2014JD021636.

T. Mitsui and M. Crucifix. Influence of external forcings on abrupt millennial-scale climate changes: a statistical modelling study. *Climate Dynamics*, 48:2729–274, 2017. doi: 10.1007/s00382-016-3235-z.

S. Mohr, U. Ehret, M. Kunz, P. Ludwig, A. Caldas-Alvarez, J. E. Daniell, F. Ehmele, H. Feldmann, M. J. Franca, C. Gattke, M. Hundhausen, P. Knippertz, K. Küpfer, B. Mühr, J. G. Pinto, J. Quinting, A. M. Schäfer, M. Scheibel, F. Seidel, and C. Wisotzky. A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe – Part 1: Event description and analysis. *Natural Hazards and Earth System Sciences*, 23(2):525–551, 2023. doi: 10.5194/nhess-23-525-2023.

M. Mudelsee. *Cliamte Time Series Analysis*. Springer, NY, USA, 2014. doi: 10.1007/978-3-319-04450-7.

NASA/GISS. NASA-GISS GISS-E2.1G model output prepared for CMIP6 CMIP historical. Version 20180830. Earth System Grid Federation [data set], 2018. URL `https://doi.org/10.22033/ESGF/CMIP6.7127`.

D. Neubauer, S. Ferrachat, C. Siegenthaler-Le Drian, J. Stoll, D. S. Folini, I. Tegen, K.-H. Wieners, T. Mauritsen, I. Stemmler, S. Barthel, I. Bey, N. Daskalakis, B. Heinold, H. Kokkola, D. Partridge, S. Rast, H. Schmidt, N. Schutgens, T. Stanelle, P. Stier, D. Watson-Parris, and U. Lohmann. HAMMOZ-Consortium MPI-ESM1.2-HAM model output prepared for CMIP6 CMIP historical. Version 20191218. Earth System Grid Federation [data set], 2019. URL `https://doi.org/10.22033/ESGF/CMIP6.5016`.

L. Niu, G. Lohmann, P. Gierz, E. J. Gowan, and G. Knorr. Coupled climate-ice sheet modelling of MIS-13 reveals a sensitive Cordilleran Ice Sheet. *Global Planet Change*, 200:103474, 2021. doi: 10.1016/j.gloplacha.2021.103474.

M. Nogaj, S. Parey, and D. Dacunha-Castelle. Non-stationary extreme models and a climatic application. *Nonlinear Processes in Geophysics*, 14(3):305–316, 2007. doi: 10.5194/npg-14-305-2007.

F. Onwuegbuche, A. Kenyatta, S. B. Affognon, E. Enock, and M. Akinade. Application of extreme value theory in predicting climate change induced extreme rainfall in Kenya. *International Journal of Statistics and Probability*, 8(4), 2019. doi: 10.5539/ijsp.v8n4p85.

T. Opitz. Extremal t-processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis*, 122:409–413, 2013. doi: 10.1016/j.jmva.2013.08.008.

B. L. Otto-Bliesner, E. C. Brady, A. Zhao, C. M. Brierley, Y. Axford, E. Capron, A. Govin, J. S. Hoffman, E. Isaacs, M. Kageyama, P. Scussolini, P. C. Tzedakis, C. J. R. Williams, E. Wolff, A. Abe-Ouchi, P. Braconnot, S. Ramos Buarque, J. Cao, A. de Vernal, M. V. Guarino, C. Guo, A. N. LeGrande, G. Lohmann, K. J. Meissner, L. Menviel, P. A. Morozova, K. H. Nisancioglu, R. O'ishi, D. Salas y Mélia, X. Shi, M. Sicard, L. Sime, C. Stepanek, R. Tomas, E. Volodin, N. K. H. Yeung, Q. Zhang, Z. Zhang, and W. Zheng. Large-scale features of Last Interglacial climate: results from evaluating the *lig127k* simulations for the Coupled Model Intercomparison Project (CMIP6)–Paleoclimate Modeling Intercomparison Project (PMIP4). *Climate of the Past*, 17(1):63–94, 2021. doi: 10.5194/cp-17-63-2021.

J. E. Overland. Causes of the Record-Breaking Pacific Northwest Heatwave, Late June 2021. *Atmosphere*, 12(11), 2021. doi: 10.3390/atmos12111434.

C. J. Paciorek and M. J. Schervish. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17(5):483–506, 2006. doi: 10.1002/env.785.

P. Pall, M. Allen, and D. Stone. Testing the Clausius–Clapeyron constraint on changes in extreme precipitation under CO_2 warming. *Climate Dynamics*, 28:351–363, 2007. doi: 10.1007/s00382-006-0180-2.

D. Panagoulia, P. Economou, and C. Caroni. Stationary and nonstationary generalized extreme value modelling of extreme precipitation over a mountainous area under climate change. *Environmetrics*, 25(1):29–43, 2014. doi: 10.1002/env.2252.

S. Parey, T. T. H. Hoang, and D. Dacunha-Castelle. The importance of mean and variance in predicting changes in temperature extremes. *Journal of Geophysical Research: Atmospheres*, 118(15):8285–8296, 2013. doi: 10.1002/jgrd.50629.

W. S. Parker. Ensemble modeling, uncertainty and robust predictions. *WIREs Climate Change*, 4(3):213–223, 2013. doi: 10.1002/wcc.220.

M. D. Penrose. Semi-Min-Stable Processes. *The Annals of Probability*, 20(3):1450–1463, 1992. doi: 10.1214/aop/1176989700.

S. E. Perkins, A. Moise, P. Whetton, and J. Katzfey. Regional changes of climate extremes over Australia – a comparison of regional dynamical downscaling and global climate model simulations. *International Journal of Climatology*, 34(12):3456–3478, 2014.

R. Pérez-Suárez and A. J. López-Menéndez. Growing green? Forecasting CO2 emissions with Environmental Kuznets Curves and Logistic Growth Models. *Environmental Science and Policy*, 54:428–437, 2015. doi: 10.1016/j.envsci.2015.07.015.

T. Rackow, H. Goessling, T. Jung, D. Sidorenko, T. Semmler, D. Barbi, and D. Handorf. Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part II: climate variability. *Climate Dynamics*, 50, 2018. doi: 10.1007/s00382-016-3192-6.

S. Rahmstorf and D. Coumou. Increase of extreme events in a warming world. *Proceedings of the National Academy of Sciences*, 108:17905–17909, 2011. doi: 10.1073/pnas.1101766108.

C. H. Reick, T. Raddatz, V. Brovkin, and V. Gayler. Representation of natural and anthropogenic land cover change in MPI-ESM. *Journal of Advances in Modeling Earth Systems*, 5 (3):459–482, 2013. doi: 10.1002/jame.20022.

K. Riahi, A. Grübler, and N. Nakicenovic. Scenarios of long-term socio-economic and environmental development under climate stabilization. *Technological Forecasting and Social Change*, 74(7):887–935, 2007. doi: 10.1016/j.techfore.2006.05.026.

K. Riahi, D. P. van Vuuren, E. Kriegler, J. Edmonds, B. C. O'Neill, S. Fujimori, N. Bauer, K. Calvin, R. Dellink, O. Fricko, W. Lutz, A. Popp, J. C. Cuaresma, S. KC, M. Leimbach, L. Jiang, T. Kram, S. Rao, J. Emmerling, K. Ebi, T. Hasegawa, P. Havlik, F. Humpenöder, L. A. Da Silva, S. Smith, E. Stehfest, V. Bosetti, J. Eom, D. Gernaat, T. Masui, J. Rogelj, J. Strefler, L. Drouet, V. Krey, G. Luderer, M. Harmsen, K. Takahashi, L. Baumstark, J. C. Doelman, M. Kainuma, Z. Klimont, G. Marangoni, H. Lotze-Campen, M. Obersteiner, A. Tabeau, and M. Tavoni. The Shared Socioeconomic Pathways and their energy, land

use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, 42:153–168, 2017. doi: 10.1016/j.gloenvcha.2016.05.009.

M. Ribatet. Modelling spatial extremes using max-stable processes. In C. L. E. Franzke and T. J. O'Kane, editors, *Nonlinear and Stochastic Climate Dynamics*, page 369–391. Cambridge University Press, 2017. doi: 10.1017/9781316339251.014.

M. Ribatet, C. Dombry, and M. Oesting. Spatial extremes and max-stable processes. *Extreme Value Modeling and Risk Analysis: Methods and Applications*, pages 179–194, 2015. doi: 10.1201/b19721-10.

F. J. Richards. A Flexible Growth Function for Empirical Use. *Journal of Experimental Botany*, 10(29):290–300, 1959.

J. Ridley, M. Menary, T. Kuhlbrodt, M. Andrews, and T. Andrews. MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP historical. Version 20190626. Earth System Grid Federation [data set], 2019. URL `https://doi.org/10.22033/ESGF/CMIP6.6109`.

R. V. Rohli and A. J. Vega. *Climatology* (*fourth edition*). Jones & Bartlett Learning, 2018. ISBN 9781284126563.

X. Rong. CAMS CAMS_CSM1.0 model output prepared for CMIP6 CMIP historical. Version 20190829. Earth System Grid Federation [data set], 2019. URL `https://doi.org/10.22033/ESGF/CMIP6.9754`.

W. F. Ruddiman. *Earth's Climate: Past and Future*. W. H. Freeman, New York, 2008. ISBN 9780716784906.

M. Rummukainen. Climate change: changing means and changing extremes. *Climatic Change*, 121:3–13, 2013. doi: 10.1007/s10584-013-0888-z.

S. Russo and A. Sterl. Global changes in seasonal means and extremes of precipitation from daily climate model data. *Journal of Geophysical Research: Atmospheres*, 117(D1), 2012. doi: 10.1029/2011JD016260.

S. Salvador and P. Chan. Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In *16th IEEE International Conference on Tools with Artificial Intelligence*, pages 576–584, 2004. doi: 10.1109/ICTAI.2004.50.

A. Sarhadi and E. Soulis. Time-varying extreme rainfall intensity-duration-frequency curves in a changing climate: Time-varying extreme rainfall IDF curves. *Geophysical Research Letters*, 2017. doi: 10.1002/2016GL072201.

K. Saunders, A. G. Stephenson, and D. J. Karoly. A regionalisation approach for rainfall based on extremal dependence. *Extremes*, 24:215–240, 2021. doi: 10.1007/s10687-020-00395-y.

M. Schlather. Models for stationary max-stable random fields. *Extremes*, 5:33–44, 2002. doi: 10.1023/A:1020977924878.

G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 1978. doi: 10.1214/aos/1176344136.

C. Schär, P. Vidale, D. Lüthi, C. Frei, C. Häberli, M. Liniger, and C. Appenzeller. The role of increasing temperature variability in European summer heatwaves. *Nature*, 427:332–336, 2004. doi: 10.1038/nature02300.

Ø. Seland, M. Bentsen, D. J. L. Oliviè, T. Toniazzo, A. Gjermundsen, L. S. Graff, J. B. Debernard, A. K. Gupta, Y. He, A. Kirkevåg, J. Schwinger, J. Tjiputra, K. S. Aas, I. Bethke, Y. Fan, J. Griesfeller, A. Grini, C. Guo, M. Ilicak, I. H. H. Karset, O. A. Landgren, J. Liakka, K. O. Moseid, A. Nummelin, C. Spensberger, H. Tang, Z. Zhang, C. Heinze, T. Iversen, and M. Schulz. NCC NorESM2-MM model output prepared for CMIP6 CMIP historical. Version 20200702. Earth System Grid Federation [data set], 2019a. URL `https://doi.org/10.22033/ESGF/CMIP6.8040`.

Ø. Seland, M. Bentsen, D. J. L. Oliviè, T. Toniazzo, A. Gjermundsen, L. S. Graff, J. B. Debernard, A. K. Gupta, Y. He, A. Kirkevåg, J. Schwinger, J. Tjiputra, K. S. Aas, I. Bethke, Y. Fan, J. Griesfeller, A. Grini, C. Guo, M. Ilicak, I. H. H. Karset, O. A. Landgren, J. Liakka, K. O. Moseid, A. Nummelin, C. Spensberger, H. Tang, Z. Zhang, C. Heinze, T. Iversen, and M. Schulz. NCC NorESM2-LM model output prepared for CMIP6 CMIP historical. Version 20190920. Earth System Grid Federation [data set], 2019b. URL `https://doi.org/10.22033/ESGF/CMIP6.8036`.

T. Semmler, S. Danilov, T. Rackow, D. Sidorenko, D. Barbi, J. Hegewald, D. Sein, Q. Wang, and T. Jung. AWI AWI-CM1.1MR model output prepared for CMIP6 CMIP historical. Version 20200511. Earth System Grid Federation [data set], 2018. URL `https://doi.org/10.22033/ESGF/CMIP6.2686`.

S. Seneviratne, N. Nicholls, D. Easterling, C. M. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi, M. Reichstein, A. Sorteberg, C. Vera, X. Zhang, M. Rusticucci, V. Semenov, L. V. Alexander, S. Allen, G. Benito, T. Cavazos, J. Clague, D. Conway, P. M. Della-Marta, M. Gerber, S. Gong, B. N. Goswami, M. Hemer, C. Huggel, B. van den Hurk, V. V. Kharin, A. Kitoh, A. M. Klein Tank, G. Li, S. Mason, W. McGuire, G. J. van Oldenborgh, B. Orlowsky, S. Smith, W. Thiaw, A. Velegrakis, P. Yiou, T. Zhang, T. Zhou, and F. W. Zwiers. Changes in climate extremes and their impacts on the natural physical environment. In C. B. Field, V. Barros, T. F. Stocker, and Q. Dahe, editors, *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*, page 109–230. Cambridge University Press, 2012. doi: 10.1017/CBO9781139177245.006.

S. Seneviratne, X. Zhang, M. Adnan, W. Badi, C. Dereczynski, A. D. Luca, S. Ghosh, I. Iskandar, J. Kossin, S. Lewis, F. Otto, I. Pinto, M. Satoh, S. Vicente-Serrano, M. Wehner, and B. Zhou. *Weather and Climate Extreme Events in a Changing Climate. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, page 1513–1766. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021. doi: 10.1017/9781009157896.013.

C. Y. Shen. Logistic growth modelling of COVID-19 proliferation in China and its international implications. *International Journal of Infectious Diseases*, 96:582–589, 2020. doi: 10.1016/j.ijid.2020.04.085.

X. Shi, G. Lohmann, D. Sidorenko, and H. Yang. Early-Holocene simulations using different forcings and resolutions in AWI-ESM. *The Holocene*, 30(7):996–1015, 2020. doi: 10.1177/0959683620908634.

D. Sidorenko, T. Rackow, T. Jung, T. Semmler, D. Barbi, S. Danilov, K. Dethloff, W. Dorn, K. Fieg, H. F. Goessling, D. Handorf, S. Harig, W. Hiller, S. Juricke, M. Losch, J. Schröter, D. V. Sein, and Q. Wang. Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part I: Model formulation and mean climate. *Climate Dynamics*, 44: 757–780, 2015. doi: 10.1007/s00382-014-2290-6.

J. Sillmann, V. V. Kharin, F. W. Zwiers, X. Zhang, and D. Bronaugh. Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *Journal of Geophysical Research: Atmospheres*, 118(6):2473–2493, 2013. doi: 10.1002/jgrd.50188.

A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8:229–231, 1959.

R. Slater, N. Freychet, and G. Hegerl. Substantial changes in the probability of future annual temperature extremes. *Atmospheric Science Letters*, 22(11):e1061, 2021. doi: 10.1002/asl.1061.

R. L. Smith. Max-stable processes and spatial extremes. Unpublished Manuscript, 1990.

A. H. Sobel, S. J. Camargo, T. M. Hall, C.-Y. Lee, M. K. Tippett, and A. A. Wing. Human influence on tropical cyclone intensity. *Science*, 353(6296):242–246, 2016. doi: 10.1126/science.aaf6574.

W. Soon and S. Baliunas. Proxy climatic and environmental changes of the past 1000 years. *Climate Research*, 23(2):89–110, 2003. doi: 10.3354/cr023089.

M. A. Stephens. Use of the Kolmogorov–Smirnov, Cramér–Von Mises and Related Statistics Without Extensive Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 32(1):115–122, 1970. doi: 10.1111/j.2517-6161.1970.tb00821.x.

B. Stevens, M. Giorgetta, M. Esch, T. Mauritsen, T. Crueger, S. Rast, M. Salzmann, H. Schmidt, J. Bader, K. Block, R. Brokopf, I. Fast, S. Kinne, L. Kornblueh, U. Lohmann, R. Pincus, T. Reichler, and E. Roeckner. Atmospheric component of the MPI-M Earth System Model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, 5(2):146–172, 2013. doi: 10.1002/jame.20015.

T. Stocker. *Introduction to Climate Modeling*. Springer Berlin, Heidelberg, 2013. doi: 10.1007/978-3-642-00773-6.

N. C. Swart, J. N. Cole, V. V. Kharin, M. Lazare, J. F. Scinocca, N. P. Gillett, J. Anstey, V. Arora, J. R. Christian, Y. Jiao, W. G. Lee, F. Majaess, O. A. Saenko, C. Seiler, C. Seinen, A. Shao, L. Solheim, K. von Salzen, D. Yang, B. Winter, and M. Sigmond. CCCma CanESM5 model output prepared for CMIP6 CMIP historical. Version 20190429. Earth System Grid Federation [data set], 2019. URL `https://doi.org/10.22033/ESGF/CMIP6.3610`.

H. Tabari, R. De Troch, O. Giot, R. Hamdi, P. Termonia, S. Saeed, E. Brisson, N. Van Lipzig, and P. Willems. Local impact analysis of climate change on precipitation extremes: are high-resolution climate models needed for realistic simulations? *Hydrology and Earth System Sciences*, 20(9):3843–3857, 2016. doi: 10.5194/hess-20-3843-2016.

F. J. Tapiador, F. Turk, W. Petersen, A. Y. Hou, E. García-Ortega, L. A. Machado, C. F. Angelis, P. Salio, C. Kidd, G. J. Huffman, and M. de Castro. Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, 104-105:70–97, 2012. doi: 10.1016/j.atmosres.2011.10.021.

K. E. Taylor, R. J. Stouffer, and G. A. Meehl. An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, 93(4):485–498, 2012. doi: 10.1175/BAMS-D-11-00094.1.

Q. Tian, Z. Li, and X. Sun. Frequency analysis of precipitation extremes under a changing climate: a case study in Heihe River basin, China. *Journal of Water and Climate Change*, 12 (3):772–786, 2020. doi: 10.2166/wcc.2020.170.

B. Timmermans, M. Wehner, D. Cooley, T. O'Brien, and H. Krishnan. An evaluation of the consistency of extremes in gridded precipitation data sets. *Climate Dynamics*, 52:1–20, 2019. doi: 10.1007/s00382-018-4537-0.

A. Toreti, O. Cronie, and M. Zampieri. Concurrent climate extremes in the key wheat producing regions of the world. *Scientific Reports*, 9, 2019. doi: 10.1038/s41598-019-41932-5.

K. Trenberth, P. Jones, P. Ambenje, R. Bojariu, D. Easterling, A. Klein Tank, D. Parker, F. Rahimzadeh, J. Renwick, M. Rusticucci, B. Soden, and P. Zhai. Observations: Surface and atmospheric climate change. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, editors, *Climate Change 2007: The Physical Science Basis. Contribution of Working Group 1 to the 4th Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, United Kingdom, 2007.

R. Twardosz, A. Walanus, and I. Guzik. Warming in Europe: Recent Trends in Annual and Seasonal temperatures. *Pure and Applied Geophysics*, 178:4021–4032, 2021. doi: 10.1007/s00024-021-02860-6.

University of East Anglia Climatic Research Unit, I. C. Harris, P. D. Jones, and T. Osborn. CRU TS4.04: Climatic Research Unit (CRU) Time-Series (TS) version 4.04 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901 - Dec. 2019). Centre for Environmental Data Analysis [data set], 2020. URL `https://catalogue.ceda.ac.uk/uuid/89e1e34ec3554dc98594a5732622bce9`.

U.S. Climate Change Program. *Strategic Plan for the U.S. Climate Change Science Program. A Report by the Climate Change Science Program and the Subcommittee on Global Change Research*. US Climate Change Science Program, Washington DC, 2003.

P. Verhulst. Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, 18:14–54, 1845.

G. Villarini, J. A. Smith, A. A. Ntelekos, and U. Schwarz. Annual maximum and peaks-over-threshold analyses of daily rainfall accumulations for Austria. *Journal of Geophysical Research: Atmospheres*, 116(D5), 2011. doi: 10.1029/2010JD015038.

A. Voldoire. CNRM-CERFACS CNRM-CM6-1 model output prepared for CMIP6 CMIP. Version 20200529. Earth System Grid Federation [data set], 2018. URL `https://doi.org/10.22033/ESGF/CMIP6.1375`.

E. Volodin, E. Mortikov, A. Gritsun, V. Lykossov, V. Galin, N. Diansky, A. Gusev, S. Kostrykin, N. Iakovlev, A. Shestakova, and S. Emelina. INM INM-CM5-0 model output prepared for CMIP6 CMIP historical. Version 20190709. Earth System Grid Federation [data set], 2019a. URL `https://doi.org/10.22033/ESGF/CMIP6.5070`.

E. Volodin, E. Mortikov, A. Gritsun, V. Lykossov, V. Galin, N. Diansky, A. Gusev, S. Kostrykin, N. Iakovlev, A. Shestakova, and S. Emelina. INM INM-CM4-8 model output prepared for CMIP6 CMIP historical. Version 20190530. Earth System Grid Federation [data set], 2019b. URL `https://doi.org/10.22033/ESGF/CMIP6.1422`.

M. P. Wand and M. C. Jones. *Kernel Smoothing (1st ed.)*. Chapman and Hall/CRC, New York, 1994. doi: 10.1201/b14876.

Q. Wang, S. Danilov, D. Sidorenko, R. Timmermann, C. Wekerle, X. Wang, T. Jung, and J. Schröter. The Finite Element Sea Ice-Ocean Model (FESOM) v.1.4: formulation of an ocean general circulation model. *Geoscientific Model Development*, 7(2):663–693, 2014. doi: 10.5194/gmd-7-663-2014.

T. Wu, L. Song, W. Li, Z. Wang, H. Zhang, X. Xin, Y. Zhang, L. Zhang, J. Li, F. Wu, Y. Liu, F. Zhang, X. Shi, M. Chu, J. Zhang, Y. Fang, F. Wang, Y. Lu, X. Liu, M. Wei, Q. Liu, W. Zhou, M. Dong, Q. Zhao, J. Ji, L. Li, and M. Zhou. An overview of BCC climate system model development and application for climate change studies. *Journal of Meteorological Research*, 28(6):34–56, 2014. doi: 10.1007/s13351-014-3041-7.

T. Wu, M. Chu, M. Dong, and Y. Fang. BCC BCC-CSM2MR model output prepared for CMIP6 CMIP historical. Version 20181126. Earth System Grid Federation [data set], 2018. URL `https://doi.org/10.22033/ESGF/CMIP6.2948`.

D. Wuebbles, G. Meehl, K. Hayhoe, T. R. Karl, K. Kunkel, B. Santer, M. Wehner, B. Colle, E. M. Fischer, R. Fu, A. Goodman, E. Janssen, V. Kharin, H. Lee, W. Li, L. N. Long, S. C. Olsen, Z. Pan, A. Seth, J. Sheffield, and L. Sun. CMIP5 Climate Model Analyses: Climate Extremes in the United States. *Bulletin of the American Meteorological Society*, 95(4):571–583, 2014. doi: 10.1175/BAMS-D-12-00172.1.

S.-P. Xie, C. Deser, G. A. Vecchi, M. Collins, T. L. Delworth, A. Hall, E. Hawkins, N. C. Johnson, C. Cassou, A. Giannini, and M. Watanabe. Towards predictive understanding of regional climate change. *Nature Climate Change*, 5(10):921–930, 2015. doi: 10.1038/nclimate2689.

B. Yang, C. Qin, A. Bräuning, T. J. Osborn, V. Trouet, F. C. Ljungqvist, J. Esper, L. Schneider, J. Grießinger, U. Büntgen, S. Rossi, G. Dong, M. Yan, L. Ning, J. Wang, X. Wang, S. Wang, J. Luterbacher, E. R. Cook, and N. C. Stenseth. Long-term decrease in Asian monsoon rainfall and abrupt climate change events over the past 6,700 years. *Proceedings of the National Academy of Sciences*, 118(30):e2102007118, 2021. doi: 10.1073/pnas.2102007118.

Y. Yano, T. Oguma, H. Nagata, and S. Sasaki. Application of logistic growth model to pharmacodynamic analysis of in vitro bactericidal kinetics. *Journal of Pharmaceutical Sciences*, 87(10):1177–83, 1998. doi: 10.1021/js9801337.

Y. Yu. CAS FGOALS-f3-L model output prepared for CMIP6 CMIP historical. Version 20190927. Earth System Grid Federation [data set], 2019. URL `https://doi.org/10.22033/ESGF/CMIP6.3355`.

S. Yukimoto, T. Koshiro, H. Kawai, N. Oshima, K. Yoshida, S. Urakawa, H. Tsujino, M. Deushi, T. Tanaka, M. Hosaka, H. Yoshimura, E. Shindo, R. Mizuta, M. Ishii, A. Obata, and Y. Adachi. MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP historical. Version 20210813. Earth System Grid Federation [data set], 2019. URL `https://doi.org/10.22033/ESGF/CMIP6.6842`.

D. D. Zhang, P. Brecke, H. F. Lee, Y.-Q. He, and J. Zhang. Global climate change, war, and population decline in recent human history. *Proceedings of the National Academy of Sciences*, 104(49):19214–19219, 2007. doi: 10.1073/pnas.0703073104.

S. Zhang and J. Chen. Uncertainty in Projection of Climate Extremes: A Comparison of CMIP5 and CMIP6. *Journal of Meteorological Research*, 35(4):646–662, 2021. doi: 10.1007/s13351-021-1012-3.

X. Zhang, L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers. Indices for monitoring changes in extremes based on daily temperature and precipitation data. *WIREs Climate Change*, 2(6):851–870, 2011. doi: 10.1002/wcc.147.

H. Zhu, Z. Jiang, J. Li, W. Li, C. Sun, and L. Li. Does CMIP6 Inspire More Confidence in Simulating Climate Extremes over China? *Advances in Atmospheric Sciences*, 37(10): 1119–1132, 2020. doi: 10.1007/s00376-020-9289-1.

T. Ziehn, M. Chamberlain, A. Lenton, R. Law, R. Bodman, M. Dix, Y. Wang, P. Dobrohotoff, J. Srbinovsky, L. Stevens, P. Vohralik, C. Mackallah, A. Sullivan, S. O'Farrell, and K. Druken. CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 CMIP historical. Version 20200529. Earth System Grid Federation [data set], 2019. URL `https://doi.org/10.22033/ESGF/CMIP6.4272`.

J. Zscheischler, M. D. Mahecha, and S. Harmeling. Climate classifications: the value of unsupervised clustering. *Procedia Computer Science*, 9:897–906, 2012. doi: 10.1016/j.procs.2012.04.096.

J. Zscheischler, A. M. Michalak, C. Schwalm, M. D. Mahecha, D. N. Huntzinger, M. Reichstein, G. Berthier, P. Ciais, R. B. Cook, B. El-Masri, M. Huang, A. Ito, A. Jain, A. King, H. Lei, C. Lu, J. Mao, S. Peng, B. Poulter, D. Ricciuto, X. Shi, B. Tao, H. Tian, N. Viovy, W. Wang, Y. Wei, J. Yang, and N. Zeng. Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data. *Global Biogeochemical Cycles*, 28(6):585–600, 2014. doi: 10.1002/2014GB004826.

A. Çiğdem Köne and T. Büke. Forecasting of CO2 emissions from fuel combustion using trend analysis. *Renewable and Sustainable Energy Reviews*, 14(9):2906–2915, 2010. doi: 10.1016/j.rser.2010.06.006.