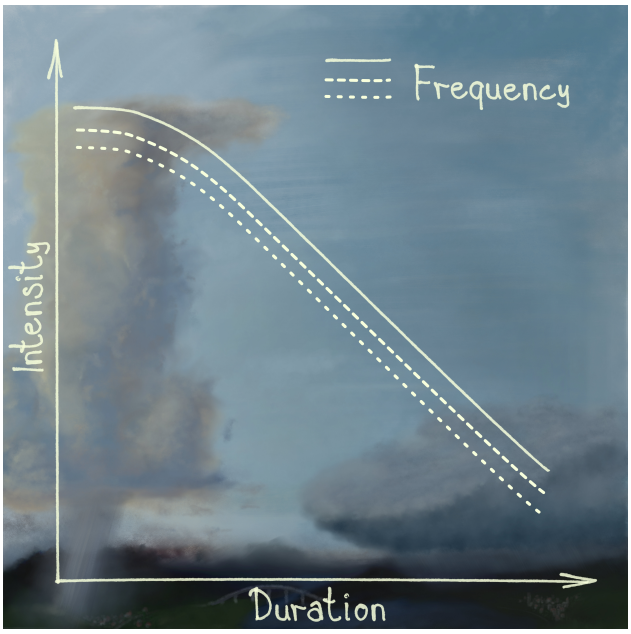
Freie Universität Berlin
Fachbereich Geowissenschaften
Institut für Meteorologie

Statistical Modeling of Extreme Precipitation on Different Timescales

Towards Consistent and More Efficient Use of Available Data



Publikationsbasierte Dissertation

vorgelegt von

Jana Ulrich

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

doctor rerum naturalium

Berlin, April 2022

Erstgutachter:	Prof. Dr. Henning Rust
Zweitgutachter:	Prof. Dr. Uwe Ulbrich
Tag der Disputation:	26. August 2022

Figure 0.1: The image on the cover page illustrates the relationship between precipitation intensity, duration and frequency, where different durations, i.e. timescales, correspond to different precipitation processes.

Long as I remember the rain been comin' down
Clouds of mystery pourin' confusion on the ground
Good men through the ages tryin' to find the sun
And I wonder, still I wonder, who'll stop the rain?

John Fogerty

Abstract

Extreme precipitation events can have serious adverse consequences for the population. Their severity and possible impacts are mainly determined by their intensity and duration. Therefore, when planning and operating water management systems or protecting infrastructure from flooding, it is important to consider the frequency at which particularly intense or long-lasting heavy rainfall events can be expected. This information is commonly summarized in intensity-duration-frequency (IDF) curves. The methods used to estimate IDF curves are based on extreme value statistics, where the challenge is modeling events of rarely occurring magnitude or those not even observed yet. A further obstacle is the availability of data, since the time series of precipitation measurements with a high temporal resolution are usually relatively short and the spatial coverage of precipitation gauges is sparse. Therefore, the estimation of IDF curves is often associated with considerable uncertainties. This is problematic, because overestimation leads to major additional construction costs, while underestimation implies an unreasonable residual risk of failure and might result in severe consequences.

Therefore, the objective of this work is to develop IDF models that allow for estimation with reduced uncertainties. We achieve this by more efficient usage of the available data. Thus, as a first step, we model the annual precipitation intensity maxima over a range of durations simultaneously within one model. To this end, we use a duration-dependent Generalized Extreme Value (d-GEV) distribution. We demonstrate that this results in a significant reduction of uncertainties, especially in the estimation of the shape parameter. This parameter determines the upper tail of the distribution and thus comprises information on very rare, severe events.

Within our first study, we extend the d-GEV model by incorporating spatial variations (in the form of covariates). This provides two advantages: on the one hand, we can combine information from different stations, which leads to a further reduction of uncertainties. On the other hand, we are able to spatially interpolate the IDF relationship, which allows us to obtain estimates at any location within the study area, including locations without observations. We analyze the performance of the model in a case study in detail. Our results show that including spatial variations into the model improves the estimation of rare events and yields reliable estimates at locations with little or no data. Using two examples of heavy rainfall events in Berlin, we further demonstrate how the spatial d-GEV

model can be applied to analyze the extremeness of events on different timescales and in space. However, another result of our first study indicates a disadvantage of the d-GEV model: it is not flexible enough to describe a wide range of durations, as it does not improve the estimates for every duration. In our second study, we therefore investigate possible extensions of the model and discuss for which applications certain enhancements might be beneficial.

Since the d-GEV distribution is commonly applied to model annual maxima, only a small fraction of the observations is used to estimate IDF curves. In our third study, we implement seasonal variations into the model, allowing us to use monthly maxima, instead. This again results in a significant reduction in uncertainties. Moreover, we can investigate the influence of seasonal variations on the IDF curves. Our results suggest that especially at stations with large differences in the seasonality of short and long-lasting extreme events, the initial d-GEV model does not provide a good approximation. At these stations, the flexible d-GEV model is clearly able to better describe the IDF relationship. Additionally, modeling monthly instead of annual maxima allows for more reliable estimation of the shape parameter, therefore, we are able to perform a more careful analysis of the dependence of the shape parameter on duration.

The methods presented in this thesis not only serve to reduce uncertainties in the estimation of IDF curves but also provide a consistent and parameter parsimonious approach to model variations of extreme rainfall. Future studies can benefit from our insights to investigate the influence of climate change on the intensity and frequency of extreme events on different timescales with greater accuracy.

Zusammenfassung

Extreme Niederschlagsereignisse können schwerwiegende Folgen für die Bevölkerung haben. Ihre Stärke und möglichen Auswirkungen werden vor allem durch ihre Intensität und Dauer bestimmt. Bei der Planung und dem Betrieb von wasserwirtschaftlichen Anlagen oder dem Schutz der Infrastruktur vor Überschwemmungen ist es daher wichtig zu berücksichtigen, mit welcher Häufigkeit besonders intensive oder lang anhaltende Starkregenereignisse zu erwarten sind. Diese Informationen werden typischerweise in Intensitäts-Dauer-Frequenz-Kurven (IDF-Kurven) zusammengefasst. Die zur Schätzung von IDF-Kurven verwendeten Methoden basieren auf der Extremwertstatistik, wobei die Herausforderung darin besteht, Ereignisse von sehr seltenem oder bisher nicht beobachtetem Ausmaß zu modellieren. Eine weitere Schwierigkeit ist die Datenverfügbarkeit, da die Zeitreihen von Niederschlagsmessungen mit einer hohen zeitlichen Auflösung in der Regel relativ kurz sind und die räumliche Abdeckung von Messstationen eher gering ist. Daher ist die Schätzung von IDF-Kurven meist mit erheblichen Unsicherheiten behaftet. Dies ist problematisch, da eine Überschätzung zu beträchtlichen zusätzlichen Baukosten führt, während eine Unterschätzung ein unangemessenes Restrisiko des Versagens bedeutet und schwerwiegende Folgen nach sich ziehen kann.

Das Ziel dieser Arbeit ist es daher, IDF-Modelle zu entwickeln, die eine Schätzung mit geringeren Unsicherheiten durch eine effizientere Nutzung der verfügbaren Daten ermöglichen. In einem ersten Schritt modellieren wir die Jahreshöchstwerte der Niederschlagsintensität über eine Reihe von Dauern gleichzeitig innerhalb eines Modells. Zu diesem Zweck verwenden wir eine dauerabhängige verallgemeinerte Extremwertverteilung (d-GEV). Wir zeigen, dass dies zu einer deutlichen Verringerung der Unsicherheiten führt, insbesondere bei der Schätzung des Formparameters. Dieser Parameter bestimmt das obere Ende der Verteilung und enthält damit Informationen über sehr seltene, schwerwiegende Ereignisse.

In unserer ersten Studie erweitern wir das d-GEV-Modell durch die Einbeziehung räumlicher Variationen (in Form von Kovariaten). Dies bietet zwei Vorteile: Zum einen können wir Informationen von verschiedenen Stationen kombinieren, was zu einer weiteren Verringerung der Unsicherheiten führt. Zum anderen sind wir in der Lage, die IDF-Beziehung räumlich zu interpolieren, was es uns ermöglicht, Schätzungen an jedem beliebigen Ort innerhalb des Untersuchungsgebiets zu erhalten, einschließlich Orten ohne Beobachtungen. Wir analysieren die Leistung des Modells in einer Fallstudie im Detail. Unsere Ergebnisse

zeigen, dass die Einbeziehung räumlicher Variationen in das Modell die Schätzung seltener Ereignisse verbessert und auch an Orten mit wenigen oder keinen Daten zuverlässige Schätzungen liefert. Anhand zweier Beispiele von Starkregenereignissen in Berlin zeigen wir außerdem, wie das räumliche d-GEV-Modell zur Analyse der Ausprägung von Ereignissen auf verschiedenen Zeitskalen und im Raum angewendet werden kann. Ein weiteres Ergebnis unserer ersten Studie deutet allerdings auch auf einen Nachteil des d-GEV-Modells hin: Es ist nicht flexibel genug, um einen weiten Dauerbereich zu beschreiben, da es die Schätzungen nicht für jede Dauer verbessern kann. In unserer zweiten Studie untersuchen wir daher mögliche Erweiterungen des Modells und erörtern, für welche spezifischen Anwendungen bestimmte Erweiterungen von Vorteil sein können.

Die d-GEV-Verteilung wird üblicherweise zur Modellierung von Jahresmaxima verwendet, wodurch nur ein kleiner Teil der beobachteten Starkregenereignisse zur Schätzung der IDF-Kurven berücksichtigt wird. In unserer dritten Studie implementieren wir saisonale Variationen in das Modell, so dass stattdessen monatliche Maxima verwendet werden können. Dies führt erneut zu einer Reduktion der Unsicherheiten. Darüber hinaus können wir den Einfluss der saisonalen Schwankungen auf die IDF-Kurven untersuchen. Unsere Ergebnisse deuten darauf hin, dass insbesondere an Stationen mit großen Unterschieden in der Saisonalität von kurz- und lang andauernden Extremereignissen das ursprüngliche d-GEV-Modell keine gute Näherung darstellt. An diesen Stationen ist das flexible d-GEV-Modell eindeutig besser geeignet, um die IDF-Beziehung zu beschreiben. Zudem ermöglicht die Modellierung von Monats- anstelle von Jahresmaxima eine zuverlässigere Schätzung des Formparameters, so dass wir eine sorgfältigere Analyse der Abhängigkeit des Formparameters von der Dauer durchführen können.

Die in dieser Arbeit vorgestellten Methoden dienen nicht nur dazu, die Unsicherheiten bei der Schätzung von IDF-Kurven zu verringern, sondern bieten auch einen konsistenten und parametersparsamen Ansatz zur Modellierung von Variationen extremer Niederschlagsintensitäten. Zukünftige Studien, die zum Beispiel das Ziel haben den Einfluss des Klimawandels auf die Intensität und Häufigkeit von Extremereignissen auf verschiedenen Zeitskalen mit größerer Genauigkeit zu erforschen, können von den erreichten Reduktionen der Modellierungsunsicherheiten profitieren.

Publications

This thesis comprises the three following published research articles, which are included in Chpts. 4-6:

I **Estimating IDF Curves Consistently over Durations with Spatial Covariates**
J. Ulrich, O. E. Jurado, M. Peter, M. Scheibel and H. W. Rust
Water, 12, 3119, 2020; doi:10.3390/w12113119

I formulated the research questions and concept in collaboration with H. W. Rust, developed the software, visualized the results, and wrote the manuscript.

II **Flexible and Consistent Quantile Estimation for Intensity-Duration-Frequency Curves**
F. S. Fauer, J. Ulrich, O. E. Jurado and H. W. Rust
Hydrol. Earth Syst. Sci., 25, 6479–6494, 2021; doi:10.5194/hess-25-6479-2021

I had the idea for the topic of the study and jointly developed the concept and methodology with H. W. Rust and F. S. Fauer. I participated in the implementation of the computer code and revised the manuscript.

III **Modeling Seasonal Variations of Extreme Rainfall on Different Timescales in Germany**
J. Ulrich, F. S. Fauer and H. W. Rust
Hydrol. Earth Syst. Sci., 25, 6133–6149, 2021; doi:10.5194/hess-25-6133-2021

I formulated the research questions in collaboration with H. W. Rust, developed the concept and methodology, implemented the computer code, visualized the results, and wrote the manuscript.

In addition, I participated in the following publications, which are closely related to the topic:

IV **Evaluating the Performance of a Max-Stable Process for Estimating Intensity-Duration-Frequency Curves**
O. E. Jurado, J. Ulrich, M. Scheibel and H. W. Rust
Water, 12, 3114, 2020; doi:10.3390/w12123314

I participated in the development of the methodology and the implementation of the computer code. I helped with the visualization of parts of the results and revised the manuscript.

V **Starkregen in Berlin: Meteorologische Ereignisrekonstruktion und Betroffenenbefragung**
L. Berghäuser, L. Schoppa, J. Ulrich, L. Dillenardt, O. E. Jurado, C. Passow, G. S. Mohor, G. S., O. Seleem, T. Petrow and A.H. Thieken
Universität Potsdam: Brandenburg, Germany, 2021; doi:10.25932/publishup-50056

This task force report was developed within the DFG Research Training Group *NatRiskChange*, and consists of a meteorological analysis of two heavy rainfall events in Berlin along with a survey addressed to affected residents. For the the statistical analysis of the events, I implemented the computer code, conducted the analysis, visualized the results and wrote the part of the manuscript. I revised the complete manuscript together with L. Berghäuser and L. Schoppa.

The methods and main results of Study IV are briefly touched upon in Chpt. 3.4. The content of Chpt. 7.1.1 is adopted from my contribution to Study V.

Contents

Chapter 1

Introduction

Extreme meteorological events can potentially cause severe harm to the population. The range of possible impacts from extreme precipitation includes flash floods and land slides, dam failures and river flooding, as well as water pollution and agricultural damage (Bronstert et al., 2018; Dotterweich, 2008; Hattermann et al., 2013; Knapp et al., 2008; Wake, 2013). Therefore, heavy rainfall events are also likely to inflict considerable economic losses (e.g., Kreibich et al., 2014). The German Insurance Association recently reported that 1.3 million damage incidents to residential buildings nationwide were caused by heavy rainfall between 2002 and 2017, resulting in total losses of around 6.7 billion euros (Schweda, 2019). A recent example is the heavy rain event of 12-19 July 2021, which affected within Germany not only the states of North Rhine-Westphalia and Rhineland-Palatinate, but also parts of Bavaria and Saxony. The destructive event resulted in 180 fatalities. Furthermore, reported property damage amounts to around seven billion euros and damage to railroads and road networks is estimated to be at least two billion euros. After the event, there was a month-long interruption in the supply of drinking water, electricity and gas in some affected areas (Fial, 2021).

There are two factors that determine the amount of precipitation: intensity and duration, as *the heaviest precipitation occurs where the rainfall rate is the highest for the longest time*. Based on this simple concept, Doswell et al. (1996) derive the conditions that favor heavy precipitation. Strong *precipitation intensity* is achieved by a high water vapor content of the air combined with a rapid rate of ascent and high rainfall efficiency. All three of these factors are amplified in warm convective precipitation processes, thus summer (or warm season) convection tends to be associated with the greatest precipitation intensities. Hence, these events are often the ones causing flash floods. The *duration* of precipitation is influenced by the size and geometry of a precipitation field, as well as by its velocity. The aforementioned extreme event of July 2021 was caused by a large precipitation field with high precipitation intensities that moved comparatively slowly (Junghänel et al., 2021; Schneider and Gebauer, 2021), ultimately leading to both flash floods and flooding of several rivers. A further example illustrating the importance of rainfall duration is that of the southern UK in the winter of 2000/2001, where high rainfall totals over a period of several months, not necessarily caused by high rainfall intensities, led to groundwater

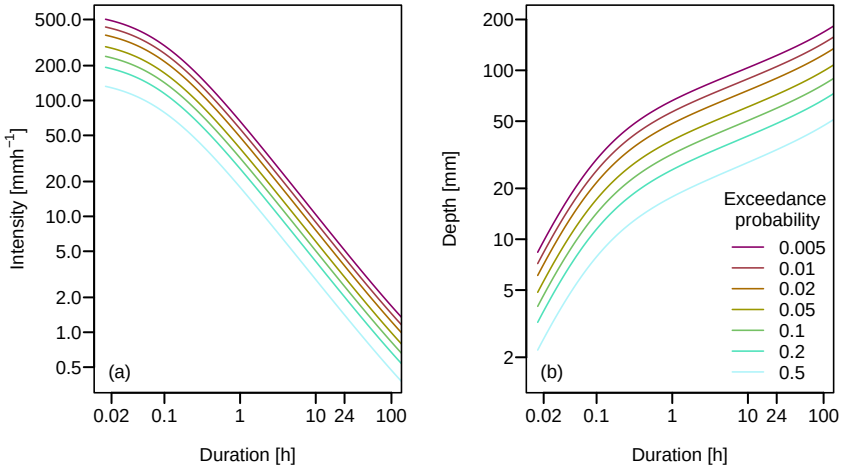


Figure 1.1: Intensity-duration-frequency curves (a) and depth-duration-frequency curves (b) for the station Berlin-Tempelhof. These are two different approaches to representing the same information.

flooding in many areas (Marsh and Dale, 2002).

Thus, information about the intensity and also the duration of a precipitation event is needed to assess its severity and potential impacts. Further relevant information, for example for the planning of buildings and infrastructure or the estimation of insurance risk, is the frequency at which certain events occur. All three factors are linked by the intensity-duration-frequency (IDF) relationship. It is often presented graphically in the form of IDF curves by plotting precipitation intensity versus duration for a family of curves representing specific frequencies or annual exceedance probabilities. Possible applications of IDF curves include the design of e.g. urban drainage networks, pumping stations, wastewater treatment plants and retention basins to estimate the risk of failure. The aim is of course to avoid a failure due to the usually severe consequences, however, overestimation of the risk is often associated with considerable additional construction costs.

To provide an example, Fig. 1.1 (a) displays the estimated IDF curves for the station Berlin Tempelhof. The representation is usually either in single or double logarithmic form. Another widely used representation are depth-duration-frequency (DDF) curves, as shown Fig. 1.1 (b). Instead of the precipitation intensity, the precipitation amount in terms of the depth per area is plotted. Since the precipitation intensity equals the amount of precipitation per time interval, it is easy to convert from the one quantity to the other and the included information is equivalent for both sets of curves. Therefore, throughout this thesis, we limit ourselves to the intensity representation. The methodology of IDF curves can also be applied to streamflow, which is usually labeled with the symbol Q, resulting in flood-duration-frequency (QDF) curves (Javelle et al., 2002). The development of methods for estimating IDF curves is an interdisciplinary field, involving different perspectives from meteorologists and hydrologists, as well as mathematicians specializing in extreme value statistics and engineers. For this reason, the literature offers many different approaches to IDF curve modeling (e.g., Ariff et al., 2012; Hosking et al., 1985; Mélèse et al., 2018;

Roksvåg et al., 2021; Stephenson et al., 2016; Van de Vyver, 2018). Consequently, different countries often have different regulations for the calculation and practical application of IDF curves (Lutz et al., 2020; Svensson and Jones, 2010).

A major challenge for estimating IDF curves is data availability, since estimations based on limited data are associated with high uncertainties. This is a general issue when modeling extreme events, as these events are by definition rare. With respect to extreme precipitation, however, the available data are further limited by the scarcity of the spatial coverage especially for observations with high temporal resolution (Courty et al., 2019). There are large global differences, but even in Germany, with one of the superior networks of gauge stations, data availability for modeling IDF curves poses difficulties. Therefore, a focus of the ongoing research is the efficient use of available data to reduce uncertainties (e.g., Koutsoyiannis et al., 1998; Lutz et al., 2020; Van de Vyver and Demarée, 2010), as well as the generation of IDF curves at locations with little or no observational data (e.g., Gaur et al., 2020; Mailhot et al., 2013; Vandeskog et al., 2021).

Another key concern with respect to modeling extreme precipitation events is the ever-growing realization that we live in a changing climate in which the frequency of extreme precipitation events is expected to increase (e.g., Fischer et al., 2014; Meredith et al., 2021; Trenberth et al., 2003). As previously addressed, the severity of precipitation events depends, among other things, on the parameters of water vapor content of the air and the movement of precipitation fields. Due to the increasing global mean temperature, the water vapor content in the air increases as a warmer atmosphere can hold more moisture. In contrast, the reduction of the meridional temperature gradient leads to a weakening of the jet stream which might cause an increase in the frequency of persistent weather situations (Davini and D'Andrea, 2020; Detring et al., 2021) where precipitation fields tend to propagate slowly. These two effects indicate that the severity and frequency of extreme precipitation events is likely to increase in the future. Many studies already suggest observed changes in the characteristics of extreme precipitation events, mainly based on daily precipitation records (Barbero et al., 2017; Scherrer et al., 2016; Westra et al., 2013; Zeder and Fischer, 2020). These possible changes need to be considered when assessing the risk of failure in infrastructure design. Furthermore, they might also generate the requirement to adapt existing structures to a future climate. Therefore, another crucial research objective is to incorporate non-stationarity and thus develop models that allow predicting the relationship between precipitation intensity, duration and frequency for the near future (Katz et al., 2002; Rootzén and Katz, 2013). Due to the high spatial and temporal variability of precipitation, however, the prediction of future changes is associated with large uncertainties (Moberg and Jones, 2005). Therefore, to address this objective, we must first improve our understanding of factors influencing the variations in precipitation intensity on different timescales.

The studies included in this thesis present consistent models for precipitation intensity, which allow reducing the number of parameters and thus decreasing the uncertainties. On this basis, it is possible to investigate the influence of additional parameters on the intensity and frequency of extreme precipitation events more accurately and in greater detail. Consequently, although throughout this thesis we assume a stationary climate,

the studies included are important steps towards predicting changes in the risks posed by extreme precipitation events.

Objectives and Research Questions

The three studies included in this thesis listed on page v address the analysis of variations of extreme precipitation on different timescales. A main objective is to improve the estimation of IDF curves by more efficient use of the available data. For this purpose, we combine different data – originating from different durations, stations or months – within one model. We use a duration-dependent extreme value model in which we include spatial or seasonal variations. Thereby, we also attempt to better understand the variations in precipitation intensity and the underlying processes.

Study I aims to include precipitation observations of different durations and from different stations simultaneously in the modeling of IDF curves. Thereby, we want to combine knowledge from records of different lengths. Our research questions are:

- To what extent can we improve the estimation of IDF curves by combining observations of different stations and durations?
- Does this method provide reliable estimates of IDF curves at locations without observations?

Study II focuses on the simultaneous use of data of different durations at a single station. We aim to improve the model which describes the dependence of precipitation intensity on duration. For this, we consider different levels of complexity to describe this relationship and investigate:

- What modifications to the model lead to an improved performance?

Furthermore, we evaluate:

- Are we able to provide meaningful uncertainties for the estimated IDF curves?

Study III examines the seasonal variations of the IDF relationship. So far, methods for estimating IDF curves have mostly neglected the seasonality of extreme events, as they are not considered relevant for planning flood protection infrastructure which exists all year round. However, there are two advantages of considering seasonality: First, it allows the use of a larger amount of data on extreme events. On the other hand, since extreme events on different timescales potentially occur in different seasons, it allows us to examine in more detail the underlying processes that influence the IDF relationship. Our research questions are:

- How does the IDF relationship evolve throughout the year?
- What differences occur among stations in this regard?

- What are the consequences for the IDF curves that result from the more complex seasonal model?
- Since using monthly maxima leads to a reduction of the uncertainties, what else can we learn about the IDF relationship via this new approach?

Outline of the Thesis

In Part I of the thesis we provide the scientific background and context. Since this thesis primarily addresses the development of methods, we focus on the introduction of statistical methods for the estimation of IDF curves. In addition to the methods used in the studies, we aim to provide an overview of the possible alternatives. We first outline the methods used for statistical modeling of extreme precipitation in Chpt. 2. They do not enable simultaneous modeling of data of different durations. The approach that we apply for this purpose will be introduced in Chpt. 3. We aim to illustrate the advantages of a consistent approach by comparing this method to the one currently used in Germany to provide the official design values. We take a brief detour in Sec. 3.4 to touch upon the issue of statistical modeling of dependent extreme values focusing on its relevance for IDF curves. Part II contains the three studies. We briefly summarize the results of the studies with respect to the formulated research questions in Part III. In addition, we provide further material regarding the possible application of the developed methods as well as further analysis to put the results of the studies into context. The thesis ends with suggestions for future studies in the context of the presented research in Chpt. 8.2, where we outline two promising ideas in detail and provide some first insights.

Part I

Scientific Background

Chapter 2

Statistical Models for Extreme Precipitation

This chapter introduces the basic concepts and methods for the statistical modeling of extreme values used in the included studies and provides an overview of alternative approaches.

The process of model building can generally be divided into the steps presented in Fig. 2.1: Even before the model building starts, the first step is the identification of the question to be addressed. The question defines what kind of observations need to be collected and which model approach might be suitable. The steps of model selection and verification are used to review the model. Model selection focuses on the choice of meaningful predictors, while verification tests the suitability of the model in answering the research question. Once a suitable model has been found, the model can finally be applied to make predictions and provide answers to the question posed initially. However, the model should not be considered as final, since new observations are constantly acquired, which make it necessary to re-evaluate the individual steps. In the following, we will discuss

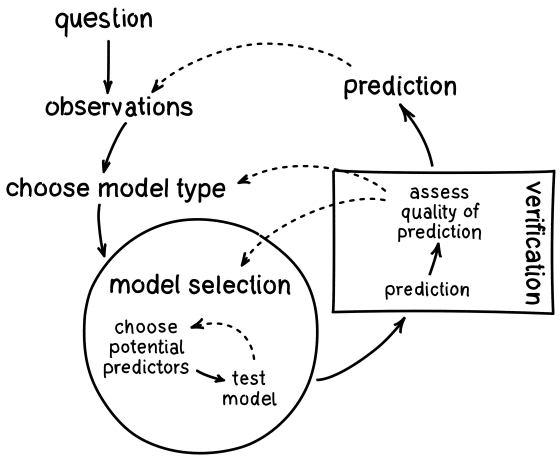


Figure 2.1: Scheme of necessary steps in the process of building a statistical model. Adapted from Rust (2021).

the individual steps in detail, focusing on the statistical modeling of extreme precipitation.

2.1 Relevant Questions Concerning Extreme Precipitation

The principal question to be answered, especially from an engineering point of view, could be formulated as:

(1) What amount of rainfall should we expect at a given location in a certain period of time?

From a meteorological perspective, we are more concerned with questions like:

(2) Which drivers influence the frequency and intensity of heavy rainfall?

(3) What changes can be expected in the future?

The studies included in this thesis present methods that allow us to answer question (1) with higher accuracy. They are, however, only a step towards addressing questions (2) and (3). Moreover, the methods applied here allow us to consider only a small aspect of assessing the hazards posed by extreme precipitation. Further relevant questions that require methods beyond those addressed in the following are for example:

(4) What amount of rainfall should we expect for an entire catchment area in a certain period of time?

(5) What are the impacts of a heavy precipitation event of a given magnitude on infrastructure or society?

2.2 Precipitation Observations

Precipitation amount or intensity can be measured in situ, as well as determined by remote sensing. Most of the studies on extreme value modeling of precipitation, as well as the studies included here, are based on station data. Precipitation amount is reported in volume per area with units of l/m^2 or mm. However, station gauges do not actually measure precipitation amount per $1\,m^2$ but have a smaller collecting area of typically $200\,cm^2$. The gauges are based on either volume or weight recording. The volume can be recorded manually, or by counting drops of a certain size, or via a tipping bucket (Löffler, 2012). The German Meteorological Service (DWD) currently operates 2238 precipitation measuring stations in Germany.[1] Figure 2.2 (a) presents their positions. At 1201 of these stations, the precipitation volume is collected with a Hellmann gauge (see Löffler, 2012) and measured manually once a day. At the remaining 1037 stations, the amount of precipitation is recorded every minute with a digital measuring device.[2] The digital gauges currently

[1] According to `https://opendata.dwd.de/climate_environment/CDC/help/RR_Tageswerte_Beschreibung_Stationen.txt`, as of 16 December 2021.

[2] According to `https://opendata.dwd.de/climate_environment/CDC/help/zehn_min_rr_Beschreibung_Stationen.txt`, as of 16 December 2021.

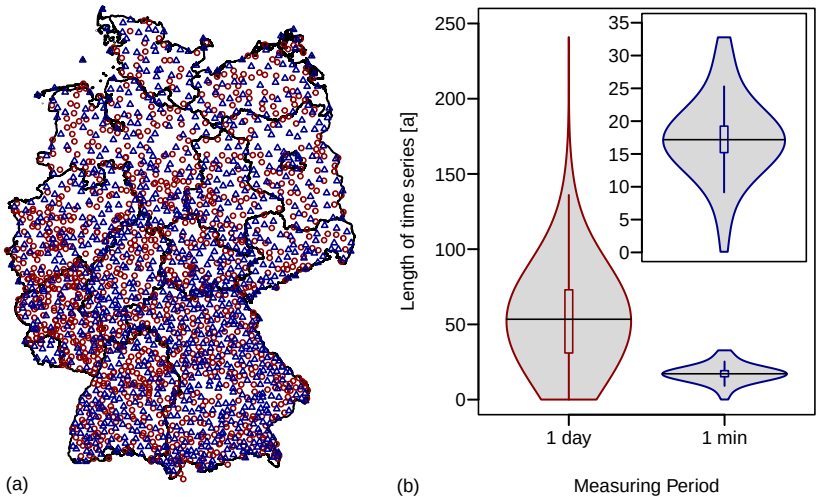


Figure 2.2: Precipitation measuring stations from the DWD: (a) locations of currently operated stations within Germany and (b) distributions of the length of all available time series with a measuring period of 1 day (red) and 1 min (blue). The inset shows a zoomed view of distribution of the 1 min data.

in use measure by determining the change in weight of the collection container after each measurement interval. The stations with daily and one-minute measurement frequencies differ in the length of their available time series. Figure 2.2 (b) shows the length of all available time series[3] as a violin plot, including stations that are no longer operated. The median of the daily stations is about 50 years, even though 379 stations with a time series of 100 years or longer exist. The station with the longest time series is Hohenpeißenberg in Bavaria, which starts in 1781. In contrast, the median of the minute time series is 17 years with time series of 30 years or slightly more available at 53 stations.

The spatial and temporal coverage of the official precipitation stations in Germany is comparable to that in other European countries[4] or in the U.S.[5]. In less accessible and sparsely populated areas, the spatial coverage is lower (e.g., Dyrrdal et al., 2015; Mekis et al., 2018). Furthermore, data availability is scarce in many low- to middle-income countries, especially in the global south and the tropics (e.g., Liew et al., 2014; Van de Vyver and Demarée, 2010). In contrast to gauging stations, remote sensing instruments such as weather radars or satellites provide precipitation estimates with high spatio-temporal resolution. Due to their limited records and the additional uncertainties associated with

[3]Missing values are not taken into account when calculating the time series length.

[4]See for example Switzerland `https://www.meteoswiss.admin.ch/home/measurement-and-forecasting-systems/land-based-stations/automatisches-messnetz.html`, Sweden `https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer` or Great Britain `https://www.metoffice.gov.uk/research/climate/maps-and-data/uk-synoptic-and-climate-stations` (last access: 21 December 2021).

[5]See `https://www.ncei.noaa.gov/maps/hourly/` (last access: 21 December 2021) for an overview of the spatial coverage of precipitation gauges in the U.S. and also worldwide.

remote sensing, the possibility of using such precipitation data sets to develop IDF curves has only recently emerged as a research topic (Goudenhoofdt et al., 2017; Marra et al., 2017; Ombadi et al., 2018). Another possible data source could be reanalysis data (e.g. Courty et al., 2019), however extreme precipitation events are typically underestimated in these data sets (Hu and Franzke, 2020).

2.3 Modeling Approaches for Extremes

After formulating a problem and collecting observations, the next step is to find a model which describes the data and is suitable for addressing the question. In some cases such a model can be derived based on basic process properties. For example, for the number of radioactive decay processes over time, the Poisson distribution is a valid model. It describes the probability that a certain number of events take place under the assumption that they occur at a constant mean rate and independently of each other (Coles, 2001, Chpt. 2). However, such a derivation is often not possible, so that more complex processes mostly require the empirical choice of models which describe the available data sufficiently well. Yet, when considering extreme events, a further possibility arises: the use of limit laws to approximate a model.

A widely used limit law in statistics is the Central Limit Theorem. It concerns the mean value $\bar{X}_n$ of n independent and identically distributed random variables X_i

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

When $n \to \infty$, the distribution of the mean value $\bar{X}_n$ converges to a Normal distribution. Therefore, for large n we can approximate the distribution of $\bar{X}_n$ with a Normal distribution, without having to know the distribution of the random variables X_i (Coles, 2001, Chpt. 2). The asymptotic theory of sample extremes can be considered as an analogue of the Central Limit Theorem for the distribution of the maximum value

$$M_n = \max\{X_1, ..., X_n\}.$$

According to the Extreme Value Theorem or Fisher-Tippett-Gnedenko Theorem, if the distribution of normalized maxima converges, then it converges to the Generalized Extreme Value (GEV) distribution

$$G(z) = \begin{cases} \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} & \text{, for } \xi \neq 0 \\ \exp\left\{-\exp\left[-\left(\frac{z-\mu}{\sigma}\right)\right]\right\} & \text{, for } \xi = 0 \end{cases}, \tag{2.1}$$

defined on $\{z : 1 + \xi(z - \mu) > 0\}$ and with location parameter $-\infty < \mu < \infty$, scale parameter $\sigma > 0$ and shape parameter $-\infty < \xi < \infty$ (Coles, 2001, Chpt. 3). The parameters μ and σ respectively control the position and the spread of the distribution, as indicated by the probability densities in Fig. 2.3 (a-b). The parameter ξ determines the behavior of the right tail. The three cases for ξ are illustrated in Fig. 2.3 (c). In case $\xi = 0$

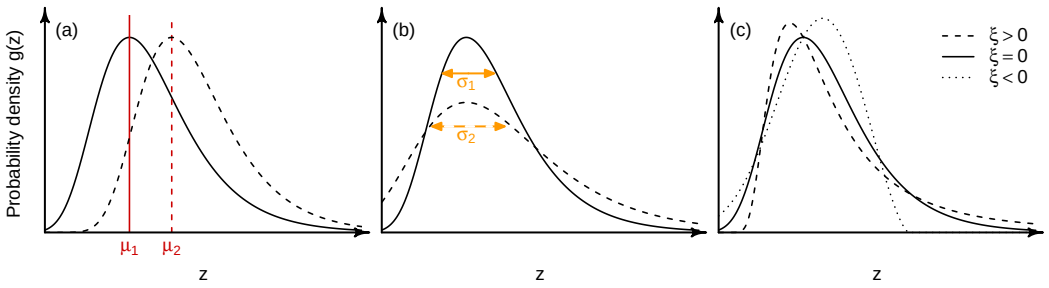


Figure 2.3: Probability density of the GEV distribution with different choices for (a) location μ, (b) scale σ and (c) shape ξ parameters.

we obtain the Gumbel distribution with exponentially decreasing probability density. For $\xi > 0$ the Fréchet distribution with polynomially decreasing tail results. In both cases, the distribution exhibits no upper bound. In contrast, in the case $\xi < 0$ we obtain the Weibull distribution, which possesses a finite upper endpoint.

Since the estimation of the shape parameter is associated with considerable uncertainties, especially for small samples, it is not uncommon to use the special case of the Gumbel distribution with a fixed parameter $\xi = 0$ when modeling extreme precipitation or runoff (Grieser et al., 2007; Svensson and Jones, 2010). However, this approach causes two complications: First, if the choice of $\xi = 0$ is inappropriate to describe the available data sufficiently well, this will lead to bias in the prediction. Second, in the subsequent inferences it is assumed that this choice for ξ is correct and the uncertainties arising from it are not taken into account. This will be discussed in more detail in Sec. 3.1.1.

Block Maxima Approach The Extreme Value Theorem holds even if the random variables are not independent, simply resulting in a slower convergence of the distribution. Thus, the following approach to model extreme values emerges: the data are separated into blocks of n observations. For large enough n, the distribution of the block maxima can be approximated with a GEV distribution. The choice of block size is a typical trade-off between bias and variance (Coles, 2001, Chpt. 3). If the blocks are too small, the GEV distribution is likely not a good approximation of the distribution and the values estimated on the basis of this model become biased. With increasing block size, the estimates are eventually based on very few data points and thus have a high variance. In the geosciences, however, the choice of block size is constrained by another consideration: Most of the variables of interest exhibit seasonality. Thus, with a block size of less than one year, the maxima of the different blocks may not have a common distribution making modeling more complicated. Therefore, in geosciences, it is common to adopt a block size of one year. Although this work is based on the use of the GEV distribution to model extreme precipitation events, in the following, we will briefly introduce two complementary approaches.

Alternative Distributions for Block Maxima Hosking and Wallis (1997, Chpt. 5) propose to test other distributions additionally to the GEV distribution, arguing that the conditions

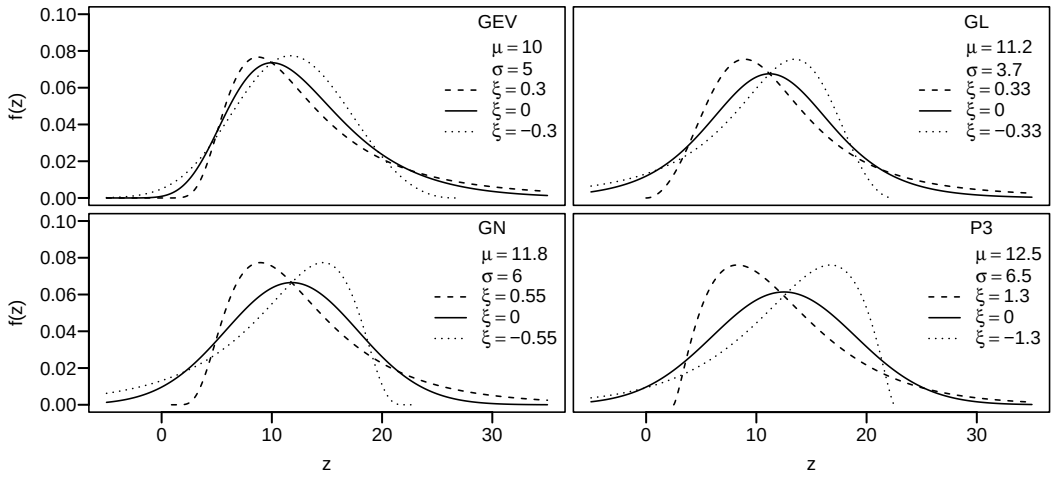


Figure 2.4: Probability density of the Generalized Extreme Value (GEV), Generalized Logistic (GL), Generalized Normal (GN) and Pearson type III (P3) distributions, for different shape parameters ξ, respectively. The parameters μ, σ and ξ are set to provide a good visualization of similarities for the case $\xi > 0$ (dashed).

necessary to approximate the distribution of annual maxima with a GEV distribution are not sufficiently fulfilled for variables such as runoff or precipitation. No asymptotic justification for these other distributions exist, but they have been identified on an empirical basis as being well suited for modeling the distribution of annual precipitation or runoff maxima. The three-parameter distributions suggested by Hosking and Wallis (1997) as possible candidates are the Generalizes Logistic (GL), Generalized Normal (GN) and Pearson type III (P3) distributions, as defined in Table A.1 (Appendix). From their respective probability densities, presented in Fig. 2.4, it becomes evident that all of these distributions can be both left-skewed and right-skewed, similar to the GEV distribution, depending on the choice of the shape parameter. In contrast, for the special case $\xi = 0$ all three distributions yield a symmetric distribution. The National Oceanic and Atmospheric Administration (NOAA) considers the distributions provided in Table A.1, among others, to produce the Precipitation-Frequency Atlas of the United States (Perica et al., 2018). In their approach, the distribution that best describes the observed data is selected based on various tests.

Peaks over Threshold Approach Modeling annual maxima results in only one extreme event per year being considered in the analysis, although several extreme events might occur within one year. The available data on extreme events can therefore be used more effectively if, instead of block maxima, the values above a certain threshold u are regarded as extreme. In case of convergence, the limiting distribution of threshold exceedances for

$u \to \infty$ is the Generalized Pareto distribution (GPD)

$$H(x-u) = \begin{cases} 1 - \left[1 + \frac{\xi(x-u)}{\sigma}\right]^{-1/\xi} & \text{, for } \xi \neq 0 \\ 1 - \exp\left(-\frac{x-u}{\sigma}\right) & \text{, for } \xi = 0 \end{cases}, \tag{2.2}$$

defined on $\{x : x > u \text{ and } \xi(x-u)/\sigma > 0\}$ with scale parameter $\sigma > 0$ and shape parameter $-\infty < \xi < \infty$ (Coles, 2001, Chpt. 4). Analogous to the GEV distribution, for $\xi \geq 0$ the GPD has no upper limit, while for $\xi < 0$ an upper bound exists. The special case $\xi = 0$ leads to an exponential distribution. As with the block maxima approach, the choice of threshold represents a trade-off between bias and variance. The aim is to select the threshold as low as possible, so that the GPD still represents a good approximation. There are different methods for threshold selection: the mean residual life plot, the stability of the estimated parameters (Coles, 2001, Chpt. 4) or the definition of an average number of exceedances per year (Mailhot et al., 2013), to name just a few. The circumstance that the choice of the threshold is to some degree subjective should not be considered as a disadvantage of the peak over threshold (POT) approach, as the problem of choosing a block size generally exists for the block maxima approach as well, except that the choice is more restricted in the case of seasonality of the signal. Conversely, however, if seasonality is present using the POT approach will still result in threshold exceedances originating only from certain seasons, unless a seasonally varying threshold is adopted. Given its higher efficiency, the POT approach is a frequently used alternative to the block maxima approach when modeling extreme precipitation.

Nevertheless, since in this thesis we aim to study the relationship of precipitation intensity and duration, the use of the POT approach is not necessarily suitable in our case. A consistent POT model for the IDF relationship would require a different threshold for each duration or rather a model for the threshold depending on duration. We will later demonstrate how the inclusion of spatial or seasonal variations into a GEV model similarly provides the opportunity to consider more data than just the annual maxima at a certain station. Therefore, here we prefer the less complex block maxima approach to the POT approach.

2.3.1 Parameter Estimation

With the exception of the GPD, the extreme value distributions presented above each have the three parameters location μ, scale σ and shape ξ. Even if we assume that a particular distribution describes the data well, the exact values for the parameters μ, σ and ξ are unknown and need to be estimated on the basis of the observed data. An estimation method should ideally yield unbiased, consistent, and efficient estimates. For the presented distributions, numerous estimation methods are offered in the literature. The most common methods used in hydrological applications are probability weighted moments (PWM) (Hosking et al., 1985), L-moments (Hosking and Wallis, 1997), maximum likelihood estimation (MLE) (Coles, 2001) or Bayesian inference (Stephenson, 2016). In the following, we will briefly discuss these four methods with focus on the GEV distribution. A comparison of estimation methods, including also methods which are rather unknown to

the engineering community, based on their application to extreme wind speed modeling is provided by Soukissian and Tsalis (2015).

PWM and L-moments

The PWM and L-moments methods are both based on the method of moments, which can be considered as the simplest method of parameter estimation for many statistical applications. Thereby, the distribution moments are mathematically equated with the sample moments (Wilks, 2011, Chpt. 4). The first three moments of the cumulative distribution function $F(Z)$ of a random variable Z are: expected value $\mathrm{E}(Z) = m_1$, variance $\mathrm{Var}(Z) = m_2$ and skewness $g = m_3/m_2^{3/2}$, defined using

$$m_1 = \mathrm{E}(Z) \quad \text{and} \quad m_r = \mathrm{E}[Z - \mathrm{E}(Z)]^r, \quad r > 1,\ r \in \mathbb{N}. \tag{2.3}$$

For any specific distribution Eq. 2.3 results in relationships between the moments and the parameters of the distribution. The most familiar example is the Normal distribution, where the moments are expressed by the distribution parameters as: $\mathrm{E}(Z) = \mu$, $\mathrm{Var}(Z) = \sigma^2$ and $g = 0$. For the GEV distribution, the moments as functions of the distribution parameters are provided in Eqs. (A.5-A.9).

The moments can be estimated from the observations z_i using, for example[6]

$$\hat{m}_1 = \bar{z} = n^{-1} \sum_{i=1}^{n} z_i \quad \text{and} \quad \hat{m}_r = n^{-1} \sum_{i=1}^{n} (z_i - \bar{z})^r. \tag{2.4}$$

Equating Eq. (2.3) with Eq. (2.4) results in a system of equations that can be solved for the parameters of the distribution yielding the moment estimators. In the simple case of the Normal distribution, the well known estimators $\hat{\mu} = \bar{z}$ and $\hat{\sigma} = \left\{n^{-1} \sum_{i=1}^{n} (z_i - \bar{z})^2\right\}^{\frac{1}{2}}$ result. The moment estimators for the Gumbel parameters are provided in Eqs. (A.10-A.10). However, for the moments estimators of the GEV parameters, there is no analytical form. They can only be obtained by equating Eqs. (A.5-A.7) with Eq. (2.4) and solving numerically for μ, σ and ξ.

Additionally, these estimators provide unreliable results for distributions with large skewness like the GEV distribution and other previously presented distributions. Therefore, the PWM and L-moments methods instead consider quantities that place a stronger weight on the tail information (Hosking and Wallis, 1997, Chpt. 2). The probability weighted moments are defined as

$$\beta_r = \mathrm{E}\{Z \cdot [F(Z)]^r\}, \quad r \in \mathbb{N}. \tag{2.5}$$

An estimator for β_r based on the observations z_i is naturally

$$\hat{\beta}_r = n^{-1} \sum_{i=1}^{n} z_i \cdot [\hat{F}(z_i)]^r, \tag{2.6}$$

[6] The presented moment estimators $\hat{m}_r$ for $r > 1$ are not unbiased. For other possible estimators see for example Hosking and Wallis (1997, Chpt. 2).

where the empirical distribution $\hat{F}(z_i)$ results from sorting the observations $z_1 \leq z_2 \leq ... \leq z_n$ with[7]

$$\hat{F}(z_i) = \frac{i}{n+1}. \tag{2.7}$$

Similar to the method of moments, expressing Eq. (2.5) for a specific distribution yield relationships between β_r and the parameters of the distribution. For an example, the analytical expressions for β_r for the GEV distribution are is given in Eqs. (A.12-A.13). Again, equating Eq. (2.5) and (2.6) yields a system of equations that can be solved either analytically or numerically for the distribution parameters to obtain their PWM estimators.

The L-moments introduced by Hosking (1990) are linear combinations of β_r, providing an easier relation of the L-moments to the parameters of the distribution. However, in terms of parameter estimation, there is no difference between PWM and L-moments, since they yield identical results (Rasmussen, 2001).

MLE

The maximum likelihood approach to parameter estimation is based on a reinterpretation of the joint probability density $f_{z_1,...,z_n}(z_1, ..., z_n)$ of the observations z_i. Instead of being considered as the probability density of the observations for given parameter vector $\mathbf{\Theta}$, the likelihood is a function of the parameters for a fixed set of observations

$$L(\mathbf{\Theta}|z_1, ..., z_n) = f_{z_1,...,z_n}(z_1, ..., z_n; \mathbf{\Theta}). \tag{2.8}$$

Hence, L is a measure of how strongly the observations support certain parameters values (Coles, 2001, Chpt. 2). The parameter values for which the likelihood is maximized serve as an estimate for the parameter vector

$$\hat{\mathbf{\Theta}} = \arg\max_{\mathbf{\Theta}} L(\Theta|z_1, ..., z_n). \tag{2.9}$$

An analytical solution to this problem is rarely available, therefore, most of the time the maximization problem has to be solved by numerical nonlinear optimization methods. Instead of the likelihood, it is convenient to optimize the logarithmic form $\ln[L(\mathbf{\Theta}|z_1, ..., z_n)]$, the so-called log-likelihood $l(\mathbf{\Theta}|z_1, ..., z_n)$. Since the logarithm is monotonous, L and l are maximized by the same parameter values. The general form of the likelihood and log-likelihood results in

$$L(\mathbf{\Theta}|z_1, ..., z_n) = \prod_{i=1}^{n} f(z_i; \mathbf{\Theta}), \tag{2.10}$$

$$l(\mathbf{\Theta}|z_1, ..., z_n) = \sum_{i=1}^{n} \ln\left[f(z_i; \mathbf{\Theta})\right] \tag{2.11}$$

[7] Numerous different estimators for the empirical distribution $\hat{F}(z_i)$, also known as empirical cumulative probability or plotting position, have been proposed in the literature (Guo, 1990).

under the assumption that the observations are independent and equally distributed. Thus, the log-likelihood of independent, GEV distributed realizations is

$$l(\mu,\sigma,\xi|z_1,...,z_n) = -n\ln(\sigma) - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{n}\ln\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right] - \sum_{i=1}^{n}\ln\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}, \tag{2.12}$$

$$l(\mu,\sigma|z_1,...,z_n) = -n\ln(\sigma) - \sum_{i=1}^{n}\left(\frac{z_i-\mu}{\sigma}\right) - \sum_{i=1}^{n}\exp\left[-\left(\frac{z_i-\mu}{\sigma}\right)\right] \tag{2.13}$$

for $\xi \neq 0$ and $\xi = 0$, respectively. Special considerations arising for example from the restrictions on the parameters are discussed in Coles (2001, Chpt. 3). We address the violation of the assumption of independent observations in Sec. 3.4.

Both the PWM and the MLE are consistent estimators and asymptotically multivariate Normal distributed. However, the PWM estimator is superior to the MLE for small samples regarding bias and mean-square error (Hosking et al., 1985). The properties for small sample sizes are especially important in extreme value statistics, where it is common that only few observations are available. Coles and Dixon (1999) argue that the superior performance of the PWM is due to the restriction on the parameter ξ, which follows from Eq. (A.12). They demonstrate that the small sample properties can be improved by penalizing the likelihood for undesirable values of $\xi > 1$.

The major advantage of the MLE compared to other methods for parameter estimation is its flexibility. Eq. 2.10 can be extended to model non-stationarity or influences of covariates, which are common for environmental processes (Coles, 2001, Chpt. 6). In general, an extension of the likelihood is possible if the distribution parameters Θ_i can be put into the following form

$$\Theta_j(\boldsymbol{X}) = h_j^{-1}(\boldsymbol{X}^T\boldsymbol{\beta}_j), \tag{2.14}$$

with a specified link function $h_j(\cdot)$, the parameter vector $\boldsymbol{\beta}_j$ and the vector $\boldsymbol{X}$ containing certain predictors. In the case of the GEV distribution we could imagine the example

$$\boldsymbol{\Theta}(\boldsymbol{X}) = \begin{pmatrix} \mu(\boldsymbol{X}) \\ \sigma(\boldsymbol{X}) \\ \xi(\boldsymbol{X}) \end{pmatrix} = \begin{pmatrix} \beta_0+\beta_1 X_1+\beta_2 X_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}, \tag{2.15}$$

where only the location parameter varies along the predictor variables X_1 and X_2. Based on the observations z_i and the corresponding vector of observed values for the predictor variables $\boldsymbol{X}_i$ we can obtain an estimate of the parameter vector $\hat{\boldsymbol{\beta}}$ simply by extending the likelihood to

$$L(\boldsymbol{\beta}|z_1,...,z_n) = \prod_{i=1}^{n} g\left[z_i;\mu(\boldsymbol{X}_i),\sigma(\boldsymbol{X}_i),\xi(\boldsymbol{X}_i)\right], \tag{2.16}$$

where g denotes the GEV density function. The extension basically implies that for every observation z_i there is now an individual set of parameters μ_i, σ_i and ξ_i. Therefore the

general form of the likelihood for any distribution with the probability density f results as

$$L(\boldsymbol{\beta}|z_1, ..., z_n) = \prod_{i=1}^{n} f\left(z_i; \boldsymbol{\Theta}_i\right), \tag{2.17}$$

where the parameter vector $\boldsymbol{\Theta}(\boldsymbol{X}; \boldsymbol{\beta})$ is a function of the predictor vector $\boldsymbol{X}$ and the parameter vector $\boldsymbol{\beta}$.

This property of the likelihood is essential for the studies included in this thesis, as it allows us to implement spatial and seasonal variations as well as the dependence of the GEV parameters on duration into the model.

Bayesian inference

Unlike the two previous approaches, the goal of Bayesian inference is not to estimate a specific value for each parameter. Rather, this method treats the elements of Θ as random variables with a specific distribution

$$\Theta_j \sim F_j(\boldsymbol{\lambda_j}), \tag{2.18}$$

where $\boldsymbol{\lambda_i}$ is a vector of hyper parameters. Bayesian inference concerning $\boldsymbol{\Theta}$ is expressed in terms of the conditional probability density for $\boldsymbol{\Theta}$ given the observations $\boldsymbol{Z}$, which is defined by Bayes' theorem as

$$f(\boldsymbol{\Theta}|\boldsymbol{Z}) = \frac{f(\boldsymbol{Z}|\boldsymbol{\Theta}) \cdot f(\boldsymbol{\Theta})}{f(\boldsymbol{Z})}. \tag{2.19}$$

In this context, $f(\boldsymbol{\Theta}|\boldsymbol{Z})$ is called the posterior density, $f(\boldsymbol{Z}|\boldsymbol{\Theta})$ is the likelihood, and $f(\boldsymbol{\Theta})$ is the prior density of $\boldsymbol{\Theta}$ (Stephenson, 2016). This prior density contains the beliefs we have about $\boldsymbol{\Theta}$ before we have seen the observations. For example, we may consider the domain of the parameter values or results from previous studies.

The divisor is called marginal likelihood, where $f(\boldsymbol{Z}) = \int f(\boldsymbol{Z}|\boldsymbol{\Theta}) f(\boldsymbol{\Theta}) d\boldsymbol{\Theta}$. This integral cannot be solved in most cases, especially in the field of extreme value statistics. However, since it does not depend on $\boldsymbol{\Theta}$, it is a constant that only affects the scaling of the posterior density, but not its overall form. The most commonly used method to obtain the posterior density is Markov chain Monte Carlo (MCMC) simulation (Dyrrdal et al., 2015; Lutz et al., 2020; Stephenson, 2016). It avoids the calculation of the term $f(\boldsymbol{Z})$ by instead generating a sample of parameter values from a converged Markov chain with the stationary distribution $f(\boldsymbol{\Theta}|\boldsymbol{Z})$. This way, a large sample of parameter values can be obtained from which $f(\boldsymbol{\Theta}|\boldsymbol{Z})$ can be estimated. Using this method we do not estimate a single set of parameter values, but instead a large sample of possible values for $\boldsymbol{\Theta}$, thus, we do not obtain a single extreme value distribution, but likewise a sample of possible distributions.

The Bayesian inference provides the major advantage that the uncertainties about the parameters are directly included in the prediction, where a predictive distribution is obtained instead of specific values. A drawback is that the MCMC method is extremely computationally demanding. In recent studies, less computationally extensive methods to

estimate $f(\boldsymbol{\Theta}|\boldsymbol{Z})$ such as ABC[8] (Erhardt and Sisson, 2016) or INLA[9] (Vandeskog et al., 2021) are also being applied for extreme value modeling.

The fact that an assumption must be made about the prior distribution of the parameters can be interpreted as both an advantage or disadvantage. Since extreme value modeling is often based on small data sets, the choice of prior can have a strong impact on the result. On the other hand, it might be considered as an opportunity to incorporate additional information external to the data into the model (Stephenson, 2016).

In the schematic representation in Fig. 2.1, the next steps in the model building process are model selection and verification. These steps focus on the model performance and thus the comparison of the predictions with the actual observations. Therefore, here it is necessary to first discuss the prediction in more detail.

2.3.2 Prediction

Our motivation for modeling the statistics of extreme events is to be able to estimate probabilities of rare and potentially damaging events, more specifically large amounts of precipitation that can lead to flooding. Conversely, we want to estimate the amount or intensity of precipitation that can be expected to occur with a specific probability within a certain period of time. Using the block maxima approach with the GEV distribution, we obtain this information by inverting Eq. 2.1. Using $G(q) = p$, we derive the quantile function $G^{-1}(p) = q_p$ as

$$q_p = \begin{cases} \mu - \frac{\sigma}{\xi}\{1 - [-\ln(p)]^{-\xi}\} & \text{, for } \xi \neq 0 \\ \mu - \sigma \ln[-\ln(p)] & \text{, for } \xi = 0 \end{cases}, \tag{2.20}$$

where q_p is the value that will not be exceeded with probability p within a block, or equivalently, that will be exceeded with probability $1 - p$ (Coles, 2001, Chpt. 3). Thus, for a block size of 1 year, p is the annual non-exceedance probability. It is common to refer to return periods instead of annual probabilities, since statistically the value q_p can be expected to be exceeded on average once within the period $T = 1/(1 - p)$. The quantile q_p might be called return level associated with a certain return period T. To clarify: An annual non-exceedance probability of 0.99 is equivalent to an annual exceedance probability of 0.01 or to a return period of 100 years. Thus, the 0.99 quantile is the same as the 100-year return level. However, the use of return periods is problematic for two reasons: First, it can mislead users into believing that such an event is expected to occur exactly once within the return period. This can lead to a false sense of security right after a flood or to a "flood is due" thinking when no flood has occurred (Grounds et al., 2018). Second, using return periods is only applicable in a stationary climate and becomes meaningless when the annual exceedance probability changes over time.

Plotting the quantile q_p (or return level) as a function of the non-exceedance probability p or the corresponding return period T according to Eq. (2.20), yields a return level plot.

[8] Approximate Bayesian Computation
[9] Integrated Nested Laplace Approximation

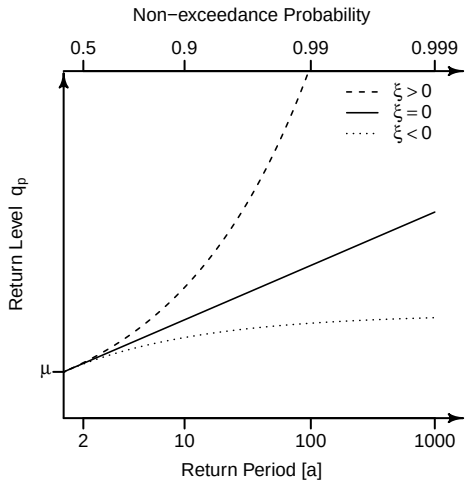


Figure 2.5: Return level plots of the GEV distribution for different cases of the shape parameter ξ.

If the p-axis is rescaled as $p' = -\ln[-\ln(p)]$, or similarly $T' = -\ln[-\ln(1-1/T)]$, the curve provides a simple indication of the sign of the shape parameter ξ: a concave curve is obtained for $\xi > 0$, a convex curve with asymptotic upper limit results from $\xi < 0$, and in the special case $\xi = 0$ a linear increase with slope σ and intercept μ occurs. Fig. 2.5 provides a schematic representation of these three cases. In addition to the presentation of the results, return level plots can be used for the verification of the model.

2.3.3 Uncertainties

Since results of statistical analyses are inherently uncertain, these results are only meaningful if they are communicated along the magnitude of their uncertainty. This is particularly relevant in extreme value statistics, where small changes of the model can have large effects on the results when extrapolating. This is especially the case for uncertainties in the shape parameter estimation, as indicated by Fig. 2.5.

In the case of Bayesian inference, a distribution is obtained for both parameter and quantile estimation, leading to a straightforward derivation of uncertainties. When using one of other presented methods for parameter estimation, uncertainty of estimated parameter values or resulting quantile values $\hat{\theta}$ can be expressed in terms of confidence intervals (Coles, 2001, Chpt. 2). A confidence interval is defined by the limits $\hat{\theta}_l$ and $\hat{\theta}_u$, so that

$$\Pr(\hat{\theta}_l < \theta < \hat{\theta}_u) = \gamma, \tag{2.21}$$

where γ corresponds to the coverage probability. The choice of γ is arbitrary, with large values yielding wider intervals. The values 0.95, 0.99 or 0.999 are commonly adopted. The definition implies only that for a large number of repetitions of the process that produced the data, $100 \cdot \gamma\,\%$ of the generated confidence intervals contain the true value θ. However, usually only one realization and therefore only one confidence interval exists. In this case, we cannot provide any information about the probability with which the true value θ lies

within the estimated confidence interval. Still, the confidence interval informs how much the estimated value $\hat{\theta}$ can vary due to sampling uncertainties. More narrow confidence intervals therefore indicate that we can be more confident in the estimate.

It is important to note that the confidence interval usually only accout for the sampling uncertainty and not for uncertainties that arise e.g., from the model assumptions. Therefore, confidence intervals should always be considered as a lower bound. In this thesis, two different methods are used to obtain confidence intervals. Namely, the delta method (Coles, 2001, Chpt. 2), which is based on the asymptotic normality of the MLE, and the bootstrap method, which is based on repeated sampling of observations (Davison and Hinkley, 1997). These methods are described in more detail in Chapter 4.2.5 and Appendix B.2. A more explicit consideration of the coverage of confidence intervals for dependent data is included in Chpt. 5.2.6.

2.4 Assessing the Model Performance

The goal of model selection is to identify relevant predictors and to avoid overfitting of the data (Wilks, 2011, Chpt. 7.4). In contrast, verification, at least in the context used here, serves to diagnose the prediction and thus to identify potential problems with the aim of improving the model (Wilks, 2011, Chpt. 9). Despite these different goals, in both cases we need to evaluate how the model performs. In other words, we are interested in how the predictions compare to the actual observations. In the following we introduce graphical methods as well as scoring rules applicable for this purpose.

Regardless of the applied method, low data availability remains the main problem when evaluating extreme value models and naturally validation of the extrapolation to not yet observed values can only be inferred on the basis of how well the model fits the already observed data.

Graphical Methods To analyze whether the model is in good agreement with the data it has been fitted to, it is advisable to use a graphical method. Therefore, (Coles, 2001, Chpt. 3) offers three techniques for model diagnosis, i.e. to compare the model predictions to the observations: p-p (probability–probability) plot, q-q (quantile-quantile) plot and return level plot. To provide an example, Figure 2.6 presents these three plots for a sampled data set and two different models (a) and (b), respectively.

In the p-p plot, the predicted non-exceedance probabilities $F(z_i)$ are plotted against the empirical probabilities $\hat{F}(z_i)$, which can be estimated from $\hat{F}(z_i) = \frac{i}{n+1}$ (Eq. 2.7) based on the sorted observations z_i. This yields the points

$$\left(\frac{i}{n+1}, F(z_i)\right), \tag{2.22}$$

with $i = 1, ..., n$ where in the context of fitting a GEV distribution $F(z_i) = G(z_i, \hat{\mu}, \hat{\sigma}, \hat{\xi})$.

In the q-q plot, on the other hand, the sorted observations z_i are plotted against the

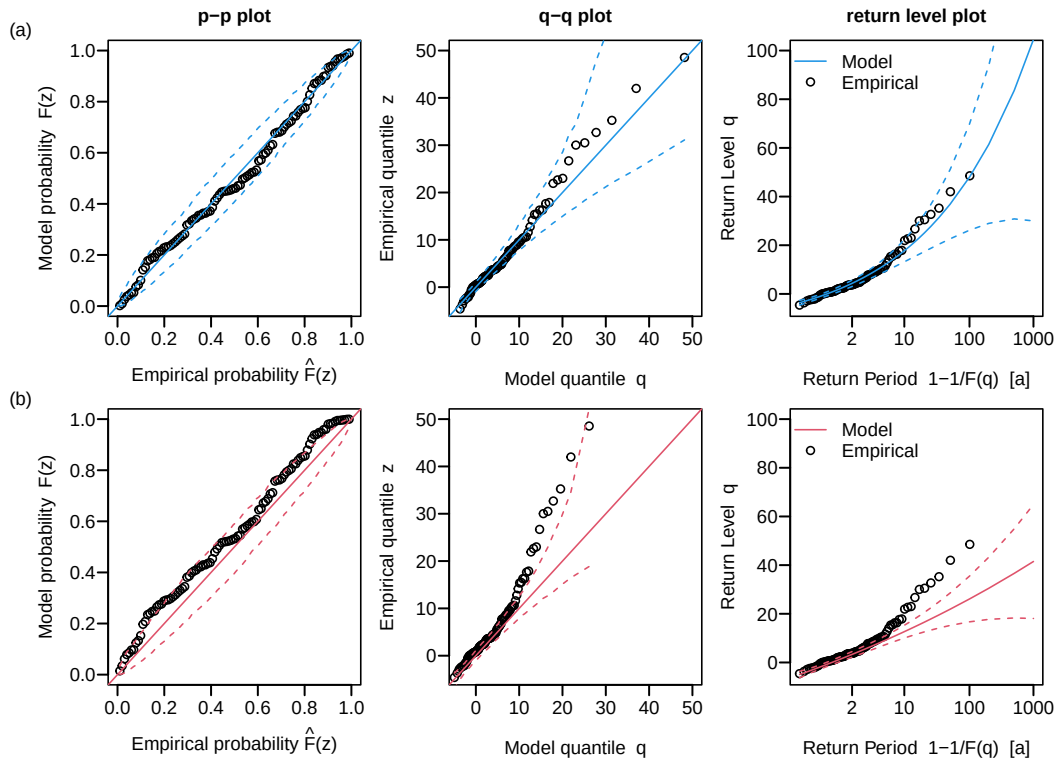


Figure 2.6: The plots for model diagnosis inform, how well two different models (a) and (b) represent the same data set. Detailed explanation follows from the text. 95 % confidence intervals are presented as dashed lines.

predicted quantiles $q_i = F^{-1}\left(\hat{F}(z_i)\right)$ resulting in

$$\left(F^{-1}\left(\frac{i}{n+1}\right), z_i\right). \tag{2.23}$$

In both cases, the goodness of fit is improved when the points are closer to the unit diagonal.

To use the return level plot for the model diagnosis, the predicted probabilities are plotted versus the return period as described in Sec. 2.3 - prediction as a line. When the empirical probabilities of the observations are additionally shown as points at the positions

$$\left(-\ln\left[-\ln\left(\frac{i}{n+1}\right)\right], z_i\right), \tag{2.24}$$

the return level plot can be used to compare the model with the observations.

All three plots contain the same information, expressed on a different scale. The q-q plot and especially the return level plot emphasize the upper tail of the distribution. Concluding from the examples in Fig. 2.6, there is no reason to doubt the validity of model (a), while the diagnostic plots for model (b) indicate a lack of fit. This is especially visible from the

q-q plot and the return level plot, since the high values of the data are outside the 95 % confidence intervals of the model predictions, which are shown as dashed lines. This is can be attributed to a poor shape parameter estimate of model (b) in the presented example.

If we assume a GEV model with non-stationary distribution parameters, see Eq. (2.14), the results can no longer be represented by a single return level plot, since different observations might correspond to different parameter values. Producing a p-p or q-q plot is still possible, but requires the transformation of the observations z_i so that they all have an identical distribution. One possibility is the transformation of the observations to the standard Gumbel distribution with parameters $\mu = 0,\ \sigma = 1$ and $\xi = 0$, which is achieved by

$$z_i^* = -\ln\left[\left(1 + \hat{\xi}_i \frac{z_i - \hat{\mu}_i}{\hat{\sigma}_i}\right)^{-1/\hat{\xi}_i}\right]. \tag{2.25}$$

The p-p and q-q plot can then be produced as described above, using z_i^* and $F(z_i^*) = G(z_i^*, 0, 1, 0)$.

Scoring Rules The presented graphical methods allow a comprehensive assessment of the agreement between one model and the observations, but they have limited applicability for the comparison of different models. Moreover, they require an individual review. As a method for comparing several models, for example within a model selection routine, it is therefore more suitable to summarize information on the model performance within a score. For this purpose, numerous scoring rules exist, which differ depending on the type of prediction (probabilistic or non-probabilistic) and observation (discrete or continuous). However, modeling an extreme value distribution does not fit into a single category, since we do obtain a probability distribution as a result, but the actual prediction is that of a quantile value (return level).

When considering the prediction of quantiles q_p, we are dealing with a non-probabilistic forecast. One measure of the accuracy of such forecasts is, for example, the mean squared error

$$\mathrm{MSE} = \frac{1}{n}\sum_{k=1}^{n}(q_{p,k} - o_k)^2, \tag{2.26}$$

where $(q_{p,k}, o_k)$ is a pair of prediction and corresponding observation. However, we cannot readily compute this or similar scores in this context because an observed value for the actual quantile o_k is not available.

Yet, there are two ways in which the MSE can still be used to evaluate quantile predictions. First, when using simulation studies to compare two different models (Roksvåg et al., 2021). In this case, data can be drawn from a known distribution. Therefore, the true value of the quantile is also known and can be used as o_k in Eq. (2.26). The second option, when using observational data, is to estimate an empirical distribution $\hat{F}$, see Eq. (2.7), in order to assign a non-exceedance probability $\hat{p}_k = \hat{F}(o_k)$ to each observation (Blanchet et al., 2016; Van de Vyver, 2018). Thus, each observation o_k can be compared to the model prediction for the corresponding quantile $q_{\hat{p}_k}$. In this case, the observations o_k could also be regarded

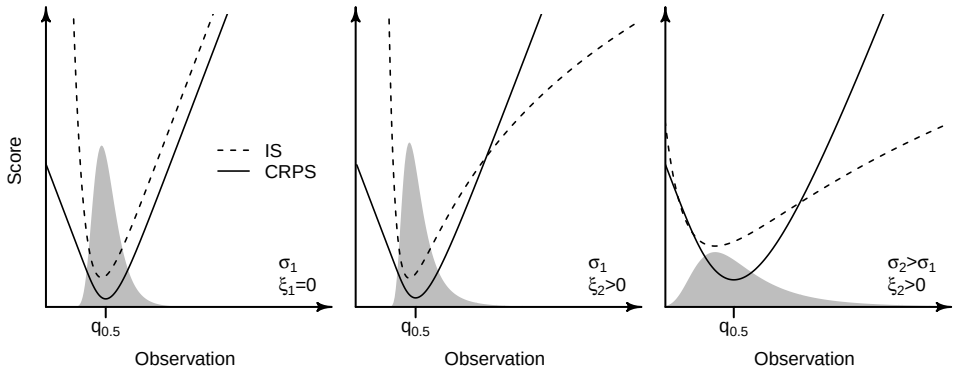


Figure 2.7: The CRPS (solid) and the IS (dashed) depending on the Observation value o for the GEV distribution with different values for ξ and σ as labeled in the figures. The (rescaled) probability density is indicated in gray. The parameter values chosen for this example are $\mu = 0$, $\sigma_1 = 1, \sigma_2 = 3, \xi_1 = 0$ and $\xi_2 = 0.2$.

as empirical quantiles. This way of calculating the MSE is directly related to the q-q plot, since it corresponds to the mean squared distance of the data points from the diagonal.

If we instead consider the complete extreme value distribution as the prediction, we can consider scoring rules designed for probabilistic forecasts. To evaluate the forecasts of a continuous distribution function $F(\cdot)$ and probability density function $f(\cdot)$, Wilks (2011, Chpt. 9) provides two scoring rules: the Ignorance Score or also called Logarithmic Score

$$\text{IS} = -\ln[f(o)] \tag{2.27}$$

and the Continuous Ranked Probability Score

$$\text{CRPS} = \int_{-\infty}^{\infty} [F(y) - F_o(y)]^2 \, dy, \tag{2.28}$$

where F_o is the step function

$$F_o(y) = \begin{cases} 0 & , y < o \\ 1 & , y \geq o \end{cases} \tag{2.29}$$

that can be considered as the cumulative distribution function of the observation o (Dyrrdal et al., 2015; Friederichs and Thorarinsdottir, 2012). Both scores are negatively oriented, i.e., smaller values represent a better prediction.

Fig. 2.7 shows both scores for a range of observations. The IS reaches its minimum when the observation o is equal to the mode, while the CRPS is minimal when o is equal to the median of the distribution. Further differences of the scores depend on the distribution as well as on the parameters. In the presented example, we can see that for $\xi > 0$, the CRPS shows a stronger increase on the right side of the minimum. Therefore, we can conclude that it penalizes observations in the upper tail of the distribution stronger than the IS.

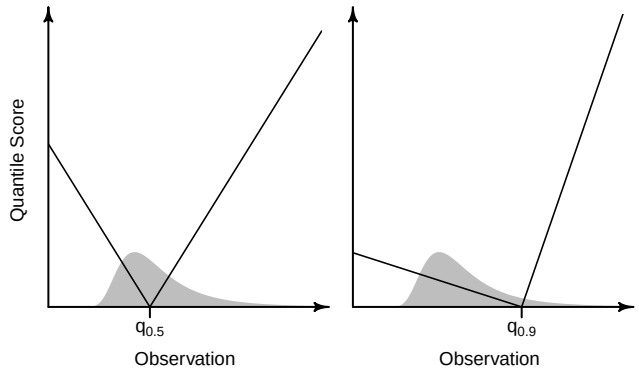


Figure 2.8: The QS depending on the value of the Observation o for the probabilities $p = 0.5$ (left) and $p = 0.9$ (right). The (rescaled) probability density of the GEV distribution is indicated in gray.

To evaluate the model performance based on n observation forecast pairs, we take the average value for both scores as follows

$$\overline{\text{IS}} = \frac{1}{n}\sum_{k=1}^{n} -\ln[f_k(o_k)], \quad \overline{\text{CRPS}} = \frac{1}{n}\sum_{k=1}^{n}\int_{-\infty}^{\infty}[F_k(y) - F_{o_k}(y)]^2 dy. \tag{2.30}$$

Here, the average Ignorance Score is proportional to the log-likelihood from Eq. (2.12), as $n \cdot \overline{\text{IS}} = l(\mu, \sigma, \xi | o_1, o_2, ..., o_n)$.

The disadvantage of using IS and CRPS in the context of extreme value statistics is that this way we can only draw a conclusion about the full probability distribution compared to the data, but not about the prediction of a specific quantile. A score that combines the advantages of the MSE and those of the IS and CRPS in this regard is the Quantile Score (Bentzien and Friederichs, 2014; Fischer et al., 2019). Both the actual observation o and the predicted quantile q_p are included in the calculation of the Quantile Score

$$\text{QS}(p) = \rho_p(o - q_p). \tag{2.31}$$

Their difference $u = o - q_p$ is weighted by the check function

$$\rho_p(u) = \begin{cases} pu & , u \geq 0 \\ (p-1)u & , u < 0. \end{cases} \tag{2.32}$$

Figure 2.8 presents QS for a range of observations for $p = 0.5$ and $p = 0.9$. It is evident that in both cases observations that are far from the value of the predicted quantile q_p are penalized more. In the case of $p = 0.5$ the weights are symmetric, while for $p = 0.9$ observations above $q_{0.9}$ are penalized stronger. The average QS is likewise used, when considering multiple observation forecast pairs[10]

$$\overline{\text{QS}}(p) = \frac{1}{n}\sum_{k=1}^{n}\rho_p(o_k - q_{p,k}). \tag{2.33}$$

[10]In the later chapters the term Quantile Score is used to refer to the average score $\overline{\text{QS}}$.

The QS is related to the CRPS, since the CRPS can be expressed as the integral

$$\text{CRPS} = 2 \int_0^1 \text{QS}(p)\, dp = 2 \int_0^1 \rho_p(o - q_p)\, dp. \tag{2.34}$$

This form of the CRPS also leads to another possibility of evaluation, where we do not only examine a certain quantile, neither the whole distribution, but a certain region of the distribution. This is interesting for extreme value statistics, since we might be particularly interested in the upper tail of the distribution. To this end, it is possible to modify the CRPS with a weighting function $\omega(p)$, in which case we obtain the so-called threshold weighted CRPS (Gneiting and Ranjan, 2011)

$$\text{twCRPS} = 2 \int_0^1 \omega(p) \cdot \rho_p(o - q_p)\, dp. \tag{2.35}$$

For example, a convenient choice for $\omega(p)$ is the step function

$$\omega(p) = \begin{cases} 0 & , p < p_0 \\ 1 & , p \geq p_0 \end{cases}, \tag{2.36}$$

since this yields a twCRPS, which considers only the upper part of the distribution (Vandeskog et al., 2021)

$$\text{twCRPS} = 2 \int_{p_0}^1 \rho_p(o - q_p) dp. \tag{2.37}$$

The derivation of the twCRPS from Eq. (2.37) for a GEV distribution is provided in the Appendix A.3.

So far, we have considered the scoring rules as a means of comparing the predictions with the observations that were used to fit the model. However, for model selection and verification, we are actually more interested in the model's ability to represent not yet observed data, or rather, observations that have not already been included in the parameter estimation. That is what is called the out-of-sample performance. For this purpose, it is necessary to split the data. The part that is used to fit the model is called the training set and an independent part upon which the model performance is evaluated – e.g., using one of the presented scores – is referred to as the test or validation set. If the data is split into exactly one training set and one test set, the method is called *hold-out*. In practice, however, this method leads to unstable results. To avoid this, the data should be split several times so that a score can be calculated for multiple test sets. A more stable result can then be obtained by averaging the scores of the individual test sets. This procedure is called cross-validation (Hastie et al., 2009, Chpt. 7.10).
Arlot and Celisse (2010) provide a detailed description of cross-validation in the context of model selection. They also give an overview about the various strategies for partitioning the data.

The presented methods for assessing the out-of-sample performance are relevant for both model selection and verification. In the following, we will briefly discuss some strategies that are unique to the individual procedures.

Model selection The aim of model selection is to choose from a set of possible predictor variables only those that are sufficient to produce a good predictive equation. For this purpose, different models need to be compared on the basis of a criterion. In the included studies, we use a cross-validated score function. Nevertheless, the cross-validation method tends to be computationally demanding, especially for a large set of possible models. In that case, there are other criteria proposed in the literature, which are based on the likelihood (Eq. 2.10) and do not involve cross-validation (Kim et al., 2017). For example, one such criterion is the Akaike Information Criterion (AIC)

$$\mathrm{AIC} = -\ln(L) + 2k, \tag{2.38}$$

where the first term is proportional to the log-likelihood and thus a measure of the in-sample performance of the model. The second term is proportional to the number of parameters k. Hence, models with many predictors are penalized without requiring explicit estimation of the out-of-sample performance.

As it is in most cases not feasible to compare all possible models, a strategy is needed to identify a reasonably good model. The most commonly used procedure is stepwise regression (Wilks, 2011, Chpt. 7.4). In forward selection, only the predictor that leads to the largest improvement is added to the model in each iteration step. Similarly, in backward elimination, the least important predictor is removed in each iteration step. The process stops when the model can no longer be improved, or after a fixed number of iterations. Hastie et al. (2009) provide an extensive overview of other possible strategies. They also present newer, more efficient methods, such as least absolute shrinkage and selection operator (LASSO) or boosting (Messner et al., 2017) , which combine parameter estimation and model selection in one step. However, the application of these methods in extreme value statistics has only recently emerged (e.g., Koh, 2021; Steinheuer and Friederichs, 2020).

Verification When determining the out-of-sample performance for a single model using a cross validated score as described above, the result is not very intuitive. Instead, it is helpful to compare the calculated score S_M with that of a reference model S_R. This comparison can be presented as a skill score. The general form of a skill score is

$$\mathrm{SS} = \frac{S_M - S_R}{S_{perf} - S_R}, \tag{2.39}$$

where S_{perf} is the value of the score that is achieved with a perfect prediction. In the case of the scores presented in Eqs. (2.26-2.37) respectively $S_{perf} = 0$. The skill score can be interpreted as the percentage of improvement of the model score over the reference score towards the perfect model, where if $S_M = S_{perf}$ then $\mathrm{SS} = 1$. For $\mathrm{SS} = 0$ there is no improvement of the model over the reference and for $\mathrm{SS} < 1$ the model even represents a deterioration (Wilks, 2011, Chpt. 9). However, due to the definition of SS no statement about the strength of the deterioration is possible, since $\mathrm{SS} \in (-\infty, 1]$.

In the context of the presented work we propose using a skill index (Ulrich et al., 2020), defined as

$$\mathrm{SI} = \begin{cases} \frac{S_M - S_R}{S_{perf} - S_R} & , S_M \leq S_R \\ -\frac{S_R - S_M}{S_{perf} - S_M} & , S_M > S_R. \end{cases} \tag{2.40}$$

This way, $\mathrm{SI} \in [-1, 1]$. Thus, the degradation of the model with respect to the reference in the case $S_M > S_R$ can be interpreted in the same way as the improvement, when $S_M \leq S_R$.

Chapter 3

Modeling the Relationship Between Precipitation Intensity and Duration

The studies included in this thesis focus on models for precipitation intensity as a function of duration and frequency, or non-exceedance probability. The extreme value models presented in Sec. 2.3 can only be used to describe extreme precipitation for one selected duration. We illustrate this concept with data from an example station: we can model the annual maxima at the station Bever-Talsperre for the different durations $d \in \{1, 4, 8, 16, 32, 60, 120, 240, 480, 960, 1440\}$ min by fitting a separate GEV distribution

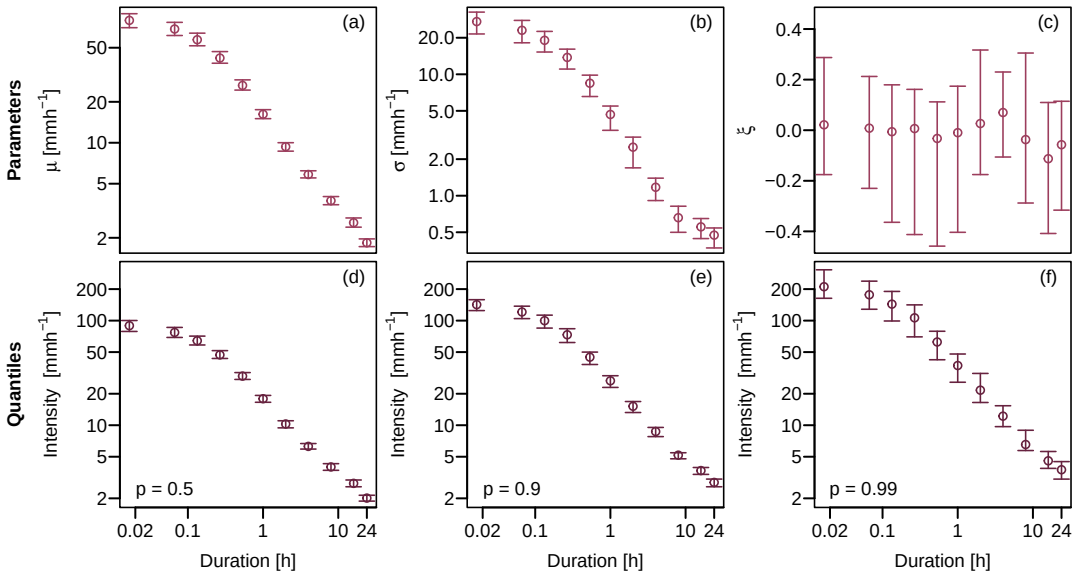


Figure 3.1: Modeling annual maxima at station Bever-Talsperre by applying a separate GEV model for each duration. The estimated GEV parameters μ, σ and ξ, as well as selected quantiles are presented depending on duration. For this example we use data from 1968-2018, i.e., 51 annual maxima are available for each duration.

for each duration. In Fig. 3.1 (a-c) the estimated GEV parameters are presented versus duration. The resulting quantiles for the non-exceedance probabilities $p \in \{0.5, 0.9, 0.99\}$ are shown in Fig. 3.1 (d-f), likewise depending on duration.

To arrive at an IDF model, we need to account for the dependence of intensity on duration. Within a univariate framework, it is possible to introduce this dependence either directly at the quantile level, or at the parameter level. However, the implementation at the quantile level requires the use of successive models to 1) estimate the quantiles of interest for a range of durations and 2) model their dependence on duration. Moreover, applying separate models for the dependence of certains p quantiles on duration can lead to inconsistent results, i.e. crossing of quantiles. In other words, for example, for certain durations, the 0.9 quantile might have a higher values than the 0.99 quantile. This can be resolved by applying a post processing procedure (Roksvåg et al., 2021).

It is therefore beneficial to implement the dependence on duration at the parameter level. Both the approach currently used to produce the official IDF curves provided by the German Meteorological Service (DWD) as well as the approach of a duration-dependent GEV (d-GEV), which is used in the studies included in this thesis, achieve the variation of precipitation intensity with duration by treating the distribution parameters as functions of duration. However, the official approach consists of several successive steps, while the d-GEV approach allows to model the maxima of different durations simultaneously within one model step. In the following, we will present and compare both approaches using precipitation observations from two selected example stations.

3.1 Application of Separate Sequential Models Following German Regulations

In Germany, the estimation of return levels (quantiles) for given durations and return periods (annual non-exceedance probabilities) used as a basis for hydraulic engineering design is provided by the German Meteorological Service (DWD) via the KOSTRA atlas (DWD). It contains estimated return levels for the durations 5 min up to 72 h and the return periods 1 year to 100 years on a $8 \times 8\,\mathrm{km}^2$ grid for entire Germany. At first, the return levels are determined station-based at locations with relatively long time series. In a subsequent step, a regionalization is carried out in order to provide full spatial coverage for Germany. The methods used to estimate return levels at individual stations are regulated by the *Merkblatt DWA A-531* of the German Association for Water, Wastewater and Waste[1] (DWA, 2017). We will abbreviate this method as DWA approach hereafter. This approach consists of three steps:

1. The parameters of the extreme value distribution are estimated separately for selected durations in the considered duration range.

2. The dependence of the parameters on the duration is modeled. For this purpose, the duration range is divided into smaller sub-ranges, for each of which an individual

[1] *Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall* (DWA)

model is applied.

3. The models for the duration sub-ranges are adjusted to prevent discontinuities at the boundaries.

A more detailed description of the methods used in each of the three steps is provided in the following sections.

3.1.1 First Step: Separate Extreme Value Model for Each Duration

To model the extreme value distribution of a selected duration, the guideline requires using either the POT approach or the block maxima approach with an annual block size. To reduce the number of parameters and the resulting uncertainties, the shape parameter of the extreme value distributions should be fixed to zero. Thus, the threshold exceedances are modeled using the exponential distribution and the annual maxima are modeled using the Gumbel distribution, which do not allow for heavy tail behavior. Since this thesis is focused on the block maxima approach, the following explanations refer only to modeling annual maxima using the Gumbel distribution. However, the subsequent model steps of the DWA approach described in Section 3.1.2 remein the same for modeling annual maxima as well as threshold exceedances.

The Gumbel distribution is obtained by estimating location μ and scale σ parameters from the observed annual maxima. The DWA guideline leaves the choice of estimation method to the user, but recommends a graphical method basen on linear regression in the return level plot (see Secs. 2.3 and 2.4). For reasons of consistency with the other approaches in this thesis we will instead use the maximum likelihood estimator (Eq. 2.9) for the examples presented in the following.

To compare the GEV and the Gumbel distribution regarding the process of parameter estimation and the uncertainties involved, we use observations from the two stations Bever-Talsperre and Berlin-Tempelhof. The observed annual maxima of precipitation intensity for different durations are shown in Figs. 3.2 (d) and 3.3 (d). At both stations, longer time series are available for the duration $d \geq 1\,\text{day}$ than for the sub-daily durations.

The estimated parameters of the GEV distribution for each duration are shown in gray in Figs. 3.2 (a-c) and 3.3 (a-c), respectively, along with their bootstrapped 95% confidence intervals. It is evident that there is considerable uncertainty associated with the estimation of the shape parameter ξ, especially for short time series.

In Figs. 3.2 (e) and 3.3 (e), the 0.99 quantiles, i.e. 100-year return levels, resulting from the estimated GEV parameters are shown along their 95% confidence intervals in gray. This highlights how the uncertainties in the parameter estimation affect the uncertainties of the derived quantiles. To give an example, we can consider the estimated 0.99 quantile for a duration of one hour. At station Berlin-Tempelhof only 25 years of observations are available for the sub-daily durations. The 95% confidence intervals of the estimated 0.99 quantile based on this data range from 28 mm/h to 120 mm/h. However, for the station Bever-Talsperre with sub-daily time series of 51 years, the 95% confidence intervals span a more acceptable range from 25 mm/h to 48 mm/h. Hence, unless sufficiently long time

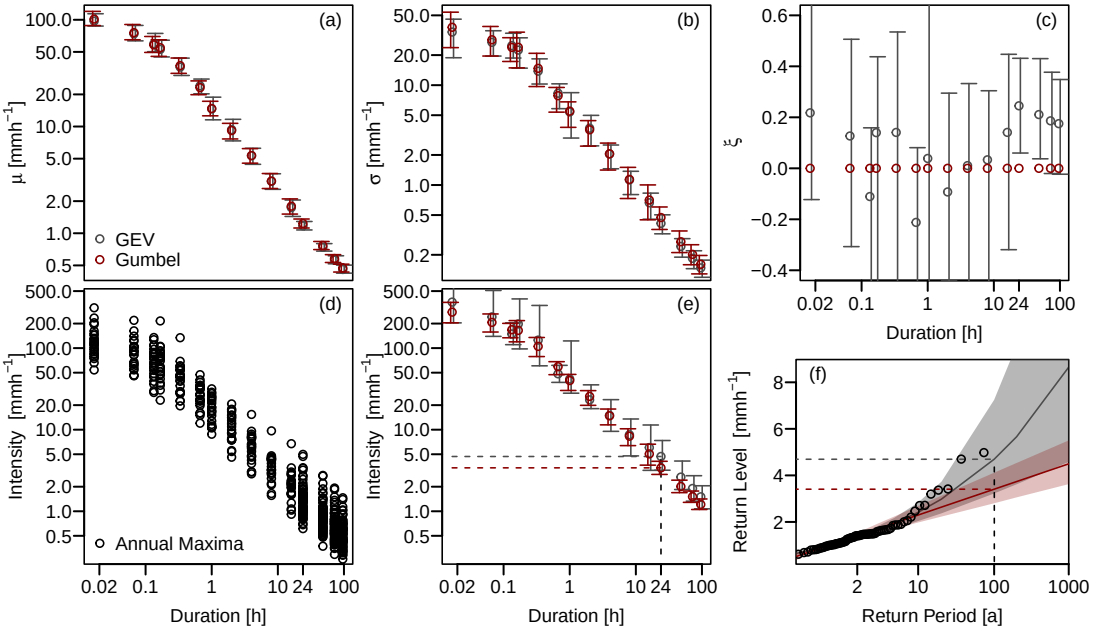


Figure 3.2: Modeling annual precipitation intensity maxima at station Berlin-Tempelhof by applying separate extreme value models for each duration. The annual maxima (d) are modeled using the GEV distribution (gray) and the Gumbel distribution (red), where $\xi = 0$. Both models are compared with respect to the estimated distribution parameters μ, σ and ξ (a-c) and 0.99 quantiles (e) for each duration. The return level plot (f) serves to asses the goodness of fit of both models for the selected duration of 24 h. Error bars and shaded areas represent the bootstrapped 95 % confidence intervals.

series are available, the separate application of the GEV for each duration does not lead to meaningful quantiles.

Therefore, the DWA guideline specifies limiting the distribution to the Gumbel distribution with $\xi = 0$. The parameter estimates as well as the 0.99 quantile along with the bootstrapped 95% confidence intervals for each duration under the assumption $\xi = 0$ are shown in red for both stations in Figs. 3.2 (a-e) and 3.3 (a-e), respectively. The restriction of the distribution mostly leads to only a minor reduction in the uncertainties in the estimation of μ and σ. However, the uncertainties of the resulting 0.99 quantile are significantly reduced due to constraining the shape parameter. The 95% confidence interval of the 0.99 quantile at a duration of one hour ranges from 30 mm/h to 48 mm/h at Berlin-Tempelhof and from 32 mm/h to 42 mm/h at Bever-Talsperre. Therefore, fixing the shape parameter seems to be a reasonable measure to reduce the uncertainties associated with the quantile estimation.

However, the clear disadvantage of this method is that the Gumbel distribution is not necessarily a suitable model for the annual maxima of a particular station. Berlin-Tempelhof serves as a good example for a station where the constraint $\xi = 0$ seems unjustified. From Fig. 3.2 (c) it is evident that the estimated values of ξ (gray) mostly deviate considerably

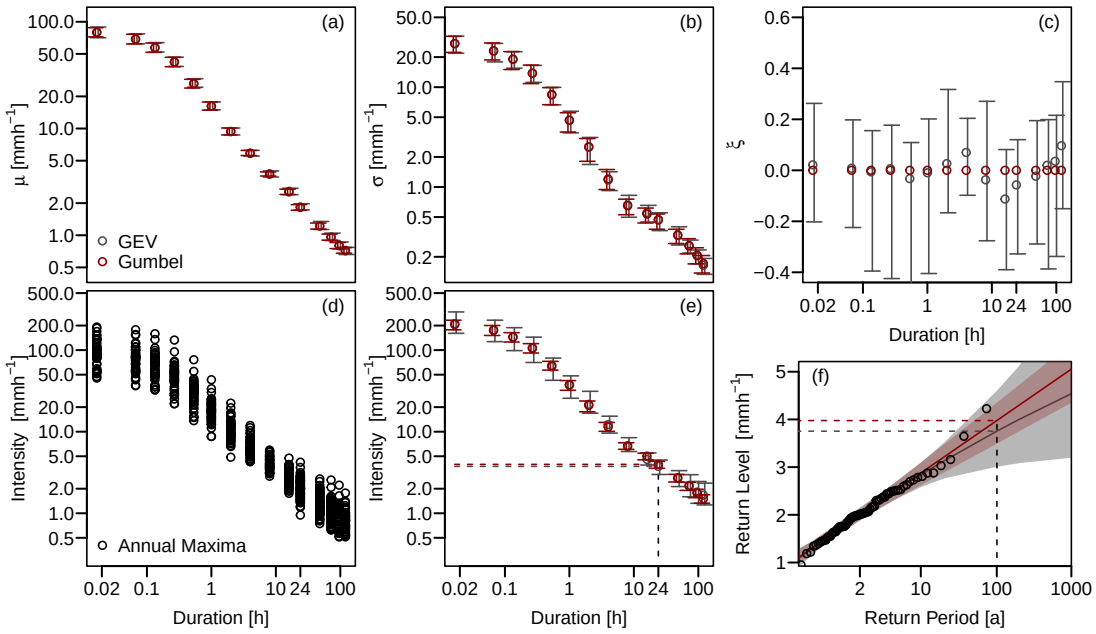


Figure 3.3: As Fig. 3.2 but for station Bever-Talsperre.

upward or downward from the fixed value of zero. As an example of how fixing the shape parameter to a possibly unsuitable value affects the quantile estimation, we can take a look at the return level plot for the duration of 24 h presented in Fig. 3.2 (f). Shown in black are the empirically estimated return periods for the observed annual maxima. Shown in gray and red are the estimated return levels as a function of return period for the GEV model and the Gumbel model, respectively. The shaded areas each represent the 95% confidence intervals. While fixing the shape parameter substantially reduces the uncertainties, the Gumbel model is no longer able to describe the upper tail of the distribution of the annual maxima. In this case, this results in a severe underestimation of the return levels for larger return periods. In contrast, the assumption $\xi = 0$ seems to be well fitting for the station Bever-Talsperre. The return level plot for $d = 24\,\mathrm{h}$ in Fig. 3.3 (f) illustrates that the constraint of the model results in a significant reduction of the uncertainties, while the Gumbel model is still able to describe the data sufficiently well in this case.

3.1.2 Modeling Duration-Dependence in a second and third step

After separately estimating the Gumbel distribution for each duration, it is necessary to model the dependence on duration. The purpose of this is partly to obtain return levels for arbitrary durations, but primarily to ensure monotonically decreasing precipitation intensities with duration and with decreasing return period. Therefore, the duration dependence is modeled on the level of the distribution parameters. To this end the DWA guideline specifies the two-step method described below.

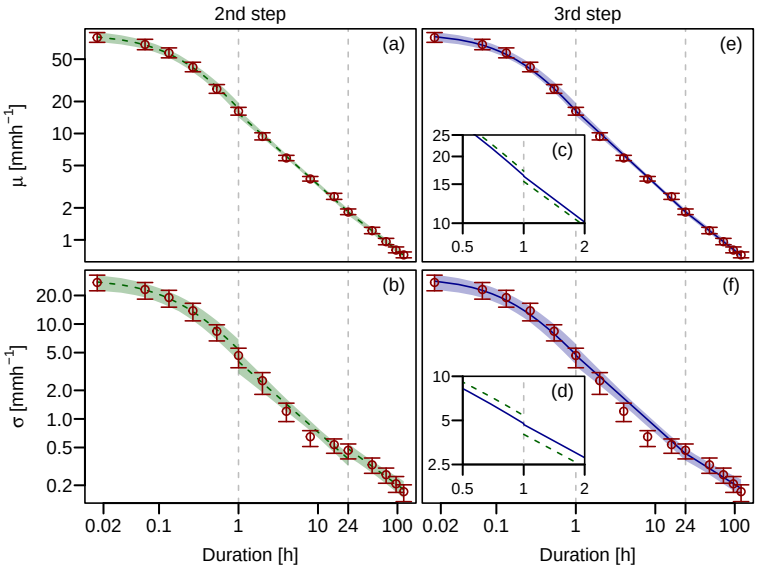


Figure 3.4: Parameter estimates for μ (top) and σ (bottom) obtained from steps one (red circles), two (left: green, dashed line) and three (right: blue line) of the DWA approach for station Bever-Talsperre. The insets (c) and (d) serve for better visibility of the discontinuity resulting from step two (green, dashed line) at $d = 1\,\mathrm{h}$. Another discontinuity occurs at $d = 24\,\mathrm{h}$. Error bars and shaded areas represent the bootstrapped 95 % confidence intervals.

Second Step: Separate Models for Duration Sub-Ranges

For modeling the duration dependence, the complete duration range is first divided into smaller sub-ranges. The DWA guideline suggests for example the following subdivision:

$$d_1 \leq 1\,\mathrm{h} \tag{3.1}$$

$$1\,\mathrm{h} < d_2 \leq 12\,\mathrm{h} \tag{3.2}$$

$$12\,\mathrm{h} < d_3. \tag{3.3}$$

In each of the three domains, a separate model is fitted for the dependence of the parameters μ and σ on duration. The following two-parameter models are considered to be suitable for both parameters $\phi \in \{\mu, \sigma\}$:

$$\phi(d) = a_1 - b_1 \cdot \ln d \tag{3.4}$$

$$\ln \phi(d) = a_2 - b_2 \cdot \ln d \tag{3.5}$$

$$\phi(d) = \frac{a_3}{d + b_3}, \tag{3.6}$$

with non-negative parameters a_i and b_i. Figure 3.4 (a-b) shows the values for the parameters μ and σ estimated in the first step (red dots) and the second step (green dashed line) for the station Bever-Talsperre. In this example, for both parameters the model from Eq. (3.6) was selected for the first duration sub-range and the model from Eq. (3.5) was chosen for

the remaining sub-ranges. The figure show that the method can lead to a good description of the dependence of the parameters on the duration. However, the uncertainties of the parameters estimated in the second step[2] remain comparable to the uncertainties of the first step, since the second step of the model builds on the results of the first step.

Third Step: Adjusted Models for Complete Duration Range

A third model step is necessary since fitting separate models for the duration sub-ranges results in discontinuities of the parameter values at the boundaries of the ranges. Figure 3.4 (c-d) illustrates the discontinuities at $d = 1\,\mathrm{h}$ resulting from the second model step (green, dashed). Accordingly, two possible values $\phi_1(d_b) \neq \phi_2(d_b)$ exist for each parameter at the boundaries of the sub-ranges $d_b \in \{1, 24\}\,\mathrm{h}$. In order to obtain a continuous model for the parameters depending on duration, first, the parameter values at the boundaries of the sub-ranges are fixed to the mean values

$$\phi(d_b) = \frac{\phi_1(d_b) + \phi_2(d_b)}{2}. \tag{3.7}$$

Subsequently, the selected models from Equations (3.4-3.6) are fitted again for each sub-range under the condition that the models must pass through the fixed points $\phi(d_b)$. Therefore, the parameters a and b are uniquely determined for the middle duration sub-range, while they must be re-estimated for the first and third range. The resulting models for the parameters μ and σ, which are continuous over the entire duration range, are shown in Fig. 3.4 (c-f) as blue solid lines. The uncertainties are slightly smaller than in the second model step because the models are constrained by the fixed points.

3.1.3 Disadvantages of Applying Separate Sequential Models

The separate use of sequential models to estimate IDF curves entails several drawbacks. A major problem of this approach is the large number of parameters that have to be estimated. Table 3.1 lists the parameters for each model step. In the given example of the Bever-Talsperre a total of 40 parameters results. Generally, such a large number of estimated parameters yields considerable uncertainties. In the presented approach, these uncertainties are substantially reduced by restricting the model through assuming the Gumbel distribution with $\xi = 0$ in the first model step. It has already been discussed that the assumption $\xi = 0$ is not appropriate for all stations.

Furthermore, the separate estimation of the distribution parameters for each duration results in inefficient use of the available data. The estimation of the distribution parameters for one duration is always based exclusively on the existing data for this duration. Especially for short time series, this leads to large uncertainties in the estimation of μ and σ in the first model step. The following model steps are then based entirely on the uncertain results of the first step. Finally, the precipitation intensity can be expected to be a smooth,

[2] In our analysis, the uncertainties presented for the second and third step include the uncertainties that already arise in previous steps.

Table 3.1: Number of parameters that require estimation in each step of the DWA approach. In the presented examples the number of durations $n_d = 15$.

Step	Number of Parameters		Description
1	$2 \cdot n_d$	$= 30$	μ, σ for each duration
2	$2 \cdot 2 \cdot 3$	$= 12$	a, b for each parameter and duration range – to obtain fixed points $\phi(d_b)$
3	$2 \cdot 2 \cdot 2$	$= 8$	a, b for each parameter and two duration ranges

i.e. continuous and differentiable function of duration for a chosen annual exceedance probability. However, this is not the case here due to the division of the duration range.

3.2 Consistent Estimation Within Single Model Step

The above-mentioned drawbacks of applying separate sequential models suggest the need to consider a consistent model for precipitation intensity as a function of return period, or rather annual non-exceedance probability p and duration d. Koutsoyiannis et al. (1998) propose using a duration-dependent extreme value distribution, where the dependence of the distribution parameters is inserted into the distribution function. For the block maxima approach, a duration-dependent GEV (d-GEV) distribution results, which enables describing the annual maxima of all durations simultaneously. Koutsoyiannis et al. (1998) propose the following assumptions about the dependence of the GEV parameters on duration:

$$\sigma(d) = \sigma_0 (d + \theta)^{-\eta}, \tag{3.8}$$

$$\mu(d) = \tilde{\mu} \cdot \sigma(d), \tag{3.9}$$

$$\xi(d) = \text{const.}, \tag{3.10}$$

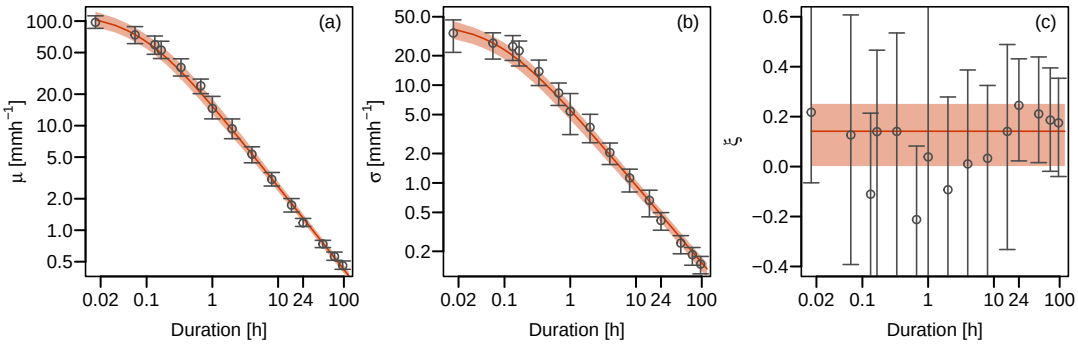


Figure 3.5: Estimates for parameters μ, σ and ξ when using a seperate GEV distribution for each duration (gray circles) and the duration-dependent GEV distribution (orange line) to model annual precipitation intensity maxima at station Berlin-Tempelhof. Error bars and shaded areas represent the bootstrapped 95 % confidence intervals.

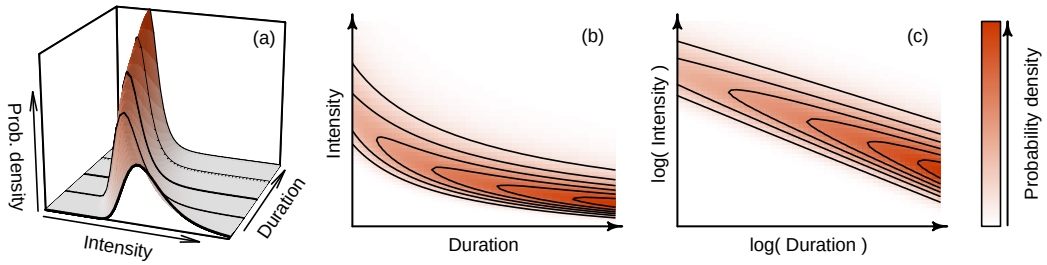


Figure 3.6: Probability density $g(z, d)$ of the duration-dependent GEV distribution (Eq. 3.12) as function of intensity z and duration d.

where the parameters $\sigma_0, \tilde{\mu}, \theta, \eta$ are non-negative and further restricting $0 < \eta < 1$. Figure 3.5 illustrates that the assumptions (3.8- 3.10) lead to a suitable description of the GEV parameters with respect to duration for the station Berlin-Tempelhof.

Inserting the assumptions into Eq. (2.1) for $\xi \neq 0$ results in the d-GEV distribution

$$G(z, d; \tilde{\mu}, \sigma_0, \xi, \theta, \eta) = \exp\left\{-\left[1 + \xi\left(\frac{z}{\sigma_0(d+\theta)^{-\eta}} - \tilde{\mu}\right)\right]^{-1/\xi}\right\}, \tag{3.11}$$

where we name the five parameters: modified location $\tilde{\mu}$, scale offset σ_0, shape ξ, duration offset θ and duration exponent η. Figure 3.6 illustrates how the resulting probability density $g(z, d)$ evolves with duration. The probability density narrows and shifts toward smaller intensity values as duration increases. The relationship between precipitation intensity I and duration d for a certain non-exceedance probability p results directly as the quantile $q_p(d)$ of the distribution:

$$I_p(d) = q_p(d) = \frac{1}{(d+\theta)^{\eta}}\left(\tilde{\mu}\sigma_0 - \frac{\sigma_0}{\xi}\left[1 - \{-\log(p)\}^{-\xi}\right]\right). \tag{3.12}$$

Rewriting Eq. 3.12 as

$$I_p(d) = \frac{\omega(p, \tilde{\mu}, \sigma_0, \xi)}{(d+\theta)^{\eta}} \tag{3.13}$$

allows to interpret how the resulting IDF curves are influenced by the parameters: while $\tilde{\mu}$, σ_0 and ξ describe the relationship between precipitation intensity and non-exceedance probability, θ and η control the dependence on duration. More specifically, $\tilde{\mu}$, σ_0 and ξ define the distance of the curves and their intensity offset, while η determines the slope for longer durations $d \geq 1\,\text{h}$ and θ the curvature for short durations $d \leq 1\,\text{h}$ when plotted in double logarithmic representation. This is visualized in Fig. 3.7.

Consequently, only five parameters need to be estimated when applying the d-GEV model. To estimate the parameter vector $\boldsymbol{\phi} = (\tilde{\mu}, \sigma_0, \xi, \theta, \eta)^T$, the maximum likelihood method given in Eq. (2.12) can easily be extended in the following way:

$$\mathcal{L}(\boldsymbol{\phi}|\boldsymbol{Z}) = \prod_{d \in D}\prod_{n \in N} g(z_{n,d}, d; \boldsymbol{\phi}), \tag{3.14}$$

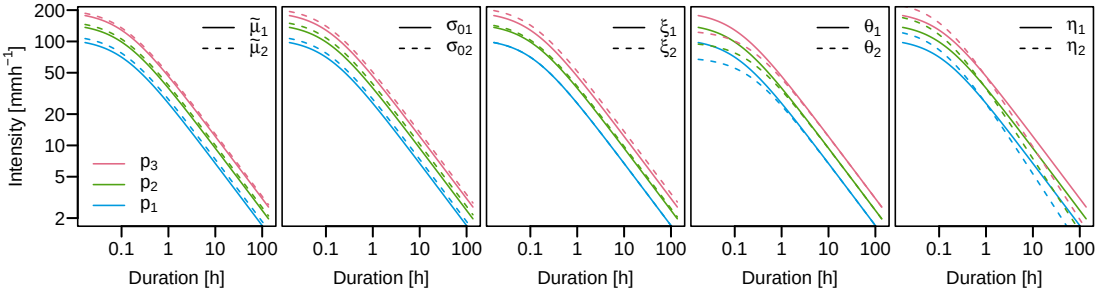


Figure 3.7: IDF curves obtained from the duration-dependent GEV for three non-exeedance probabilities p_1, p_2 and p_3 width $p_1 < p_2 < p_3$ and different values of the parameters $\phi \in \{\tilde{\mu}, \sigma_0, \xi, \theta, \eta\}$, where $\phi_1 < \phi_2$.

with the vector $\boldsymbol{Z}$ containing all observed maxima $z_{n,d}$ for different years n and durations d. Thus, the annual maxima of all durations are considered collectively to estimate the parameters through $\hat{\boldsymbol{\phi}} = \arg\max\{\mathcal{L}(\boldsymbol{\phi}|\boldsymbol{Z})\}$. This leads to a distinct reduction in the uncertainties of the estimated values $\hat{\mu}(d)$, $\hat{\sigma}(d)$ and in particular $\hat{\xi}$ compared to applying a separate GEV distribution for each duration, as can be observed in Fig. 3.5. This is especially beneficial since it allows us to include information from the daily observations, for which longer time series are available, to model the sub-daily durations with more limited data availability.

However, we need to remark that Eq. 3.14 is only valid under the assumption of independent observations. This assumption is clearly violated, since a dependence between the maxima of different duration within one year exists. The problem of considering the dependence between durations when applying the d-GEV approach will be discussed in more detail in Sec. 3.4.

3.3 Comparison Between Both Approaches

From the previous sections we can conclude that the two described methods for the estimation of IDF curves are rather different. To compare them, we present the 0.99 quantiles (i.e. 100-year return levels) estimated by both methods for the example stations Bever-Talsperre and Berlin-Tempelhof in Fig. 3.8. Furthermore, we compare the results of both models with values from the GEV model estimated separately for each duration, which are indicated by gray dots. Considering the uncertainties (shaded areas and bars), we observe that both IDF models, the DWA approach and the d-GEV approach, result in a reduction of the uncertainties with respect to the separately applied GEV model.

The application of separate successive models in the DWA approach requires the estimation of 40 parameters. On the contrary, results of the d-GEV approach are based on the estimation of only 5 parameters. Despite what could be expected from the large difference in the number of parameters between the two models, they yield somewhat comparable uncertainties. This is due to their different treatment of the shape parameter. As we have seen, the estimation of the shape parameter is associated with considerable uncertainties,

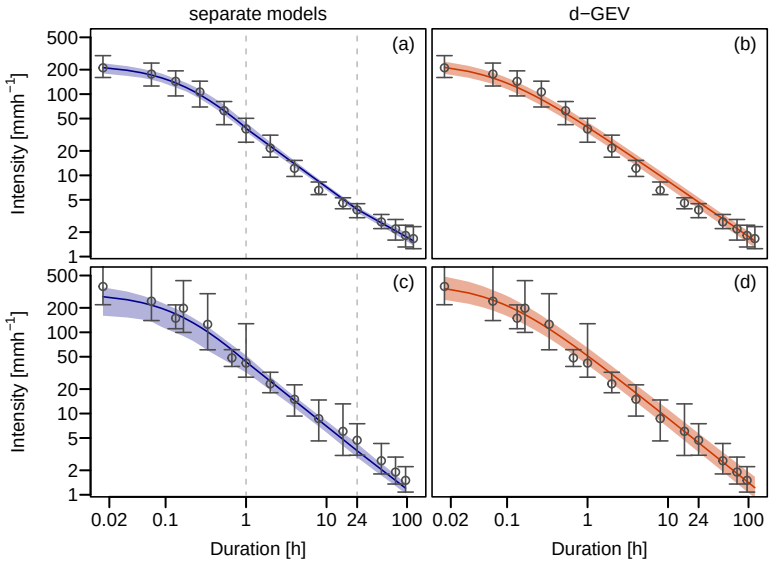


Figure 3.8: Estimates 0.99 quantile depending on duration using a separate GEV model for each duration (gray circles), the DWA approach (blue line) and the d-GEV approach (orange line) at the stations Bever-Talsperre (a-b) and Berlin-Tempelhof (c-d). Error bars and shaded areas represent the bootstrapped 95 % confidence intervals.

especially for short time series, since it determines the upper tail of the distribution and thus the very rare events. In the DWA approach, the shape parameter is fixed to $\xi = 0$. As a result, the uncertainties in the quantile estimation are drastically reduced. In contrast, in the d-GEV model the shape parameter is estimated taking into account the annual maxima of all durations. Using more data similarly leads to a major reduction of uncertainties.

However, the reduction of uncertainties in the DWA approach can be considered more or less artificial. By defining the shape parameter, other possible models for describing the data are excluded, while it is not always guaranteed that the choice of $\xi = 0$ leads to a suitable model. As we can see from Fig. 3.5 (c), considering the data of all durations results in a positive value for ξ. Therefore fixing $\xi = 0$ leads to a severe underestimation of the 0.99 quantile for most durations at this station when compared to the GEV estimates (gray circles), as shown in Fig. 3.8 (c). For this station, the d-GEV model presented in Fig. 3.8 (d) appears to be suitable, as it yields results similar to those obtained by modeling each duration separately with a GEV distribution (gray).

For the station Bever-Talsperre on the other hand, the DWA approach provides the better results, see Fig. 3.8 (a) and (b). The choice of $\xi = 0$ seems to be appropriate for this station and the DWA approach is able to sufficiently model the dependence of the 0.99 quantile on the duration. A clear deficit of the d-GEV approach can be identified in this case: the model is less flexible, which causes underestimation in the range of small durations $d \leq 15\,\text{min}$ and overestimation in the range between $1\,\text{h} < d \geq 24\,\text{h}$. This limitation of the d-GEV model results from the assumptions for the dependence of the

Table 3.2: Advantages and disadvantages of the d-GEV approach for modeling annual precipitation maxima of a range of durations when compared to the DWA approach.

Advantages	**Disadvantages**
• parameter parsimony • efficient use of data • pool information between durations • reduced uncertainties for $\hat{\xi}$ • allow $\xi \neq 0$, i.e. other than Gumbel • straightforward extension to include spatial variations	• strong restrictions for $\mu(d)$ and $\sigma(d)$ • less flexible than DWA approach • assumption of independence between maxima of different durations

parameters on the duration (Eqs. 3.8-3.10), which severely restrict the model. However, the higher flexibility of the DWA approach is achieved by dividing the duration range into three smaller sub-ranges. This contradicts the idea of a consistent model for all durations. To still be able to use the data efficiently and pool information between durations it would be more reasonable to revise the assumptions in Eqs. (3.8-3.10) and to extend the number of parameters. This idea is explored in Study II (Chpt. 5) of this thesis.

Finally, there is one additional advantage of the d-GEV approach compared to the DWA approach: In order to obtain spatial coverage for the entire country, the station-based values resulting from the DWA approach are spatially interpolated using yet another model. Therefore, these values are based on four successively applied models. In contrast, the d-GEV model can be extended relatively straightforward to model the spatial variations and the duration dependence simultaneously within one model, as presented in Study I (Chpt. 4) of this thesis. The summary in Table 3.2 clearly indicates that the benefits of the d-GEV approach outweigh the disadvantages.

3.4 Dependence Between Maxima of Different Durations

In the previous section, the d-GEV (Eq. 3.11) was introduced for simultaneous modeling of precipitation intensity maxima of a range of durations. We can estimate the parameters of this distribution using the maximum likelihood method as shown in Eq. (3.14), based on the assumption that the maxima of different durations are independent. However, this assumption is not justified since, given a measurement interval d_i, the time series for durations $n \cdot d_i$ with $n \in \mathbb{N}$ are generated by aggregating the original measurement time series. Therefore, we can expect that the block maxima M of neighboring durations, for example $d_1 = 1\,\text{min}$ and $d_2 = 2\,\text{min}$, are strongly dependent, as they were very likely generated by the same precipitation event and therefore

$$\Pr\{M(d_2) \leq z \,|\, M(d_1) \leq z\} \neq \Pr\{M(d_2) \leq z\}. \tag{3.15}$$

To account for this dependence we need to use a multivariate model, such as the Copula approach (Ariff et al., 2012) or a max-stable process (Tyralis and Langousis, 2019). We

investigated the implications of considering the dependence between maxima of different durations using a max-stable process on the model performance (Jurado et al., 2020). As our study is not comprised in this thesis, but is closely related to the discussed papers, its methods and main findings are briefly touched upon in this section. The essential question of the study is illustrated in the graphical abstract in Fig. 3.9.

3.4.1 Max-Stable Processes

Max-stable processes are commonly used to model spatial extremes. Although Jurado et al. (2020) use a max-stable process in a temporal context, to explain the basic concepts we will refer to the original spatial idea, as it offers a somewhat simpler interpretation. As mentioned before, Eq. (3.14) or Eq. (2.16), which can be used to model spatial variations of GEV parameters, are based on the assumption of independent maxima. To demonstrate how this independence can be understood, we provide the following example: Let us assume a random GEV distributed variable $Z(x, y) \sim G(\mu(x, y), \sigma, \xi)$, where the location parameter varies in space with $\mu = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 y + \beta_4 y^2$. The point-wise 100-year return levels, or 0.99 quantiles, resulting from this model are presented in Fig. 3.10 (a). Although the parameters and thus the quantiles of the distribution change smoothly in space, this is not the case when considering a single realization $z(x, y)$ of the model as shown in Fig. 3.10 (b). At each point (x, y), the value of z is independent of the value of the surrounding points, which does not lead to a smooth spatial process. When we employ such a model for precipitation, we must be aware that individual realizations of the model do not provide realistic spatial structures resembling precipitation fields.

We can consider max-stable processes as the equivalent of the GEV distribution for the multivariate case. As described in Sec. 2.3, the distribution of block maxima converges to the GEV distribution in the case of convergence. We can motivate this with the argument that only the GEV distribution is max-stable, which means that the operation of taking maxima from samples of size $n > 1$ leads to an identical distribution apart from a change in scale and location (Coles, 2001, Chpt. 3). This relationship is illustrated in Fig. 3.11 (a-b),

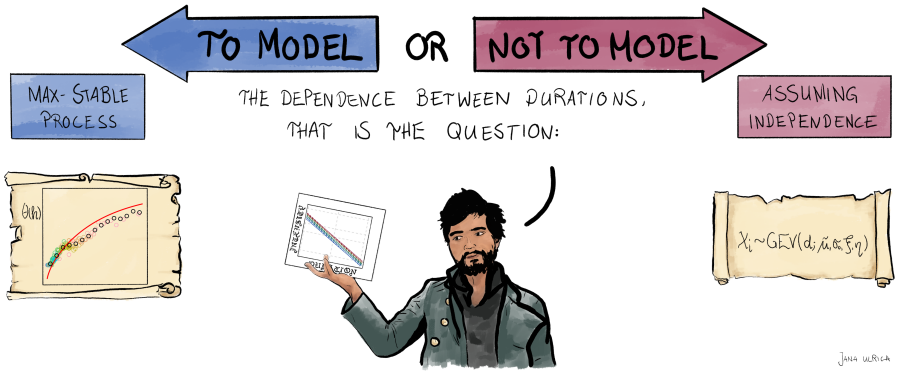


Figure 3.9: Graphical abstract of the study by Jurado et al. (2020), comparing the performance of a max-stable process with that of a d-GEV when estimating IDF curves.

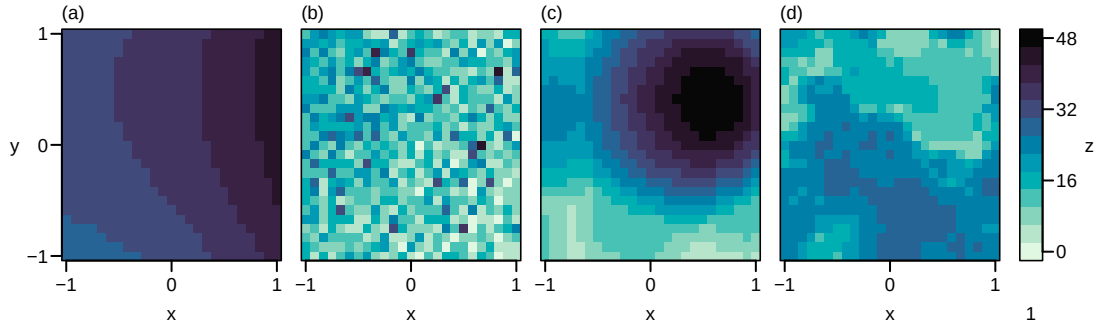


Figure 3.10: Realizations of different spatial models, all providing the same point-wise 100-year return levels, or 0.99 quantiles presented in (a). The models are: (b) GEV distribution with spatial covariates assuming independence, (c) Smith process and (d) Schlather process.

where each set of gray circles represents a sample of the GEV distribution presented in (a). The maxima of these samples (red circles) likewise follow a GEV distribution (b). We can extend this approach analogously to stochastic processes. A stochastic process is max-stable if the maximum of $n > 1$ independent copies of this process is again a stochastic process from the same family. This is illustrated in Fig. 3.11 (c-e): (c) shows a realization of a max-stable process. The same realization is presented together with four other realizations (dashed) of the same process in (d). Finally, (e) shows the maximum of all five realizations at each location x. If we compare (c) and (e), it is not difficult to assume that both are realizations of processes from the same family. By analogy with the Extreme Value Theorem, it follows that the maximum of a sequence of independent copies $Y_1(x), Y_2(x), ...$ of a stochastic process converges to a max-stable process in the case of convergence. Therefore, we can assume that max-stable processes are likely to be good candidates to model $\max\{Y_1(x), Y_2(x), ..., Y_n(x)\}$, provided n is large enough (Ribatet et al., 2016).

A relevant difference to the Extreme Value Theorem is, however, that the GEV distribution is the only univariate distribution which is max-stable, or expressed differently: the GEV family includes all three univariate max-stable distribution families Gumbel, Fréchet and Weibull. No such general formulation exists for max-stable processes. It is therefore necessary to select a family of max-stable processes for modeling, which is appropriate for the problem at hand (see e.g., Davison et al., 2012; Sebille et al., 2017). Figure 3.10 (c-d) shows a realization of two different max-stable processes, the Smith process (c) and the Schlather process (d), where the point-wise return levels of both processes are identical to those of the independent model depicted in (a). Contrary to the independent model, the values $z(x, y)$ of the realizations of the max-stable processes vary smoothly in space. However, the Smith model is too simplistic to provide a good representation of precipitation fields. Realizations of the Schlather model, on the other hand, seem to be capable to resemble precipitation fields.

To model environmental extremes (Asadi et al., 2015; Thibaud et al., 2016), and more specifically extreme precipitation (Davison et al., 2012; Sebille et al., 2017), a commonly

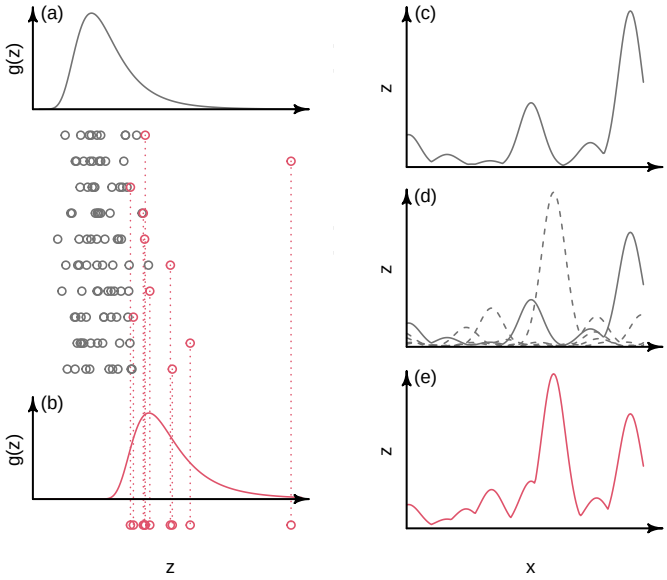


Figure 3.11: Illustrating the property of max-stability: (a-b) of a distribution and (c-e) of a stochastic process. More detailed explanation is provided in the text.

used max-stable process is the Brown-Resnick process. In Figs. 3.12 (a) and (c) one realization of two different Brown-Resnick processes in one dimension is shown. The two processes differ regarding their dependence characteristics. To describe the dependence we need to find a dependence measure. One possible measure is the extremal coefficient θ, which can take values $\theta \in [1, 2]$ in the bivariate case. The meaning of θ is best illustrated in the following definition using the conditional probability of two events, one at the location x and one in a certain distance h (Ribatet et al., 2016)

$$\begin{aligned} \Pr\{Z(x+h) \leq z \,|\, Z(x) \leq z\} &= \Pr\{Z(x+h)\}^{\theta(h)-1} \\ &= \begin{cases} 1 & \text{, complete dependence} \\ \Pr\{Z(x+h)\} & \text{, independence.} \end{cases} \end{aligned} \tag{3.16}$$

The extremal coefficient for the Brown-Resnick process is (Tyralis and Langousis, 2019)

$$\theta_{\mathrm{BR}}(h) = 2\,\Phi\left[\sqrt{\frac{1}{2}\left(\frac{h}{\rho}\right)^{\alpha}}\right], \tag{3.17}$$

where α and ρ are the smooth and range parameters, respectively and $\Phi(\cdot)$ is the standard Normal distribution (see Eq. A.3). The extremal coefficients of the Brown-Resnick processes used in the example are shown in Fig. 3.12 (b) and (d) as function of the distance h, respectively. In both cases, θ reaches approximately the same value for large distances h, but θ increases much more rapidly in case (b) than in (d). Thus, for locations x and $x + h$ closer together, there is a stronger dependence in case (d). Considering the realizations of the max-stable processes, this is reflected in stronger fluctuations of the realization shown in (a).

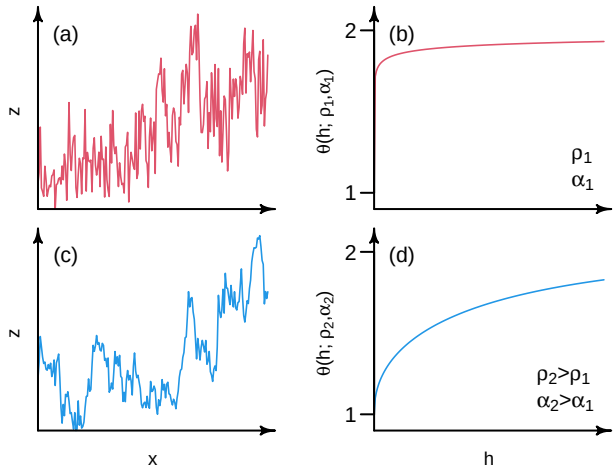


Figure 3.12: Dependence measure exemplified by two different Brown-Resnick process (red and blue): (a) and (c) present one realization of the respective processes, (b) and (d) show the extremal coefficient θ as function of the distance h of two locations x. Apart from the distance, θ depends on the smooth α and range ρ parameter of the Brown-Resnick process.

3.4.2 Applying a Max-Stable Process as IDF Model

So far, we have considered max-stable processes in the spatial context. However, the variables (x, y) can likewise describe other coordinates, such as time. Tyralis and Langousis (2019) have proposed a so-called duration space, which uses the duration d as a unidimensional coordinate analog to x. This way, we can use max-stable processes to model the maxima of different durations with explicit consideration of their dependence. In the study by Jurado et al. (2020), we investigated whether this method is suitable for the estimation of IDF curves and compared its performance with that of the independent d-GEV model.

Main Findings The results of the conducted simulation study show that, if the assumptions of the max-stable process model are fulfilled, accounting for the dependence leads to an improvement especially when estimating quantiles with high non-exceedance probabilities and in the case of strong dependence. However, when applied to observations the explicit modeling of the dependence had only a minor impact on the model performance, despite the higher complexity of the model. For the 6 stations considered in the case study, we found that modeling the dependence mostly led to minor improvement for durations $1\,\mathrm{h} < d \lesssim 48\,\mathrm{h}$ but to a minor disadvantage for longer durations. The study does not provide any information about sub-hourly durations since these were not included in the analysis.

Considering the distance h in the duration space, we found that the euclidean distance $h_e = |d_j - d_i|$ does not provide an adequate distance measure, as it led to different dependence characteristics for shorter durations compared to longer durations. As an alternative measure we suggest to use the logarithmic distance $h_l = \ln(d_j) - \ln(d_i)$ under which the assumption of isotropic dependence seems to be more reasonable. Even though the logarithmic distance appears to be the better choice for modeling the dependence

between durations, the usage of both distance measures resulted in very similar quantile estimates. This again leads us to the conclusion that the model for the dependence does not considerably affect the quantile estimates. We suspect that it has a stronger impact on the resulting uncertainties, however, comparing the uncertainty estimates was beyond the scope of the study.

In conclusion, the assumption of independent maxima appears to be an acceptable choice when the aim is merely to provide pointwise quantile estimates. Davison et al. (2012) came to a similar conclusion when analyzing different models for the spatial dependence of precipitation extremes.

Part II

Advances in IDF Modeling

Chapter 4

Estimating IDF Curves Consistently over Durations with Spatial Covariates

Authors:
Jana Ulrich
Oscar E. Jurado
Madlen Peter
Marc Scheibel
Henning W. Rust

Published in:
Water, 12, 3119, 2020;
doi:10.3390/w12113119

Author contributions:
Conceptualization: H.W.R. and J.U.
Data curation: J.U., O.E.J. and M.S.
Formal analysis: J.U.
Funding acquisition: H.W.R.
Methodology: J.U. and M.P.
Software: J.U., O.E.J. and M.P.
Visualization: J.U.
Supervision: H.W.R.
Writing—original draft: J.U.
Writing—review & editing: O.E.J., M.P., H.W.R. and M.S.

Abstract

Given that long time series for temporally highly resolved precipitation observations are rarely available, it is necessary to pool information to obtain reliable estimates of the distribution of extreme precipitation, especially for short durations. In this study, we use a duration-dependent Generalized Extreme Value distribution (d-GEV) with orthogonal polynomials of longitude and latitude as spatial covariates, allowing us to pool information between durations and stations. We determine the polynomial orders with step-wise forward regression and cross-validated likelihood as a model selection criterion. The Wupper River catchment in the West of Germany serves as a case study area. It allows us to estimate return level maps for arbitrary durations, as well as intensity-duration-frequency curves at any location—also ungauged—in the research area. The main focus of the study is evaluating the model performance in detail using the Quantile Skill Index, a measure derived from the popular Quantile Skill Score. We find that the d-GEV with spatial covariates is an improvement for the modeling of rare events. However, the model shows limitations concerning the modeling of short durations $d \leq 30\,\text{min}$. For ungauged sites, the model performs on average as good as a Generalized Extreme Value distribution with parameters estimated individually at the gauged stations with observation time series of 30–35 years available.

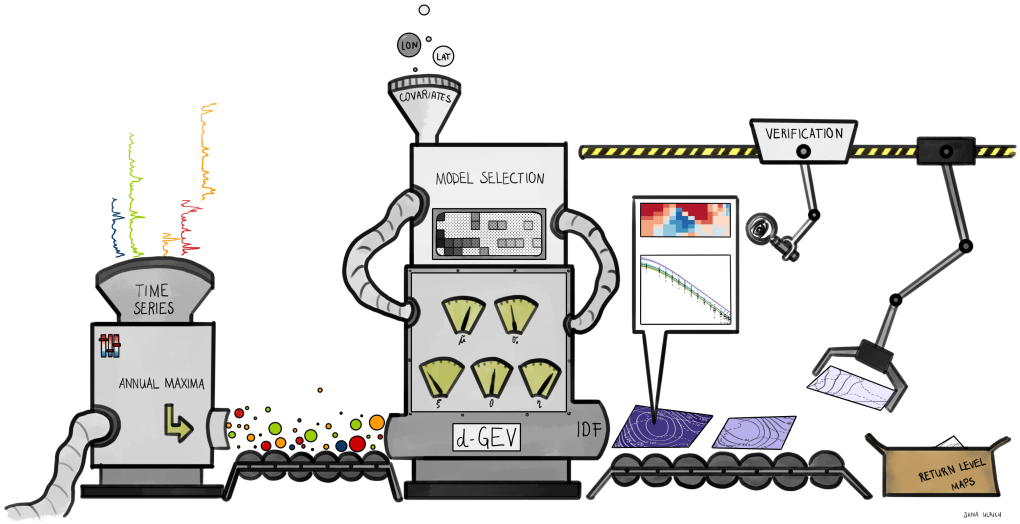


Figure 4.1: Graphical abstract illustrating the order of the methods used in this study to obtain, e.g., return level maps of various durations.

4.1 Introduction

Extreme precipitation events are often associated with hazards such as flooding and the resulting damage. In Germany, many destructive floods have occurred in recent decades, i.e., the Elbe floods in 2002 and 2013. Weather conditions favoring the occurrence of heavy rainfall events are likely to increase with global warming (Hattermann et al., 2013) in Germany as well as in many places worldwide (Seneviratne et al., 2012). Therefore, it becomes even more important to adequately estimate occurrence probabilities of precipitation amounts and intensities, as this information is needed for the design of water management systems. These range from urban drainage systems to river and creek design and retention basins. Therefore different stakeholders need information on the occurrence of extreme precipitation for different durations. Consequently, it is necessary to understand the relationship between precipitation intensity, duration, and exceedance-probability.

For a single location, this relationship can be represented graphically in intensity-duration-frequency (IDF) curves, a commonly used tool for the design of hydrological structures (Chow, 1953). However, there is no uniform procedure for estimating IDF curves, and different countries have different regulations for which method to use. In Germany, IDF curves for the entire state region are currently provided by KOSTRA-DWD (DWD, online), a project of the German Meteorological Service. The KOSTRA-DWD IDF curves are the results of a multi-step procedure and a set of different strategies for different ranges of durations (Junghänel et al., 2017). In the USA, the National Weather Service provides estimates of precipitation frequency via an online portal (NOAA, online). These estimates are based on a regional frequency analysis (Hosking and Wallis, 1997; Perica et al., 2018). The model used by the Swiss Weather Service is based on a seasonal Bayesian approach

(Fukutome et al., 2018). The results are also made available online (MeteoSwiss, online). Recent developments also suggest a wide range of methods, such as the use of radar data (Goudenhoofdt et al., 2017), cluster analysis to group stations (Olsson et al., 2019) or support-vector machines to estimate extreme events based on reanalysis data (Gaur et al., 2020).

In statistics, the definition of extreme events is based on their rare occurrence. Their statistical analysis is therefore based on small samples and it is necessary to use those efficiently in order to extrapolate from observed to unobserved levels of intensity. Extreme value theory provides several approaches to this problem (for an introduction, see Coles (2001)). In geosciences, the block-maxima approach is popular. This approach is based on modeling the probability distribution of block-maxima (e.g., monthly or annual maxima) with a Generalized Extreme Value (GEV) distribution. The longer the time series are, the more reliable the estimates are. Even if relatively long time series of 50 years or more exist at many places in Germany for daily precipitation sums, similarly long time series for observations at shorter durations are still an exception, since recording at such high frequencies is based on relatively new technology. Therefore, pooling existing information across duration can be beneficial.

In this study, we model both spatial variations of the probability distribution as well as its dependence on the accumulation duration in a consistent way. This approach allows us to include data of several gauge stations and a range of durations simultaneously in our estimation and hence makes efficient use of the available data. Instead of modeling the probability distribution individually for different precipitation durations, Koutsoyiannis et al. (1998) proposed a duration-dependent distribution based on empirical dependencies of distribution parameters on duration. This approach provides the advantages of parameter parsimony and the direct availability of estimates for all durations within the interval considered. This was already employed in previous studies (Lehmann et al., 2013; Ritschel et al., 2017; Van de Vyver and Demarée, 2010). Similar to the studies of Blanchet et al. (2016); Lehmann et al. (2013); Stephenson et al. (2016), who used a single model for a wide range of durations, we model a duration range spanning from one minute to five days. Thereby, we aim to transfer knowledge from the long durations, for which long time series exist, to the short durations.

Extending the model to include spatial variations not only provides the opportunity to estimate the IDF relationship for several locations simultaneously, but we expect that pooling information from several stations will reduce the uncertainties of parameter estimation, especially for stations with short observation time series. Many different statistical methods are used to model the spatial variation of the IDF relationship. The most straightforward way would be the spatial interpolation of the estimated distribution parameters, as done in (Blanchet et al., 2016). A commonly used approach is regional frequency analysis, which combines data from stations with similar characteristics (Hosking and Wallis, 1997). In contrast, spatial variations can be modeled in a single step, using Bayesian Hierarchical Models (BHM) (Davison et al., 2012; Dyrrdal et al., 2015; Lehmann et al., 2013), Vector Generalized Linear Models (VGLM) (Fischer et al., 2019; Van de Vyver, 2012), or Vector Generalized Additive Models (VGAM) (Yee and Stephenson, 2007), which simplifies

the estimation of uncertainties. The BHM's provide the uncertainty estimates directly, while for VGLM's and VGAM's, they can be obtained using, for example, the bootstrap method (Mélèse et al., 2018). Fischer et al. (2019) used a GEV to model daily precipitation sums and showed that the inclusion of Legendre polynomials for longitude, latitude, and altitude as covariates in location, scale and shape parameter contributed to a considerable improvement of the model compared to station-wise modeling.

Here we use the idea of Koutsoyiannis et al. (1998) in the framework of VGLMs to combine the modeling of multiple durations and spatial variations by integrating orthogonal polynomials of longitude and latitude as covariates to describe the spatial variability of the parameters of a duration-dependent GEV (d-GEV).

We expect that this will allow us to provide estimates for all durations within the range that is used for parameter estimation and, to a certain extend, also to extrapolate beyond. Furthermore, we obtain IDF relations at ungauged sites and improve the estimates for locations and durations with existing but short time series. To verify these assumptions, we test the approach in the study area of the Wupper catchment in the West of Germany and use the Quantile Skill Score (Bentzien and Friederichs, 2014) in a cross-validation setting (Wilks, 2011) to evaluate the model performance for a range of return periods and individually resolved for all durations. We focus on two research questions:

1. Under which conditions is the spatial d-GEV approach an improvement compared to the separate application of the GEV for each duration and station?

2. Does the spatial d-GEV approach provide reliable estimates at ungauged sites?

In Section 4.2, we describe the data on which the study is based and the methods used for modeling, i.e., parameter estimation, model selection, estimation of confidence intervals and verification. The verification results are presented in Section 4.3.1. Return level maps and IDF curves are provided in Section 4.3.2. The results are discussed in Section 4.4, the last section summarizes methods, results and conclusions.

4.2 Methods

We integrate spatial covariates for the parameters of a duration-dependent GEV (d-GEV) to model extreme precipitation, both in space and over a range of durations. As a case study, we use data form the area of the Wupper River catchment. The covariates for the d-GEV parameters are selected through step wise forward regression and the model results are then verified using the Quantile Skill Score. Finally, confidence intervals for the IDF curves can be obtained using the bootstrap method. This section presents the data and describes the methods used throughout this study.

4.2.1 Data

We carry out a case study in an area surrounding the catchment of the Wupper River in North Rhine-Westphalia in western Germany. For this purpose we use precipitation

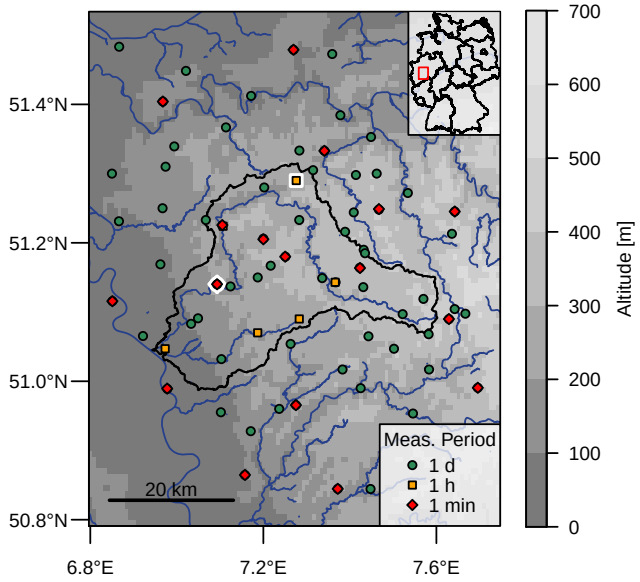


Figure 4.2: Study area containing 92 gauge stations with different measurement periods. The black line borders the Wupper catchment. Gauges marked white are those used as example stations (Schwelm (square) and Solingen-Hohenscheid (diamond)). The altitude is coded along a grey scale and stems from `http://www.diva-gis.org/gdata`, river shapes come from `https://www.openstreetmap.org`.

measurements from 92 gauge stations, shown in Figure 4.2. The Wupper River is a right tributary of the Rhine with a length of 116 km, whereby the Wupper catchment, represented by a black line, has a moderate area of 813 km^2. As the area extends from the Cologne-Bonn lowlands in the west to the Bergisches Land in the east, different altitudes are well represented by the stations and a great variability in topographic shapes is covered.

The used gauge stations are operated by two different institutions: the German Meteorological Service (DWD) (`https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/`; DWD CDC, 2019, 2021) and the Wupperverband (WV) (`https://www.wupperverband.de/`). The station properties are summarized in Table 4.1. Notably, the measuring intervals at the respective stations differ.

Observations were accumulated to multiples of their original measuring interval, resulting in time series for 15 different durations: $1, 4, 8, 16, 32$ min, $1, 2, 3, 8, 16$ h and $1, 2, 3, 4, 5$ days. For each station the annual maxima of these respective time series were considered. Whereby, years with more than a total of 20 days of missing values were discarded. For gauge stations that are within 250 m of each other and do not vary more than 10 m in height, the measurements were grouped, to avoid very high correlations of the annual maxima. For stations that are grouped together, only those available values per year resulting from the higher measurement frequency were taken into account. Hence our data set contains a total of 24,304 annual precipitation maxima for all stations and durations combined. We provide the annual maxima as a data set online as supplementary material.

Table 4.1: Properties of precipitation gauge stations.

Provider	Number of Stations	Measuring Interval	Device	Length of Time Series
DWD	69	1 day	Hellmann	9–121 years
DWD	17	1 min	Pluvio	5–14 years
WV	6	1 hour	Pluvio	38 years

4.2.2 d-GEV as a Model for Annual Maxima for Different Durations

Extreme value theory (EVT) provides methods for the statistical description of the tail of probability distributions and thus allows to estimate probabilities of very rare or even not yet observed events. The following descriptions are based on the introduction by Coles (2001). The basis of EVT is the Fischer-Tippett-Gnedenko Theorem, which essentially states that under certain assumptions the probability distribution of block maxima can be approximated by the Generalized Extreme Value (GEV) distribution. More precisely, for n independent and identically distributed copies X_i of a random variable X, we define the block-maximum as

$$M_n = \max\{X_1, ..., X_n\}. \tag{4.1}$$

If for block-size $n \to \infty$ the distribution of the properly normalized M_n converges to a non-degenerate distribution, then for a finite but large n the non-exceedance probability

$$\Pr\{M_n \leq z\} \approx G(z), \tag{4.2}$$

can be modeled with the Generalized Extreme Value distribution (GEV)

$$G(z; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{4.3}$$

defined on $\{z : 1 + \xi(z-\mu)/\sigma > 0\}$, with location parameter $\infty < \mu < \infty$, scale parameter $\sigma > 0$ and shape parameter $-\infty < \xi < \infty$, $\xi \neq 0$. Therefore the GEV can be used to model the annual precipitation intensity maxima for a certain precipitation duration d, e.g., daily precipitation sums.

Station-Wise Model for a Range of Durations (d-GEV)

In order to describe the relationship between precipitation intensity, duration and frequency, (i.e., non-exceedance probability), it is necessary to model the precipitation intensity maxima over a range of durations. The classical approach consists of two sequential model steps (Chow et al., 1988; Singh, 1992). The first step is to separately estimate an extreme value distribution, e.g., the GEV, for a certain number of durations. In the next step, certain selected quantiles of the individual distributions are fitted using an empirical function with two to three parameters, which describes the relationship between intensity and duration. A detailed summary of frequently used empirical functions can be found

in (García-Bartual and Schneider, 2001). Koutsoyiannis et al. (1998) demonstrated that all these empirical models are special cases of the more general form with 4 parameters

$$q_p = \frac{\omega}{(d^\nu + \theta)^\eta}, \tag{4.4}$$

with intensity quantile q_p corresponding to the non-exceedance probability p, duration d and with the non-negative coefficients $\omega, \nu, \theta, \eta$, where $\nu\eta \geq 1$. They showed furthermore that the assumption $\nu = 1$ is a sufficiently good approximation, resulting in a three parameter IDF model.

Their proposal is to implement Equation (4.4) with $\nu = 1$ directly into the parameters of the used extreme value distribution, to estimate the IDF curves in a considerably more consistent approach. Thus, an extreme value distribution is obtained in one step that is valid for a whole range of durations. We follow the ideas of Koutsoyiannis et al. (1998) for the dependence of the GEV (Equation (4.3)) parameters on duration:

$$\sigma(d) = \sigma_0(d + \theta)^{-\eta}, \tag{4.5}$$

$$\mu(d) = \tilde{\mu} \cdot \sigma(d), \tag{4.6}$$

$$\xi(d) = \text{const.} \tag{4.7}$$

The dependence of location and scale parameters on duration is described using duration offset $\theta \geq 0$ and duration exponent $0 < \eta \leq 1$. Furthermore, $\sigma_0 > 0$ can be interpreted as a scale offset, since it indicates the scale parameter of the GEV distribution at $d = 1 - \theta$. Re-parameterizing the location $\tilde{\mu} = \mu(d)/\sigma(d)$ and inserting relation Equation (4.5) into Equation (4.3) results in a duration-dependent Generalized Extreme Value distribution (abbreviated as d-GEV)

$$G(z, d; \tilde{\mu}, \sigma_0, \xi, \theta, \eta) = \exp\left\{-\left[1 + \xi\left(\frac{z}{\sigma_0(d + \theta)^{-\eta}} - \tilde{\mu}\right)\right]^{-1/\xi}\right\}. \tag{4.8}$$

This model describes the distribution of precipitation block maxima for a whole range of durations, with only two additional parameters than the GEV for one single duration.

The distribution's quantiles $q_{d,p}$ for a duration d, corresponding to the non-exceedance probabilities p, are equal to intensity-duration-frequency (IDF) relationships

$$q_{d,p} = \frac{\tilde{\mu}\sigma_0}{(d + \theta)^\eta} + \frac{\sigma_0}{\xi(d + \theta)^\eta}\left[1 - \{-\ln(p)\}^{-\xi}\right]. \tag{4.9}$$

Whereby the parameters θ and η, respectively, describe the curvature for short durations and the slope for longer durations of the resulting IDF curves shown in a double-logarithmic plot. Hence, we can use the d-GEV to model annual precipitation intensity maxima at a single station over a range of durations simultaneously. In the following we will call this approach station-wise d-GEV. Its advantage is the reduction of the number of parameters needed to be estimated. More precisely, $3n_d + 3n_q$ parameters would be required to model n_q IDF curves using Equation (4.4) with $\nu = 1$. This is because we would first have to estimate the three parameters of the GEV distribution (Equation (4.3)) for a number of

durations n_d and then the three parameters of Equation (4.4) for each quantile. In contrast, we only need to estimate five parameters to model the distribution for all durations as well as the dependency of any quantile on duration in one step using the d-GEV (Equation (4.8))

Additionally, we consider a special case for time series with sampling interval $d \geq 1\,\mathrm{h}$ (e.g., hourly or daily): As the curvature (departure of a straight line) of the IDF relation shown in a double-logarithmic plot is only visible for small durations $d < 1\,\mathrm{h}$, we assume $\theta = 0$ (i.e., no curvature) for durations $d \geq 1\,\mathrm{h}$. Consequently, for gauges with observations sampled at hourly or longer sampling interval, θ is not estimated but set to zero.

Adding Spatial Covariates

The station-wise d-GEV approach already enables interpolation and pooling of information across durations. We further extend this approach in the framework of vector generalized linear models (VGLM) (Yee and Stephenson, 2007), to additionally allow for interpolation and pooling of information between gauge stations. We therefore model the spatial variations of every d-GEV parameter $\phi \in \{\tilde{\mu}, \sigma_0, \xi, \theta, \eta\}$ using a generalized linear model (GLM) of the form

$$l^{\phi}(\phi) = \phi_0 + \sum_{i=1}^{I} \beta_i^{\phi}\, x_i, \tag{4.10}$$

with the parameter specific link function $l^{\phi}(\cdot)$, intercept ϕ_0 and regression coefficients β_i^{ϕ} and the covariates x_i. We implemented a function for parameter estimation based on maximizing the likelihood for the d-GEV with spatial covariates. This function is available as package `IDF` for the `R` environment (R Core Team, 2020; Ulrich and Ritschel, 2019). Typically the choice of a link function ensures parameters to be positive or be within a predefined range. Here, we implemented intervals for parameter directly into the optimizer and thus used the identity $l^{\phi}(\phi) = \phi$ as link function for all parameters.

Following Fischer et al. (2019), we intended to use orthogonal polynomials of longitude, latitude and altitude as covariates for the d-GEV parameters to model the spatial variations. However, since the area of our investigation is small and longitude and altitude are highly correlated in this area, we only use orthogonal polynomials of longitude and latitude. These polynomials are produced using the function `poly` from the package `stats` in the `R` environment (R Core Team, 2020). We also add interactions resulting from the products of the respective terms. This yields the following model for each d-GEV parameter

$$\phi = \phi_0 + \sum_{j=1}^{J} \beta_j^{\phi}\, P_j(\mathrm{lon}) + \sum_{k=1}^{K} \gamma_k^{\phi}\, P_k(\mathrm{lat}) + \sum_{j=1}^{J}\sum_{k=1}^{K} \delta_{j,k}^{\phi}\, P_j(\mathrm{lon}) P_k(\mathrm{lat}). \tag{4.11}$$

Considering the size of the study area, we expect that a maximum order of $J = K = 6$ is sufficient to model the spatial variations within. We maximize the likelihood to obtain estimates of the intercepts ϕ_0 and regression coefficients $\beta_j^{\phi}, \gamma_k^{\phi}, \delta_{j,k}^{\phi}$.

Equation (4.11) models the spatial variation (with longitude and latitude) of d-GEV parameters; in general, it takes on different values for different stations s as longitude and

latitude vary with station location. This results in a different d-GEV distribution at every station s (and in between stations at arbitrary locations (lon,lat)).

Assuming independent observations (even across durations) leads to a factorization of the likelihood for the parameters given the observations $\boldsymbol{Z}$

$$\mathcal{L}(\tilde{\boldsymbol{\mu}}, \boldsymbol{\sigma_0}, \boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\eta} \mid \boldsymbol{Z}) = \prod_{s \in S} \prod_{d \in D} \prod_{n \in N} g(z_{s,d,n}, d \mid \tilde{\mu}_s, \sigma_{0_s}, \xi_s, \theta_s, \eta_s), \tag{4.12}$$

where g is the probability density function of the d-GEV, $\boldsymbol{Z}$ is a vector containing annual maxima for different stations s, durations d and years n and $\tilde{\boldsymbol{\mu}}$, $\boldsymbol{\sigma_0}$,$\boldsymbol{\xi}$,$\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are vectors containing the d-GEV parameters at each station. Strictly speaking, Equation (4.12) is only valid under the assumption that the observations in $\boldsymbol{Z}$ are independent. We are aware that there is some asymptotical dependence between the annual maxima at nearby stations (for an overview on the topic, see (Davison et al., 2012)), as well as between the annual maxima for different durations (Jurado et al., 2020; Tyralis and Langousis, 2019). However, Jurado et al. (2020) showed that the use of a model for IDF curves that accounts for asymptotical dependence between durations had a limited improvement on the performance of predicted return levels, with the increased model complexity of a max-stable process. Davison et al. (2012) also suggests that spatial dependencies can be neglected for the estimation of point-wise return levels. Throughout our study, we will assume independence between the annual maxima of different durations and neighboring stations.

To obtain a parsimonious model, we use a selection procedure to determine which of the regression coefficients in Equation (4.11) are actually needed.

4.2.3 Model Selection

Choosing $J = K = 6$ in Equation (4.11) as the maximum order of the orthogonal polynomials for modeling the five d-GEV parameters results in 48 terms in the predictor (covariates) and thus for the 5 parameters yielding $2^{5 \cdot 48}$ possible models. Hence, the model selection is a challenging and also crucial task. We apply a step-wise forward regression, where we iteratively add one covariate to each of the parameters in a predefined order. If the addition of the covariate to the parameter model results in a better score, the model is augmented with this covariate. Since the d-GEV parameters are not independent of each other, the order in which the covariates are added strongly influences the result of the model selection. We chose the sequence $\theta \rightarrow \eta \rightarrow \sigma_0 \rightarrow \tilde{\mu} \rightarrow \xi$, according to the order in which the parameters occur in Equations (4.5)–(4.7). However, this is just one of many possible options that could be considered. A more efficient strategy for model selection based on boosting (Pasternack et al., 2021) is currently under investigation for this use case.

We use the cross-validated likelihood as a criterion for model selection. For this purpose, we carry out a k-fold cross-validation with a small number of folds $k = 2$, as suggested by Arlot and Celisse (2010). To ensure independence of the cross-validation sets and equal distribution of data from different stations and durations, the cross-validation subsets are

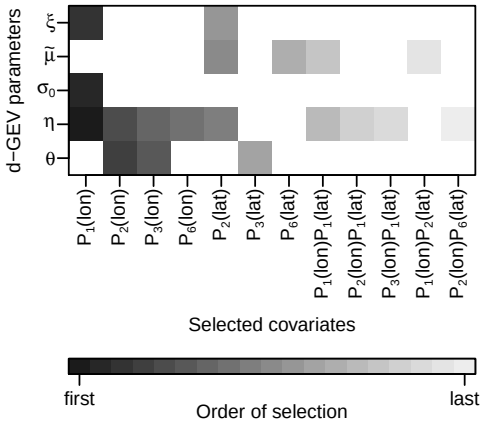


Figure 4.3: Final model selection result. For each d-GEV parameter, the added covariates are shown as colored boxes, according to the order of their selection.

drawn from every other year. The forward selection is stopped when the cross-validated likelihood stops increasing.

This approach results in a model with 24 coefficients in total. The covariates selected for each of the d-GEV parameters, are presented in Figure 4.3 along with the order of their selection. The parameters differ greatly in the number of covariates. From the large number of covariates in η we conclude that this parameter varies particularly strongly in space.

4.2.4 Verification

Since estimates of the maxima distributions' tail (upper quantiles) are particularly important in this case, we evaluate the performance of the d-GEV model through cross-validation using the Quantile Score (QS) (Bentzien and Friederichs, 2014). This score allows a detailed analysis for individual non-exceedance probabilities. The QS associated with exceedance probability p is the weighted mean difference between observations o_n and the modeled quantile q_p

$$\mathrm{QS}(p) = \frac{1}{N} \sum_{n=1}^{N} \rho_p(o_n - q_p), \tag{4.13}$$

where the check loss function ρ_p is defined as

$$\rho_p(u) = \begin{cases} pu & , u \geq 0 \\ (p-1)u & , u < 0. \end{cases} \tag{4.14}$$

Therefore, $\mathrm{QS} \geq 0$, is negatively oriented with optimal value at zero. Thus the model performance can be examined for any non-exeedance probability $p \in (0, 1)$. However, the result may be less reliable when only few data are available for the verification of high probabilities (upper quantiles).

We use two different verification strategies: (1) to assess the performance in detail, we calculate QS at each station s for a certain duration d and probability p

$$\mathrm{QS}_{s,d}(p) = \frac{1}{|Y_{s,d}|} \sum_{y \in Y_{s,d}} \rho_p(o_{s,d,y} - q_{s,d}(p)) \,, \tag{4.15}$$

where $|Y_{s,d}|$ is the cardinality of $Y_{s,d}$, the set of years the score is evaluated for; (2) for an overview of model performance in the study area, we average QS over stations

$$\overline{\mathrm{QS}}_d(p) = \frac{1}{S_d} \sum_{s=1}^{S_d} \mathrm{QS}_{s,d}(p) \,, \tag{4.16}$$

where S_d is the number of stations, for which a time series for duration d exists. With a k-fold cross-validation experiment, we assess the model's out-of-sample performance using Equation (4.15) (Hastie et al., 2009; Wilks, 2011). Hence, we partition the data into k sets $Y_{s,d}$ containing each $n_{\mathrm{val}} = |Y_{s,d}| = 3$ years of data. This results in a varying number of sets depending on the length of the time series. Successively, each set is used once for validation, the remaining for training. Therefore, the number of years used for training also depends on the length of the time series. For the station-wise model the training set consists of the remaining years of data at the station under investigation; for the spatial model the training set additionally contains all observations from all other stations.

Finally, we compare the cross-validated score of the model QS^M with the cross-validated score of a reference model QS^R using the Quantile Skill Score (QSS) (Wilks, 2011)

$$\mathrm{QSS}^M(p) = \frac{\mathrm{QS}^M(p) - \mathrm{QS}^R(p)}{0 - \mathrm{QS}^R(p)} = 1 - \frac{\mathrm{QS}^M(p)}{\mathrm{QS}^R(p)}. \tag{4.17}$$

with $\mathrm{QS}_{s,d}(p)$ (cf. Equation (4.15)) to evaluate QSS at each station s and $\overline{\mathrm{QS}}_d(p)$ (cf. Equation (4.16)) to assess performance across the study area. $\mathrm{QSS}^M(p) \in (-\infty, 1]$ is positively oriented and optimal at 1, representing the gain in performance ($\mathrm{QS}^M(p) - \mathrm{QS}^R(p)$) relative to the difference between a perfect and the reference model ($0 - \mathrm{QS}^R(p)$).

If, however, the model performs worse than the reference, QSS is negative and the interpretation less intuitive. We thus define a Quantile Skill Index (QSI) as a combination of the model's skill with respect to a reference $\mathrm{QSS}^M(p)$ and the skill of the reference with respect to the model $\mathrm{QSS}^R(p)$

$$\mathrm{QSI}(p) = \begin{cases} \mathrm{QSS}^M(p) & , \mathrm{QS}^M(p) \leq \mathrm{QS}^R(p) \\ -\mathrm{QSS}^R(p) & , \mathrm{QS}^M(p) > \mathrm{QS}^R(p) \,. \end{cases} \tag{4.18}$$

Positive values still indicate a gain with respect to the reference while negative values now indicate a gain of the reference with respect to the model. Now, we have $\mathrm{QSI} \in [-1, 1]$. In the following, we distinguish between the QSI calculated using the QS for each station $\mathrm{QS}_{s,d}(p)$ and the QS averaged over stations $\overline{\mathrm{QS}}_d(p)$ by referring to the later one as average QSI and using the notations $\mathrm{QSI}_{s,d}(p)$ and $\overline{\mathrm{QSI}}_d(p)$, respectively. As reference we always use the GEV for annual maxima at each station and duration separately.

4.2.5 Confidence Intervals

We obtain 95% confidence intervals for the estimated IDF relation at a given station s, by applying the ordinary non-parametric bootstrap percentile method (Davison and Hinkley, 1997). In a simulation study (see Appendix B.2), we compare the coverage of the 95% confidence intervals for the estimated quantiles derived through both the bootstrap percentile and the delta method (Coles, 2001) under the assumption of no dependence between block maxima. We find that the coverage for the delta method intervals depends on the duration and probability of the quantile, for which the confidence intervals are estimated. In most cases, the coverage deviates strongly (upwards or downwards) from the nominal 95%. The bootstrapped confidence intervals, on the other hand, show a consistent behavior for different durations and probabilities with a reasonable coverage for large enough sample sizes (see Figure B.1).

4.3 Results

We will first present the results of the verification. We use different variations of the methods presented in Section 4.2.4 to assess different aspects of the model performance. We then address the estimation of quantiles for spatial maps of return levels, IDF curves at selected stations; both with their associated uncertainty.

4.3.1 Model Performance

Addressing the two questions posed in Section 4.1, we investigate model performance using variations of the cross-validation experiment described in Section 4.2.4. An overview is given in Table B.1.

Overall Performance

We calculate $\overline{\mathrm{QSI}}_d(p)$ (cf. Equation (4.18)) using the mean quantile score over all stations $\overline{\mathrm{QS}}_d(p)$ (Equation (4.16)) to asses the overall performance in the whole study area. We furthermore obtain $\overline{\mathrm{QSI}}_d(p)$ for both, the station-wise d-GEV approach and the spatial d-GEV approach, to be able to compare them.
For a detailed assessment, we use $\mathrm{QSI}_d(p)$ for all durations used and for the range of non-exceedance probabilities $p \in \{0.5, 0.8, 0.9, 0.95, 0.98, 0.99, 0.995\}$ associated with return periods $T = \{2, 5, 10, 20, 50, 100, 200\}$ years. Results are presented in Figure 4.4 for both, the station-wise d-GEV (upper panel) and the spatial d-GEV (lower panel). Both approaches show similar results. Overall, for probabilities $p \in \{0.5, 0.8, 0.9\}$ we see no or only a small improvement, while higher probabilities show improvements in model performance. This suggests that, since the station-wise and the spatial d-GEV approach make more use of the available data, they are both better for modeling rare events than the approach of modeling each duration at every station separately. Deviations from this pattern occur: First, strongly negative QSI values are observed for short durations $d < 1\,\mathrm{h}$ in a sequence of probabilities. Since this effect is apparent for both approaches, it is thought

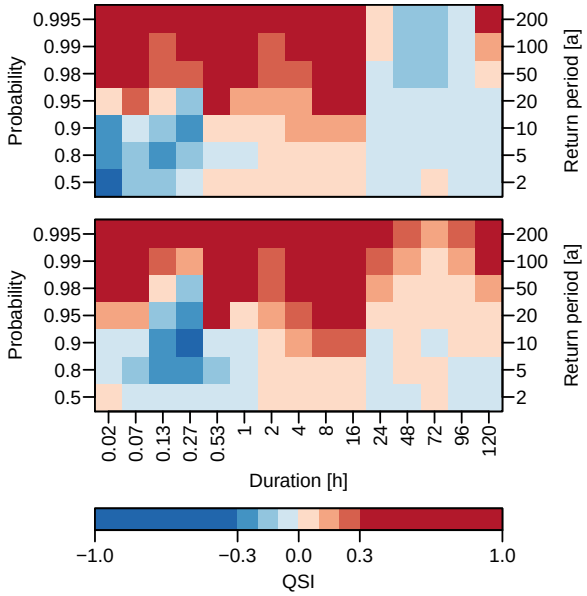


Figure 4.4: Average Quantile Skill Index $\overline{\mathrm{QSI}}_d(p)$ for different durations d and probabilities p. Upper panel: station-wise d-GEV. Lower panel: spatial d-GEV. Positive values (red) indicate an improvement compared to the quantile estimates obtained by modeling each station and duration separately.

to be due to insufficient modeling of the curvature of the IDF curves in this range using the d-GEV. Second, for durations $d \geq 24\,\mathrm{h}$ the station-wise d-GEV approach leads to a loss of skill at almost all probabilities, and the spatial approach also shows a lower skill compared to the shorter durations. As mentioned in Section 4.2.1, the length of the time series varies strongly for different measurement intervals: where most stations with daily measurements have longer time series than stations with a shorter measurement intervals. We thus assume that the different behavior for longer durations $d \geq 24\,\mathrm{h}$ stems from the longer time series of daily observations used to train the model. In the subsequent section we examine this assumption in more detail.

Dependence on Time Series Length

A slightly modified verification method investigates the effect of time series length used for training. We divide each times series such that the series available for training at a particular station all have a fixed length n_{train}. These are then used to train the model at this station. To this end, each station's data is split into blocks of years containing $n_{\mathrm{train}} + n_{\mathrm{val}}$ years each. Depending on the length of the time series for each station and duration, this will result in a varying number of blocks $b_{s,d}$. Again, k-fold cross-validation is used in each block to assess performance as previously described. The resulting $b_{s,d}$ cross validated scores at each station are then averaged to obtain $\mathrm{QS}_{s,d}(p)$. Thereafter, the analysis again follows as described in Section 4.2.4.

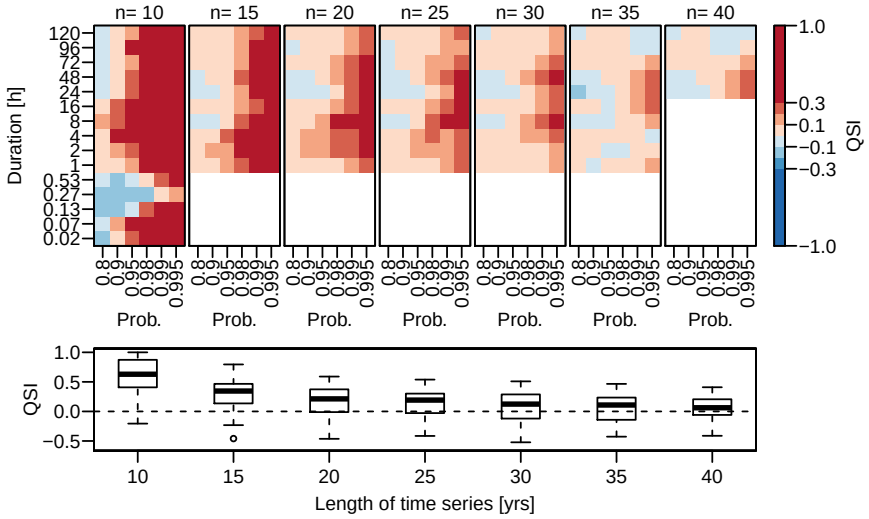


Figure 4.5: Dependence of the spatial d-GEV model performance on length of the training time series. Upper panel: Average Quantile Skill Index $\overline{\text{QSI}}_d(p)$ for different durations d and probabilities p as seen in Figure 4.4 but rotated. Different columns represent $\overline{\text{QSI}}_d(p)$ for different numbers of years in the training set n. Lower panel: Boxplot of the Quantile Skill Index $\text{QSI}_{s,d}(p)$ with probability $p = 0.99$ dependent on the length of the training time series.

The resulting $\overline{\text{QSI}}_d(p)$, obtained for different numbers of years n_t available to train the model at a certain station, is presented in the upper panel of Figure 4.5. Advantage and disadvantage of the spatial d-GEV is represented through colors red and blue, respectively; white indicates a situation with insufficient data to evaluate $\overline{\text{QSI}}_d(p)$. The two main features identified in Figure 4.4 recur: (i) an increase of $\overline{\text{QSI}}_d(p)$ with larger p for all n_t and (ii) a disadvantage of the spatial d-GEV for durations $d < 1\,\text{h}$ for a range of probabilities. Additionally, we observe a gradual decrease of the $\overline{\text{QSI}}_d(p)$ with n_t. Leading to the advantage of the spatial model in modeling rare events becoming smaller in each step until the average QSI fluctuates around zero. This indicates that there is approximately the same number of stations with gain as with loss in skill. This approach therefore provides us with the length of the time series of a station, up to which it is beneficial to use the spatial d-GEV model, instead of a separate GEV model for each time series. As an example, the $\text{QSI}_{s,d}(p)$ for $p = 0.99$ is shown in Figure 4.5 in the lower panel as a boxplot for various n_t. From this we can see that for $n_t = 35$ there are about as many stations with the spatial d-GEV being superior as stations where it is inferior to the reference model.

An alternative representation of the $\text{QSI}_{s,d}(p)$ is shown in Figure 4.6 for $p = 0.99$ and selected values for d and n_t. Since the values for the QSI vary strongly, it is more difficult to detect the relationships between $\text{QSI}_{s,d}(p)\,, d\,, p$ and n_t in this form of presentation. However, this way it is easier to observe that even with a large positive average QSI, negative QSI values can occur at individual stations. We further notice that for individual stations the sign of the QSI can be opposite for different durations. The average QSI decreases with increasing n_t, as more stations with negative QSI appear and the values at

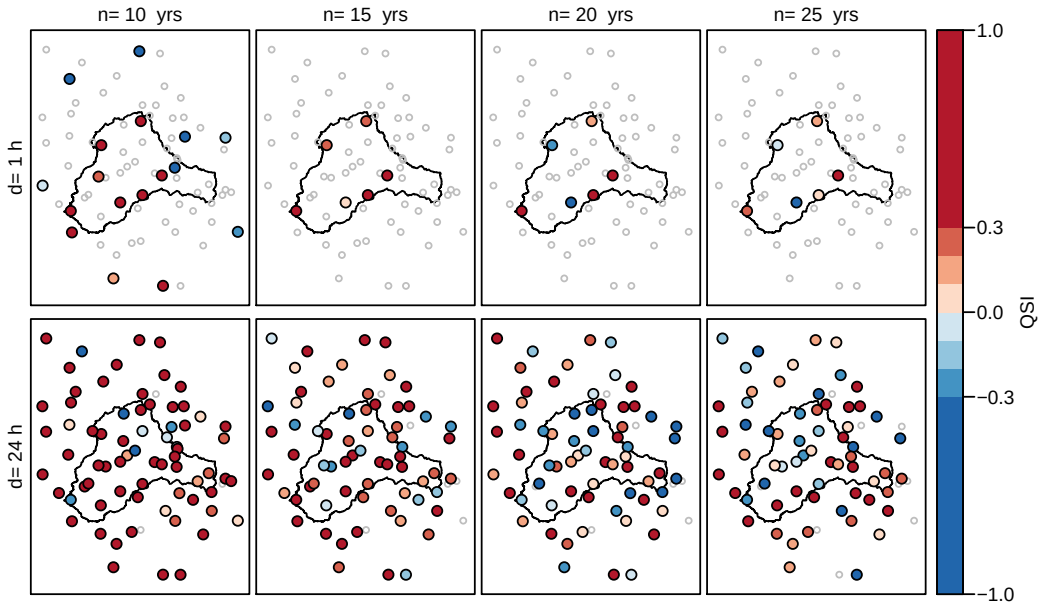


Figure 4.6: Quantile Skill Index $\mathrm{QSI}_{s,d}(p)$ for the spatial d-GEV approach with probability $p = 0.99$ for durations $d = 1\,\mathrm{h}$ (upper panel) and $d = 24\,\mathrm{h}$ (lower panel) at all stations, for the number of training years $n = \{10, 15, 20, 25\}$ (different columns). Colored dots indicate superiority of the spatial d-GEV (red) or inferiority (blue), while gray circles show stations without an estimate.

the stations with positive QSI become smaller.

Ungauged Sites

Model performance at ungauged sites can be assessed by disregarding all values at a certain station for training. Afterwards, the estimates at the location of the station excluded are compared to the actual observations. This procedure is carried out for each of the stations in the research area. However, we can only use this approach for the spatial model and not for the station-based reference model. Even though one might consider to compare the spatial d-GEV model to another spatial modeling approach, here we continue to use the GEV applied separately for each duration and station as reference model to keep the performance of the reference model as consistent as possible. We use the method described in the previous section to train the reference model with a fixed number of years at the station under investigation. Although the spatial d-GEV model is always trained on all available data except for the excluded station, we compare the model and reference with the same validation data set in each cross-validation step. Figure 4.7 shows the results for the average QSI with every column representing a different reference model associated to the number of years used for training, but the spatial d-GEV model in each column is trained on identical data.

This provides us with an approximate estimate of what is comparable to the performance of the spatial model at ungauged sites. Therefore we conclude that in our study area

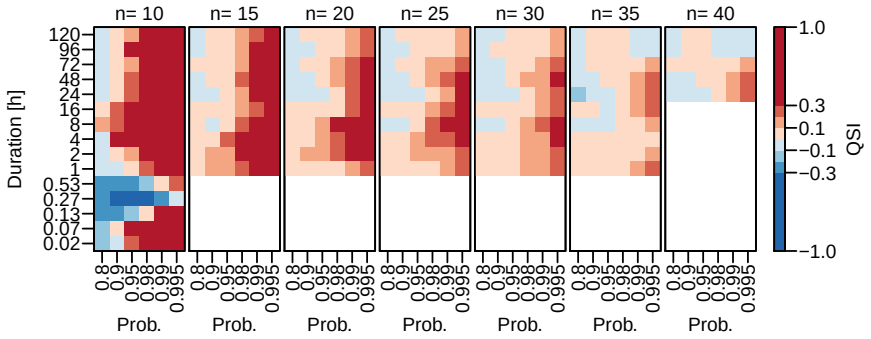


Figure 4.7: Average Quantile Skill Index $\overline{\mathrm{QSI}}_d(p)$ at ungauged sites for different durations d and probabilities p, similar to Figure 4.5. For the spatial d-GEV the data at the respective stations are omitted for fitting, while the reference model uses a different number of years n in the training set (different columns.)

the estimates of higher quantiles $p \geq 0.98$ at ungauged sites is on average as good as the estimates based on the GEV for a station with measurements for 30–35 years. The comparison with Figure 4.5 reveals a strong similarity of the results. From this we conclude that adding n_t years of data at one station only has a minor effect on the estimates of the spatial model on average. The differences between the columns in Figures 4.5 and 4.7 thus originate mainly from differences in the reference model.

4.3.2 Quantile Estimation and Uncertainty

The spatial d-GEV allows for estimating the quantiles of the extreme value distribution at arbitrary locations in our research area for arbitrary durations. Hence we are able to provide return level maps for any desired duration. The 100-year and 20-year return level maps for the durations 5, 30 min and 1 h are presented in Figure 4.8. With increasing duration, we observe a shift of the spatial patterns from a minimum in the center of the catchment to a west-east gradient ($d \geq 1\,\mathrm{h}$). In the Wupper catchment area there are three prominent weather conditions that occur together with precipitation: convective conditions from south-east and south-west and advective conditions from north-west. The gradient in the intensity of extreme precipitation of 1 h or longer is plausible regarding the main direction of advective weather conditions and the increase in elevation towards the east in the study area. In contrast, the intensity patterns of the shorter events appear to be unrelated to the weather conditions mentioned above. The spatial variations in the precipitation intensity are larger for the short durations and the more rare events. Even though a number of stations placed in the surroundings of the catchment area were included into the modeling, high gradients can still be detected at the boundaries, resulting from extrapolation.

In addition to the return level maps, classical IDF curves can be obtained at any location in the study area. Exemplary, Figure 4.9 shows the IDF curves for the station Solingen-Hohenscheid. The curves represent $p \in \{0.5, 0.75, 0.99\}$, corresponding to return periods $T \in \{2, 4, 100\}$ years, together with their 95% confidence intervals. The observations—

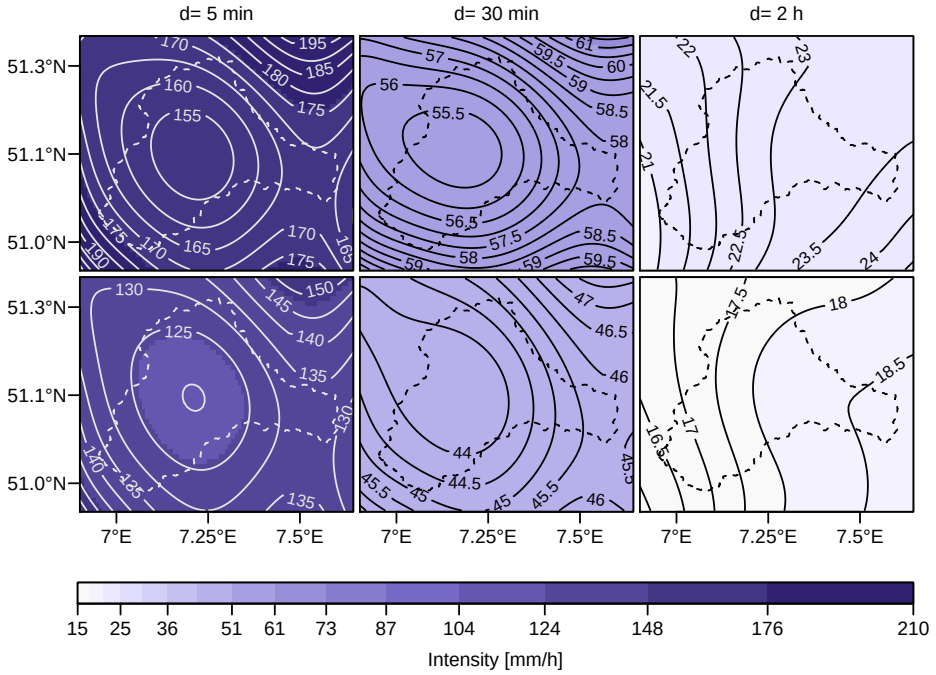


Figure 4.8: Point-wise return level maps for durations $d \in \{5, 30, 60\}$min (different columns) and probabilities $p = 0.99$ (upper panel) and $p = 0.95$ (lower panel) corresponding to return periods of 100 years and 20 years, respectively. The colors are provided as a general reference between the plots.

shown as boxplots—are well represented by the modeled IDF curves. To directly relate the estimated quantiles to the verification, the values of the QSI for this station are also shown in the upper panel. The QSI at this station exhibits similar structures as the average QSI. In the case of this station, the negative QSI values in the range 2 min to 4 h coincide with an underestimation of the respective quantiles by the spatial d-GEV. This suggests that the model lacks flexibility for this range of durations. However, the Quantile Score for these durations might not be very meaningful due to the limited amount of data.

The bootstrap confidence intervals are rather narrow across all durations. They represent the sampling uncertainty, assuming an adequate model has been selected for the data. The width of the intervals indicates that the uncertainty of the estimates is larger for shorter durations. Figure 4.10 compares the confidence intervals at small durations for the station-wise d-GEV and the spatial d-GEV at two example stations. At station Solingen-Hohenscheid data is measured at one minute intervals and at station Schwelm at hourly intervals. As a result of the increased data availability the confidence intervals based only on the uncertainties of the parameter estimation are smaller for the spatial model than for the station-wise model. The spatial d-GEV is additionally able to estimate the quantiles even for durations smaller than the measurement interval, based on the data of neighboring stations with relatively low uncertainties. This information is not available

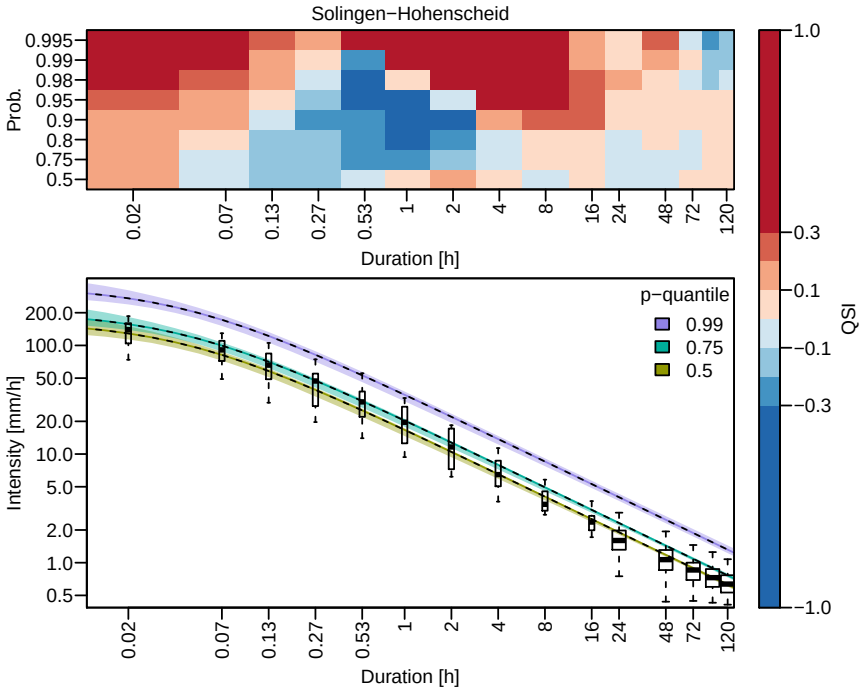


Figure 4.9: IDF-curve estimate for Solingen-Hohenscheid (marked with white square in Figure 4.2) (lower panel) obtained by using the spatial d-GEV (black dashed lines) and their 95% confidence intervals (colored areas). Observations are shown as boxplots, where the width of the box is proportional to the number of data points available at a certain duration. The upper panel shows the corresponding QSI values at that station following the presentation of Figure 4.4.

using the station-wise model.

4.4 Discussion

Since the QSI varies strongly between individual stations and individual durations, the assessment of the model performance and its presentation is challenging. However, the averaged QSI presented in Figure 4.4 allows for some conclusion. We find that the average QSI for the spatial d-GEV is strongly positive for upper quantiles (large non-exceedance probabilities $p \geq 0.98$). From this, we conclude that the spatial d-GEV approach is an improvement for modeling rare events since it benefits from the increased data availability at neighboring sites. However, for a range of smaller durations $d \leq 32\,\text{min}$, the skill decreases for both the station-wise d-GEV and the spatial d-GEV model, compared to the reference which is based on individual GEV for stations and durations. This suggests that the d-GEV does not describe the variations in this range of small durations sufficiently well. Figure 4.9 presents the QSI together with the IDF estimates for gauge Solingen-Hohenscheid; this figure suggests that the negative QSI values in the range of 2 min to 4 h are related to an underestimation of the quantiles in this range. This supports the

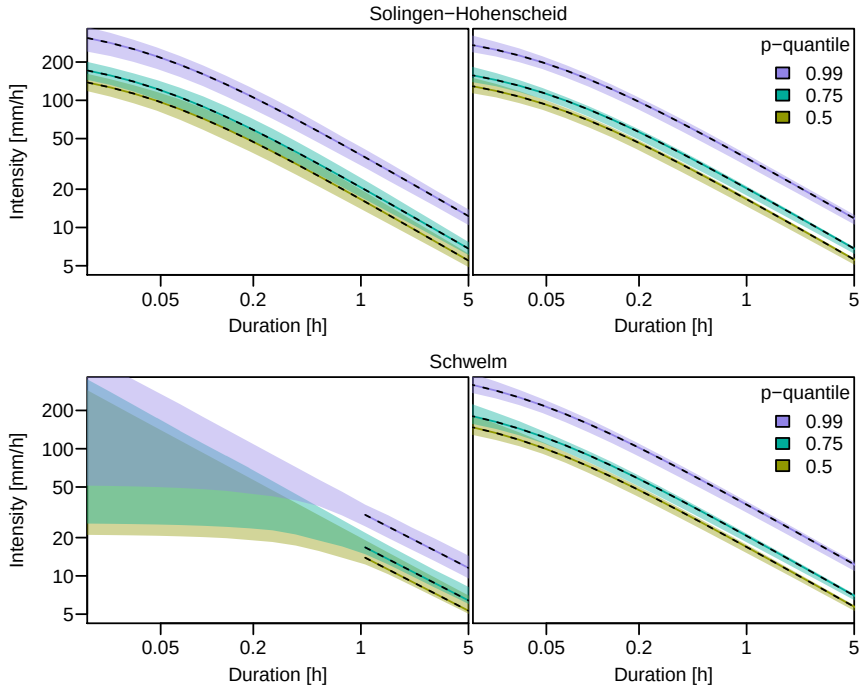


Figure 4.10: Bootstrapped 95% confidence intervals for IDF-estimates at the example stations Solingen-Hohenscheid and Schwelm (marked in Figure 4.2), using the station-wise d-GEV (**left column**) and the spatial d-GEV approach (**right column**).

assumption that the d-GEV is not sufficiently flexible in this range and a more complex model might be necessary, as suggested in (e.g., Van de Vyver (2018)). Nevertheless, we only used time series with a maximum of 14 years to investigate the model performance for sub-daily durations and thus cannot exclude the possibility that the effect may occur due to insufficient data.

Moreover, the average QSI is lower for durations $d \geq 1\,\mathrm{h}$. We believe this is due to a larger data availability for these durations and thus longer time series are available to train the reference model. This is supported by investigating the influence of time series length, where a fixed number of years n_t is used to train the model for each duration and at each station (cf. Figure 4.5). We observe a gradual decrease in the average QSI with the length of the training time series. We conclude that the advantage of the spatial d-GEV model over the station-wise GEV model is reduced for longer time series; The pooling of information becomes less important. We find that for a length of about 35 years, there are about as many stations with the spatial d-GEV being superior as stations where it is inferior to the reference model. This implies that in case of a single gauge with a long time series, the spatial d-GEV approach cannot outperform single site estimates for individual durations. However, due to the lack of data, we are unable to make any statements about the behavior of the estimates for $d < 1\,\mathrm{h}$ with n_t. This information would be particularly helpful for these durations, as here often only short time series are available. Moreover, even

with a strongly positive average QSI, negative values of the QSI for individual stations, and durations, occur. Yet, Fischer et al. (2019) showed that even for long time series with more than 50 years of observations, a GEV model with spatial and seasonal covariates performed better than separate models for each month and station at almost all stations investigated. Therefore, the major improvement of the model by adding spatial and seasonal covariates is not directly applicable to the spatial d-GEV model.

However, a large advantage of the d-GEV model is its ability to interpolate between durations and stations at the level of distribution parameters and it is therefore possible to obtain estimates for durations and sites for which no measurements exists in a consistent way. This advantage has been disregarded in the verification process, where we used a separate model for individual durations and stations as a reference model. From the results presented in Figure 4.7, we infer that the model performance at ungauged sites is comparable to that of a separately applied GEV for an available time series of 30–35 years, at least for high quantiles $p \geq 0.98$. Therefore we conclude that the spatial d-GEV model provides reliable estimates at ungauged sites. However, the available time series for $d \leq 1\,\mathrm{h}$ are again not long enough to investigate the model performance at ungauged sites for this range of small durations in this way.

4.5 Summary

In this study, we model annual precipitation maxima simultaneously in space and across durations. To this end, we integrate orthogonal polynomials of longitude and latitude as spatial covariates into the duration-dependent GEV proposed by Koutsoyiannis et al. (1998). This allows for a parameter parsimonious description compared to modeling of individual stations and durations, efficient use of existing data, and the pooling of information between stations and durations. We specifically model a wide range of durations from one minute to 5 days in order to investigate to what extent knowledge can be transferred from long observation time series and whether estimates for stations or durations with fewer observations benefit from this. We investigate this model in the Wupper catchment with the main focus on evaluating the model performance. Model validation is based on techniques from forecast verification: we use a variant of the Quantile Skill Score, the Quantile Skill Index (QSI). In the presentation used here, this score allows a detailed analysis of the model performance for different non-exceedance probabilities and durations, simultaneously. As a reference model that is not based on any empirical relationship between intensity and duration, the GEV is used to model precipitation maxima independently at individual stations and durations.

We find that using the spatial d-GEV improves the modeling of rare events of all durations, as it benefits from greater data availability. Accordingly, this model is advantageous for stations with short time series and does not necessarily improve the estimation if a longer time series is available. We also find that the d-GEV model is most likely not flexible enough to model the whole range of durations sufficiently and that a model with additional parameters (e.g., (Van de Vyver, 2018)) might be necessary. Therefore we recommend reducing the duration range in cases where the aim is exclusively the description of short

durations $d \leq 1\,\mathrm{h}$. Future studies will also explore the use of more flexible models to describe the whole range of durations. We expect that the estimation of further additional parameters for the duration dependence in these more flexible models will benefit from a spatial covariates setting.

Since this approach allows us to interpolate between stations and durations, spatial maps of return levels can be readily obtained for any duration, as well as IDF curves for any location in the research area. The bootstrap method provides 95% confidence intervals representing the sampling uncertainty. For the d-GEV with spatial covaiates, these uncertainties are smaller than for the station-wise d-GEV, since the spatial model can draw information from both neighboring sites and durations. Uncertainties from the model selection are not considered. For a reliable estimation of the uncertainties, Mélèse et al. (2018) suggests a Bayesian Hierarchical Model. In the return level maps we observe that the spatial patterns change from a minimum in the center of the catchment for short durations to a west-east gradient for long durations. This is likely related to the main north-west direction of advective weather conditions in the study area and might be also linked to the orography.

In this work, we assume that there is no dependence between observations of different durations. This seems to be reasonably well justified as reported in another study which investigates the effect of including this dependence explicitly using a max-stable process for six stations in the same research area ((Jurado et al., 2020) in this issue). We also assume that there is no dependence between observations at neighboring stations; this dependence could also be modeled using a max-stable process Davison et al. (2012); Stephenson et al. (2016). We use the assumption that the IDF relationship does not vary in time. However, we plan to account for the temporal variations of the IDF relationship by a straightforward extension of the spatial d-GEV model with further covariates in future studies. Nevertheless we could demonstrate that the approach presented here allows obtaining reasonable estimates of return levels for any arbitrary duration or location within the study domain, performing particularly well for rare events.

Supplementary Materials: The following are available online at `http://www.mdpi.com/2073-4441/12/11/3119/s1` .

Funding: This study was developed within the framework of the research training program *NatRiskChange* funded by the Deutsche Forschungsgemeinschaft (DFG; GRK2043/1 and GRK2043/2) at Potsdam University and partially supported through grant CRC 1114 "Scaling Cascades in Complex Systems", Project A01 "Coupling a multiscale stochastic precipitation model to large scale flow". O.E.J. acknowledges support by the Mexican National Council for Science and Technology (CONACyT) and the German Academic Exchange Service (DAAD).

Acknowledgments: The authors would like to thank the Wupperverband as well as the Climate Data Center of the DWD, for providing and maintaining the precipitation time series. We would also like to acknowledge the constructive comments of the three

anonymous reviewers and the academic editor that have contributed to improving the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Chapter 5

Flexible and Consistent Quantile Estimation for Intensity-Duration-Frequency Curves

Abstract

Assessing the relationship between the intensity, duration, and frequency (IDF) of extreme precipitation is required for the design of water management systems. However, when modeling sub-daily precipitation extremes, there are commonly only short observation time series available. This problem can be overcome by applying the duration-dependent formulation of the Generalized Extreme Value (GEV) distribution which fits an IDF model with a range of durations simultaneously. The originally proposed duration-dependent GEV model exhibits a power-law-like behaviour of the quantiles and takes care of a deviation from this scaling relation (curvature) for sub-hourly durations (Koutsoyiannis et al., 1998). We suggest that a more flexible model might be required to model a wide range of durations (1 min to 5 days). Therefore, we extend the model with the following two features: i) different slopes for different quantiles (multiscaling) and ii), the deviation from the power law for large durations (flattening), which is newly introduced in this study. Based on the quantile skill score, we investigate the performance of the resulting flexible model with respect to the benefit of the individual features (curvature, multiscaling and flattening) with simulated and empirical data. We provide detailed information on the duration and probability ranges for which specific features or a systematic combination of features leads to improvements for stations in a case study area in the Wupper catchment (Germany). Our results show that allowing curvature or multiscaling improves the model only for very short or long durations, respectively, but leads to disadvantages in modeling the other duration ranges. In contrast, allowing flattening on average leads to an improvement for medium durations between 1 h and 1 d, without affecting other duration regimes. Overall, the new parametric form offers a flexible and enhanced performance model for consistently describing IDF relations over a wide range of durations, which has not been done before as most existing studies focus on durations longer than one hour or day and do not address the deviation from the power law for very long durations (2-5 days).

Authors:

Felix S. Fauer
Jana Ulrich
Oscar E. Jurado
Henning W. Rust

Published in:

Hydrol. Earth Syst. Sci., 25, 6479–6494, 2021; doi:10.5194/hess-25-6479-2021

Author contributions:

Conceptualization: H.W.R., F.S.F. and J.U.
Data curation: F.S.F. and J.U.
Formal analysis: F.S.F.
Funding acquisition: H.W.R.
Methodology: F.S.F. and J.U.
Software: F.S.F., J.U. and O.E.J.
Visualization: F.S.F.
Supervision: H.W.R.
Writing—original draft: F.S.F.
Writing—review & editing: J.U., O.E.J. and H.W.R.

5.1 Introduction

The number of heavy precipitation events has increased significantly in Europe (Kundzewicz et al., 2006; Tank and Können, 2003) and worldwide (Hartmann et al., 2013). Such events are related to flooding and other hazards which can cause severe damage to agriculture and infrastructure (Brémond et al., 2013). The impact of extreme precipitation depends on the temporal scale of the event. Short intense convective precipitation exhibits different characteristic consequences than long-lasting, mostly stratiform, precipitation. Examples for events on different timescales of minutes to hours, days, and weeks are pluvial or flash floods (Braunsbach, Germany, May 2016), river flooding (Elbe, Germany, 2013), and groundwater flooding (Leicestershire, UK, March 2017), respectively.

The definition of precipitation extremes is based on the occurrence probability and is quantified using quantiles (return levels) and associated occurrence probabilities, often expressed as return periods in a stationary interpretation. Quantitative estimations of quantiles and associated probabilities mostly follow one of two popular methods, namely (1) block maxima and their description with the Generalized Extreme Value (GEV) distribution — a heavy-tailed and asymmetric distribution — or (2) threshold exceedances and a description with the generalized Pareto distribution (GPD; e.g., Coles, 2001). Typically, annual precipitation maxima of different timescales are used to describe extreme rainfall events. Both GEV and GPD can be used to model the extreme precipitation of a certain timescale. A common problem for short timescales, i.e., the scarce availability of data with high measurement frequency, can be overcome by modeling several timescales at once. Different timescales can be represented as durations over which the precipitation rate is aggregated and averaged. Most frequently, daily precipitation sums are reported but hourly or 5-minute aggregation are also common.

A way to describe the characteristics of extremes for various durations (timescales) are intensity-duration-frequency (IDF) curves, which describe the relationship between extreme precipitation intensities, their duration (timescale), and frequency (occurrence probability or average return period). These relations have been known since the mid-20th century (Chow, 1953) and have become popular among hydrologists and meteorologists. In estimating these curves, one tries to exploit the assumption of a smooth intensity-frequency relationship across different durations. This helps for interpolating between durations, and it can also improve the estimation for short durations, as shorter time series are often available for those.

Historically, a set of GEV distributions is sought individually for a set of durations (e.g., 5min, 1h, 12h, 24h, and 48h) leading to quantile (return level) estimates for specified probabilities (return periods) for all durations considered. In a second step, for a given probability a smooth function is estimated by interpolating associated quantiles across durations (García-Bartual and Schneider, 2001). However, by using a duration-dependent extreme value distribution (d-GEV) (see the first examples in Menabde et al., 1999; Nguyen et al., 1998), IDF estimation can be carried out in one step within a single model. To achieve this, GEV parameters are defined as functions of duration. This approach prevents the crossing of quantiles across durations and is, thus, considered consistent. Moreover, the

data are used more efficiently, which simplifies the extension of the model, e.g., the inclusion of large-scale covariates at a later stage. This will enable first insights into physical effects beyond the pure statistical evaluation of precipitation.

It is widely agreed that precipitation intensities for given exceedance probabilities follow a power-law-like function (scaling) across duration (Burlando and Rosso, 1996; Gupta and Waymire, 1990; Veneziano and Furcolo, 2002) with higher intensities for short durations. For a range of very short durations ($d \leq 1\,\mathrm{h}$), the scaling assumption does not hold, because maxima from different durations often originate from the same event. This leads to a curvature of IDF curves, where intensity no longer follows a power law with decreasing duration. Bougadis and Adamowski (2006) approached this issue by using two different duration exponents for small durations and for other durations. A more smooth transition can be achieved by including the curvature as a duration offset in the parameters of the GEV distribution without explicitly distinguishing between short and long durations. Koutsoyiannis et al. (1998) used a model with five parameters to describe the complete IDF relation for different probabilities (return periods) and across durations. The underlying idea is based on a reparameterization of the GEV and its three characteristic parameters of location μ, scale σ, and shape ξ. While shape ξ is held constant for all durations d, it is further assumed that the ratio between location and scale μ/σ remains constant across duration d (for details, see Sect. 5.2.3).

Ulrich et al. (2020) built on the approach of Koutsoyiannis et al. (1998) and extended it to a spatial setting with covariates for the d-GEV parameters. Although using both consistent modeling and spatial pooling improves model performance, the need for more flexibility of the IDF curves in longer durations is emphasized and will be addressed in the present study. Therefore, we aim to look for new parameterization of the IDF curve's duration dependence by combining the existing approaches of multiscaling and duration offset and also extending it by a new parameter, namely the intensity offset.

The commonly used variant of the d-GEV with five parameters (Koutsoyiannis et al., 1998) might not be flexible enough for a wide range of durations from minutes to several days. A first approach for extending the d-GEV addresses the simple scaling relation. This model assumes a scaling that is independent of the exceedance probability (return period). However, relaxing this assumption leads to so-called multiscaling, which allows for different scaling-like behavior for different exceedance probabilities (return periods). This is achieved by introducing another parameter η_2, as in Eqs. 5.8 and 5.10 in Sect. 5.2.3. Then, the ratio between location and scale is not constant anymore. Multiscaling is found to be effective for durations longer than 1 h (Burlando and Rosso, 1996; Veneziano and Furcolo, 2002). Van de Vyver (2018) employs the multiscaling approach in a Bayesian setting. On a global scale different scale parameters have been investigated by Courty et al. (2019). None of the named studies combine multiscaling with curvature for short durations but focus on only one of these aspects, while our study is aiming for a combination and analysis of three different features.

In this study, we compare different ways to parameterize IDF curves, including the features of multiscaling and duration offset. In addition, we present a new d-GEV parameter, the intensity offset, which accounts for the deviation from the power law and the flattening

of IDF curves for long durations. To our knowledge, this comprehensive analysis of different features has not been conducted before. Section 5.2 lists the data sources and introduces the different features and their modeling equations, as well as the verification methods, to analyze modeling performance. In Sect. 5.3, the cross-validated verification results of all features are shown with respect to modeling performance of different return periods and durations. For verification, we perform a case study using rainfall gauge data from a catchment in Germany. IDF curves that include all analyzed features are presented for selected stations.

5.2 Data and Methods

We use precipitation measurements in an area in and around the catchment of the river Wupper in western Germany. In order to compare different models for the d-GEV, we use the quantile skill index (QSI) introduced in Ulrich et al. (2020) within a cross-validation setting. For the resulting IDF curves, we obtain confidence intervals using a bootstrapping method and test their coverage with artificial data with known dependence levels in a simulation study. The data and all necessary methods are explained in the following section.

5.2.1 Station-Based Precipitation Data

Precipitation sums for the minute, hour, and day are provided by Wupperverband and the German Meteorological Service (DWD CDC, 2019, 2021). Rain gauges are located in and around the catchment of the river Wupper in North Rhine-Westphalia, Germany. In total, 115 stations are used. Data from two measuring devices with a distance below 250 m are combined into one station each in order to obtain a longer time series, thus resulting in a total of 92 grouped stations. However, in cases where measurement series are grouped together, it is common to have measurements from both instruments for a certain period of time. Thus, when merging, we decided to use only the observations with the higher measuring frequency for the analysis. For example, when combining two time series with hourly and daily data, respectively, the time series of all aggregation levels (durations) are obtained from the hourly data for the overlapping period. This choice is made because 24-h values from an hourly measurement frequency might show higher intensities since the 24-hour window is shifted hour by hour in order to find the annual maximum. On the other hand, 24-h values from the daily measurement frequency are recorded for a fixed day block, i.e., from 8:00 to 7:59 LT (local time). A test of robustness was performed (not shown) to assess how the estimated IDF curves are affected by the choice of measurement frequency which is preserved during the time of overlap when merging the time series. For this purpose, two IDF curves were created, i.e., (1) choosing the maxima from a higher measurement frequency and (2) choosing the maxima from a lower measurement frequency. In most cases, there was no relevant difference between both methods.

Years with more than 10% of missing values are disregarded. Some years contain measurement artifacts, where identical rainfall values were repeated over several time

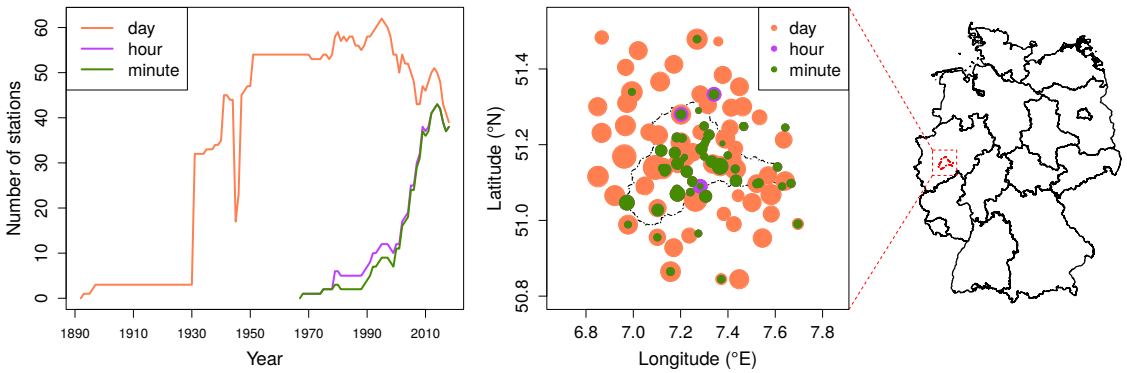


Figure 5.1: Left: Number of stations according to the temporal resolution. Right: Station location (circles) with data availability (circle area size) by temporal resolution (color). The red line within the map of Germany indicates the Wupper catchment boundary.

steps. After consulting the data maintainers, these years are removed before the analysis. The data exhibit heterogeneity in terms of the temporal frequency and the length of the resulting time series. Figure 5.1 presents the availability of data over time (left) and space (right) for three possible temporal resolutions. More specifically, minute data have been available since 1968, whereas daily records range back to 1893.

Time series for different durations are obtained from an accumulation over a sample of durations as follows:

$$d \in \{1\,\text{min}, 4\,\text{min}, 8\,\text{min}, 16\,\text{min}, 32\,\text{min}, 1\,\text{h}, 2\,\text{h}, 4\,\text{h}, 8\,\text{h}, 16\,\text{h}, 24\,\text{h}, 48\,\text{h}, 72\,\text{h}, 96\,\text{h}, 120\,\text{h}\} \tag{5.1}$$

respectively, as done by Ulrich et al. (2020). In the following, numerical values are presented in hours. We use only the annual maxima of these time series to model the distribution of extreme precipitation. The original precipitation time series are accumulated with the R package `IDF` (Ulrich and Ritschel, 2021). The data set with annual maxima can be found online and supports the findings of this study (Fauer et al., 2021b).

The set of durations d is chosen such that small durations are presented with smaller increments than larger durations. In a simulation study, we tested whether using a different set of durations with a stronger focus on short durations affects the results, but no such effect could be found (see Appendix C.3).

Minute measurements might be less accurate when only a small number of rain drops is recorded and measured intensity is affected by sampling uncertainty. However, for events that are identified as annual maxima, we expect the rain amount to be large enough so that a higher sampling uncertainty compared to larger measurement accumulation sums can be neglected.

5.2.2 Generalized Extreme Value Distribution

One of the most prominent ideas of extreme value statistics is based on the Fisher-Tippett-Gnedenko theorem, which states that under suitable assumptions, maxima drawn from

sufficiently large blocks follow one of three distributions. These distributions differ in their tail behavior. The GEV distribution comprises all three cases in one parametric family and is widely used in extreme precipitation analysis as follows (Coles, 2001):

$$G(z) = \exp\left\{ -\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}. \tag{5.2}$$

Here, the non-exceedance probability G for precipitation intensity z depends on the parameters location μ, scale $\sigma > 0$, and shape $\xi \neq 0$, with z restricted to $1+\xi(z-\mu)/\sigma > 0$. The non-exceedance probability can be expressed as a return period, e.g., for an annual block size $T(z) = 1/(1-G(z))$ years. Consequently, return values for a given non-exceedance probability $0 < p < 1$ can be calculated by solving Eq. (5.2) for z, as follows:

$$z = ([-\ln\{p\}]^{-\xi} - 1)\frac{\sigma}{\xi} + \mu. \tag{5.3}$$

Water management authorities and other institutions rely on return values for different durations. However, the GEV distribution in the form of Eq. (5.2) is limited to one selected duration at a time. One way to account for that need is to model each duration separately and then, in an independent second step, interpolate the resulting quantiles (return levels) across duration d, as done in the KOSTRA atlas (DWD, online) of the German Meteorological Service (DWD). One huge disadvantage of this method is that quantile crossing can occur, meaning that quantiles (intensities) associated with smaller exceedance probabilities can have higher values than quantiles from larger exceedance probabilities in some duration regimes. To solve this problem, Nguyen et al. (1998), Koutsoyiannis et al. (1998), and Menabde et al. (1999) proposed a distribution with parameters depending on duration d; there is thus only one single model required to obtain consistent (i.e., non-crossing) duration dependent quantiles (return values). Another advantage is the involvement of data from neighboring durations in the estimation of GEV parameters. For the modeling of short-duration rainfall, often very little data are available than for longer durations $d \geq 24$ (one day). Thus, in this setting, information from long durations has the potential to increase modeling performance for short durations as well.

5.2.3 Duration Dependence

There are multiple empirical formulations for the relationship between intensity z and duration d. Koutsoyiannis et al. (1998) proposed a general form with five parameters for IDF curves. Therefore, a reparameterization and extension of the GEV is needed with the following:

$$\sigma(d) = \sigma_0(d+\theta)^{-\eta} \tag{5.4}$$

$$\mu(d) = \tilde{\mu}\sigma(d) \tag{5.5}$$

$$\xi = \text{const.} \tag{5.6}$$

$$G(z) = \exp\left\{ -\left[1 + \xi\left(\frac{z}{\sigma_0(d+\theta)^{-\eta}} - \tilde{\mu}\right)\right]^{-1/\xi}\right\}. \tag{5.7}$$

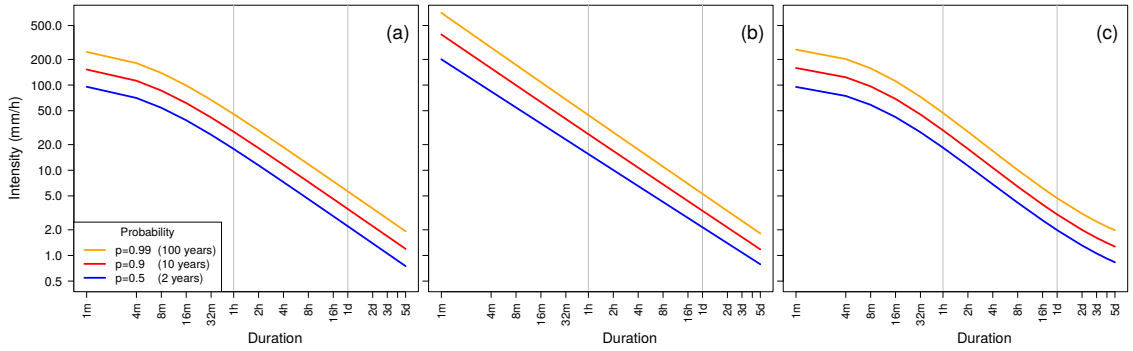


Figure 5.2: IDF curve examples showing the visualization of different IDF curve features. (**a**) Curvature for short durations. (**b**) Multiscaling. (**c**) Curvature for short durations, flattening for long durations, and multiscaling.

Here, $\tilde{\mu}$ is the rescaled location parameter, θ is the duration offset, and η the duration exponent. Scale σ follows a two-parameter power law (scaling relation) of duration d, with scale offset σ_0 being constant for all durations. For $d \gg \theta$, it is justified to disable the duration offset feature by setting duration offset $\theta = 0$. The resulting IDF curves (Fig. 5.2a) have two main features. (1) The curves follow a power law for a wide range of durations $d > 1$ (1 h), and the power law exponent (slope in a double logarithmic plot) is described by a duration exponent η and is equal for all probabilities. (2) The deviation from the power law (or curvature) for $d < 1$ (one hour) is described by the duration offset θ.

When applying the GEV separately to every duration out of a set of durations and interpolating in a second independent modeling step, the number of parameters equals three GEV parameters times the number of selected durations plus at least three parameters for interpolating every quantile. For the set of durations chosen here, and for evaluating five quantiles, this implies estimating $15 \times 3 + 5 \times 3 = 60$ parameters. For the consistent approach, estimation is reduced to only five d-GEV parameters, i.e., $\tilde{\mu}, \sigma_0, \xi, \theta$, and η. A smaller parameter set is less likely to overfit the data and enables us to better investigate underlying physical processes.

Models can be further improved by adding the multiscaling feature (Burlando and Rosso, 1996; Gupta and Waymire, 1990; Van de Vyver, 2018) which introduces different slopes, depending on the quantile (or associated probability). Figure 5.2b) presents how this feature affects the IDF curves. In order to highlight the effect of the multiscaling, we set $\theta = 0$, resulting in no deviation from a power law (curvature) for short durations. The multiscaling feature is added at the cost of one additional parameter η_2 (second duration exponent) in the model, as follows:

$$\sigma(d) = \sigma_0 d^{-(\eta+\eta_2)} \tag{5.8}$$

$$\mu(d) = \tilde{\mu}\sigma_0 d^{-\eta}. \tag{5.9}$$

Using only the curvature feature, Ulrich et al. (2020) reported decreasing performance in consistent modeling for longer durations $d \geq 24$ (1 d) when compared to using separate

GEV models for each duration. Our attempt aims at more flexibility in the IDF curve in the long-duration regime. Therefore, we combine both features, curvature (duration offset) and multiscaling (second duration exponent), and add a new parameter τ, the intensity offset, which allows for a slower decrease of intensity for very long durations $d \gg 24$ (Fig. 5.2c). This effect will be called flattening in the following:

$$\sigma(d) = \sigma_0(d+\theta)^{-(\eta+\eta_2)} + \tau, \tag{5.10}$$

$$\mu(d) = \tilde{\mu}(\sigma_0(d+\theta)^{-\eta} + \tau). \tag{5.11}$$

Summarizing this section, the following three different features were presented: (1) curvature, described by the duration offset parameter θ, (2) multiscaling, using a second duration exponent η_2, and (3) flattening with the intensity offset τ, which is introduced in this study. In the following, we use the notation IDF_c for a model including only the curvature feature, i.e., $\eta_2 = 0$ and $\tau = 0$, IDF_m for a model including only the multiscaling feature, i.e., $\theta = 0$ and $\tau = 0$, and IDF_f for a model including only the flattening feature, i.e., $\theta = 0$ and $\eta_2 = 0$. Feature combinations are denoted as, e.g., $\mathrm{IDF}_\mathrm{cmf}$ for all features. The plain duration-dependent model without curvature, multiscaling, and flattening is denoted as IDF.

5.2.4 Parameter Estimation

Parameters of the d-GEV distribution are estimated by maximizing the likelihood (maximum likelihood estimation – MLE) under the assumption of independent annual maxima. Jurado et al. (2020) showed that independence is a reasonable assumption in many cases, especially for long durations. Their study was performed on an earlier version of the same data set that was used in this study. Moreover, Rust (2009) showed that in strongly dependent time series convergence towards the GEV distribution is slower. But, assuming an appropriate choice of model, dependence does not play a large role. We use the negative log-likelihood L, to avoid using products of small numbers, as follows:

$$L(\tilde{\mu}, \sigma_0, \xi, \theta, \eta, \eta_2, \tau \mid \boldsymbol{Z}) = \sum_{n=1}^{N} \sum_{d \in D} -\ln\left(G(z_{n,d}, d \mid \tilde{\mu}, \sigma_0, \xi, \theta, \eta, \eta_2, \tau)\right), \tag{5.12}$$

with the number of data points N, the duration set D, data points for each duration $z_{n,d}$, and the characteristic parameters of the modified d-GEV $\tilde{\mu}, \sigma_0, \xi, \theta, \eta, \eta_2, \tau$. The sum of L over all data points $\boldsymbol{Z}$ is minimized with the R package `optim()` function (R Core Team, 2020). The R package `IDF` (Ulrich and Ritschel, 2021), which was already used for the accumulation, also provides functions for fitting and plotting IDF curves. Its functionality was extended in the context of this study and now provides options for both multiscaling and flattening in IDF curves.

Finding reasonable initial values for d-GEV parameters in the optimization process was a major challenge during parameter estimation, because optimization stability strongly depends on the choice of initial values. Details about this procedure can be found in Appendix C.1.

5.2.5 Quantile Skill Index

After estimating GEV parameters, quantiles q (return levels) can be predicted for a chosen non-exceedance probability p (or return period T, with $T = 1/(1-p)$), using Eq. 5.3. To verify how well a modeled quantile q of a given probability p represents extremes in the data, we use the quantile score (QS) as follows (Bentzien and Friederichs, 2014; Koenker and Machado, 1999):

$$QS(p) = \sum_{n}^{N} \rho_p(z_n - q), \tag{5.13}$$

with a small score indicating a good model. Here, ρ_p is the tilted absolute value function, also known as the so-called check function. For high non-exceedance probabilities p, it leads to a strong penalty for data points that are still higher than the modeled quantile ($z_n > q$), as follows:

$$\rho_p(u) = \begin{cases} pu & , u > 0 \\ (p-1)u & , u \leq 0 \end{cases} \tag{5.14}$$

with $u = z_n - q$. Using this approach, the QS allows for detailed verification for each probability p and duration d separately, by predicting a quantile intensity for a given p and d and comparing it with data points $z_{n,d}$ of duration d.

To compare different IDF models in terms of the QS, we require another verification measure. The quantile skill score QSS compares the quantile score QS_M of a new IDF model M with the quantile score QS_R of a reference IDF model R as follows:

$$QSS_{M|R} = 1 - QS_M/QS_R. \tag{5.15}$$

The QSS takes values $-\infty < QSS \leq 1$ with $QSS = 1$ for a perfect model. Positive values $QSS_{M|R} > 0$ are associated with an improvement of M over R. In case the model M is outperformed by the reference R, the resulting QSS is negative $QSS_{M|R} < 0$. In this case, its value is not easily interpretable. This issue is acknowledged by the quantile skill index (QSI) suggested by Ulrich et al. (2020). In the case of $QSS_{M|R} < 0$, reference R and model M are exchanged and $-QSS_{R|M}$ is used for negative values of the QSI, as follows:

$$QSI = \begin{cases} 1 - QS_M/QS_R & , QS_M \leq QS_R \\ QS_R/QS_M - 1 & , QS_M > QS_R. \end{cases} \tag{5.16}$$

The QSI has a symmetric range and indicates either (1) a good skill over the reference when leaning clearly towards 1, (2) little or no skill when being close to 0, or (3) worse performance than the reference when leaning clearly towards -1.

In this study, the quantile score was calculated in a cross-validation setting. For each station, the available years with maxima are divided into n_{cv} non-overlapping blocks of 3 consecutive years. Then, for each cross-validation step i, one block is chosen as testing set, and all the other blocks are used as training data set. For the remaining cross-validation steps, this procedure is repeated with another block chosen as testing set in each step until

all blocks have been used as testing sets exactly once. The cross-validated QS is obtained by averaging the score of all cross-validation steps as follows:

$$\mathrm{QS}^{\mathrm{cv}} = \frac{1}{n_{\mathrm{cv}}} \sum_{i=1}^{n_{\mathrm{cv}}} \mathrm{QS}_i. \tag{5.17}$$

Then, the QSI is derived from the averaged cross-validated QS of the model, QS_M^{cv}, and the averaged cross-validated QS of the reference, QS_R^{cv}, according to Eq. (5.16). If a year was assigned to the training or testing data set, then all available accumulation durations are used for training or testing, respectively, to avoid dependence between the test and validation set.

In order to compare individual model features, we will use the mentioned models without this specific feature as a reference in the following.

5.2.6 Bootstrapping and Coverage

To provide an estimate of the uncertainty of the intensity quantile estimates in IDF curves, we obtain 95% confidence intervals using a bootstrapping method. To account for dependence between annual maxima of different durations we apply the ordinary non-parametric bootstrap percentile method (Davison and Hinkley, 1997) as follows.

Please note that, in this paragraph, the empirical quantiles used for the confidence intervals should not be confused with the intensity quantiles, which describe the return level and are referred to as quantiles as well. For each station, we draw a sample of years (with replacement) from the set of years with available data. This way, for a chosen year, all maxima from this year are used, and we expect that sampling in this way maintains the dependence structure of the data. We then estimate the parameters of the d-GEV, which is used to calculate the intensity quantile that is connected to a certain non-exceedance probability (see Eq. 5.3). We obtain a distribution of the estimated intensity quantiles by repeating this process 500 times. From this distribution, we use the empirical 0.025 and 0.975 quantiles to obtain the upper and lower bounds of the 95% confidence interval of intensity quantiles.

We conduct a simulation study to examine whether the derived confidence intervals provide reasonable coverage despite the dependence between the annual maxima of different durations. Therefore, we simulate 500 samples of data, each with a size of $n = 50$ years, with a known dependence between durations. In a first case, samples with no dependence between durations are obtained by drawing random values from a d-GEV distribution. Further details on the simulated data can be found in Appendix Sect. C.2. In a second case, to obtain data with dependence between durations, we use the R package `SpatialExtremes` to simulate values from a Brown-Resnick simple max-stable process with known dependence parameters. We use the range and smooth parameter $(\rho, \alpha) \in \{(1, 0.2), (120, 1), (60, 1)\}$ for (1) a weak dependence, (2) a strong dependence, and (3) a dependence found for Wupper catchment (Jurado et al., 2020), respectively. We transform the simulated data from having Fréchet margins to d-GEV margins with the chosen parameters, as done by Jurado et al. (2020), and adjust them to the hourly scale used in this study. Then, for the artificial

data, confidence intervals are obtained by bootstrapping with 500 repetitions, as described above. For better understanding, this results in a total of 500x500x50 data points and 500 confidence intervals. The ratio of samples in which the confidence intervals cover the true intensity z and the total number of samples (500) is called the coverage. It can be calculated for each duration and probability separately.

5.3 Results

Results are presented in the following order: (1) modeling performance is verified with the QSI for the three different IDF curve features, i.e., curvature, multiscaling and flattening. (2) IDF curves with all three features are shown for two rain gauges. Curves are presented with a 95% confidence interval, as created by a bootstrapping method. (3) The trustworthiness of this bootstrapping method applied to the new model with all three features is investigated with a coverage analysis, based on simulated data.

5.3.1 Model Validation

The QSI is used to compare the quantile score of a model with that of a reference. In order to specifically investigate the influence of a single model feature, we use these features in a model and compare with a reference without this specific feature; e.g., $QSI_{\mathrm{IDF}_c|\mathrm{IDF}}$ gives the performance for a model including only curvature against the plain reference without curvature, or $QSI_{\mathrm{IDF}_{cmf}|\mathrm{IDF}_{\mathrm{mf}}}$ gives the performance for the full model including curvature, multiscaling, and flattening against a reference with multiscaling and flattening and without curvature (see Sect. 5.2.3).

Figure 5.3 shows the QSI for an IDF model including each of the three features of curvature, multiscaling and flattening (columns) combined with no other, one other, or both other features (rows) against a reference model which differs only in the one feature under investigation (labels on top of the columns). For each $QSI_{\mathrm{IDF}_1|\mathrm{IDF}_2}$, one panel is provided in Fig. 5.3. Models and references are listed in Table C.2 in Appendix C.5. In this way, the potential performance of each feature, e.g. curvature, is analyzed and is denoted as e.g., curvature skill. QSI values between -0.05 and 0.05 are considered as being an indicator of no relevant difference between model performances.

The curvature (duration offset $\theta \neq 0$) for short durations can be explained by a stronger connection between the annual maxima of different durations, which tend to originate from the same event. Usually, the most intense phase of a heavy precipitation event lasts for several minutes, and aggregated maxima do not differ much on this scale. Based on this idea, curvature influences the IDF curve's shape only for very short durations below one hour (Fig. 5.2a). The consistently positive QSI values for $d = 1/60$ (one min) for the curvature skill support this theory. These results show that this duration regime $d = 1/60$ is much better modeled with the curvature, compared to models without this feature. However, the slope for medium durations, described by the duration exponent (see Fig. 5.2b), is steeper when using curvature compared to models that do not use curvature. So, for medium and long durations, models perform equally well or worse when curvature

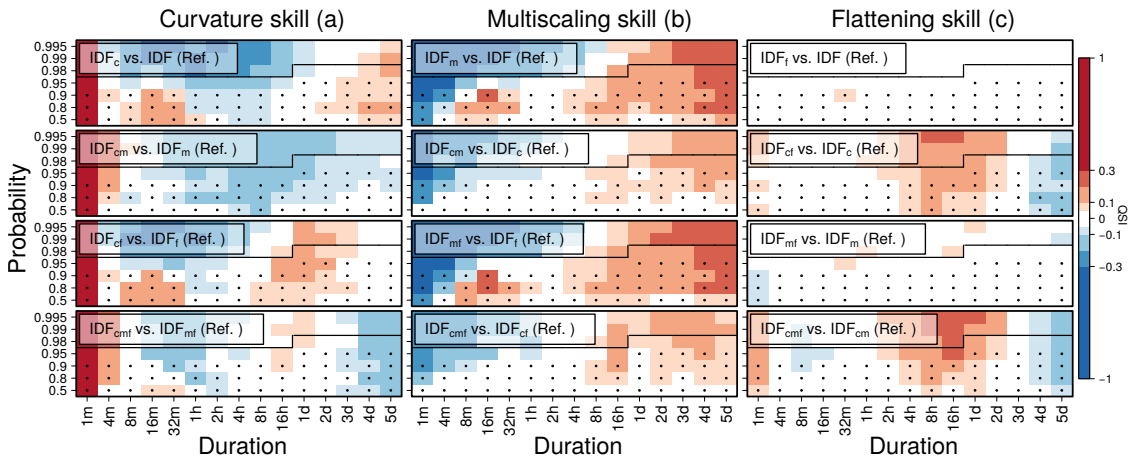


Figure 5.3: Quantile skill index (QSI) of the three features (columns) for four different cases (rows) where the investigated feature is combined with no other (upper row), one other (second and third row), or both other features (lower row) in model and reference. Column titles indicate the feature switched on in the model and switched off in the reference. The slightly opaque labels in the panels indicate which model and reference is used (see also Table C.2). Dots show whether the average length of the time series over all stations is longer than the return period $T = 1/(1-p)$ (here shown as probability p) and indicate the verification trustworthiness. Black lines are derived from the number of years of the station with the longest time series. The verification for rare events (upper part of each panel) above the black line has to be treated carefully because the data do not cover this time period. For this verification plot, only stations that provide data on a minute scale were used.

is used than reference models without curvature, in most cases, on average (see the blue regimes in Fig. 5.3a). In the absence of multiscaling (rows 1 and 3), a further performance increase could be found for durations between 8 hours and 5 days.

Multiscaling allows for different slopes of different p quantiles on a double logarithmic scale. Figure 5.3b shows that this feature increases modeling performance mainly for long, but also for some sub-hourly, durations when estimating quantiles for small non-exceedance probabilities. Not using curvature enables a small multiscaling skill gain for sub-hourly durations (rows 1 and 3 in Fig. 5.3b). An explanation could be that multiscaling tends to let IDF curves associated with different return periods diverge for short durations and converge for long durations. This behavior might interfere with the duration offset's introduction of curvature in short durations. Furthermore, the presence of curvature leads to a slightly smaller skill increase for durations longer than 16 hours (rows 2 and 4 in Fig. 5.3b). This effect agrees with the results from the curvature skill verification (Fig. 5.3a), where it was shown that curvature improves modeling performance only for very short durations and has no or a negative effect on medium and long durations.

The intensity offset τ is a new feature, first introduced in this study, which addresses the empirically observed slower decrease in intensity for very long durations, called flattening. In the case where curvature is enabled in both model and reference (rows 2 and 4 in Fig. 5.3c), the flattening feature improves modeling performance slightly for the shortest

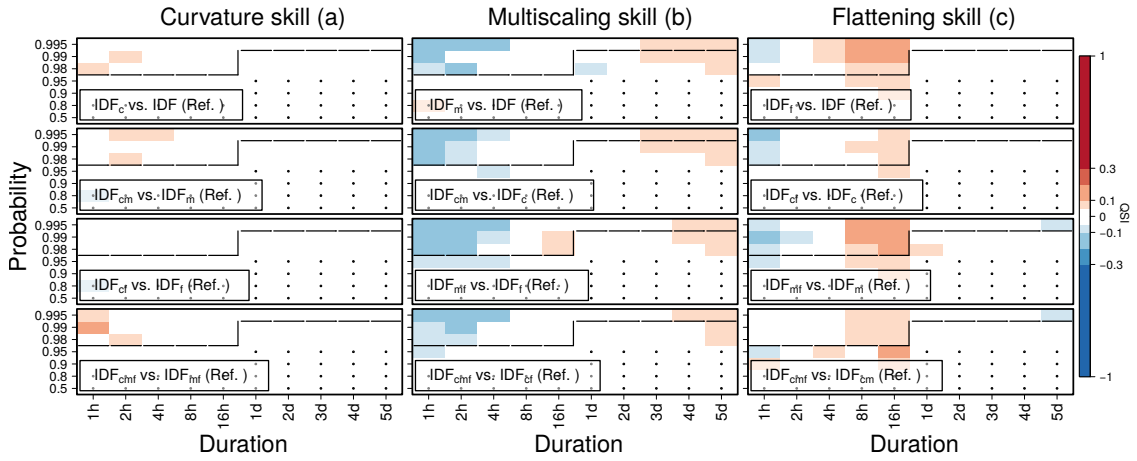


Figure 5.4: Quantile skill index for data with an hourly resolution. The visualization scheme follows that of Fig. 5.3. Here, all models were trained and tested for durations $d \geq 1$. Here, all stations were considered, regardless of the temporal resolution.

duration of 1 minute and strongly for medium durations between 2 hours and 1 day. Here, the flattening might compensate for the loss in skill that we observe for medium durations for models with curvature. In these cases, there is a slight loss in skill for very long durations. In cases where curvature is not used, flattening is not needed as it provides no clear skill. An explanation for the flattening of the IDF curves in long durations could be seasonal effects, with annual maxima of short or long durations occurring more often in the summer or winter months, respectively. These effects are currently under further investigation.

When modeling only the durations $d \geq 1$ (one hour) of all available stations, the models are rather indifferent towards parameterization (Fig. 5.4). Here, multiscaling and flattening show some skill improvements for long and medium durations, respectively, similar to that in Fig. 5.3, but to a much smaller extent when compared to a data set which uses the whole range of durations from 1 minute to 5 days for both training and testing (Fig. 5.3).

We conclude that the choice of parameters depends on the study purpose. When focusing on long ranges of durations, we recommend using features like curvature, multiscaling and flattening. If the focus lies on long durations, or the data do not provide a sub-hourly resolution, simple scaling models might be sufficient. These recommendations are further elaborated on in the discussion in Sect. 5.4.

5.3.2 IDF Curves

Figure 5.5 shows IDF curves for the stations Bever and Buchenhofen, where a long precipitation series is available (51 years with minute resolution and 76 years with daily resolution in Bever and 19 and 77 years in Buchenhofen, respectively). The difference in available years for different durations has an impact on the width of 95% confidence intervals, with uncertainty being larger when little data are available. Noticeably, confidence

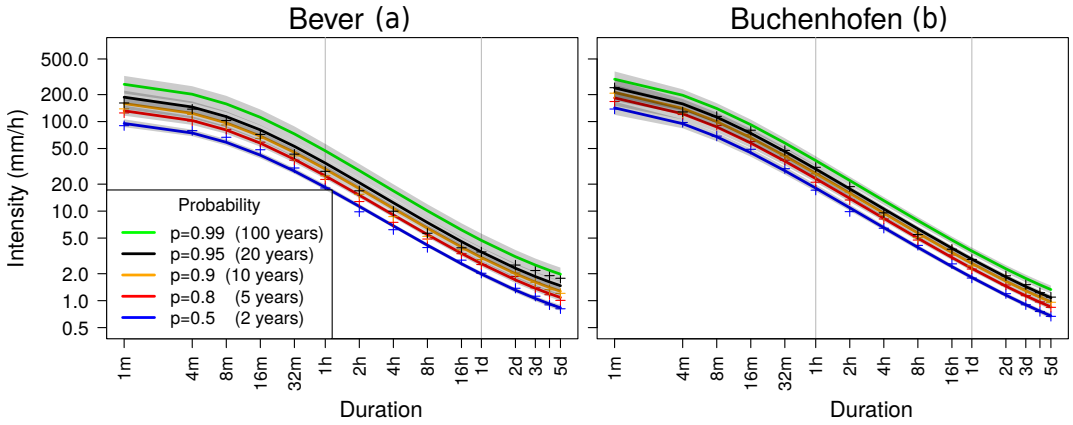


Figure 5.5: IDF curves for two example stations within the Wupper catchment. Empirical quantile estimates are denoted with the plus signs (+). Confidence intervals are obtained from a bootstrapping procedure.

intervals for both stations for $p = 0.99$ and $d = 1/60$ have a wide range over more than 100 mm/h. Considering that the 100-year return level was not observed in either station, a wide confidence interval range was expected. For $p \leq 0.8$ in Bever, the confidence intervals remain narrow, even on a minute scale.

5.3.3 Coverage

Confidence intervals in Fig. 5.5 are obtained from a bootstrapping procedure. In the formulation of the likelihood, we assume the maxima of different durations to be independent. This assumption might not be justified especially for short durations (see Jurado et al., 2020), and thus, this dependence must be taken into account when estimating uncertainties. Disregarding the dependence would result in an underestimation of the uncertainty. To account for this effect, all annual maxima of each year — for all considered durations — are always included jointly into a bootstrapping sample. We assume that this procedure preserves the dependence structure between durations. To investigate this assumption, we calculate the coverage of simulated data (see Appendix C.2) from (1) a d-GEV distribution without dependence and a Brown-Resnick max-stable process with (2) a typical dependence for the Wupper station (Jurado et al., 2020) and (3) a rather weak and (4) strong dependence between durations (Fig. 5.6). In the first case without any dependence, the displayed coverage does completely agree with the 95% confidence interval, without any respect to duration or frequency (probability). When using dependence on a weak or strong level, the coverage is smaller but still around 90%. This can be interpreted as an underestimation of uncertainty to a small extent, by the confidence intervals in case of a high dependence. The true dependence of durations was not investigated in this study and could be lower. That said, these results suggest that bootstrapping is a suitable tool for estimating confidence intervals in the presented context.

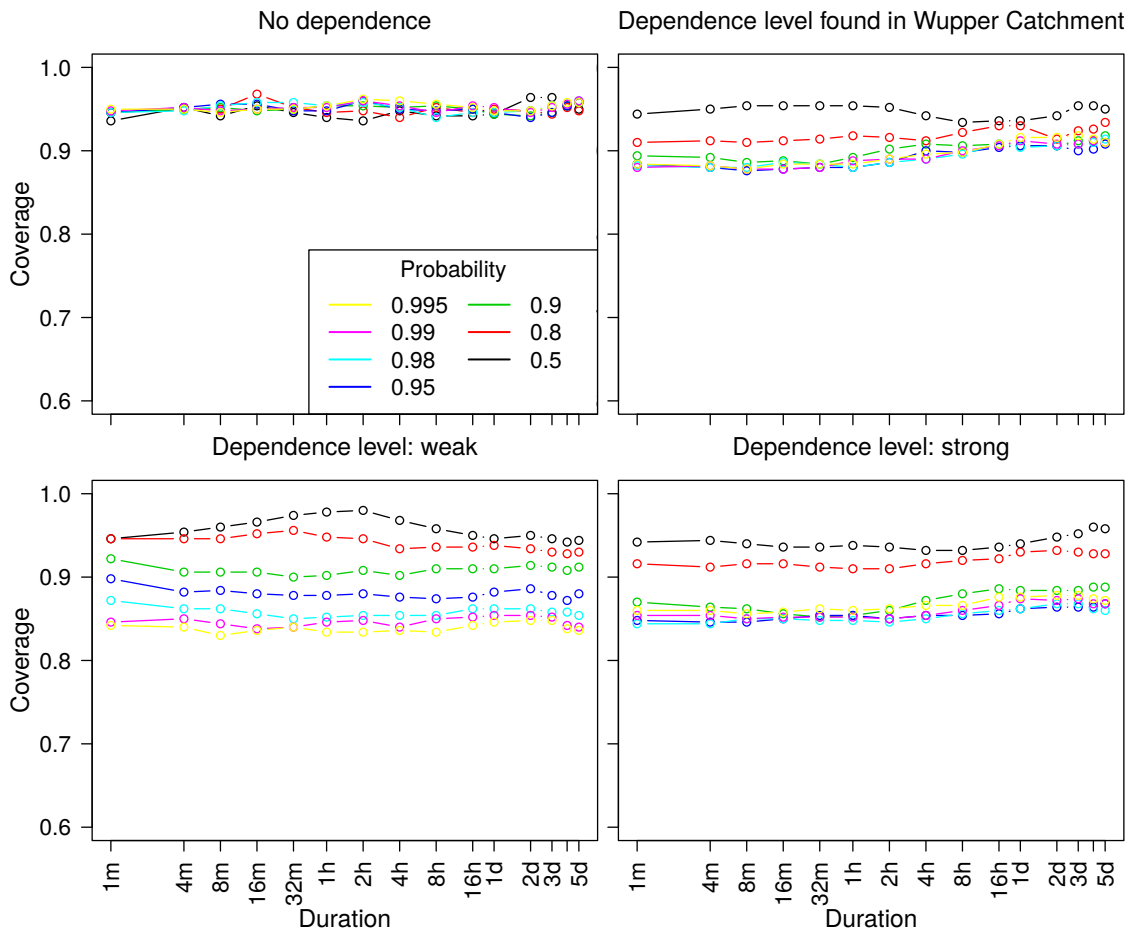


Figure 5.6: Bootstrapping coverage. Using a Brown-Resnick max-stable process, the coverage was determined in order to investigate the reliability of 95% confidence intervals from bootstrapping. A total of three different levels of dependence were used.

5.4 Discussion

In this study, we show that model performance can be increased when the flattening of IDF curves in the long-duration regime is taken into account. We assume that this behavior arises from seasonal effects. That means that the annual maxima of different durations may not follow the same scaling process. However, this topic is currently under further investigation (Ulrich et al., 2021b).

The analyzed features — curvature, multiscaling, and flattening — were seen in the results to have a different impact on modeling performance, depending on the duration and return period. All features are able to improve the model for certain regimes, but depending on the problem that is approached, features should be chosen accordingly. If the focus is on a small timescales of minutes, then the curvature skill is important for a

good modeling result. When curvature is used and medium to long timescales are also of importance, then the flattening feature should be used. This helps to compensate for the deterioration due to curvature over longer durations. Multiscaling is a good choice, if a loss in skill for short durations can be accepted in exchange fora simultaneous improvement at long durations, regardless of which other features are requested.

The skills of the features depend on another feature's presence. This dependence is strongest for the flattening, which can only improve the model when curvature is used. The modeling performance of the curvature depends less on the presence of other features. The same applies to the multiscaling feature.

These suggestions hold for models that are supposed to cover a wide range of timescales from minutes to days. For data with hourly or more coarse temporal resolution, the skill gain from using the features is much smaller. Here, flattening can improve the model slightly on a daily timescale and multiscaling only improves modeling long durations a little bit but leads to a slight reduction in skill for the hourly timescale.

Additional parameters give the model more flexibility. Including τ in the model allows one to reduce deviations between model and data points particularly for long durations. This, in turn, opens the possibility to vary the remaining parameters such that deviations between model and data points can be reduced in other (e.g. short) duration regimes. Conversely, this also holds when including θ. In this way, a parameter that changes the curve in long durations can increase the modeling performance in other durations and even slightly decrease model performance in long durations in certain cases. In Fig. 5.7, IDF curves for two models are compared for a chosen station (Bever). The model that includes flattening ($\mathrm{IDF}_{\mathrm{cmf}}$) is able to follow the empirical quantiles in long durations as well as the model without flattening ($\mathrm{IDF}_{\mathrm{cm}}$). However, flattening gives the model the opportunity to better follow the empirical quantiles in medium durations between 4 h and 1 d which is in accordance with the results in Fig. 5.3c.

The parametric form of the IDF relation is based on three modifications to a simple power law which are motivated by our understanding of the rainfall process:, namely curvature (e.g. Koutsoyiannis et al., 1998) for small durations addressing limits to rainfall intensity, multiscaling (e.g. Van de Vyver, 2018) taking care of a varying scaling behavior for events of different strength, and flattening (suggested here) resulting from a mixing of convective and stratiform generated precipitation extremes from different seasons in the climate regime under study. Our contribution is to combine these modifications into a flexible parametric form capable of describing various effects related to rainfall processes. The resulting model is based on empirical grounds and it can be shown to improve the description over simple power law models. Other modifications are possible. To our knowledge, there is no theoretical justification for these forms which can be derived from first principles for rainfall processes. Since our results might apply only to the geographical region under investigation, further studies are necessary to find out whether the found model performances of the different features are generally applicable. The character of the shape parameter ξ with respect to duration is still unclear, since no linear or log-linear duration dependence could be found as with the location and scale parameters μ and σ. A possible approach would be to let the shape parameter vary smoothly across durations in

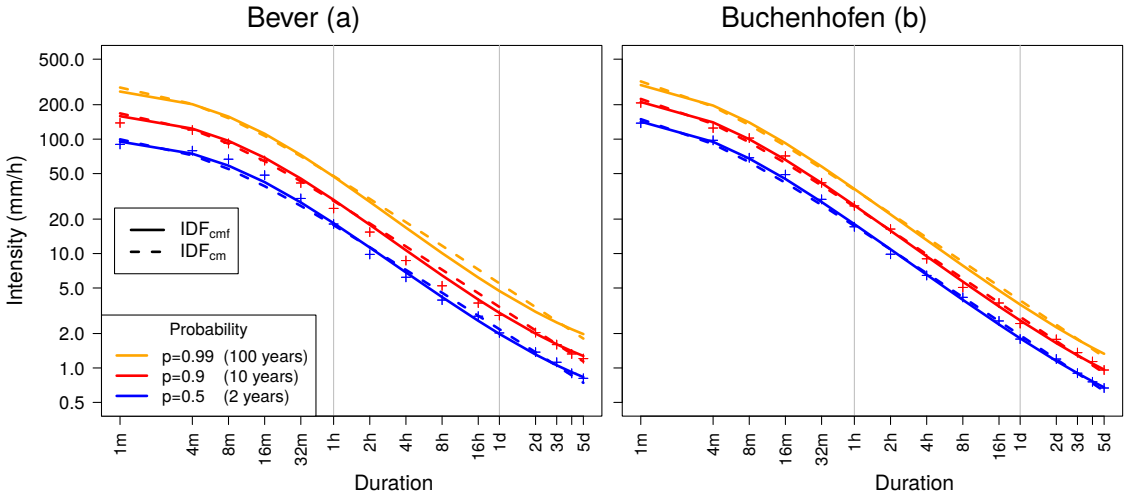


Figure 5.7: IDF curve for Bever. A comparison of a model with flattening (IDF_{cmf}) and a model without flattening (IDF_{cm}). Empirical quantile estimates are denoted with the plus signs (+).

a non-parametric manner but to penalize its deviation from the median over all durations (Bücher et al., 2021). However, the scarce data availability hampers a more complex estimation of the shape parameter.

5.5 Summary and Outlook

The aim of this study is to compare and suggest new parametric forms of consistent IDF curves that are applicable to a large range of durations from minutes to several days and, therefore, cover events from short-lived convective storms to long-lasting synoptic events. The dependence on duration is implemented in the location and scale parameter and allows for three features, i.e., curvature, multiscaling, and flattening. The analysis of these features enables us to understand more about the underlying physical effects beyond the subject of return periods and provides more flexible IDF curves that are suitable for a wide range of durations. The results of our simulation study show that we are able to provide reasonable estimates of uncertainty using bootstrapping and also with regard to dependence between durations.

Our findings agree with Veneziano and Furcolo (2002), who found that simple scaling was adequate for modeling short durations, and multiscaling was adequate for long durations. Moreover, our conclusion that curvature improves the modeling of short durations indirectly agrees with Bougadis and Adamowski (2006), who used different slopes for durations longer or shorter than 1 hour, respectively, and concluded that linear scaling does not hold for small durations.

Consistent modeling using the d-GEV enables the use of fewer parameters. In this way, the model can be easily extended, e.g. using physically relevant atmospheric covariates. Thus, improving the parameterization of the d-GEV is crucial to leading the path for

further steps. In future studies we plan to include spatial covariates into the estimation of the newly proposed d-GEV parameters, including intensity offset, in order to use data from different locations more efficiently. Also, the concept of non-stationary precipitation with respect to IDF curves is important to consider (see Cheng and AghaKouchak, 2014; Ganguli and Coulibaly, 2017; Yan et al., 2021) since extremes are expected to vary due to climate change. For example, Benestad et al. (2021) found a model that enables downscaling of 24-h measurement data to shorter durations without assuming stationarity. However, we think that implementing atmospheric large-scale covariates (as in Agilan and Umamahesh, 2017) into the flexible d-GEV model proposed here would allow for a better understanding of the underlying processes. We plan to use this approach to investigate the change in characteristics of extreme precipitation due to climate change in future studies. While the choice of method depends on the study target, different approaches have been taken to create IDF curves, as in Bezak et al. (2016), who used copula-based IDF curves and reported that IDF curves might be sensitive to the choice of method. This is important to consider when deciding on the appropriate way to create IDF curves. Moreover, the origin of flattening in annual maxima for long durations is currently investigated in more detail (Ulrich et al., 2021b).

The analysis of the performance shows that the new parametric form of the duration-dependent GEV suggested here, together with the bootstrap-based confidence intervals, offers a consistent, flexible, and powerful approach to describing the intensity-duration-frequency (IDF) relationships for various applications in hydrology, meteorology, and other fields.

Code and data availability: Parts of the rainfall data are freely available from the German Meteorological Service (DWD CDC, 2019, 2021). Annual precipitation maxima of stations from Wupperverband and DWD for 15 different durations are available in Fauer et al. (2021b). The d-GEV analysis is based on the R package `IDF` and is available on CRAN (`https://CRAN.R-project.org/package=IDF`; R Core Team, 2020; Ulrich and Ritschel, 2021; Ulrich et al., 2020).

Competing interests: The contact author has declared that neither they nor their co-authors have any competing interests.

Acknowledgements: We would like to thank the German Meteorological Service (DWD) and the Wupperverband, especially Marc Scheibel, for maintaining the station-based rainfall gauge and providing us with data.

Financial support: This research has been supported by the Bundesministerium für Bildung und Forschung (grant no. 01LP1902H) and the Deutsche Forschungsgemeinschaft (grant nos. GRK2043/1, GRK2043/2, and DFG CRC 1114).
We acknowledge support from the Open Access Publication Initiative of Freie Universität Berlin.

Chapter 6

Modeling Seasonal Variations of Extreme Rainfall on Different Timescales in Germany

Authors:

Jana Ulrich
Felix S. Fauer
Henning W. Rust

Published in:

Hydrol. Earth Syst. Sci., 25, 6133–6149, 2021;
doi:10.5194/hess-25-6133-2021

Author contributions:

Conceptualization: H.W.R. and J.U.
Data curation: J.U.
Formal analysis: J.U.
Funding acquisition: H.W.R.
Methodology: J.U.
Software: J.U. and F.S.F.
Visualization: J.U.
Supervision: H.W.R.
Writing—original draft: J.U.
Writing—review & editing: H.W.R. and F.S.F.

Abstract

We model monthly precipitation maxima at 132 stations in Germany for a wide range of durations from 1 min to about 6 d using a duration-dependent Generalized Extreme Value (d-GEV) distribution with monthly varying parameters. This allows for the estimation of both monthly and annual intensity–duration–frequency (IDF) curves: (1) The monthly IDF curves of the summer months exhibit a more rapid decrease of intensity with duration, as well as higher intensities for short durations than the IDF curves for the remaining months of the year. Thus, when short convective extreme events occur, they are very likely to occur in summer everywhere in Germany. In contrast, extreme events with a duration of several hours up to about 1 d are conditionally more likely to occur within a longer period or even spread throughout the whole year, depending on the station. There are major differences within Germany with respect to the months in which long-lasting stratiform extreme events are more likely to occur. At some stations the IDF curves (for a given quantile) for different months intersect. The meteorological interpretation of this intersection is that the season in which a certain extreme event is most likely to occur shifts from summer towards autumn or winter for longer durations. (2) We compare the annual IDF curves resulting from the monthly model with those estimated conventionally, that is, based on modeling annual maxima. We find that adding information in the form of smooth variations during the year leads to a considerable reduction of uncertainties. We additionally observe that at some stations, the annual IDF curves obtained by modeling monthly maxima deviate from the assumption of scale invariance, resulting in a flattening in the slope of the IDF curves for long durations.

6.1 Introduction

Extreme precipitation events can potentially cause significant damage (Barredo, 2009; Davenport et al., 2021; Linnerooth-Bayer and Amendola, 2003), depending on their duration and spatial extent: extreme convective events can lead to flash floods, while long-lasting stratiform precipitation may lead to river flooding. In recent years, floods and landslides following heavy precipitation have become increasingly frequent in many European countries (Bronstert, 2003; Paprotny et al., 2018), and weather conditions favoring the occurrence of heavy rainfall events are expected to further increase due to anthropogenic climate change (Hartmann et al., 2013; Hattermann et al., 2013, and references therein). However, in addition to regional differences, changes in the frequency and intensity of extreme precipitation in Europe have been found to also differ between different storm types, namely convective and stratiform events (Berg et al., 2013), as well as between different seasons (e.g., Kunz et al., 2017; Moberg and Jones, 2005; Łupikasza, 2017, and references therein). Hence, it is critical to research and improve our understanding of the occurrence of extreme precipitation events on different timescales as well as in different seasons in order to detect and interpret changes in seasonality in a consistent way.

The characteristics of extreme precipitation on different timescales can be summarized in terms of intensity–duration–frequency (IDF) curves. These are a standard tool in hydrology for designing hydrological structures and managing water supplies (Durrans, 2010). IDF curves are basically probability distributions for extreme values of precipitation intensity for a range of durations, or more precisely aggregation times. Thus, they provide the relationship between precipitation intensity and duration for selected occurrence frequencies (i.e. exceedance probabilities or return periods). Since differences exist in the storm characteristics of different seasons, it is essential to provide information on precipitation extremes on a seasonal basis. Even though a seasonal resolution may not be relevant for planning or adjusting hydrological structures, it could be beneficial for stakeholders managing water storage. In addition, a seasonal approach allows for a more detailed examination of the underlying mechanisms that influence the IDF relationship, considering that extreme events with different durations may occur in different seasons. However, while studies exist that investigate seasonality in extreme precipitation of a selected duration (e.g., Fischer et al., 2018, 2019; Maraun et al., 2009; Rust et al., 2009) or also in flood frequency (e.g., Durrans et al., 2003; Kochanek et al., 2012; Rottler et al., 2020), there are few studies regarding seasonal IDF curves (Durrans, 2010; Willems, 2000).

Extreme value theory offers several approaches to describe the occurrence probability of extreme events (for an introduction, see Coles, 2001). There are numerous applications of extreme value statistics in hydrology and climatology (e.g., Davison and Gholamrezaee, 2012; Friederichs, 2010; Katz et al., 2002; Lazoglou et al., 2019; Papalexiou and Koutsoyiannis, 2013; Sebille et al., 2017), making use of two commonly accepted concepts: the block maxima approach and the peaks over threshold (POT) approach. For the block maxima approach, the observed time series is divided into blocks of equal length and the probability distribution of the maxima of these blocks is modeled using a Generalized Extreme Value (GEV) distribution. For the peaks over threshold (POT) approach, on the other hand,

the distribution of exceedances above a chosen threshold is modeled using a generalized Pareto distribution (GPD), potentially allowing for the use of more data. However, while a sufficient block size has to be selected for the block maxima approach, the POT approach requires the choice of a suitable threshold. In the context of this study, it would be necessary to choose a threshold that varies seasonally as well as with duration, therefore we consider the block maxima approach to be the more suitable choice.

Since extreme events are by definition rare, the estimation of quantiles (return levels) corresponding to small exceedance probabilities (return periods) is always associated with the problem of limited data. When modeling IDF curves, the limitations of the observations are firstly the spatial coverage and secondly the temporal resolution (Courty et al., 2019). For example, the German Meteorological Service (DWD) operates a relatively dense weather station network, so that many long observation time series exist for daily precipitation measurements. However, fewer stations provide sub-daily measurements and, in addition, considerably shorter time series are available at these stations, since operating instruments with hourly or minute by minute measurement intervals has only been feasible without considerable maintenance for a few decades. The situation is similar in many other countries (see e.g., Dyrrdal et al., 2015; Olsson et al., 2019). The objective in modeling sub-daily extreme precipitation events is therefore to use the available data most efficiently, i.e. to pool the information where possible. Hence, in this study we aim to combine different information on extreme precipitation within one model, namely information on different durations as well as seasonal variations.

In order to assess extreme precipitation observations of different aggregation times simultaneously, it is possible to use a duration-dependent extreme value distribution (Koutsoyiannis et al., 1998; Lehmann et al., 2013; Van de Vyver, 2018; Van de Vyver and Demarée, 2010). In the context of the block maxima approach, a duration-dependent GEV (d-GEV) distribution is derived by implementing empirical dependencies of the GEV parameters on duration. Thus, we are able to directly obtain quantile estimates for all durations within the considered interval while additionally reducing the uncertainties of the estimation by combining information of different durations (Ulrich et al., 2020). To the best of our knowledge, this approach has so far only been used with an annual block size. This means that only the annual maxima of each aggregation time are used, and therefore large amounts of data are neglected for the analysis. However, when modeling daily precipitation sums, monthly block sizes have been shown to be sufficient to model extreme precipitation in the midlatitudes (Coles, 2001; Maraun et al., 2009; Rust et al., 2009). Naturally, it would be possible to model the block maxima separately according to the month of their occurrence, but the choice of a more complex model that explicitly includes the intra-annual variations results in a substantial reduction in the number of parameters that need to be estimated. This can be accomplished by adding smooth periodic functions as covariates for the GEV parameters (Fischer et al., 2018, 2019). Fischer et al. (2018) demonstrated that this approach provides more precise quantile estimates than using an annual block size as it allows for the use of more data.

In this study, we implement monthly covariates analogously for the parameters of the d-GEV distribution. Hence, we model intra-annual variations of extreme precipitation for a

wide range of durations from one 1 min to approximately 6 days at 132 stations in Germany. This not only allows us to estimate and compare IDF curves of different months, but we also expect to obtain more reliable annual IDF curves due to the more efficient use of the available data. Furthermore, we anticipate that accounting for seasonality and reducing uncertainties in parameter estimation will provide a better understanding of the underlying processes. Hence, we expect to gain new insights into the empirical dependencies of GEV parameters on duration, which are in turn relevant for the modeling of annual maxima. This study addresses the following research questions:

- How does the IDF relationship at different stations in Germany evolve throughout the year?
- To what extent do the annual IDF curves based on monthly and annual maxima differ?
- Does explicit modeling of seasonal variations allow allow us to draw conclusions aimed at improving the modeling of annual maxima?

The remainder of this study is organized as follows: In Sect. 6.2, we present the data and methods on which this study is based. We address both the methods used for modeling as well as for comparing the different models. We then present and discuss the respective results regarding our research questions in Sect. 6.3. We close with our conclusions in Sect. 6.4.

6.2 Methods

We aim to model the intra-annual variations of extreme precipitation on different timescales. For this purpose, we use observations with high temporal resolution from stations in Germany. We use a duration-dependent GEV (d-GEV) distribution with monthly covariates to describe the monthly maxima over a range of durations collectively in one model. Thereby, appropriate models for the intra-annual variations of the d-GEV parameters are selected through stepwise forward regression. This approach allows us to examine how the IDF curves vary throughout the year in different areas of Germany. From this seasonal model, we can derive annual IDF curves as well. We compare these annual IDF curves with those resulting from directly modeling the annual maxima via a verification procedure using the quantile skill index. Finally, we verify whether modeling monthly maxima allows for a more precise estimate of the relationships between GEV parameters and duration. Therefore, we model each duration separately using the GEV distribution with monthly covariates. Details of the data as well as all methods involved are described in the following section.

6.2.1 Data

We use precipitation measurements at 132 stations in Germany that provide a temporal resolution of 1 min. Their locations are presented in Fig. 6.1. The majority (129) of these stations are operated by the German Meteorological Service (DWD). The data were

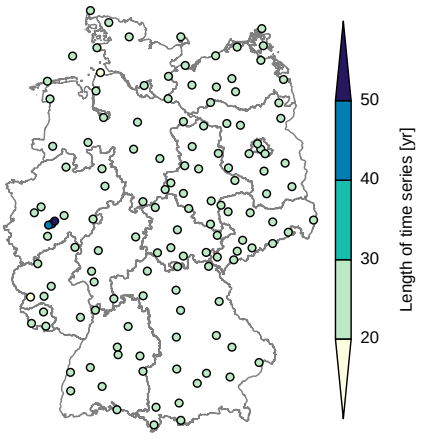


Figure 6.1: Map of Germany with positions of all 132 stations considered. Colors represent the length of the available time series with minute resolution. The longest observation period of 51 years exists for station Bever-Talsperre (dark blue).

obtained via the Climate Data Center (`https://opendata.dwd.de/climate_environment/CDC/ observations_germany/climate`, last access: 5 March 2021). The available time series at these stations range from 19 to 28 years (Fig. 6.1 yellow and green). Additionally we use three stations operated by the Wupperverband (`https://www.wupperverband.de`, last access: 11 June 2021) with time series $\geq$ 44 years (Fig. 6.1 blue). The station Bever-Talsperre with the longest observation period of 51 years is used as example station.

The observations were accumulated to the following durations: $d \in 2^{\{0,1,2,..,13\}}$ min $= \{1, 2, 4, ..., 8192\}$ min, with the longest duration 8192 min $\approx$ 5.7 days, thus, resulting in 14 time series per station. Of each time series, we consider both the monthly and annual maxima. Blocks are excluded from the analysis if they contain more than 10% missing values.

6.2.2 Modeling Annual Maxima of Different Durations

The challenge in modeling extremes is to estimate probabilities of very rare events or those not even observed yet. Here, we apply the block maxima approach which is commonly used for this purpose. It is based on the Fisher–Tippett–Gnedenko theorem, which essentially states that under certain assumptions the probability distribution of block maxima can be modeled by the Generalized Extreme Value (GEV) distribution (Coles, 2001).

More precisely, let $X_1, ...X_n$ be a sequence of n random variables which are independent and identically distributed (iid), with an unknown distribution. We denote the maximum of this sequence as

$$M_n = \max\{X_1, ..., X_n\}. \tag{6.1}$$

In the limit of large block sizes n, the non-exceedance probability can be approximated by

the Generalized Extreme Value (GEV) distribution

$$\Pr\{M_n \leq z\} \approx G(z), \tag{6.2}$$

if for $n \to \infty$ the distribution of properly rescaled M_n converges to a nondegenerate distribution. The GEV distribution

$$G(z; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{6.3}$$

is defined on $\{z : 1 + \xi(z-\mu)/\sigma > 0\}$ and has three parameters: location parameter $-\infty < \mu < \infty$, scale parameter $\sigma > 0$, and a shape parameter $-\infty < \xi < \infty$. Thus, the position and width of the distribution are specified by μ and σ, respectively, whereas ξ determines the right tail behavior, resulting in bounded right tails for $\xi < 0$ and polynomial decay for $\xi > 0$. In the case $\xi = 0$ Eq. (6.3) is interpreted in the limit of $\xi \to 0$, leading to the Gumbel distribution, with an exponentially decaying tail.

The GEV distribution is thus likely to be a well suited model for the distribution of annual precipitation intensity maxima of one selected aggregation duration. In order to model the distribution for different durations simultaneously, Koutsoyiannis et al. (1998) proposed that the empirical relationship between precipitation intensity and duration can be directly used to model the parameters of the GEV distribution depending on duration, which leads to a duration-dependent GEV (d-GEV) distribution $G(z, d; \mu(d), \sigma(d), \xi(d))$. The relationship between precipitation intensity I and duration d for a chosen non-exceedance probability p corresponds to the quantile $q_p(d)$ of the d-GEV distribution:

$$I_p(d) = q_p(d) = \mu(d) - \frac{\sigma(d)}{\xi(d)}\left[1 - \{-\ln(p)\}^{-\xi(d)}\right]; \tag{6.4}$$

hence IDF curves can be estimated in a consistent way (Ulrich et al., 2020). For the empirical dependence of the parameters on duration, we follow the assumptions of Koutsoyiannis et al. (1998):

$$\sigma(d) = \sigma_0\left(\frac{d}{1\,\mathrm{h}} + \theta\right)^{-\eta}, \tag{6.5}$$

$$\mu(d) = \tilde{\mu} \cdot \sigma(d), \tag{6.6}$$

$$\xi(d) = \text{const.}, \tag{6.7}$$

with re-parameterized location parameter $-\infty < \tilde{\mu} < \infty$, scale offset $\sigma_0 > 0$, duration offset $\theta \geq 0$ and duration exponent $0 < \eta \leq 1$. These assumptions are commonly used (Lehmann et al., 2013; Ritschel et al., 2017; Stephenson et al., 2016; Van de Vyver, 2015), however, it may be beneficial to introduce additional parameters (Fauer et al., 2021a; Van de Vyver, 2018). By inserting assumptions (6.5)-(6.7) into Eq. (6.3), we obtain the d-GEV distribution with five parameters

$$G(z, d; \tilde{\mu}, \sigma_0, \xi, \theta, \eta) = \exp\left\{-\left[1 + \xi\left(\frac{z}{\sigma_0(d+\theta)^{-\eta}} - \tilde{\mu}\right)\right]^{-1/\xi}\right\}, \tag{6.8}$$

which constitutes a model for the distribution of annual precipitation maxima for a range of durations.

6.2.3 Modeling Monthly Maxima

According to the Fisher–Tippett–Gnedenko-Theorem, the GEV distribution is an adequate model for block maxima if the block size is sufficiently large. For geophysical applications, such as modeling extreme precipitation, it is common to choose a block size of 1 year, as explicit modeling of seasonality is thereby avoided. However, this results in two major disadvantages: large portions of the data are lost for the analysis if only the annual maxima are used, and the assumption that precipitation events originate from an identical distribution is violated if a distinct intra-annual cycle exists. Therefore, the use of a smaller block size is worth considering. Multiple studies suggest that the GEV distribution is well suited to model monthly block maxima of daily precipitation sums in the midlatitudes (Fischer et al., 2018; Maraun et al., 2009; Rust et al., 2009). Similarly, we use monthly maxima to model extreme precipitation of different durations: either with separate models for each duration using the GEV (Eq. 6.3) or simultaneously by using the d-GEV distribution (Eq. 6.8). Inspection of the quantile-quantile (q-q) plots indicates that the d-GEV distribution is a reasonable approximation for the distribution of monthly maxima at the regarded stations. The q-q plots for station Bever-Talsperre with respect to each month are shown in Fig. D.1.

To account for any form of variability in the GEV model (Eq. 6.3), the GEV parameters $\varphi \in \{\mu, \sigma, \xi\}$ can be modeled as linear functions of covariates x_i within the framework of vector generalized linear models (VGLMs) (Yee and Stephenson, 2007)

$$l^{\varphi}\left(\varphi(x_i)\right) = \varphi_0 + \sum_{i=1}^{I} \beta_i^{\varphi}\, x_i, \tag{6.9}$$

where φ_0 represents the intercept and β_i^{φ} are the regression coefficients. The choice of the parameter specific link function $l^{\varphi}(\cdot)$ can ensure that parameters stay within a predefined range. However, we employ the identity $l^{\varphi}(\varphi) = \varphi$ as link function for all parameters. Following Fischer et al. (2018), the intra-annual variations of the GEV parameters can be modeled as a periodic functions of the day of the year (doy) using a series of harmonic functions with a fundamental period of 1 year:

$$\varphi(\mathrm{doy}) = \varphi_0 + \sum_{j=1}^{J} \left[\alpha_j^{\varphi} \cos\left(\frac{2\pi j \cdot \mathrm{doy}}{365.25}\right) + \beta_j^{\varphi} \sin\left(\frac{2\pi j \cdot \mathrm{doy}}{365.25}\right) \right], \tag{6.10}$$

where J is the maximum order of harmonic functions. To obtain the parameters for each month, Eq. (6.10) is evaluated at the corresponding center days of each month. We model the seasonal variations of the d-GEV distribution in exactly the same way. Essentially, this means that each of the parameters $\varphi_{\text{d-GEV}} \in \{\tilde{\mu}, \sigma_0, \xi, \theta, \eta\}$ can be expressed in the form of Eq. (6.10).

6.2.4 Parameter Estimation

The parameters of the GEV distribution can be estimated from a time series of observed block maxima. For this purpose, we apply the widely used maximum likelihood estimator

(MLE) (Coles, 2001). Thus, the parameters are chosen by optimizing the likelihood

$$\mathcal{L}(\boldsymbol{\phi} \mid \boldsymbol{Z}) = \prod_{n \in N} g(z_n; \boldsymbol{\phi}), \tag{6.11}$$

where the parameter vector $\boldsymbol{\phi} = (\mu, \sigma, \xi)^T$ contains the unknown GEV parameters, the vector $\boldsymbol{Z}$ consists of the observed maxima z_n for different blocks (years/ months) n and $g(z_n; \boldsymbol{\phi})$ is the probability density function of the GEV distribution. This can be applied analogously for the d-GEV distribution:

$$\mathcal{L}(\boldsymbol{\phi} \mid \boldsymbol{Z}) = \prod_{d \in D} \prod_{n \in N} g(z_{n,d}, d; \boldsymbol{\phi}). \tag{6.12}$$

Whereas in this case the parameter vector $\boldsymbol{\phi} = (\tilde{\mu}, \sigma_0, \xi, \theta, \eta)^T$, $\boldsymbol{Z}$ now contains all observed maxima z_n for different blocks (years/ months) n and durations d and $g(z_{n,d}, d; \boldsymbol{\phi})$ is the probability density function of the d-GEV distribution. A benefit of the MLE is that it can be easily extended in the case of using covariates to model the parameters (Eq. 6.10). The parameter vector then contains the parameter intercepts φ_0 and regression coefficients α_j^{φ} and β_j^{φ} for each parameter in the case of both GEV and d-GEV distribution.

Since the logarithm of the likelihood reaches the maximum at the same value, but is easier to calculate, the parameters are estimated by optimizing the log-likelihood numerically

$$\hat{\boldsymbol{\phi}} = \arg\max_{\boldsymbol{\phi}} \{ \ln[\mathcal{L}(\boldsymbol{\phi} \mid \boldsymbol{Z})] \}. \tag{6.13}$$

It is possible to derive the uncertainty of the parameter estimates, i.e., the variance-covariance matrix, via the Fisher information matrix estimated in this process.

Nevertheless, Eqs. (6.11) and (6.12) are only valid if the block maxima are independent of each other. The assumption that maxima of different years or also months are independent is reasonable. However, a dependency exists between the maxima of different durations. Jurado et al. (2020) have shown that accounting for asymptotic dependence between durations yields a modest improvement in the estimation of quantiles of short durations $d \leq 10\,\mathrm{h}$, but comes at the cost of increased model complexity. We therefore decide to neglect the dependence between durations when estimating the d-GEV parameters using Eq. (6.12). Yet, the dependence between durations is taken into account when estimating the uncertainties of the quantiles using the bootstrap method (see Sect. 6.2.6).

6.2.5 Model Selection

To obtain a parsimonious model, we use a selection procedure consisting of two steps: in the first step, we determine for which of the GEV/ d-GEV parameters the modeling of the intra-annual variations is not appropriate and which should therefore remain constant. In the second step, we select which terms of the harmonic series in Eq. (6.10) are actually needed in order to model the nonconstant parameters.

When modeling intra-annual variations of GEV parameters, the shape parameter ξ is often assumed to be constant (Fischer et al., 2018; Maraun et al., 2009; Rust et al.,

2009). This is justified by the fact that ξ controls the tail of the distribution, and thus the estimation of ξ is already associated with large uncertainties. Hence, adding additional coefficients to the estimation of ξ is only reasonable if there are sufficient data available. Fischer et al. (2019) demonstrated that modeling the intra-annual variations of ξ can indeed improve the GEV model. However, their model is able to combine the observations of many stations – due to additional spatial covariates – and therefore the amount of data on which the estimation is based is increased. Since in contrast we employ separate models for each station, we choose the shape parameter to remain constant $\xi(\text{doy}) = \xi_0$. For the parameters $\mu(\text{doy})$ and $\sigma(\text{doy})$, a variation in the form of Eq. (6.10) is adopted.

To be consistent, we also use a constant shape parameter in the d-GEV case. The estimation of the duration offset parameter θ is likewise associated with considerable uncertainties, because it is strongly influenced by the estimation of the parameters η and σ_0. Eq. (6.5) clearly indicates this effect. Therefore, we choose for θ to remain constant $\theta(\text{doy}) = \theta_0$ as well. The parameters $\tilde{\mu}(\text{doy}), \sigma_0(\text{doy})$ and $\eta(\text{doy})$ are allowed to vary periodically throughout the year according to Eq. (6.10).

For the maximum order of the harmonic series in Eq. (6.10), we choose $J = 4$. This results in a maximum of eight regression coefficients α_j^φ and β_j^φ for each nonconstant parameter. Thus, in the GEV case one would obtain one model per duration containing $3 + 2 \cdot 8 = 19$ parameters and in the d-GEV case one model describing all durations simultaneously with $5 + 3 \cdot 8 = 29$ parameters to be estimated. To reduce this number to a level where the model describes the variations sufficiently well without overfitting, we apply a stepwise forward regression. For both the GEV model and the d-GEV model, we use the same methods to select the necessary predictor terms: as model selection criterion we use the cross-validated log-likelihood. For this purpose, the observations are divided into a training set and a test set, and $\ln[\mathcal{L}(\boldsymbol{\phi} \mid \boldsymbol{Z})]$ is computed as in Eqs. (6.11) and (6.12), where the parameter vector $\boldsymbol{\phi}$ is estimated based on the training set and the observations $\boldsymbol{Z}$ originate from the test set. We choose a small number of folds $k = 2$, as recommended for cross-validation with the aim of model selection (Arlot and Celisse, 2010). Thus, the data are divided into two sets, each of which is used once as a training and once as a test set. In the first step of the stepwise regression, we compare all possible models that can result from the addition of the cosine term with $j = 1$ as covariate for the nonconstant parameters. This yields four possible models for the GEV case, listed in Table 6.1, and analogously eight models for the d-GEV case. The model resulting in the maximum cross-validated likelihood is retained for the next model selection step. In this next step we similarly identify the

Table 6.1: All possible models in the first step of the stepwise regression (GEV case) with $\omega = \frac{2\pi}{365.25}$

$\mu(\text{doy})$	$\sigma(\text{doy})$	$\xi(\text{doy})$
μ_0	σ_0	ξ_0
$\mu_0 + \alpha_1^\mu \cos(\omega \cdot \text{doy})$	σ_0	ξ_0
μ_0	$\sigma_0 + \alpha_1^\sigma \cos(\omega \cdot \text{doy})$	ξ_0
$\mu_0 + \alpha_1^\mu \cos(\omega \cdot \text{doy})$	$\sigma_0 + \alpha_1^\sigma \cos(\omega \cdot \text{doy})$	ξ_0

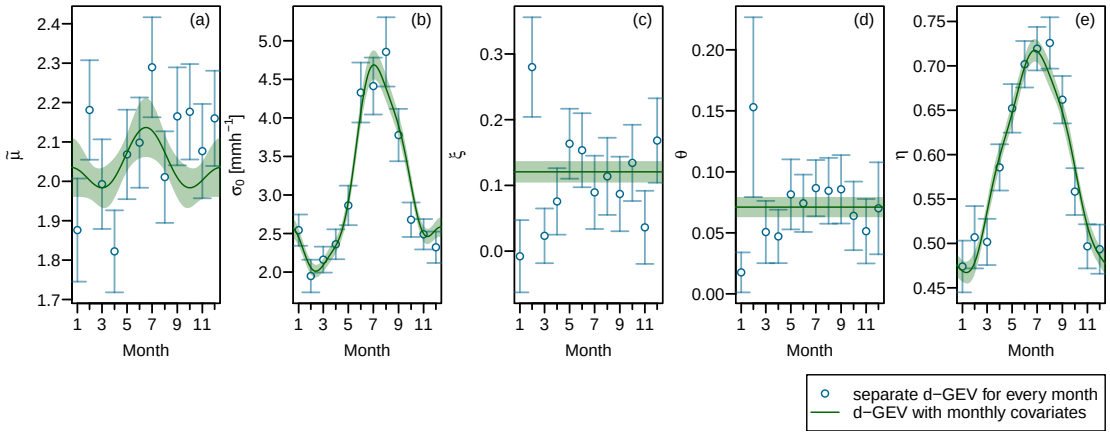


Figure 6.2: Estimated d-GEV parameters $\tilde{\mu}, \sigma_0, \xi, \theta$ and η (a-e) for station Bever-Talsperre: through applying one separate d-GEV model for each month (blue dots) and by modeling all month simultaneously using a d-GEV model with monthly covariates (green lines). The error bars and shaded areas show the 95% confidence intervals obtained via the estimated Fisher information matrix.

nonconstant parameters for which the addition of the sine term with $j = 1$ results in an improvement of the model. We proceed to the maximum order $J = 4$ accordingly.

For the station Bever-Talsperre, the resulting estimated d-GEV parameters are presented in Fig. 6.2. For comparison, the estimated parameters resulting from using a separate model for each month are shown as well. In the case of the station Bever-Talsperre, model selection yields a model with 16 parameters to be estimated. This represents a large parameter reduction compared to using one separate d-GEV model per month with $5 \cdot 12 = 60$ parameters. From Fig. 6.2 we can note that the choice to keep the parameters ξ and θ constant seems to be justified. In addition, the parameters σ_0 and η show a clear variation throughout the year. The estimates of the separate models per month and the d-GEV model with covariates agree well for these two parameters. In the case of the modified location parameter $\tilde{\mu}$, the variations are not as pronounced as for σ_0 and η, so it might be possible to model this parameter as constant as well. However, following Eq. (6.6), setting $\tilde{\mu}(\text{doy}) = \text{const.}$ would enforce the annual cycle of the location parameter $\mu(d)$ and the scale parameter $\sigma(d)$ to be in phase for any fixed duration d. Based on the results of an exploratory analysis, see Supplement D.3, we conclude that the assumption of phase equality of $\mu(d)$ and $\sigma(d)$ for each duration may be too restrictive. Therefore, we decide to allow variations in $\tilde{\mu}(\text{doy})$ throughout the year.

6.2.6 Obtaining IDF Curves

When modeling the annual maxima with the d-GEV distribution according to Eq. (6.8), we can derive IDF curves that correspond to the annual exceedance probabilities using Eq. (6.4). Likewise, when modeling monthly maxima using the d-GEV distribution with monthly covariates (Eq. 6.10), Eq. (6.4) yields separate IDF curves for each month of the

year. These correspond to the probabilities that a certain intensity will not be exceeded within a specific month. To distinguish between these two types of IDF curves, we refer to them as annual and monthly IDF curves, respectively.

However, we can also derive annual IDF curves, i. e. quantiles of the distribution of annual maxima, from modeling the monthly maxima. Assuming the maxima of all months in a year as independent, the non-exceedance probability p of an intensity level $q_{p,d}$ within 1 year is derived from its monthly non-exceedance probabilities as

$$p = \prod_{m=1}^{12} G_m(q_{p,d}; \mu_m, \sigma_m, \xi_m), \tag{6.14}$$

for a fixed duration d. Therefore, to obtain the quantiles of the distribution of annual maxima, we numerically solve Eq. (6.14) for $q_{p,d}$, where $q_{p,d}$ is the quantile corresponding to the exceedance probability $1 - p$, sometimes interpreted as the return level associated with the return period $1/(1 - p)$. We compute $q_{p,d}$ for the entire duration range to yield the annual IDF curves.

We determine the uncertainties of the estimated IDF curves using the ordinary non-parametric bootstrap percentile method (Davison and Hinkley, 1997). In this process, a sample is first created from the data by drawing with replacement. This sample is used to estimate the model parameters, from which a certain intensity quantile dependent on duration $I_p(d)$ is then calculated using Eq. (6.4). By repeating this process $R = 500$ times, we obtain a distribution of intensity quantiles. Finally, the empirical 0.025 and 0.975 quantiles of the bootstrap distribution are used as the lower and upper bounds of the 95% confidence interval for $I_p(d)$. We assume that considering an appropriate sampling strategy, this method accounts for the dependence between maxima of different durations. For this purpose, all maxima from a particular year are jointly sampled, thus we obtain a sample of n years matching the length of the station time series. Fauer et al. (2021a) demonstrated in a sampling experiment that the coverage of the 95% confidence intervals obtained in this manner stays adequate, even when the dependence of the maxima of different durations is increased.

To visualize the differences between the annual IDF curves resulting from the different models, it can be useful to compare the parameters of the respective distributions of annual maxima. Unfortunately, these are only directly available when modeling the annual maxima. However, we can assume that the distribution of the annual maxima resulting from modeling the monthly maxima is for each duration again a GEV distribution, due to its max stability property. Thus, we estimate the GEV parameters of the distribution of annual maxima by firstly using Eq. (6.14) to estimate the quantiles in the range $p \in [0, 1]$ and through inversion obtaining $p(q_{p,d})$. We then fit the GEV distribution to $p(q_{p,d})$ using the nonlinear least-squares method to estimate μ_d, σ_d and ξ_d

$$p(q_{p,d}) \sim \exp\left\{-\left[1 + \xi_d \left(\frac{q_{p,d} - \mu_d}{\sigma_d}\right)\right]^{-1/\xi_d}\right\}. \tag{6.15}$$

The uncertainties of the estimates of μ_d, σ_d and ξ_d are likewise determined using the described bootstrap method.

6.2.7 Verification

We apply a verification procedure in order to assess the estimated quantiles, i.e. IDF curves. At a given station, we aim to compare the annual IDF curves obtained by modeling the monthly maxima with those obtained by modeling the annual maxima. We model the monthly maxima using the d-GEV distribution with monthly covariates according to Eqs. (6.8) and (6.10). We abbreviate this model as monthly d-GEV in the following. For modeling the annual maxima, we use the d-GEV distribution (Eq. 6.8) and abbreviate this model as annual d-GEV.

To provide a detailed analysis we follow Ulrich et al. (2020), who suggest a verification strategy that allows the estimated quantiles for each duration d and probability p to be examined separately. The approach is based on the comparison of the observations o_n with the modeled quantile q_p via the quantile score (QS) (Bentzien and Friederichs, 2014):

$$\mathrm{QS}(p) = \frac{1}{N}\sum_{n=1}^{N}\rho_p(o_n - q_p), \quad \text{where } \rho_p(u) = \begin{cases} pu & ,u \geq 0 \\ (p-1)u & ,u < 0, \end{cases} \tag{6.16}$$

where the check-loss function $\rho_p(u)$ is evaluated at $u = o_n - q_p$. To obtain the out-of-sample performance of the model, QS is evaluated in a cross-validation setting (Wilks, 2011). For this purpose, we split the available time series at a station into n_y sets, corresponding to the length of the time series in years, by removing the maxima of all durations of a specific year y for each set. The model parameters and thus the quantile $q_{p,d}$ are estimated based on the remaining data. The quantile score for a cross-validation set is calculated from $q_{p,d}$ and the respective omitted observed annual maximum $o_{d,y}$. Therefore, the cross-validated quantile score results in:

$$\mathrm{QS}_{\mathrm{cv}}(p,d) = \frac{1}{n_y}\sum_{y\in Y}\rho_p(o_{d,y} - q_{p,d}). \tag{6.17}$$

To compare the score of the monthly d-GEV $\mathrm{QS}^m_{\mathrm{cv}}$ with that of the annual d-GEV $\mathrm{QS}^a_{\mathrm{cv}}$, we use the quantile skill index (QSI) (Ulrich et al., 2020) that is based on the quantile skill score (Wilks, 2011):

$$\mathrm{QSI}(p,d) = \begin{cases} 1 - \mathrm{QS}^m_{\mathrm{cv}}(p,d)/\mathrm{QS}^a_{\mathrm{cv}}(p,d) & ,\mathrm{QS}^m_{\mathrm{cv}}(p,d) \leq \mathrm{QS}^a_{\mathrm{cv}}(p,d) \\ -1 + \mathrm{QS}^a_{\mathrm{cv}}(p,d)/\mathrm{QS}^m_{\mathrm{cv}}(p,d) & ,\mathrm{QS}^m_{\mathrm{cv}}(p,d) > \mathrm{QS}^a_{\mathrm{cv}}(p,d)\,. \end{cases} \tag{6.18}$$

Therefore, $\mathrm{QSI} \in [-1, 1]$, where negative values indicate a superior performance of the annual d-GEV model, whereas positive values indicate a superior performance of the monthly d-GEV model.

6.3 Results and Discussion

We first show the results for the monthly IDF curves at the station Bever-Talsperre. Based on the probability that the annual p quantile is exceeded within a certain month, we

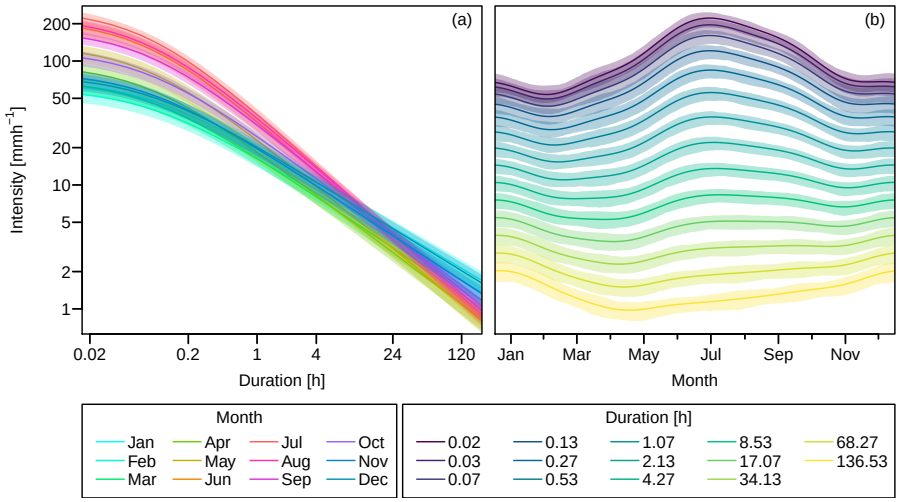


Figure 6.3: The 0.9 quantiles for station Bever-Talsperre for each month (a) and for various durations (b). The durations shown correspond to the durations $d \in 2^{\{0,1,2,\ldots,13\}}$ min discussed in Sect. 6.2.1. Shaded areas represent the 95% confidence intervals obtained via the bootstrap method.

investigate the seasonal variations of the IDF relationship across Germany. We present the results in detail for six selected stations. Furthermore, we examine the annual IDF curves resulting from modeling monthly maxima and compare them to those obtained from modeling annual maxima. We present the resulting annual IDF curves together with the verification results for three example stations. Since the annual IDF curves derived from the monthly maxima deviate from our original assumptions about the duration dependence, we finally investigate the dependence of the estimated GEV parameters on duration using annual and monthly maxima, respectively. We focus in detail on the shape parameter.

6.3.1 Intra-Annual Variations

We obtain quantile estimates, i.e. IDF curves, for each month using the d-GEV distribution (Eq. 6.8) with monthly covariates (Eq. 6.10), through Eq. (6.4). In the case of the monthly IDF curves, p indicates the probability that the value of the p quantile will not be exceeded within a given month. The 0.9 quantile for each month dependent on duration is shown in Fig. 6.3 (a) for the station Bever-Talsperre. The IDF curves exhibit a steeper slope in the summer months (pink) than in the autumn and winter months (blue). This is related to the duration exponent η, which exhibits higher values in summer than in winter (Fig. 6.2 (e)). This result matches the findings of Willems (2000) who determined the rate of exponential decline β of intensity with duration (equivalent to η) for different seasons and likewise found that β exhibits the smallest values in the winter season. For short durations $d \leq 1\,\mathrm{h}$ the intensities reach their maximum in the summer months and their minimum in the winter months. This corresponds to the scale offset parameter σ_0, which similarly peaks in summer (Fig. 6.2 (b)). In contrast, the intensity maximum of long durations $d \geq 24\,\mathrm{h}$ occurs in winter and the minimum in spring and summer, since the curves for different

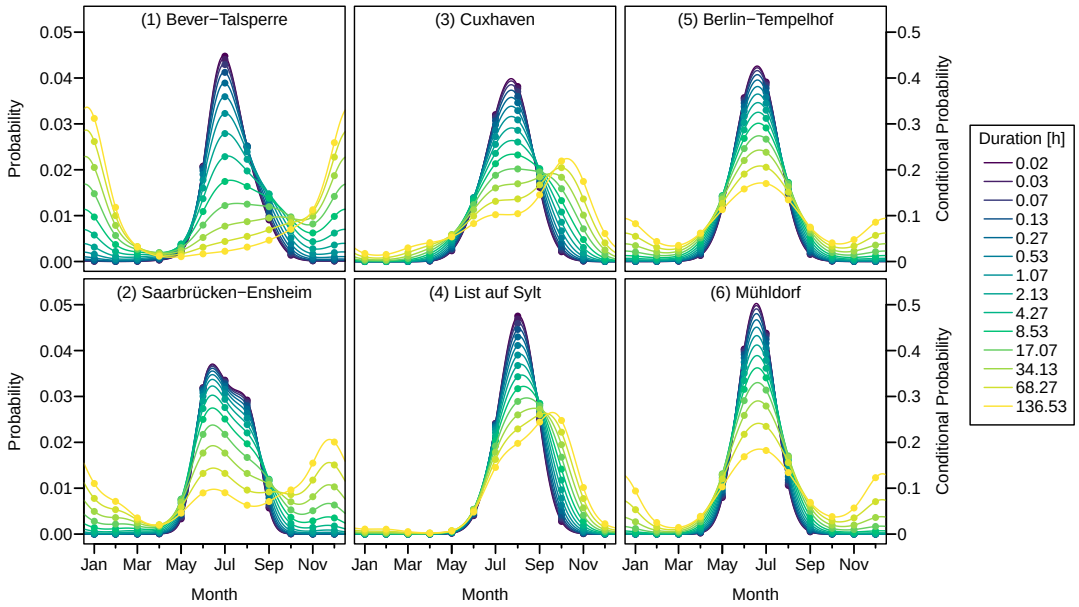


Figure 6.4: Probability of the annual 0.9 quantile $q_{0.9,d}$ being exceeded in a given month. Dots indicate the probability that $q_{0.9,d}$ is exceeded in a certain month, while each point in a line can be interpreted as the probability that $q_{0.9,d}$ is exceeded within the surrounding block of 30 days. To illustrate the interpretation: the product of all probability values presented as dots for a chosen duration results in the annual exeedance probability $1 - p = 0.1$. Therefore, dividing the probability values by 0.1 yields the conditional probability of exceeding $q_{0.9,d}$ in a given month. For reasons of visual clarity, confidence intervals are not presented. Station names are listed at the top of each plot, while the numbers indicate their positions in Fig. 6.5.

months intersect at $d \approx 8\,\mathrm{h}$. The annual variation of the intensity for the different durations is presented in Fig. 6.3 (b). It is evident that the intensity maximum shifts from summer for the short durations (purple/ blue) through autumn into winter for the long durations (light green/ yellow). Here, only the quantiles of the monthly distributions for $p = 0.9$ are shown. However, the monthly quantiles for other probability values exhibit the same behavior. The intersection of the IDF curves and the resulting change in seasonality for different durations is not generally present at all stations in the investigated duration range. However, since different duration exponents of the curves in summer and winter occur at all stations, we suspect that at all stations, given sufficiently long durations, the intensity maximum moves into winter eventually.

Due to the exponential decrease of intensity with duration, a comparison of the p quantiles of different durations is only possible on a logarithmic intensity scale. However, the interpretation of a logarithmic axis is often difficult. Therefore, in addition to the monthly 0.9 quantiles presented in Fig. 6.3, we will also consider the probabilities that the annual 0.9 quantile is exceeded within a given month. To do this, we first use Eq. (6.14) to calculate the annual quantiles $q_{p,d}$ (return values) from the monthly non-exceedance probabilities. Based on this, we calculate the probability that $q_{p,d}$ will be exceeded within

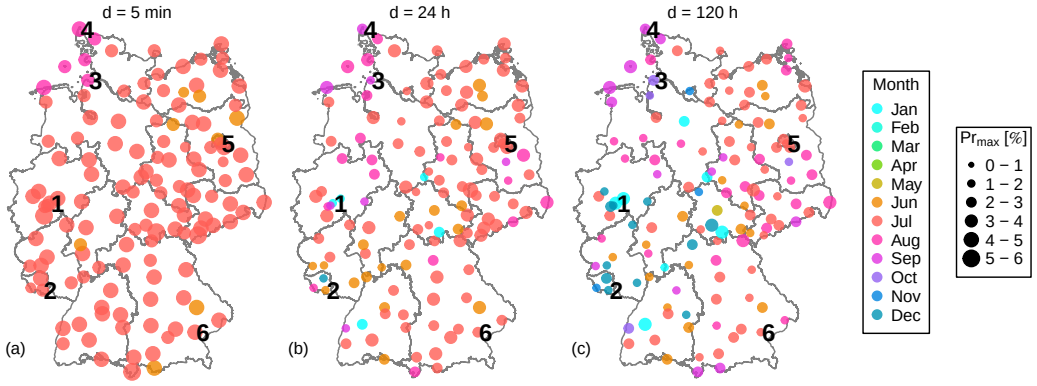


Figure 6.5: Maximum probability of the annual 0.9 quantile $q_{p,d}$ being exceeded in 1 month and months when the maximum occurs at each of the considered stations for three different durations. Dividing the probability values by the annual exceedance probability $1 - p = 0.1$ yields the conditional probability. Therefore a value of $\mathrm{Pr}_{\max} = 5\%$ can be interpreted as 50% of the exceedances of $q_{p,d}$ occurring in a single month.
Numbers indicate the locations of example stations, presented in more detail in Fig. 6.4.

a given month. Fig. 6.4 (upper left) presents the probabilities that the annual 0.9 quantile $q_{0.9,d}$ (10-year return value) for different durations will be exceeded within a given month at the Bever station. This depiction is a useful complement to Fig. 6.3, since the probabilities for different durations vary on a linear scale, unlike the intensities. The monthly exceedance probability for short durations $d < 1\,\mathrm{h}$ (purple/ blue) exhibits a sharp peak with a maximum in July. The probability that $q_{0.9,<1\,\mathrm{h}}$ is exceeded in the months November to April is approximately zero. This means that extreme events of short duration, i.e. caused by convective precipitation cells, are likely to occur in summer while the probability of these events occurring in the months October to April is very small. This is consistent with the results for one station in Belgium (Willems, 2000). In the transition to longer durations, the probability decreases in July, while a second maximum occurs in December to January. For durations of about 8 h to 17 h, this results in an extended period of time ranging from June to February, during which the probability shows similarly elevated values. For the long durations $d > 48\,\mathrm{h}$ (light green/ yellow), the probability again has one clear maximum, which occurs in December to January. The probability of $q_{0.9,>48\,\mathrm{h}}$ being exceeded in the months April to June is relatively low in this case. Therefore, these long-lasting extreme events, i.e. frontal events, are more likely to occur in late autumn or winter.

To investigate the intra-annual variations across Germany, we calculate the probability that the annual p quantile is exceeded in a given month for each station. We present the results for $p = 0.9$ in Fig. 6.5, whereby a different choice of p yields very similar results. We summarize the information on a map by indicating the maximum probability (size of the dots), as well as the month in which the maximum occurs (color of the dots). To a certain extent, the maximum probability provides information about the shape of the curve: A high maximum probability is associated with a narrow probability peak. This implies that the probabilities in the remaining months of the year are comparatively small.

In contrast, a small maximum probability suggests that there are other months in the year with similar probability values.

From Fig. 6.5 (a) it is evident that for short durations the probability peaks in summer at every station, specifically in July at most stations. There are also some stations where the maximum occurs in June or August. Noticeably, the stations where the maximum is in August are all located on the North Sea coast. Furthermore, at most stations the maximum probability is greater than 3%, which indicates a narrow probability peak in summer for short durations. When comparing the maximum probabilities with those at $d = 24\,\mathrm{h}$ (Fig. 6.5 (b)), we find that the maximum probability decreases considerably at almost all stations, thus, broadening the time window within which the annual 0.9 quantile is more likely to be exceeded. At most stations, the maximum occurs in the period from June to September. At six stations, however, the maximum is reached in December or January. These stations are all located at higher altitudes (Mittelgebirge). Figure 6.5 (c) for $d = 120\,\mathrm{h}$ reveals a comparatively heterogeneous spatial distribution, both in terms of the maximum probability and the months in which the maximum occurs. At most stations in Germany the maximum still lies between June and September, however, the number of stations with maximum probability between November and January is considerably increased in the west of Germany, especially in higher regions.

Regarding the month with the highest probability in the case of long durations, we derive a rough division of the stations within Germany into three types. For each type we choose two stations, for which we present the probabilities in detail in Fig. 6.4. We separate the stations into those with maximum in late autumn and winter (1,2), between September and October (3,4), and in summer (5,6). The locations of the stations are indicated by numbers in Fig. 6.5.

Station (1) is Bever-Talsperre, which has already been discussed in detail. Similarly to (1), the maximum at the station Saarbrücken-Ensheim (2) also shifts from summer for short durations into late autumn to winter for long durations. The stations differ, however, insofar as the maximum probability for short durations at station (2) in July varies only slightly from the probability in the months of June and September. In addition, the probability for long durations remains rather high in summer, in contrast to station (1). Stations that show a similar shift of the maximum exceedance probability, from summer for short durations to late autumn or winter for long durations, are located exclusively in the western half of Germany and also occur mostly at higher altitudes. The exception are two stations in northern Germany. The location of these stations coincides well with the results of Fischer et al. (2018). They explain the increased occurrence of longer-lasting extreme events in these regions in late autumn and winter with the stronger westerly winds during these months, which cause particularly high precipitation amounts on the windward sides of the mountain chains (Mittelgebirge) due to the forced uplift of the air. However, this study is based on daily precipitation sums. The fact that when modeling several durations simultaneously, the seasonal variations are observed at longer durations than when modeling a single duration might be due to the smoothing of the seasonal signal.

As examples for stations where the maximum occurs in September or October, for long durations, we present the monthly exceedance probabilities for Cuxhaven (3) and List auf

Sylt (4). At both stations the width of the probability peak increases for long durations while its maximum shifts. This shift is more pronounced at station (3). The probabilities in the interval between December and May are relatively low at both stations for all durations. Some stations of this type are located scattered throughout Germany. However, a clear cluster of these stations exists on the North Sea coast. This group is also characterized by extreme convective precipitation events occurring most likely in August, which could be related to the water temperature in this region reaching its maximum during this month. Accordingly, a possible explanation for the high probability of long-lasting heavy precipitation in the following months might be that extratropical cyclones transport air, which was warmed over the North Sea and thus features a high water content, into this region.

Examples of stations where the probability maximum for all durations occurs in summer are Berlin-Tempelhof (5) and Mühldorf (6). An essential difference to the stations (3) and (4) is that in this case a second maximum occurs in winter for long durations. At station (6) this second maximum is even almost as high as the one in summer for $d > 120\,\mathrm{h}$. Stations of this type occur everywhere in Germany, but are the prominent station type in the eastern half of Germany. The example stations (5) and (6) show a very distinct behavior for the probabilities with increasing duration. At most of the other stations of this type the signal for longer durations is less clear, as sometimes several maxima occur, or the summer maximum might be shifted by 1 or 2 months. However, the common characteristics of these stations continue to be the maximum for all durations occurring in the period between May and October and the probability for long durations showing similarly increased values throughout several months of the year.

The monthly exceedance probability is a useful indicator of the months from which the annual maxima of different durations originate. For all stations, the peak of the probability is relatively narrow for short durations, with the maximum in summer. The probability that the annual 0.9 quantile occurs in one of the other seasons is negligible. This fact contradicts the assumption of the block maxima approach that precipitation intensities are identically distributed within the block of 1 year. In other words, a block size of 1 year for short durations results in a much smaller effective block size of about 4 to 6 months. With respect to longer durations, the stations differ greatly, but it can be generally stated that the effective block size increases for long durations. Thus, the annual maxima at a station for different durations originate from effective blocks of different sizes, which might even be in different seasons, depending on the station's location. This effect is further emphasized in Fig. D.2. Modeling the monthly maxima, on the other hand, avoids this problem. Therefore, in the following section we compare the annual IDF curves derived from annual maxima with those derived from monthly maxima.

6.3.2 Annual IDF Curves

We obtain the annual IDF curves and their confidence intervals from modeling annual and monthly maxima, using the respective methods described in Sect. 6.2.6. We compare the estimated quantiles from both models along with the QSI described in Sect. 6.2.7. Figure 6.6 presents the IDF curves (lower panels) together with the QSI (upper panels) for

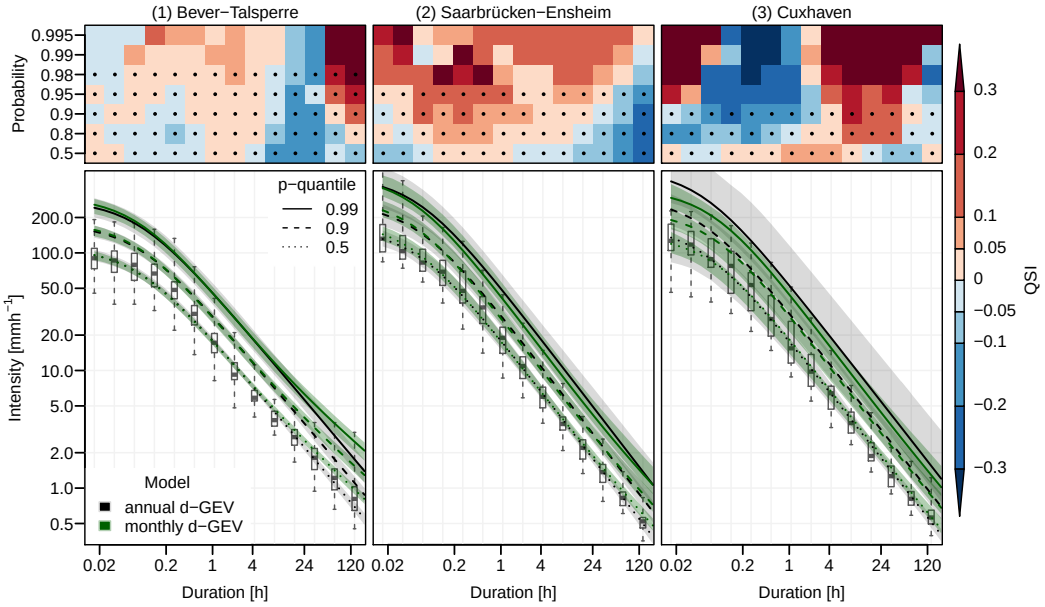


Figure 6.6: Annual IDF curves for three example stations estimated via two models: modeling annual maxima with the d-GEV distribution and modeling monthly maxima using the d-GEV distribution with monthly covariates. The shaded areas represent the respective 95% confidence intervals. The distributions of the observed annual maxima are shown as box-and-whisker plots, where the whiskers cover the complete data range. In the upper panels the corresponding QSI values are presented, indicating the comparison of the models' performances, where positive values indicate an increase in the skill of the monthly d-GEV model compared to the annual d-GEV model.

three example stations. In addition to the station Bever-Talsperre (1), with the longest time series, we present the results for the stations Saarbrücken-Ensheim (2) and Cuxhaven (3), since these three stations cover a broad spectrum in terms of differences between the quantile estimates obtained by both models as well as their uncertainties. The QSI is used to compare the performance of both models. Positive values (red) indicate an increase in the skill of the monthly d-GEV model compared to the annual d-GEV model, while negative values (blue) indicate that the annual model is superior. The result of the QSI may be less reliable if the length of the time series T is shorter than the period corresponding to the non-exceedance probability $p = 1 - 1/T$ being verified. Therefore, we indicate the length of the time series available for verification by dots in the upper panels in Fig. 6.6.

For the station Bever-Talsperre, the IDF curves resulting from the two different models are almost identical over a long duration range $d < 8\,\mathrm{h}$. Therefore, in this duration range, the models' performances differ only marginally, indicated by $|\mathrm{QSI}| \leq 0.05$ for almost all probabilities. For high probabilities $p \geq 0.98$, the QSI suggests a slightly better performance of the monthly model for durations $4\,\mathrm{min} \leq d \lesssim 2\,\mathrm{h}$. At $d \approx 8\,\mathrm{h}$ the IDF curves of the monthly model start to deviate from those of the annual model. More precisely, the IDF curves of the monthly model no longer exhibit a power law behavior for $d > 8\,\mathrm{h}$ but decrease

more gradually. Due to the larger differences in the IDF curves, the models performances vary considerably ($|\mathrm{QSI}| > 0.05$) in this range. However, the sign of the QSI differs for $d \lesssim 34\,\mathrm{h}$ and $d \gtrsim 68\,\mathrm{h}$: for $d \lesssim 34\,\mathrm{h}$, the data are better represented by the annual model, as indicated by the negative values of the QSI in this range. For $d \gtrsim 68\,\mathrm{h}$ and $p \geq 0.95$, however, the monthly model is a considerable improvement over the annual model. This is evident both from the strongly positive QSI values in this range as well as directly from the data, shown as box-and-whisker-plots, as the maximum of the observations extends above the modeled 0.99 quantile. The models do not differ much with respect to the width of the 95% confidence intervals.

At station Saarbrücken-Ensheim, the differences in the IDF curves of the two models are more pronounced throughout the entire duration range: the estimated quantiles of the monthly model are higher for very small and very large durations, but lower in the range $8\,\mathrm{min} \lesssim d \lesssim 48\,\mathrm{h}$ than those of the annual model. Thus, again the monthly model does not comply with a power law. The QSI indicates that the monthly model is mostly an improvement, except for smaller probabilities at longer durations. Since this station provides a shorter time series than station Bever-Talsperre, i.e. only 24 years, the 95% confidence intervals of the annual model are wider, especially for short durations and high probabilities. This indicates that the monthly model benefits from utilizing more data regarding the uncertainties.

This appears even more prominent at the Cuxhaven station. Since the high-resolution time series at this station covers only 19 years, the uncertainties of the annual model are considerably wider than those of the monthly model. The estimates of the monthly model for the 0.9 quantile and the 0.99 quantile are below the respective estimates of the annual model. The estimates for the 0.5 quantile differ only slightly. The quantiles of the monthly model roughly parallel those of the annual model for longer durations. Thus, the monthly model does not deviate essentially from a power law at this station. The QSI does not provide a clear indication regarding which model better represents the data but fluctuates between positive and negative values. This seems to be in agreement with the observations. The spread of the boxes and whiskers first increases and then decreases over duration. As a result, in the duration ranges with narrower box-and-whisker plots, the monthly model better represents the data, especially for higher quantiles, while the annual model is more suitable particularly for $8\,\mathrm{min} \leq d \lesssim 1\,\mathrm{h}$, where the boxes and whiskers are rather broad.

Overall, we find that the differences between the annual and the monthly model are very heterogeneous for individual stations. However, two general statements can be made:

1. Modeling monthly maxima provides a clear improvement in terms of the quantile estimates' uncertainties, especially for stations with short observational time series.

2. Although a power-law behavior for long durations is assumed for the monthly IDF relations, the resulting annual IDF curves can deviate from this behavior and are therefore more flexible.

This deviation from a simple power law behavior for long durations is consistent with the findings of Willems (2000), who observes a decrease in the rate of exponential decline of intensity with increasing duration. We observe this decrease to be particularly pronounced

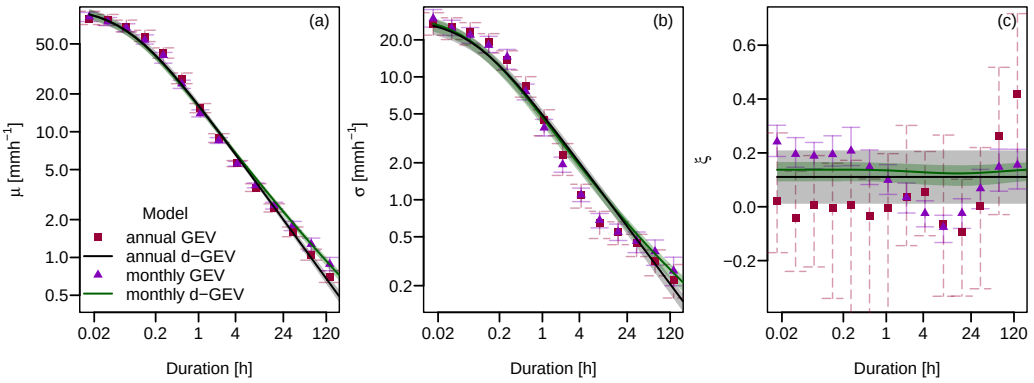


Figure 6.7: Annual GEV parameters μ, σ and ξ for station Bever-Talsperre estimated via four different models along with bootstrapped 95% confidence intervals shown as error bars and shaded areas.

at the station Bever-Talsperre, where we also find a clear shift of the seasons in which extreme events of different durations occur. We therefore suspect that the lower slope of the annual IDF curves at long durations is related to this shift. Since we aim to better understand this deviation from the original assumptions in Eqs. (6.5)-(6.7), in the following section we examine the relationships between the GEV parameters and duration that follow from modeling monthly maxima.

In terms of model performance, we likewise cannot draw general conclusions. At some stations, such as Cuxhaven, we find that modeling monthly maxima improves the estimates of the annual IDF curves for almost all probabilities and durations. However, at many stations the improvement in the estimation is limited to a selected range of probabilities and durations, and there are also stations at which the estimated quantiles of the monthly model are always worse than those of the annual model. Since the objective of this study is to model the seasonal variations at all stations by applying a uniform framework, the model selection was performed identically for all stations, e.g. choosing $\theta(\text{doy}) = \text{const.}$ and $\xi(\text{doy}) = \text{const.}$. This results in varying quality of representation of the parameters at different stations. We expect that parameter estimation for rare events should generally improve, as the introduction of smooth variation during the year allows for the inclusion of additional information. Therefore we assume that with more focus on estimating the IDF curves of a single station, and thus a more targeted choice of model selection approach and initial conditions, the model performance of the monthly model at this station can be considerably improved.

6.3.3 Duration Dependence

To model the IDF relationship, we have so far assumed that the GEV parameters depend on duration according to Eqs. (6.5)-(6.7). This results in a power law behavior, or so-called simple scaling, of intensity with duration except for short durations $d < 1\,\text{h}$. The curvature of the IDF curves for short durations (in a double-logarithmic plot) is controlled by the

parameter θ. Fig. 6.3 (a) illustrates that the IDF curves for each month follow this imposed pattern. However, the resulting annual IDF curves, shown in Fig. 6.6 in green, deviate from this behavior and thus from the assumptions in Eqs. (6.5)-(6.7). To investigate this deviation in more detail, we estimate the annual GEV parameters resulting from modeling monthly maxima, using nonlinear regression according to Eq. (6.15) separately for each duration. We use the obtained parameters only to compare them with those estimated directly from modeling the annual maxima. They are not intended as a basis of any further analysis. We compare the following four models:

- modeling annual maxima using
  - a separate GEV distribution for each duration (annual GEV)
  - one d-GEV distribution (annual d-GEV)
- modeling monthly maxima using
  - a separate GEV distribution with monthly covariates for each duration (monthly GEV)
  - one d-GEV distribution with monthly covariates (monthly d-GEV).

The annual GEV parameters μ, σ and ξ estimated via these models are presented depending on duration for the station Bever-Talsperre in Fig. 6.7 including their respective 95% confidence intervals. Although the uncertainties of the parameters estimated directly from the annual maxima can be derived using the Fisher information matrix, for comparability, the uncertainties of the parameter estimates are obtained using the bootstrap method described in Sect. 6.2.6 for all models.

For the location parameter μ (Fig. 6.7 (a)), the estimates of both annual models (red squares and black line) agree well, i.e. μ follows a power-law for durations $d \geq 1\,\mathrm{h}$, while the curve decreases more gradually for shorter durations. The estimates of the monthly models (purple triangles and green line) are consistent with those of the annual models for durations $d \lesssim 24\,\mathrm{h}$, however, both monthly models agree on a slower decline of μ for longer durations $d \gtrsim 24\,\mathrm{h}$ and thus a deviation from simple scaling. We observe quite a similar behavior of the model estimates for the scale parameter σ (Fig. 6.7 (b)): the estimates of all models agree relatively well for short durations $d \lesssim 1\,\mathrm{h}$ and both monthly models show an upwards deviation from simple scaling for longer durations $d \gtrsim 24\,\mathrm{h}$. However, the estimates of both GEV models (squares and triangles) deviate noticeably from simple scaling towards smaller values in the range $1\,\mathrm{h} \lesssim d \lesssim 24\,\mathrm{h}$. Regarding the uncertainties for the estimates of μ and σ, the annual GEV model (red) is associated with the largest uncertainties. By considering more data, the parameters can be estimated more accurately using the monthly GEV model (purple). Similarly, the annual d-GEV model (black) exhibits considerably smaller uncertainties than the annual GEV model because here the addition of data from other aggregation levels leads to a more confident estimate. Consequently, the joint use of data from all months and durations in the monthly d-GEV model results in the smallest uncertainties regarding parameter estimation. The estimates for μ and σ obtained from all four models agree relatively well at most of the other stations

(not presented). Similar to the station Bever-Talsperre, at several stations the estimates of the monthly models show an upward deviation from simple scaling for long durations. The estimates are associated with considerably larger uncertainties for most other stations, where shorter time series are available.

The shape parameter ξ determines the tail of the distribution and its estimation is therefore subject to relatively large uncertainties depending on the length of the time series. This is similarly the case for the station Bever-Talsperre (Fig. 6.7 (c)) with a time series of 51 years, where the estimates of ξ based on annual maxima of a single duration (red squares) vary in the range 0.09 to 0.4, but since the 95% confidence intervals are very broad, it is challenging to derive a relationship between ξ and duration. For the d-GEV we therefore assume a constant shape parameter (Eq. 6.7), which for the annual d-GEV model (black line) is estimated to be 0.11. Compared to the separate estimates of the annual GEV for each duration (red squares), this value appears reasonable. Additionally, the uncertainties decrease significantly when modeling the annual maxima of all durations simultaneously. For the monthly d-GEV, ξ is also assumed to remain constant over duration for each month. In addition, we hypothesize that ξ does not change throughout the year either, as described in Sect. 6.2.5. Interestingly, $\xi(d)$ resulting from the monthly d-GEV model (green line) nevertheless deviates slightly from being constant. The estimated values are slightly higher than those of the annual d-GEV model and vary between 0.12 and 0.14, with the minimum at about $d \approx 17\,\mathrm{h}$. Similarly to μ and σ, the uncertainties of the monthly d-GEV model (green) for the estimation of ξ are even smaller than those of the annual d-GEV model (black). When we model the monthly maxima for each duration separately, we benefit from being able to estimate the shape parameter with smaller uncertainties without constraining the relation between ξ and duration. Thus, we observe a distinct variation of ξ estimated via the monthly GEV model (purple triangles) over duration. The estimated values exhibit a minimum, similar to those of the monthly d-GEV model (green line), at $d \approx 8\,\mathrm{h}$ and vary in the range -0.08 to 0.24. To evaluate this result, we can visually examine the distribution of annual maxima for different durations, seen in Fig. D.2 and in Fig. 6.6 (bottom left) as box-and-whisker plots. It is noticeable that the observations for $d \approx 8\,h$ actually cover the smallest range, thus agreeing with the results for the shape parameter.

Since the shape parameter estimates of the four different models vary substantially at the individual stations, we summarize the information for all stations in Fig. 6.8. For each model, the distribution of the estimated value over all stations is plotted with respect to duration. We observe that ξ estimated using the annual GEV model (red) varies widely among stations. The model seems to only provide estimates in a reasonable range at 75% of the stations based on the small amount of data. The median of all stations appears to oscillate around a constant value. For the annual GEV model (black), the range in which the estimates vary among stations is narrower, but implausibly high values still occur at some stations. The median is consistent with the median of the annual GEV model (red). In comparison, the range in which the estimated values of the monthly d-GEV model (green) vary is much smaller and the values are within a reasonable range at all stations. Although in this model ξ is assumed to be constant over duration and months, we see a subtle decrease in the median with duration. Likewise, the variation for

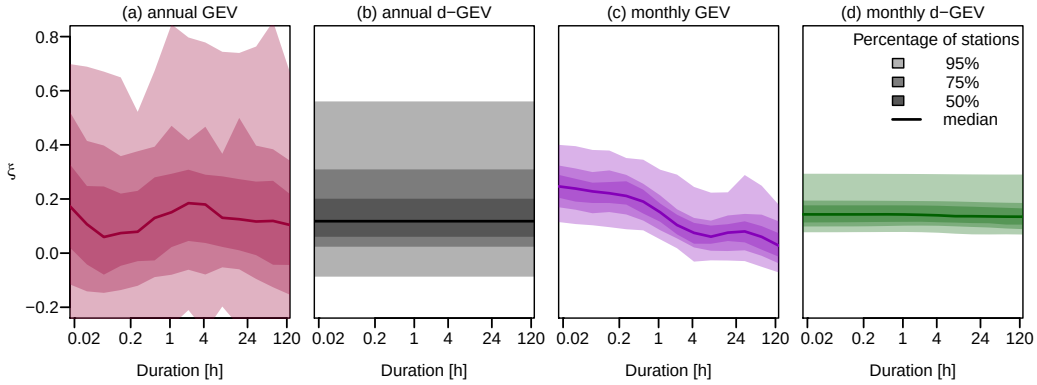


Figure 6.8: Distribution of shape parameter ξ estimated using four different models (columns) at all 132 stations. The shaded areas present the percentage of stations for which the estimated value lies within the respective range while the line depicts the median value.

ξ between individual stations is smaller for the monthly GEV model (purple) than for either of the annual models (red and black). Additionally, we can see a clear decrease of the shape parameter with duration for this model, since in this case no relation between ξ and duration is predefined.

To summarize, we find that modeling the monthly maxima allows new conclusions to be drawn about the behavior of the parameters of the distribution of annual maxima depending on duration. Instead of using the more complex modeling of monthly maxima to estimate annual IDF curves, one might also try to implement the resulting characteristics directly into the model for annual maxima. We find that for some stations the location μ and scale σ parameters deviate from the assumption of simple scaling toward higher values for long durations. Fauer et al. (2021a) showed that this behavior of the parameters can be modeled by an additional parameter τ, called intensity offset. They report that the addition of this parameter for the stations of the Wupper catchment, in which the example station Bever-Talsperre is located, leads on average to an improved estimation of the annual IDF curves for medium to long durations. Regarding the shape parameter ξ, we observe that the reduced uncertainties in the estimation of ξ, resulting from modeling monthly maxima, allow for further investigation of the dependence of ξ on duration. We find that ξ decreases with duration when taking the average of the investigated stations in Germany, reaching values around zero for most stations at long durations. We believe that this finding provides a good basis to explore a potentially more suitable formulation of $\xi(d)$ in future studies. We could imagine that the explicit modeling of this decrease of ξ yields similar results to assuming a different duration exponent for the parameters μ and σ of the d-GEV, so called multiscaling (Gupta and Waymire, 1990; Van de Vyver, 2018). Possibly the latter implementation could be beneficial, since the estimation of these parameters is associated with less uncertainty than that of ξ.

6.4 Conclusions

This study focuses on modeling the intra-annual variations of extreme precipitation on different timescales. For this purpose, we employ a duration-dependent Generalized Extreme Value (d-GEV) distribution with monthly covariates. Using this approach allows for the following:

- investigation of seasonal variations in the intensity–duration–frequency (IDF) relationship,
- the obtaining of more reliable estimates for the annual IDF curves by utilizing information on extreme events more efficiently,
- a better understanding of the underlying processes, i.e., the dependence of the parameters on the duration.

Regarding the seasonal variations, we find that everywhere in Germany the short convective extreme events are most likely to occur in the summer months, whereas there are regional differences for the seasonality of long-lasting stratiform extreme events. Our findings will allow future studies to identify meaningful factors accounting for these regional differences.

Furthermore, our results show that the annual IDF curves based on the monthly maxima constitute a major improvement in terms of uncertainties of the estimates. Using the quantile skill index (QSI), we compare the performance of the models based on the annual and monthly maxima and show that, for some stations, modeling the monthly maxima also leads to a considerable improvement in this regard. A limitation of this study is the strict assumptions that are imposed on the seasonal variations of the distribution parameters. Subsequent studies should therefore investigate the degree to which relaxing these assumptions might further improve the performance of the model based on monthly data. For example, in the framework of a vector generalized additive model (Yee and Stephenson, 2007) it would be possible to model these smooth variations in a nonparametric form. Based on our results, it might be beneficial to model the monthly maxima for obtaining annual IDF curves, when there are large differences in the seasonality of extreme events on different timescales, such as at the station Bever-Talsperre, or for stations where only short observation time series are available. However, it must be considered that a misspecification of the seasonal variations of the parameters can lead to poor results. Moreover, modeling monthly precipitation maxima with the GEV may not be possible in regions with very small precipitation amounts during some months of the year. Therefore, the applicability of the model to the data should always be verified.

Finally, we can demonstrate that at some stations the annual IDF curves based on the monthly maxima deviate from the assumption of scale invariance for long durations. We illustrate that this behavior can be captured by a different parameterization of the location and scale parameter. For future research, it might be of interest to compare the monthly model employed in this study with an annual model that uses different parameterization, e.g., the one proposed by Fauer et al. (2021a). Moreover, by including

additional information in the form of smooth variations during the year, we observe that the shape parameter decreases with duration when averaged over all stations. Based on this result, future research should investigate whether the assumption of a constant shape parameter is appropriate for a wide range of durations from minutes to several days, or whether a more appropriate explicit relationship can be identified.

In conclusion, the use of monthly maxima can be beneficial in several respects when estimating IDF curves, even when information on seasonal variations is not required.

Code and data availability: The station data are mostly publicly available via the Climate Data Center of the DWD (`https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/1_minute/precipitation/historical/`; DWD CDC, 2019). We provide the monthly maxima for 14 aggregation times at all 132 stations, serving as the basis for the analysis, online (`https://doi.org/10.5281/zenodo.5025657`; Ulrich et al., 2021a). The statistical analysis was performed using the package `IDF` for the `R` environment (R Core Team, 2020; Ulrich et al., 2020). The package is available for download at `https://cran.r-project.org/web/packages/IDF` (Ulrich and Ritschel, 2021).

Supplement: The supplement related to this article is available online at: `https://doi.org/10.5194/hess-25-6133-2021-supplement` (see Appendix D).

Competing interests: The contact author has declared that neither they nor their co-authors have any competing interests.

Acknowledgements: The authors would like to thank the Wupperverband as well as the Climate Data Center of the DWD, for providing and maintaining the precipitation time series. Jana Ulrich kindly appreciates the support and motivation offered by Lisa Berghäuser and Oscar E. Jurado. Finally, we would like to thank Juliette Blanchet and two anonymous referees for their valuable feedback and suggestions to improve the manuscript.

Financial support: This research has been supported by the Deutsche Forschungsgemeinschaft (grant nos. GRK2043/1, GRK2043/2, and CRC 1114) and the Bundesministerium für Bildung und Forschung (grant no. 01LP1902H).
We acknowledge support from the Open Access Publication Initiative of Freie Universität Berlin.

Part III

Conclusion

Chapter 7

Synthesis

The main focus of the presented studies was the combined use of observations from different stations or months over a range of durations, and therefore using available data more efficiently. In the following, we will summarize the findings regarding our research questions, starting with the results of Study I. Additionally, we present the application of the methods developed in this study to the analysis of selected extreme events. Subsequently, we summarize the results of Studies II and III and discuss conclusions that arise for modeling the relationship between precipitation intensity, duration and frequency.

7.1 Study I: Summary and Application

The objective of this study was to incorporate precipitation data of different durations and different stations simultaneously in the modeling of IDF curves. For this purpose, we used the d-GEV distribution with spatial covariates. We expected that this would enable the transfer of knowledge from longer to shorter records. We tested the approach in the study area of the Wupper River Catchment in the West of Germany. Our first research question was:

Under which conditions is the spatial d-GEV approach an improvement compared to the separate application of the GEV distribution for each duration and station?

To address this question, we compared the out-of-sample performance of the spatial d-GEV model with that of separately fitted GEV distributions for individual stations and durations based on the Quantile Score. We found that, on average, the skill of the model depends on both the non-exceedance probability p and the duration. The skill increases with increasing p. This is consistent with our expectations, since events with higher p occur less frequently. It is therefore beneficial to apply models which utilizes more data on these rare events. The dependence of skill on duration can be explained in two ways: Firstly, the d-GEV model appears to lack the flexibility to model the precipitation intensity equally well over the entire duration range. This leads to a loss of skill in the middle of the duration range at

most stations. Based on this result, we investigated possible parameterizations to improve the flexibility of the d-GEV model in Study II. Secondly, we observed a decrease in skill for daily durations. We could show that this is due to the longer time series available for these durations. This mainly results in an increase of the performance of the reference model. If a long enough time series is available, the separate application of the GEV model can be advantageous, since this model is based on fewer assumptions. For stations with short time series, however, the advantage of combining different data using the spatial d-GEV model is beneficial.

Using the spatial d-GEV model is not only convenient for pooling information, but it also allows us to spatially interpolate and thus estimate IDF curves at any location within the study area. Therefore, we also aimed to investigate:

Does the spatial d-GEV approach provide reliable estimates at ungauged sites?

We tested the model's ability to predict quantiles for ungauged locations by comparing its performance with that of a seprarate GEV model for each duration at these locations. We found that the average skill of the d-GEV model at ungauged locations depends on the length of the time series available for the reference model. Thus, we could assess that the estimated quantile values of the spatial d-GEV model at ungauged locations are on average comparable with those of applying a GEV model to 30 to 35 annual maxima. We can therefore assume that the model provides reliable estimates for ungauged locations, and use it to provide IDF curves for each location, or spatial return level maps for each duration. On the other hand, we can also associate an exceedance probability (or return period) to the magnitude of an observed event for each duration and at each point in space based on this model. We provide an example of this type of application hereafter.

7.1.1 Spatial Analysis of Selected Heavy Rainfall Events as an Application of the d-GEV with Spatial Covariates

In this section, we focus on Berlin as in the summers of 2017 and 2019 the city was repeatedly hit by heavy rain, which led to pluvial flooding. In both years, disruptions to traffic and everyday life as well as major property damage occurred. The following event analysis is published as part of the report of a task force within the DFG Research Training Group *Natural hazards and risks in a changing world (NatRiskChange)* to investigate the hazards and vulnerability of the Berlin population to heavy rainfall (Berghäuser et al., 2021). In addition to the assessment from a meteorological perspective, the report contains an evaluation on the impacts on citizens and their impairments due to heavy rain as well as the topics of warning, emergency, precautionary measures and risk perception. For the following analysis of heavy rainfall events in Berlin, we model annual maxima of precipitation intensity for different durations from stations in Berlin and Brandenburg using the methods from Chpt. 4. Complementary details on the data and methods used are provided in Appendix E. Based on the spatial model, we can estimate return level maps for arbitrary durations, or IDF curves for each point, as described in Section 4.3.2, for Berlin

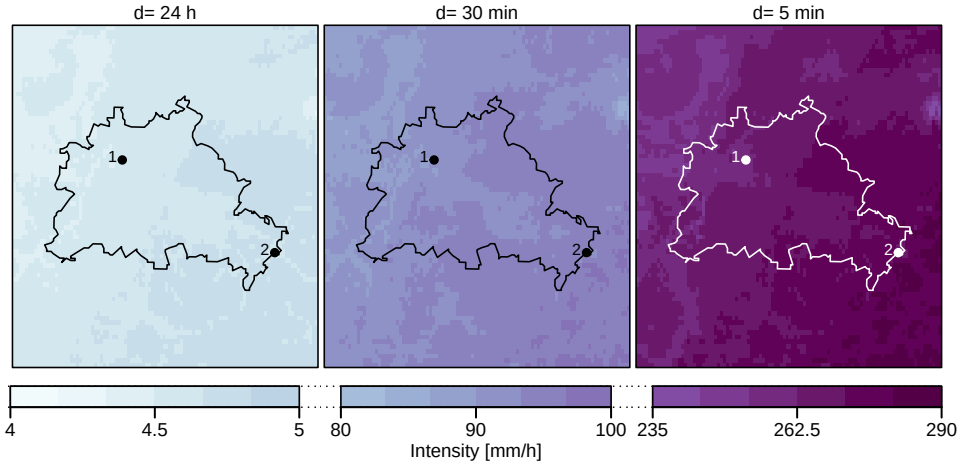


Figure 7.1: Estimated point-wise return level maps for durations $d \in \{24\,\text{h}, 30\,\text{min}, 5\,\text{min}\}$ (different columns) for the non-exceedance probability $p = 0.99$ corresponding to the average return period of 100 years. Dots indicate the positions of the stations (1) Berlin-Tegel and (2) Berlin-Kaniswall.

and the surrounding area. These are presented in the following section. However, we can further apply the spatial model to estimate annual exceedance probabilities, i.e. average return periods, of specific events. We analyze two selected events from 2017 and 2019, which are described in Section 7.1.1 using radar-based and station data. A comparison of the selected events based on their estimated annual exceedance probabilities for different durations follows in Section 7.1.1.

Estimated Return Levels for the Berlin Area

The 100-year return level maps for the durations 24 h, 30 min and 5 min for Berlin and the surrounding area are provided in Figure 7.1. In contrast to the Wupper catchment (compare Figure 4.8), for Berlin we can see that all durations exhibit a similar spatial pattern with intensity increasing towards the southeast. We can further illustrate the differences in precipitation intensity for different durations within Berlin by comparing the example stations Berlin-Tegel (1) and Berlin Kaniswall (2), which are located about 34 km apart. The positions of both stations are marked by dots in Figure 7.1. The IDF curves for both stations are shown in Figure 7.2. Differences between the stations occur mainly for high non-exceedance probabilities and short durations. However, the differences are smaller than the margins of the respective 95% confidence intervals. Additionally, a comparison with the results for the Wupper catchment in Section 4.3.2 indicates that precipitation intensities in the Berlin area for short durations are considerably higher than in the Wupper catchment area.

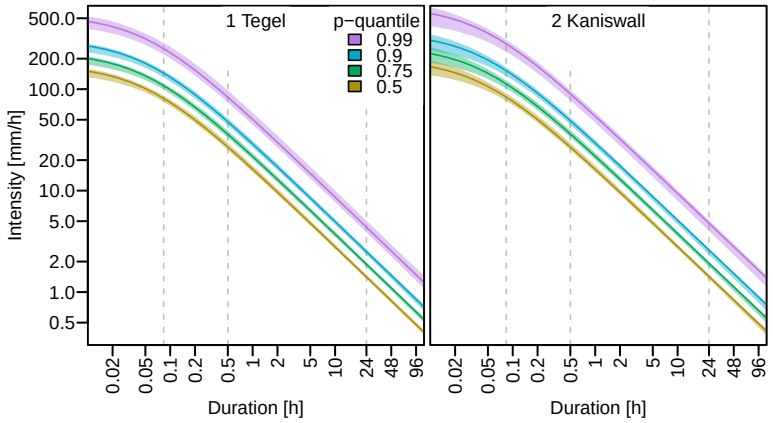


Figure 7.2: Estimated IDF-curves for Berlin-Tegel and Berlin-Kaniswall with 95% confidence intervals (shaded areas). Vertical dashed lines indicate the durations for which return level maps are provided in Fig. 7.1.

Description of the Selected Events

In Berlin, flooding due to heavy precipitation that exceeds the capacity of the city's sewage systems often occurs as a result of convective rainfall in summer. Convective events are caused by the rapid rise and cooling of warm and humid air masses, leading to the formation of high clouds. As a result, heavy rain occurs on a rather small scale and for a short period of time, but with very high precipitation intensities. Throughout the summer in 2019, several thunderstorms, all caused by single convective cells, led to local flooding in Berlin. The most intense event that year occurred on 2 August 2019 and mainly affected the city center. In 2017, a notable heavy rain event happened from 29 to 30 June 2017. However, this event was not a typical convective event with local heavy rainfall. Instead, the collision of cold air from the north and warm moist air from the Mediterranean region led to a widespread rise of the warm moist air and thus to the formation of extensive rain fields. Apart from Berlin, Potsdam, Oranienburg, Leegebruch and other towns in the north of Berlin were also affected by the event. We describe both events[1] with respect to the following questions:

- Which weather situation led to the extreme precipitation in each case?
- How much precipitation occurred in Berlin in which period of time?
- What were the impacts of the events?

For this, we consider precipitation intensities derived from radar reflectivities and precipitation measurements at stations of the DWD (see Appendix E.1). Additionally, the description follows the analyses of the Berliner Wetterkarte (Gebauer, 2017; Gebauer et al., 2017; Niketta, 2017, 2019).

[1]The event description is in non-chronological order, as the August 2019 event could be considered a more typical convective heavy rainfall event than the event in June 2017.

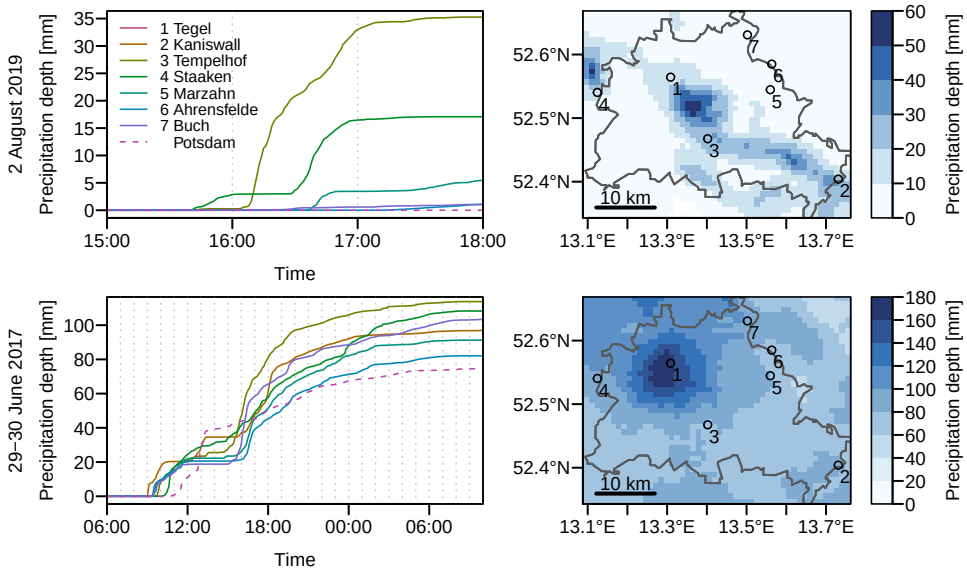


Figure 7.3: Accumulated precipitation for the convective heavy rainfall event (2 August 2019, top row) for the duration of 3 h and the large-scale heavy rainfall event (29-30 June 2017, bottom row) for the duration of 24 h. Presented are, in each case, the cumulative precipitation amounts at individual gauging stations (left), and the amount of total precipitation for the respective duration of the event, derived from radar-based data (right). The location of the stations used are indicated by dots in the maps on the right. The station in Potsdam is outside the area of the map. The station Berlin-Tegel did not record during both events.

Convective Heavy Rain Event in the Afternoon of 2 August 2019 Prior to this event, in the morning of 2 August 2019, an upper-level trough moved from the North Sea in a southeasterly direction over the Netherlands and interacted with an existing low-pressure system over northern Russia. The resulting rise of warm, moist air masses in central and northern Germany led to favorable conditions for the formation of local thunderstorms.

In the afternoon, strong thunderstorms developed over the Berlin area, locally causing intense rainfall, with inner-city areas being particularly affected. The strong spatial heterogeneity can be observed in Figure 7.3 (top right), which presents the three-hour accumulated precipitation (based on radar data) in the period between 3 and 6 p.m. The event was not well recorded by the measuring stations of the DWD, shown in Figure 7.3 (top left), because the highest precipitation amounts occurred in between the stations. Apart from Tempelhof (light green) and Staaken (green), the stations recorded almost no precipitation. At the Tempelhof station, a total of 35.3 mm rain occurred, the majority of it within half an hour. Even greater amounts were reported by measuring stations of other providers. For example, 49.5 mm were recorded by station of the Institute of Meteorology of FU Berlin in the district of Tiergarten. The Berliner Wasserbetriebe even reported almost 59 mm (Niketta, 2019) within one hour at a measuring station in Wedding. Both stations are located in the dark blue region in central Berlin in Figure 7.3 (top right).

The intense rainfall led to flooding of the streets and disruption of traffic in the center

of Berlin. In Berlin Tiergarten, the water height reached up to half a meter in some areas. The Berlin fire department was sent out on more than 150 weather-related operations, mostly due to water damage in buildings and on roads.[2]

Large-scale heavy rainfall event on 29 and 30 June 2017 In the days before the event, an upper-level trough was located over western Europe, which on the one hand transported very warm moist air masses from the Mediterranean region towards central eastern Europe and, on the other hand, brought cooler maritime air from the North Atlantic to western Europe. On the night of the 29 June, a small low-pressure system formed on the ground level over the Czech Republic at the air mass boundary between very warm and humid air over Poland and cooler maritime air over Germany. Within the next 12 hours, the low pressure area strengthened considerably and moved to northwestern Poland. The resulting severe uplift of warm and humid air led to the formation of a widespread precipitation field with embedded thunderstorms moving from Poland over northeastern Germany around noon.

The event affected many areas in Berlin and was a lot less heterogeneous than the event in August 2019. It lasted from 29 June to 2 July 2017, with about 90% of the total precipitation falling within the first 24 hours, i.e. from 9 a.m. on 29 June to 9 a.m. on the following day. The development of these 24 hours can be reconstructed well from the cumulative precipitation sums at the Berlin measuring stations of the DWD, shown in Figure 7.3 (bottom left). At 9 a.m. on 29 June, the first thunderstorm cells from the direction of Poland hit the southeastern border of Berlin. In the following four hours, this first of two successive precipitation fields moved over most parts of Berlin. Already at noon, precipitation values of more than 20 mm were measured at several stations within Berlin. After a break of a few hours around 4 p.m., another large precipitation field reached the north of Berlin and Brandenburg. It was not only larger, but also more persistent than the previous one, and is responsible for the majority of the measured precipitation amounts. In the early morning hours of 30 June 2017 the precipitation finally weakened into a light drizzle. In total, the most precipitation in Berlin occurred in the north-west of the city, as can be seen in Fig. 7.3 (bottom right). However, the maximum values were in this case also not recorded by one of the stations of the DWD, since the station Tegel was not recording during the event. The nearby FU station Tegel-Forstamt measured 221.4 mm within the first 24 hours (Gebauer et al., 2017). Although within Berlin, high precipitation sums occurred mostly due to the long duration of the precipitation, the event also exhibited characteristics of convective events similar to the August 2019 event in some locations. For example, at the station in Potsdam 40 mm of precipitation occurred in the two hours between 11 a.m. and 1 p.m.

The persistent heavy rain led to flooding of numerous streets, allotments as well as subway stations and basements in the center of Berlin. The Berlin fire department reported more than 1,800 weather-related operations. The region in and around Oranienburg in the north of Berlin was also affected severely. Moreover, in addition to the precipitation falling

[2] As reported by `https://www.berlin.de/wetter/nachrichten/5854048-3722621-starke-gewitter-feuerwehr-rueckt-zu-150-.html`, last accessed: 4 November 2021.

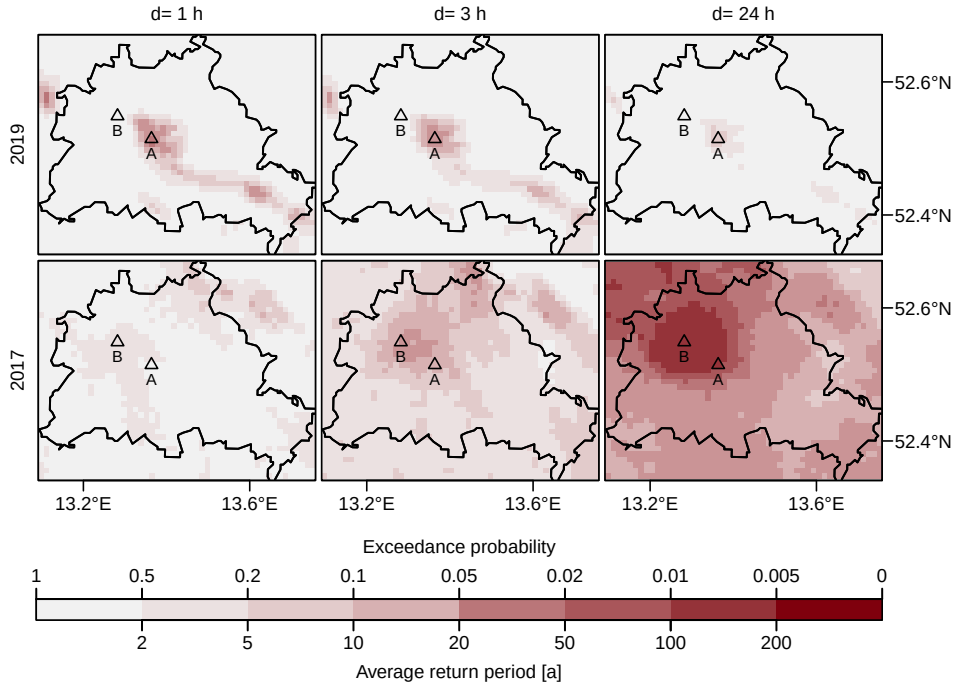


Figure 7.4: Annual exceedance probabilities or average return periods for the two heavy rainfall events from 2 August 2019 (top row) and from 29-30 June 2017 (bottom row) for different durations of the events (different columns). The marked locations A (Tiergarten) and B (Volkspark Jungfernheide) are the locations where maximum precipitation intensities occurred in 2019 and 2017, respectively.

directly into the water bodies, the wastewater from sewer systems caused rivers such as the Spree and Havel to rise considerably (Gebauer et al., 2017).

Estimated annual exceedance probabilities of both events

In the following section, the selected extreme precipitation events described in Sec. 7.1.1 are compared based on their estimated annual exceedance probabilities in the Berlin area for different durations.

Figure 7.4 presents the estimated annual exceedance probabilities or average return periods of the maximum precipitation intensities that occurred during the events in the Berlin area for the three different durations $d = \{1, 3, 24\}$ h. For the event of 2 August 2019 (upper panel), we can conclude that extreme precipitation intensities occurred only for short durations and, moreover, only very localized. Precipitation intensities with exceedance probabilities of 1% or less can be observed at only two grid cells within Berlin. The strong spatial heterogeneity is illustrated by the fact that just few kilometers away from the location with maximum intensities, the annual exceedance probabilities reach merely 50% or less. The characteristics of the event from 29-30 June 2017 (lower panel) are quite opposite. In this event, no extreme precipitation intensities occurred on short timescales.

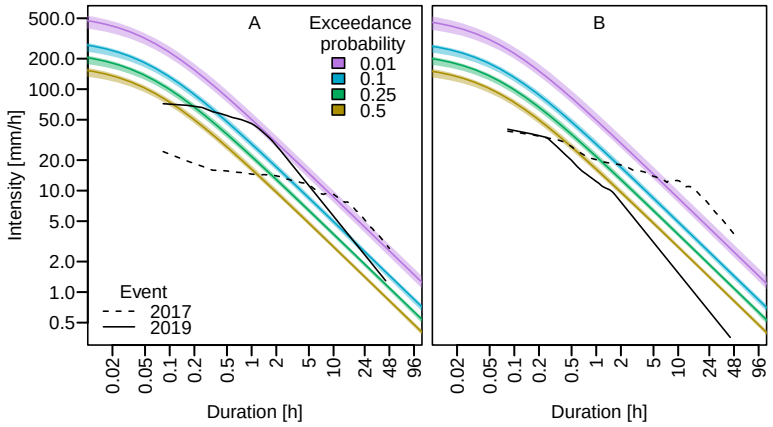


Figure 7.5: Estimated IDF curves at locations A (left) and B (right) (see Fig. 7.4). The black lines show the maximum precipitation intensities dependent on duration that occurred during the heavy rainfall events of 29-30 June 2017 (dashed) and 2 August 2019 (solid) at both locations.

In contrast, for the duration of 24 h in a large area in the north of Berlin, the maximum precipitation intensities that occurred are exceeded with a probability of 1% or less within one year. Hence, the average return periods in this area are 100 years or more. This reflects that the event caused extreme precipitation amounts in a large area due to its persistence.

When focusing on individual locations, we can analyze the events in more detail with regard to different timescales. The different characteristics of the two events, in terms of duration and spatial extent, are especially highlighted when comparing the annual exceedance probabilities of the precipitation intensities that occurred at points A and B (marked in Figure 7.4). These are the locations where the greatest precipitation intensities occurred in 2019 (A) and 2017 (B), respectively, within Berlin when considering the total duration of the events. The distance between the locations is 6.7 km. Figure 7.5 depicts the annual exceedance probabilities of the precipitation intensities that occurred during the events for both locations along with the IDF curves. From the intersection of the event intensities with the IDF curves we can identify for which durations the occuring precipitation intensities were extreme. Again, we find opposite characteristics for both events: For the 2019 event, the exceedance probabilities of the precipitation intensities for a duration of $1 - 2$ h are about 1% at location A. For longer durations, the exceedance probability decreases considerably. On the other hand, at nearby location B, the event does not reach an exceedance probability of 50 % for any duration. This corresponds to an average return period of less than two years. On the contrary, the 2017 event was not extreme on short timescales at both locations. For a duration of two hours, the precipitation intensities reach an exceedance probability of 10%, which corresponds to an average return period of 10 years. However, for durations $d \geq 5$ h precipitation intensities with an exceedance probability of less than 1% occured at location B. Likewise, at location A the exceedance probabilities for the longer durations $d > 8$ h are about 1%.

To summarize, the two events were very different in terms of their spatial extent and duration. Locally, both events led to comparable flooding and damage in Berlin. However,

in the 2017 event, the affected area was much larger. The fact that both events exhibit the most extreme precipitation intensities on sub-daily timescales emphasizes the relevance of generating reliable estimates of IDF curves for these durations. In addition, the example of the event analysis presented can be used to illustrate the need to include spatial variations into IDF models: Both events were not recorded properly within Berlin by the measuring stations of the DWD. This is consistent with the results of Lengfeld et al. (2020). The study indicates that only a small proportion of the hourly heavy rainfall events that occurred in Germany in the period from 2001 to 2018 were captured by the rain gauge station network. In contrast, a large proportion of the daily events were recorded. The radar-based data allow us to understand the spatial and temporal evolution of the events. Although it would also be possible to use radar data to estimate the point-wise return levels, this poses new challenges, notably the limited record lengths. Therefore, for these estimates we rely on station data, whereby the shorter time series of sub-daily observations are combined with the longer time series of daily observations. Thus, a spatial model is necessary to analyze the occurred precipitation intensities at each point in the area of interest. The second and more crucial reason is that, as we have seen, especially smaller thunderstorm cells tend to pass between precipitation gauge stations. The spatial pooling of data makes it possible that extreme events that were recorded by one station are also included in the estimation of IDF curves of surrounding stations.

7.2 Studies II and III: Summary and Comparison

As seen in Chpt. 3.2 and the results of Study I, for some stations the d-GEV model, as defined in Eq. (3.11), lacks flexibility to describe the precipitation intensity over a wide duration range from minutes to several days sufficiently well. Therefore, we investigated possible extensions to this model and their impact on model performance in Study II. Study III analyzes seasonal variations in the IDF relationship and their impact on IDF curves. Study III thus provides possible explanations for the deviations of the observations from the d-GEV model. Both studies are related in this respect. We will start by summarizing the results of the individual studies and then discuss the implications arising jointly from both studies.

Study II considers the extension of the d-GEV model with two additional parameters to provide a more flexible dependence of intensity on duration. The d-GEV model is based on the assumption of a power law, i.e. simple scaling, for the decrease of intensity with duration, although a deviation from this for sub-daily durations is allowed. We refer to this feature as *curvature*. We studied two additional parameters that enable a different scaling of the intensity for events of different magnitude and a deviation from simple scaling for long durations. These additional features are referred to as *multiscaling* and *flattening*. In this respect, the study address the following research question:

Which of the features *curvature, multiscaling* and *flattening* leads to an improved performance of the d-GEV model?

We examined the average model performance in the study area of the Wupper Catchment resulting from the use of the three different model features. We have found that all three features lead to an improvement of the model in a specific duration regime. As expected, allowing *curvature* improves the modeling of very short durations of a few minutes. However, it also results in a decrease in model performance in the middle of the duration range. We could show that the *flattening* feature is able to compensate this decrease in performance. A model with *flattening* but without *curvature* leads to no change in model performance. Therefore, it seems appropriate to use these two features in combination. Surprisingly, allowing *flattening* does not improve the model performance in the range of long durations, although this is the range where this feature is expected to alter the model. We could show that this is caused by the joint estimation of all parameters. The use of *flattening* thus leads to a change in the estimated values of the other parameters as well. The result being that the model performance is improved in the middle duration range, while it slightly decreases in the long duration regime. With respect to *multiscaling*, we found that while this feature significantly improves the model performance in the long duration regime, it leads to a deterioration of the same magnitude in the sub-hourly duration range. When the sub-hourly observations are not included in the model, none of the features lead to a relevant improvement of the model. Therefore, the use of the features depends on the particular objective of the model.

Another result of the study is related to the coverage of the estimated confidence intervals. In Study I, we compared two different approaches to estimating confidence intervals for quantiles of the d-GEV distribution, namely the delta method and the non-parametric bootstrap method. We were able to show in a sampling experiment that the bootstrap method, in contrast to the delta method, provides reasonable coverage. However, we only investigated the coverage under the assumption of independence between the maxima of different durations. In Study II, we therefore investigated the coverage for sampled data with known dependence, and asked:

Does the non-parametric bootstrap method yield confidence intervals with the assumed coverage in the case of dependent data when an appropriate sampling strategy is applied?

We assumed that by sampling all data of one year collectively, the dependence structure of the data would be taken into account in the bootstrap method. We tested this assumption in a sampling experiment, where we generated random data with known dependence using a max-stable Brown-Resnick process. The confidence intervals estimated on the basis of these data using the described sampling strategy exhibit slightly lower coverage than the assumed 95 %. Thus, we conclude that the provided confidence intervals underestimate the uncertainties due to the dependence of the data to a small extent.

Study III explores seasonal changes in the IDF relationship by modeling monthly maxima using harmonic functions of the month as covariates in the parameters of the d-GEV distribution to represent the seasonal cycle. First, this approach allows the use of more data on extreme events. In addition, we expected to gain insights into how the different processes that cause extreme precipitation in different seasons affect the modeling of IDF curves. Therefore, we first investigated:

How does the IDF relationship at different stations in Germany evolve throughout the year?

We could show that primarily the parameters σ_0 and η of the d-GEV distribution vary throughout the year. These parameters determine the intercept and the slope of the IDF curves in double logarithmic representation. We found that at all investigated stations in Germany the IDF curves are steeper in summer than in the remaining months, i.e. in summer the strongest decrease of intensity with duration exists. The different intercepts of the curves provide information about the months with maximum precipitation intensities for short durations. We observed that the intensity maximum for short durations occurs in summer at all stations. Regarding the intensity maximum for long durations, however, we found differences between the stations. For a better understanding of how the monthly IDF relationships differ among stations, we studied this topic from another perspective, by asking:

During which months of the year are annual maxima of different durations more likely to occur?

To investigate this, we calculated the probability that the annual 0.9 quantile is exceeded within a given month. As expected, we found that the short convective extreme events occur almost exclusively in summer everywhere in Germany. At the majority of stations, the long-lasting extreme events also occur most likely in summer. However, often these events are similarly likely to occur in other seasons, as the exceedance probabilities possess comparable values during other months of the year. Stations where long-lasting extreme events are most likely to occur in winter were found to be mainly located in high altitude regions in the western half of Germany. We suspect that this is related to the stronger westerly winds in these months, which can lead to increased precipitation on the windward sides of the mountains. At stations along the North Sea coast, long-lasting extreme events are most likely to occur in late summer to autumn, which is likely to be related to the warm surface water temperatures interacting with increasingly colder air masses and stronger winds at the beginning of autumn.

We have found that the IDF relationship varies throughout the year. Additionally, similar to Study I, the use of covariates in this seasonal context allows including more data into the model. This led us to the question:

To what extent do the annual IDF curves based on monthly and annual maxima differ?

As expected, we found that the use of monthly maxima significantly reduces the uncertainties of the estimated quantiles. This is particularly relevant for stations with shorter time series. In addition, the seasonal d-GEV model allows a more flexible relationship of precipitation intensity and duration. In this context, we observed a deviation of the IDF curves from a power law at some stations, especially for long durations. This deviation is similar to the one caused by the *flattening* feature investigated in Study II. We observed this deviation primarily at the previously mentioned stations at higher elevations, where maximum precipitation intensities shift from summer for short durations into winter for long durations. We therefore assume that the *flattening* feature is caused by pronounced differences in the seasonality of short and long extreme events. When comparing the seasonal and the annual d-GEV models in terms of model performance, we find that the seasonal model leads to an improvement only at some stations. We suspect that for stations, where the seasonal model results in a worse description of the annual data, the model assumptions for the dependence of the d-GEV parameters on the month might be to strict or not valid.

Since we found that accounting for seasonality can lead to a deviation from the d-GEV model as defined in Eq. (3.11), we also investigated what differences result for the dependence of the GEV parameters on duration. Thus, our final research question stated:

Does explicit modeling of seasonal variations allow us to draw conclusions aimed at improving the modeling of annual maxima?

Concerning location and scale parameters to describe annual maxima, we found a deviation from the assumptions of the d-GEV distribution given in Eqs. (3.8-3.9) at some stations. It may be possible that these deviations can be described using the extensions to the d-GEV model discussed in Study II. The estimation of the shape parameter ξ is associated with large uncertainties, especially for short time series. It is therefore difficult to detect a relationship between ξ and duration. However, taking seasonality into account leads to a significant reduction of these uncertainties. This allowed us to investigate the relationship between ξ and duration with more clarity. Our results indicate that the assumption of a constant shape parameter does not seem to be justified. Instead, on average over all stations, ξ is positive decreases with duration.

7.2.1 Comparison of Results and Implications for Flexible IDF Curves

In Study III, we observed a deviation of the IDF curves from the simple scaling assumption for long durations, especially at stations exhibiting substantial differences in the seasonality of short and long lasting extreme events. At these stations, the different slope and intercept of the IDF curves for different months causes the curves to cross. Therefore, we assume that the deviation for long durations can be explained by the intersection of the IDF curves for different seasons.

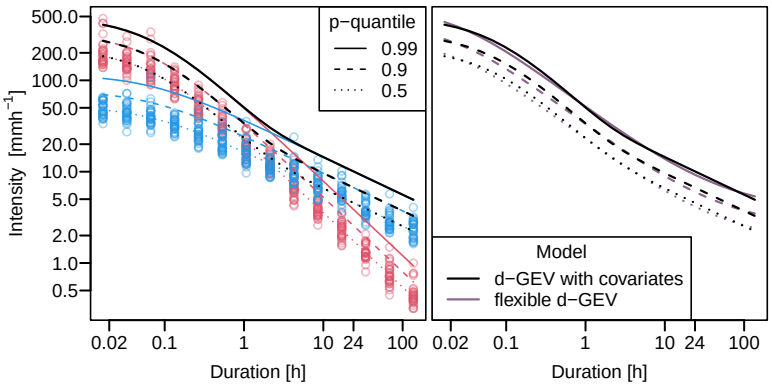


Figure 7.6: Example data consisting of sampled maxima from two groups (red and blue circles) with estimated IDF curves for each group (red and blue lines). IDF curves for the maxima of the joined groups were estimated from all data using the d-GEV distribution with covariates (black lines) and from only the respective maxima of both groups using the flexible d-GEV model (purple).

Do pronounced seasonal differences in the IDF relationship cause a deviation from the simple scaling assumption for long durations?

To examine this idea in a simplified form, we take a basic example where block maxima can originate from only two different groups. The groups could be interpreted as two seasons or two different processes causing extreme events, i.e. convective and stratiform precipitation. To do this, we specify an IDF relationship for each group such that the IDF curves of the groups intersect analogously to the results from Study III. For each Group, we draw a random sample for a range of durations. In Fig. 7.6, these samples are shown as red and blue circles. We can apply one model to both groups by treating the group as a covariate, comparable to the d-GEV with monthly covariates. The estimated IDF curves for each group are shown as red and blue lines in Fig. 7.6. The IDF curves that result jointly for both groups, effectively the annual IDF curves, are shown in black. For the upper and lower end of the duration range, the annual IDF curves correspond to those of the individual groups, because in both regimes only one of the two groups contributes to the annual maxima. This results in a bending of the annual IDF curves in the middle of the duration range. This example confirms our assumption that a deviation from the d-GEV model can be caused by such seasonal differences.

In Study III, we proposed the use of the flexible d-GEV model as an alternative to explicitly account for seasonality. Therefore, we briefly assess:

Does modeling annual maxima with a flexible d-GEV model produce comparable results to modeling monthly maxima using the d-GEV distribution with monthly covariates?

We first explore this using the example shown in Fig. 7.6. For this purpose, we take the maximum value of both groups for each index. This corresponds to the annual maxima.

From these data, we estimate the parameters of the flexible d-GEV distribution (Eqs. 5.3, 5.10-5.11) under the assumptions $\tau \neq 0$ and $\eta_2 = 0$. The resulting IDF curves are shown in purple in Fig. 7.6 (right panel). We can see that the IDF curves of the monthly model and those of the flexible annual model are very similar. The flexible d-GEV model allows a smoother transition between the regions of shorter and longer durations. The curves can be expected to diverge when extrapolating to longer durations. Since this example is a very exaggerated representation of the seasonal variations found in Study III, we can assume that the flexible d-GEV model may provide even better agreement for actual observations.

To investigate this assumption, we compare the annual flexible d-GEV model with the monthly d-GEV model at three stations for which we found pronounced seasonal differences between short and long lasting events in Study III. These are the stations Bever-Talsperre and Saarbrücken-Ensheim, which already served as example stations in the study, and additionally the station Schmücke, which is located in the Thuringian Forest, a mountain range in central Germany. Fig. 7.7 shows the estimated annual IDF curves of both models together with their respective 95 % confidence intervals for the three stations. The top row presents the QSI, see Eq. (6.18), for a range of non-exceedance probabilities and durations,

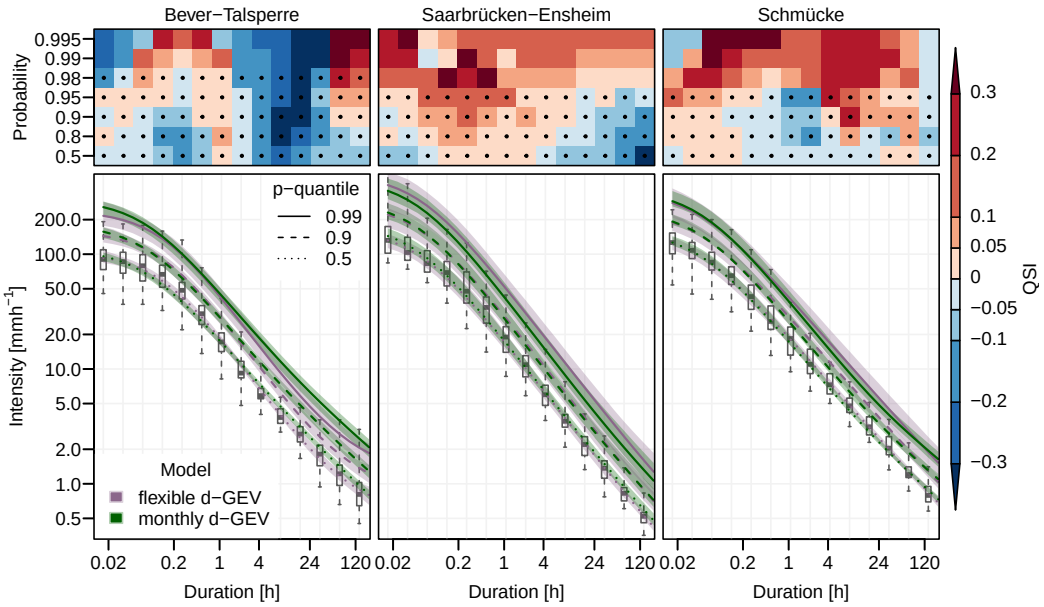


Figure 7.7: Comparing the annual IDF curves estimated by modeling annual maxima with the flexible d-GEV model (purple) and by modeling monthly maxima using the d-GEV distribution with monthly covariates (green) for three example stations. The shaded areas represent the respective 95 % confidence intervals. The distributions of the observed annual maxima are shown as box-and-whisker plots, where the whiskers cover the complete data range. In the upper panels the corresponding QSI values are presented for comparison of the models' performances, where positive values indicate an increase in the skill of the monthly d-GEV model compared to the annual flexible d-GEV model. Dots represent the non-exceedance probability p corresponding to length T of the available time series used for verification where $T = 1/(1-p)$.

with positive values indicating the monthly model outperforming the annual model and negative values representing the opposite. For station Bever-Talsperre, we observed the strongest deviation of the monthly model (green) from simple scaling at long durations in Study III. Here, the flexible d-GEV model (purple) shows an even stronger deviation in the form of pronounced bending of the IDF curves in the range between 1 h and 120 h. From the QSI we can conclude that the deviation in this range has a positive effect on the model performance in the mid duration range. This result is consistent with the results on the effect of the *flattening* feature on model performance from Study II. In the short duration range, the IDF curves of the flexible d-GEV model also exhibit stronger *curvature*. In this range, the QSI does not provide clear information regarding which model performs better. In contrast to Bever-Talsperre, the flexible d-GEV model shows weaker *flattening* than the monthly model at the stations Saarbrücken-Ensheim and Schmücke. At both stations, the regions where the monthly model outperforms the annual model predominate. In conclusion, this implies that although the flexible d-GEV model is generally capable to reproduce the shape of the curves described by the monthly model, the parameter estimation based on annual maxima at the three example stations leads to different results.

Chapter 8

Summary and Outlook

8.1 Summary

The aim of this thesis was the development of methods for a more efficient use of available data in the context of extreme precipitation modeling. In this regard, the studies included in this thesis identify possibilities to pool knowledge from different sources to assess the relationship between precipitation intensity, duration and frequency. This is especially relevant when considering extreme precipitation of short duration, since measurement data are usually scarce for the analysis of these events.

To this end, in a first step we have demonstrated the feasibility of modeling annual maxima of different durations simultaneously using a duration-dependent GEV distribution. We used the approach of Koutsoyiannis et al. (1998) in which assumptions about the dependence of the GEV parameters on the duration are directly implemented into the distribution to obtain one duration-dependent GEV (d-GEV) distribution. We presented the advantages of this approach in comparison to applying separate models in sequence. The main strengths of the d-GEV model are its parameter parsimony and the resulting reduction of uncertainties in the parameter estimation, especially for the shape parameter. However, we also found that for some stations the initial assumptions of Koutsoyiannis et al. (1998) do not provide a sufficiently flexible description of precipitation intensity as a function of duration, when considering a wide range of durations from minutes to several days. We addressed this problem in Study II of this thesis. For this purpose, we extended the d-GEV model by two parameters and tested how increased flexibility affects model performance. We were able to show that the inclusion of additional parameters improves the model in specific regions of the duration range and concluded that the choice of parameterization should depend on the objective of the model.

In Studies I and III, we demonstrated how the d-GEV model can be extended using covariates. We applied this approach to pool knowledge not only between different durations but also between different stations (Study I) or between different months of the year (Study III). Both covariate models therefore allow the more efficient use of larger data sets and result in a reduction of uncertainties in the estimation of quantiles. In both cases, their use has been identified as particularly beneficial in situations of low data availability,

i.e., when modeling rare events as well as durations or stations with only short time series available.

In addition, both models provide further advantages: The d-GEV model with spatial covariates presented in Study I allows the estimation of IDF curves at any location within the research area and the generation of return level maps for arbitrary durations. We were able to verify that the model provides reliable estimates at ungauged sites. Furthermore, we could apply the model to analyze two recent extreme events that led to flooding and damage in the Berlin area. Thus, we have demonstrated that it is possible to use the spatial d-GEV model in combination with radar-based estimates of event intensities for a detailed analysis of their frequency both in space and over the duration of an event. In the cases of the selected events, the extreme precipitation intensities were to a large extent not recorded by the DWD stations, which further illustrates the main advantages of the spatial d-GEV model, namely the possibility of spatial interpolation and the consideration of measurements at neighboring stations.

The seasonal d-GEV model applied in Study III enables the modeling of monthly maxima rather than annual maxima, using covariates to explicitly account for intra-annual variations. This allowed us to investigate the IDF relationship in different seasons and thus to some extent with respect to different precipitation generating processes. Our results indicate that short convective extreme events occur almost exclusively in summer everywhere in Germany, while the seasonality of long-lasting, mostly stratiform events depends on location. Especially the altitude, the distance to the coast and the mean wind direction seem to be influential parameters. We also found that the seasonal d-GEV model results in a notable deviation from the assumption of simple scaling for long durations, especially at stations with strong differences in the seasonality of short and long-lasting extreme events. We could attribute this to the crossing of the monthly IDF curves at these stations. We suspected that modeling monthly maxima with the seasonal d-GEV model as well as modeling annual maxima with the flexible d-GEV could lead to similar results in these cases. However, testing at three example stations did not confirm our hypothesis. Finally, modeling monthly maxima by taking into account the intra-annual variations leads to a significant reduction of uncertainties in the estimation of the shape parameter. We could use this advantage to analyze the dependence of the shape parameter on duration. Our results suggest that the assumption of a constant shape parameter over duration may not be reasonable and a different parametrization for this parameter appears to be required.

In conclusion, pooling information does not only allow for more reliable estimates but also provides a deeper understanding of the characteristics of extreme precipitation events on different time scales and possible influences on their variability.

8.2 Outlook

This work constitutes a step towards the goal of developing efficient statistical methods to model extreme precipitation events and assessing their possible consequences. As expected, many research questions remain and likewise new questions emerge from the acquired

insights. Therefore, in the following section, we present some directions in which the research presented here may be continued.

Alternatives for Modeling the Duration Dependence A main objective of the presented work has been the pooling of information to provide more reliable estimates where data is scarce, such as for sub-daily extreme precipitation events. In order to combine information within one model, it is necessary to specify certain model assumptions, e.g. assumptions for the relationship between intensity and duration. We have demonstrated that model assumptions that are too strict prevent an improvement in model performance, even though they allow the model to utilize more data. We have identified one possible solution, which is to adopt more complex models with additional parameters. Another interesting option could be to describe the change in intensity with duration using a non parametric model. One way to do this would be via a generalized additive model (Yee and Stephenson, 2007). This approach is similar to the presented d-GEV model, but instead of assuming empirical relationships for the dependence of GEV parameters $\theta \in \mu, \sigma, \xi$ on duration d, sums of unspecified smooth functions $f_i^\theta(d)$ are adopted

$$\theta(d) = \theta_0 + \sum_i f_i^\theta(d). \tag{8.1}$$

The challenges here are the uncertain degree of smoothness, extrapolation and interpretation.

Another possibility involves the use of a penalized maximum likelihood estimator (PMLE) (Bücher et al., 2021; Coles and Dixon, 1999). This parameter estimation method results from subtracting a penalty term $\Omega(\boldsymbol{\Theta}) \geq 0$ from the log-likelihood l (Eq.2.11) yielding

$$\hat{\boldsymbol{\Theta}}_\Omega = \arg\max_{\boldsymbol{\Theta}} \left[l(\boldsymbol{\Theta}|z_1, ..., z_n) - \Omega(\boldsymbol{\Theta}) \right]. \tag{8.2}$$

The penalty term can be seen as a certain assumption or so called expert knowledge, that has to be considered within the parameter estimation. A simple example adopted by Bücher et al. (2021) to illustrate the applicability of the estimation method to flood frequency analysis is based on the assumption of a certain value for the shape parameter $\xi = \xi_c$. We can incorporate this assumption into the parameter estimation by choosing the penalization term $\Omega(\boldsymbol{\Theta}) = \lambda(\xi - \xi_c)^2$, where the hyperparameter $\lambda \geq 0$ determines how much confidence is placed in the assumption. The estimator then yields

$$\hat{\boldsymbol{\Theta}}_\Omega(\lambda) = \arg\max_{\boldsymbol{\Theta}} \left[l(\boldsymbol{\Theta}|z_1, ..., z_n) - \lambda(\xi - \xi_c)^2 \right], \tag{8.3}$$

where for $\lambda = 0$ the ordinary ML estimator results, while $\lambda = \infty$ yields $\hat{\xi} = \xi_c$. An optimal value for λ can be determined based on the out-of-sample performance of the model. The PMLE approach therefore offers the possibility to similarly implement assumptions for the dependence of the parameters on the duration. In this respect, the straightforward way would be to use the d-GEV assumptions from Eqs. (3.8-3.10). We could for example

implement these assumptions into the estimation of the GEV parameters μ_d, σ_d and ξ_d for a range of durations d by using the penalty term $\Omega(\mathbf{\Theta}) = \Omega_\sigma + \Omega_\mu + \Omega_\xi$ with

$$\Omega_\sigma = \lambda_\sigma \sum_d \left\{\sigma_d - \left[\sigma_0 (d+\theta)^{-\eta}\right]\right\}^2 \tag{8.4}$$

$$\Omega_\mu = \lambda_\mu \sum_d \left\{\mu_d - \tilde{\mu}\sigma_d\right\}^2 \tag{8.5}$$

$$\Omega_\xi = \lambda_\xi \sum_d \left\{\xi_d - \xi_c\right\}^2 \tag{8.6}$$

in Eq. (8.2), where $\tilde{\mu}, \sigma_0, \theta, \eta$ and ξ_c are values that have to be previously defined. If those values are set to the maximum likelihood estimates of the d-GEV model and in the limit $\lambda \to \infty$, we again obtain the d-GEV model. A reduction of λ therefore leads to a relaxation of the d-GEV assumptions.

In a preliminary analysis, we tested this type of penalization using the annual maxima of the station Bever-Talsperre. In the example presented in Fig. 8.1 we choose a constant value for the hyperparameter $\lambda_\xi = 100$ and thus only assess the influence of the strength of the penalization terms Ω_σ and Ω_μ by increasing $\lambda_\sigma = \lambda_\mu = \lambda_1$. The plots (b-d) respectively present the estimated parameters μ_d, σ_d and ξ_d for a range of durations d and different values of λ_1. The dashed lines indicate the parameter values for which the penalization terms would reduce to zero. Since the estimates for μ_d are already fairly close to the dashed line for $\lambda_1 = 0$ (purple dots), we observe little change for this parameter when increasing λ_1. In contrast, for σ_d, this leads to a reduction at short durations $d < 1\,\mathrm{h}$ and an increase at intermediate durations $1\,\mathrm{h} < d \lesssim 24\,\mathrm{h}$. Interestingly, this changes in σ_d affect the estimated values for ξ_d. For this parameter, we can see a slight fluctuation around the constant value ξ_c (dashed line) for $\lambda_1 = 0$ (purple dots), with stronger deviations occurring at longer durations. Penalization of σ_d leads to an increase of ξ_d at short durations and a reduction at medium durations and thus to a compensation of the behavior of σ_d. Nevertheless, we obtain a slight change in the resulting quantiles for the ranges of short and medium durations, as can be seen from Fig. 8.1 (e). We investigate the influence of λ_1 on the model performance by comparing the QS of the model to the reference model with $\lambda_1 = 0$ for a range of durations and non-exceedance probabilities. The resulting QSI values are displayed in Fig. 8.1 (a). We find that increasing λ_1 leads to an improvement of the model (QSI>0) in the range of short durations especially for higher non-exceedance probabilities. For medium durations, however, increasing λ_1 leads to a deterioration of the model. In the presented example, it would be possible to choose a smaller value of $\lambda_1 \leq 2$ so that the improvement of the model for short durations outweighs the deterioration for medium durations.

We can conclude that it is of interest to further investigate the use of the PMLE method both for the estimation of IDF curves and to better understand the dependence of GEV parameters on duration. In the presented example the penalization terms are based on assumptions of parametric equations for the parameters. However, less restrictive assumptions, such as smoothness or a monotonic behavior, could also be implemented and might even lead to more promising results. As seen in Study III, the assumption ξ =const.

might not be justified and a better assumption could be that ξ decreases with duration.

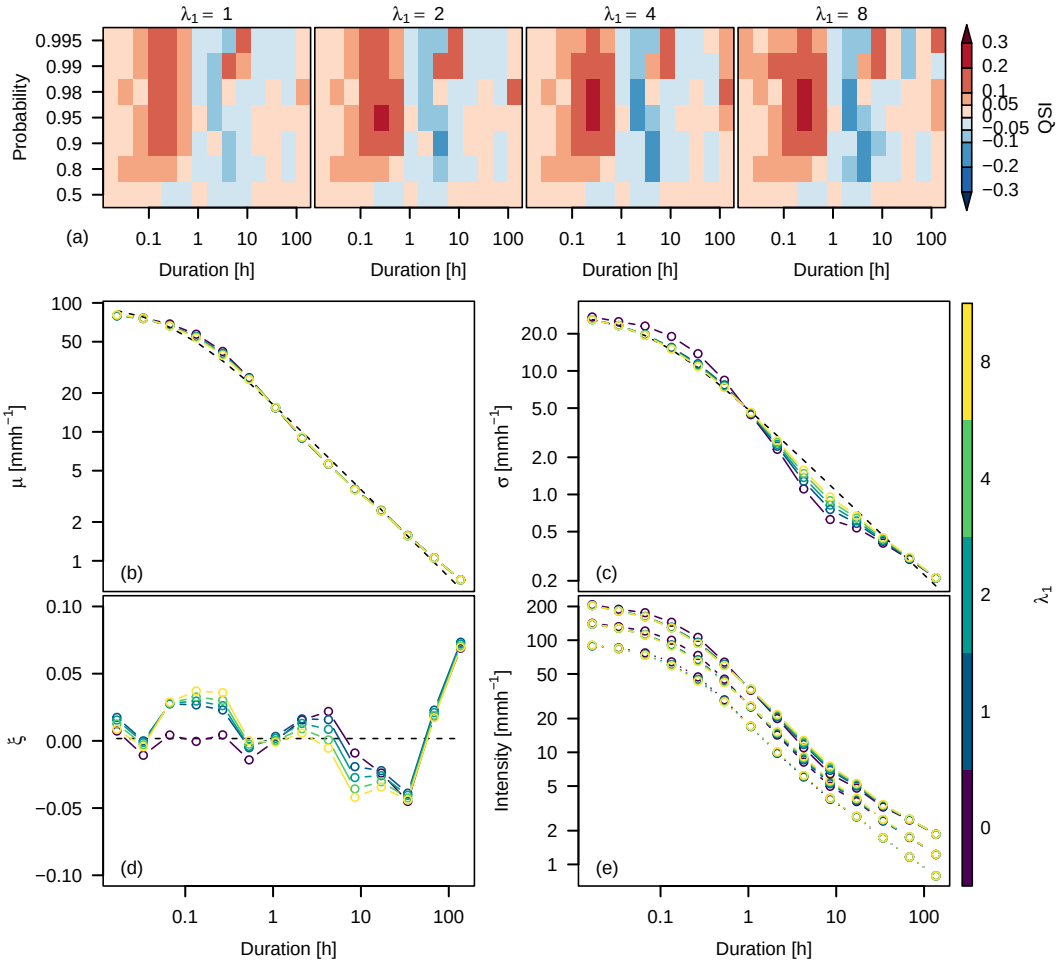


Figure 8.1: First analysis of the applicability of the PMLE method for the estimation of IDF curves using the penalization terms in Eqs. (8.4-8.6) with $\lambda_\sigma = \lambda\mu = \lambda_1$ and $\lambda_\xi = 100$. Parameter estimates (b-c) and quantiles (e) for non-exceedance probabilities $p \in 0.99, 0.9, 0.5$ (solid, dashed, dotted) are presented for different choices of λ_1. Panel (a) provides the QSI as measure of the model performance for different values of λ_1, when compared to the reference model with $\lambda_1 = 0$.

Therefore, this kind of non parametric penalization seems especially interesting for the shape parameter. Disadvantages of the PMLE method for the estimation of IDF curves are that the obtained IDF curves are not necessarily consistent, i.e. a crossing of quantiles may occur, and an interpolation between the durations is no longer possible.

Combination of Station Data and Radar-Based Estimates The d-GEV model allows simultaneous modeling of long daily and shorter sub-daily measurement time series. We demonstrated that this more efficient use of data leads to more reliable estimates, especially for rare events. However, we have seen from two case studies in Berlin that extreme

precipitation events are not necessarily captured by gauging stations. The study of (Lengfeld et al., 2020) points out that especially the short convective events are to a large extent not recorded by measuring stations. Several studies on the estimation of IDF curves based on radar data were recently performed (Goudenhoofdt et al., 2017; Haberlandt and Berndt, 2016; Marra et al., 2017), as radar data have the advantage of a high spatial coverage. Yet, the available time series are relatively short. Therefore, it would be considerably advantageous to combine the two data sources for the estimation of IDF curves to benefit from both the spatial coverage of radar data and the long measurement time series especially for daily observations at gauging stations. Thus, future studies should investigate the suitability of models for the spatial IDF relationship based on such a combined data set.

Consideration of Non-Stationarity A major aspect that has not been addressed in the present work is the extent to which the intensity and frequency of extreme precipitation events evolve in a changing climate. However, we expect that the monthly d-GEV model proposed in Study III is well suited for detailed future investigations. We have performed a preliminary analysis to suggest how to proceed further in this direction using the developed methods. For this purpose, we have divided the time series of the station data used in Study III into running time windows of 15 years. Fitting the monthly d-GEV model separately to the data within each time window can provide us with a first idea of how the IDF curves change over time. For this type of analysis, the monthly d-GEV model is advantageous, since dividing the total available time period into smaller time windows represents a reduction in the data available to estimate the parameters. Thus, the uncertainty of the estimates increases. But as we observed in Study III, modeling monthly maxima represents a significant reduction in uncertainties when modeling shorter time series compared to modeling annual maxima. In addition, the monthly d-GEV model allows us to analyze potential changes in seasonality, or monthly exceedance probabilities, over time.

Our exploratory analysis indicates an increase in the frequency of extreme precipitation events within the last 30 years at many of the considered stations. Only for a small part of the stations a decrease in frequency can be observed. In Fig. 8.2 we provide the results of the analysis for three example stations to illustrate the different behavior at individual stations. We present the IDF curves for the time windows in different colors in the left column. The remaining columns show the monthly exceedance probabilities of the annual 0.9 quantile calculated on the basis of the full available time period for the time windows and for different durations. At the station Berlin-Tempelhof (a) we observe an increase in intensity of the quantiles over the whole duration range. Likewise, we find a steady increase in the monthly exceedance probabilities. The seasonality of the extreme events remains unchanged over time. Also for the station Schmücke in the Thuringian Forest (b) we find an increase in intensity of the quantiles for short durations. For longer durations, however, a decrease in intensity occurs at this station, so that the IDF curves as a whole become steeper over time. Looking at the monthly exceedance probabilities, it is noticeable that the decrease in intensity for long durations is mainly due to a change in seasonality.

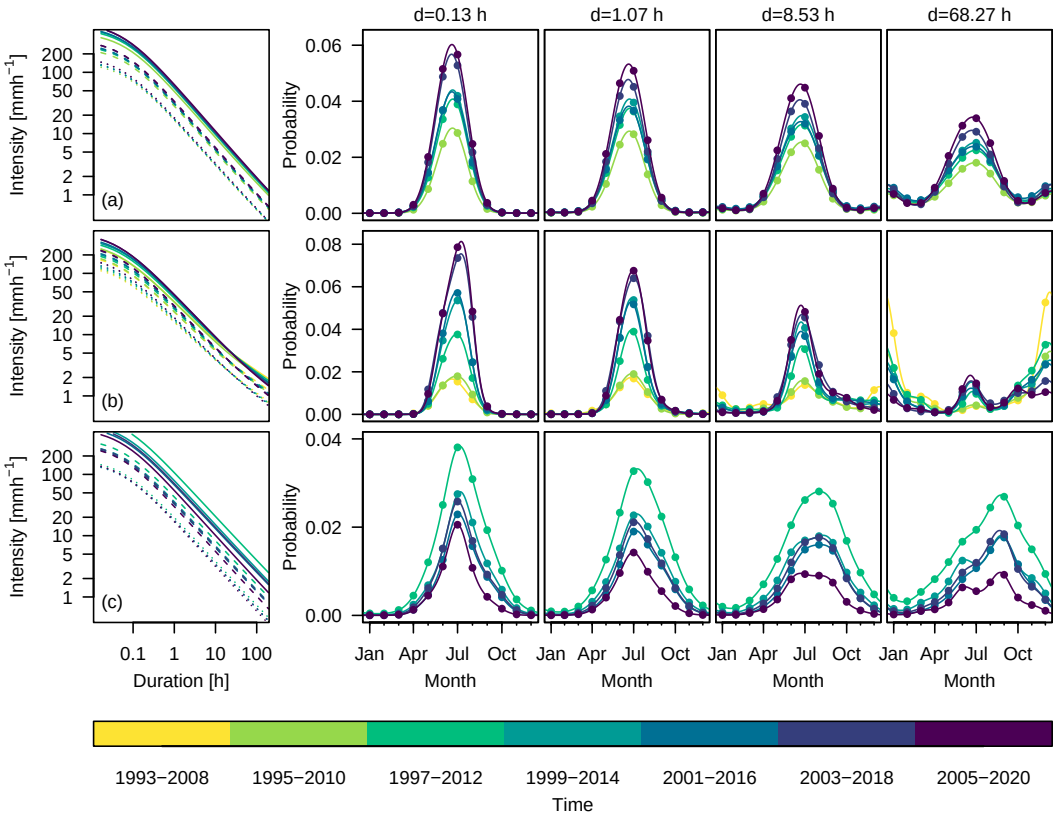


Figure 8.2: IDF curves and monthly exceedance probabilities of the annual 0.9 quantile for different 15 year time windows (colors) and different durations (as indicated in the top) at the stations Berlin-Tempelhof (a), Schmücke (b) and Ummendorf (c). For the monthly exceedance probabilities, the annual 0.9 quantile is calculated on the basis of the full available time period. The available time series at the three stations differ, only the station Berlin-Tempelhof covers the period of 1993-2020.

In this context, extreme events of longer duration occur less frequently in winter over time. Likewise, we observe a slight increase in the exceedance probabilities in summer. For the station Ummendorf in Saxony-Anhalt (c) we observe a decrease in the intensities, as well as the monthly exceedance probabilities over time for all durations.

Continuing this analysis appears rather promising. Next steps should deal with the analysis of longer time series. In addition, it is important to investigate to what extent the changes over time are within the sampling uncertainties. Two other questions that could be studied through this approach are: What is the spatial distribution of stations with different temporal trends? Can the trends be explained by the occurrence of single events of large magnitude? A subsequent step should explore explanatory predictors for the temporal changes in the distribution parameters. This would finally allow the temporal variations to be implemented directly into the model in the form of further covariates to not only assess the observed changes but instead also allow predictions about future developments.

The issue of non-stationarity also raises questions that go beyond the field of statistical modeling of IDF curves. So far, the planning of hydrological structures has been based on return levels associated with specified average return periods. However, in the case of non-stationary intensity and frequency of extreme precipitation events, this concept is no longer applicable. Therefore, future studies must address the development of concepts to communicate changing annual exceedance probabilities on an interdisciplinary basis. More research is required on how to apply the results of non-stationary models and their associated uncertainties to the design of hydrologic structures. Although some concepts have been developed (e.g., Read and Vogel, 2015; Rootzén and Katz, 2013), they are still far from being implemented in operational use.

Appendices

Appendix A

Supporting Information for Chapter 2

A.1 Possible Distribution Functions for Maxima

Hoskin and Wallis suggest the Generalized Logistic (GL), Generalized Normal or sometimes called Lognormal (GN) and Pearson type III (P3) distributions in addition to the GEV distribution for modeling annual maxima of precipitation and streamflow. Each of these distributions has three parameters: location $-\infty < \mu < \infty$, $\sigma < 0$ scale, and $-\infty < \xi < \infty$ shape. The cumulative distribution functions $F(x)$, as well as the name of the distribution in the special case $\xi = 0$ are provided in Table A.1. We define

$$y_1 = \begin{cases} \xi^{-1} \ln[1 + \xi(x-\mu)/\sigma] & ,\xi \neq 0 \\ (x-\mu)/\sigma & ,\xi = 0 \end{cases} \tag{A.1}$$

on the range $\{x : 1 + \xi(x-\mu) > 0\}$ and

$$y_2 = \sqrt{\alpha}(x-\mu)/\sigma - \alpha, \quad \text{with } \alpha = 4/\xi^2 \tag{A.2}$$

on the range $\{x : y_2 \cdot \xi > 0\}$. Furthermore $\Phi(\cdot)$ is the standard Normal distribution

$$\Phi(x) = \int_{-\infty}^{x} \exp(-t^2/2)/\sqrt{2\pi}\, dt, \tag{A.3}$$

$\Gamma(\cdot)$ denotes the gamma function and $\Gamma_i(\cdot,\cdot)$ the incomplete gamma function

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} \exp(-t)\, dt, \qquad \Gamma_i(\alpha, x) = \int_0^x t^{\alpha-1} \exp(-t)\, dt. \tag{A.4}$$

Table A.1: Definition of cumulative distribution functions according to Hoskin and Wallis.

Name	**Distribution Function**	**case** $\xi = 0$
GEV	$F(x) = \exp[-\exp(-y_1)]$	Gumbel
GL	$F(x) = 1/[1+\exp(-y_1)]$	Logistic
GN	$F(x) = \Phi(y_1)$	Normal
P3	$F(x) = \begin{cases} \Gamma_i(\alpha, y_2)/\Gamma(\alpha) & ,\xi > 0 \\ \Phi[(x-\mu)/\sigma] & ,\xi = 0 \\ 1 - \Gamma_i(\alpha, -y_2)/\Gamma(\alpha) & ,\xi < 0 \end{cases}$	Normal

A.2 Moment and PWM Estimators for the GEV Parameters

For the GEV distribution the following relationships between the moments and the parameters result for $\xi \neq 0$:

$$\mathrm{E}(Z) = \mu - \frac{\sigma}{\xi}[1 - \Gamma(1-\xi)], \quad \xi < 1, \tag{A.5}$$

$$\mathrm{Var}(X) = \frac{\sigma^2}{\xi^2}[\Gamma(1-2\xi) - \Gamma^2(1-\xi)], \quad \xi < \frac{1}{2}, \tag{A.6}$$

$$g = \mathrm{sign}(\xi)\frac{\Gamma(1-3\xi) - 3\Gamma(1-\xi)\Gamma(1-2\xi) - \Gamma^3(1-\xi)}{[\Gamma(1-2\xi) - \Gamma^2(1-\xi)]^{3/2}} \tag{A.7}$$

and for the special case of $\xi = 0$:

$$\mathrm{E}(Z) = \mu + \sigma\gamma \tag{A.8}$$

$$\mathrm{Var}(X) = \frac{\pi^2}{6}\sigma^2, \tag{A.9}$$

with the gamma function $\Gamma(\cdot)$ as defined in Eq. (A.4) and Eulers' constant γ (Soukissian and Tsalis, 2015; Wilks, 2011, Chpt. 4).

Equating Eqs. (A.8-A.9) with Eq. (2.4) results the moment estimators for the parameters of the Gumbel distribution:

$$\hat{\sigma} = \frac{\sqrt{6\,\hat{m}_2}}{\pi} \tag{A.10}$$

$$\hat{\mu} = \hat{m}_1 - \gamma\hat{\sigma}. \tag{A.11}$$

The analytical expressions for the probability weighted moments of the GEV are in the case $\xi \neq 0$ (Coles and Dixon, 1999)

$$\beta_r = (r+1)^{-1}\left\{\mu - \frac{\sigma}{\xi}\left[-(r+1)^{\xi}\Gamma(1-\xi)\right]\right\}, \quad \xi < 1 \tag{A.12}$$

and in the case $\xi = 0$, i.e., the Gumbel distribution (Mahdi and Cenac, 2005)

$$\beta_r = (r+1)^{-1}\left\{\mu - \sigma\left[\ln(r+1) + \gamma\right]\right\}. \tag{A.13}$$

A.3 Threshold weighted CRPS for a GEV distribution

We show the derivation the threshold weighted CRPS for a GEV distribution only for the case $\xi \neq 0$.
Friederichs and Thorarinsdottir (2012) derive the CRPS for a GEV distribution, starting with Eq. (2.34) and the definition of the QS (Eq. 2.31)

$$\begin{aligned}
\mathrm{CRPS}_{\mathrm{GEV}} &= 2\int_0^1 \mathrm{QS}(p)\,dp \\
&= 2\int_0^1 \rho_p\left[z - F^{-1}_{\mathrm{GEV}}(p)\right]\,dp \\
&= -\left\{z - \mu + \frac{\sigma}{\xi}\right\}\{1 - 2F_{\mathrm{GEV}}(z)\} \\
&\qquad - \frac{\sigma}{\xi}\left\{2^{\xi}\,\Gamma\left[1-\xi\right] - 2\,\Gamma_i\left[1-\xi, -\ln\left(F_{\mathrm{GEV}}(z)\right)\right]\right\},
\end{aligned} \tag{A.14}$$

where z is the observation, $F_{\mathrm{GEV}}(z)$ and $F^{-1}_{\mathrm{GEV}}(p)$ are the GEV distribution (see Eq. 2.1) and its quantile function (see Eq. 2.20) and $\Gamma(\cdot)$ and $\Gamma_i(\cdot,\cdot)$ are the gamma and incomplete gamma functions as defined in Eq. (A.4). An important step in their derivation is solving the two integrals

$$\int_{x_1}^1 \left[-\ln(p)\right]^{-\xi}\,dp = \Gamma_i\left[1-\xi, -\ln(x_1)\right] \tag{A.15}$$

$$\int_{x_2}^1 p\,\left[-\ln(p)\right]^{-\xi}\,dp = 2^{\xi-1}\Gamma_i\left[1-\xi, -2\ln(x_2)\right]. \tag{A.16}$$

Eq. (A.15) can be shown by substituting $t = -\ln(p)$ and for Eq. (A.16) the necessary substitution is $t = -2\ln(p)$. When inserting both relationships to obtain Eq. (A.14), the lower limits of the integrals are $x_1 = F_{\mathrm{GEV}}(z)$ and $x_2 = 0$, respectively.

To derive the threshold weighted twCRPS (Eq. 2.37) for a GEV distribution, we follow their derivation, first, writing the QS as

$$\mathrm{QS}(p) = p\,\left[z - F^{-1}(p)\right] - \mathbb{H}\left[p - F(z)\right]\cdot\left[z - F^{-1}(p)\right], \tag{A.17}$$

using the Heaviside step function

$$\mathbb{H}(x) = \begin{cases} 0 & , x < 0 \\ 1 & , x \geq 0. \end{cases} \tag{A.18}$$

Inserting this into Eq. (2.37) for the case of a GEV distribution results to

$$\begin{aligned}
\mathrm{twCRPS}_{\mathrm{GEV}}(\phi) &= 2\int_{\phi}^1 \rho_p\left[z - F^{-1}_{\mathrm{GEV}}(p)\right]\,dp \\
&= 2\underbrace{\int_{\phi}^1 p\,\left[z - F^{-1}_{\mathrm{GEV}}(p)\right]\,dp}_{A} - 2\underbrace{\int_{\phi}^1 \mathbb{H}\left[p - F_{\mathrm{GEV}}(z)\right]\cdot\left[z - F^{-1}_{\mathrm{GEV}}(p)\right]\,dp}_{B}.
\end{aligned} \tag{A.19}$$

We separately solve the terms A and B. Starting with A, expanding, inserting the GEV quantile function (Eq. 2.20) and using Eq. (A.16), where $x_2 = \phi$ leads to

$$A = z\left(1-\phi^2\right) - 2\int_{\phi}^{1} p\, F_{\mathrm{GEV}}^{-1}(p)\, dp \tag{A.20}$$

$$= z\left(1-\phi^2\right) - 2\int_{\phi}^{1} p\left\{\mu - \frac{\sigma}{\xi}\left[1-(-\ln(p))^{-\xi}\right]\right\} dp \tag{A.21}$$

$$= z\left(1-\phi^2\right) - \left(\mu - \frac{\sigma}{\xi}\right)\left(1-\phi^2\right) - \frac{\sigma}{\xi} 2^{\xi}\, \Gamma_i\left[1-\xi, -2\ln(\phi)\right]. \tag{A.22}$$

To solve term B we define

$$\phi' = \max\left\{\phi, F_{\mathrm{GEV}}(z)\right\}. \tag{A.23}$$

Inserting Eq. (A.18), Eq. (A.23), Eq. (2.20) and using Eq. (A.15) with $x_1 = \phi'$ results in

$$B = 2\int_{\phi'}^{1} z - F_{\mathrm{GEV}}^{-1}(z)\, dp \tag{A.24}$$

$$= z(2-2\phi') + 2\int_{\phi'}^{1} F_{\mathrm{GEV}}^{-1}(z)\, dp \tag{A.25}$$

$$= z(2-2\phi') + 2\int_{\phi'}^{1} \mu - \frac{\sigma}{\xi}\left[1-(-\ln(p))^{-\xi}\right] dp \tag{A.26}$$

$$= z(2-2\phi') + \left(\mu - \frac{\sigma}{\xi}\right)(2-2\phi') + 2\frac{\sigma}{\xi}\,\Gamma_i\left[1-\xi, -\ln(\phi')\right]. \tag{A.27}$$

Finally, inserting A (Eq. A.22) and B (Eq. A.27) into Eq. (A.19) leads to the twCRPS for the GEV distribution

$$\begin{aligned}\mathrm{twCRPS}_{\mathrm{GEV}}(\phi) = &\left\{z - \mu + \frac{\sigma}{\xi}\right\}\left\{-1 + 2\phi' - \phi^2\right\} \\ &- \frac{\sigma}{\xi}\left\{2^{\xi}\,\Gamma_i\left[1-\xi, -2\ln(\phi)\right] + 2\,\Gamma_i\left[1-\xi, -\ln\left(\phi'\right)\right]\right\}.\end{aligned} \tag{A.28}$$

Appendix B

Supporting Information for Study I (Chpt. 4)

B.1 Overview of Verification Variations

We performed the model verification with small variations in the cross-validation methods, to asses different aspects of the model performance:

1. the overall performance
2. the dependence of the model performance on the length of the time series used for training the model
3. the model performance at ungauged sites.

Table B.1 provides an overview of the differences in the cross-validation sets used for training and validation in all three cases for the spatial d-GEV and the reference model, the GEV applied separately for each station and duration.

Table B.1: Cross-validation sets used for training and validating the spatial d-GEV and the reference model (GEV). N represents the complete length of the time series and varies for different durations and stations, while n_t is a fixed number of years for each time series.

		Overall Performance		**Dependence on Time Series Length**		**Ungauged Sites**	
		Training	**Validation**	**Training**	**Validation**	**Training**	**Validation**
spatial d-GEV	station	$(N-3)$ years	3 years	n_t years	3 years	-	3 years
	remaining stations	all data	-	all data	-	all data	-
GEV	station	$(N-3)$ years	3 years	n_t years	3 years	n_t years	3 years
	remaining stations	-	-	-	-	-	-

B.2 Coverage of Confidence Intervals

We conduct a simulation study to investigate the coverage of the 95% confidence intervals computed with the delta method and the bootstrap method. Both methods are briefly explained in the following.

Using the delta method, the variance of an estimated quantile $q_{d,p}$ (see Equation (4.9) can be calculated as follows (Coles, 2001):

$$\mathrm{Var}(\hat{q}_{d,p}) \approx \nabla q_{d,p}^T V \nabla q_{d,p}, \tag{B.1}$$

where V is the variance-covariance matrix of the parameter estimations $(\hat{\tilde{\mu}}, \hat{\sigma}_0, \hat{\xi}, \hat{\theta}, \hat{\eta})$ and $\nabla q_{d,p}$ is the gradient

$$\nabla q_{d,p} = \left(\frac{\partial q_{d,p}}{\partial \tilde{\mu}}, \frac{\partial q_{d,p}}{\partial \sigma_0}, \frac{\partial q_{d,p}}{\partial \xi}, \frac{\partial q_{d,p}}{\partial \theta}, \frac{\partial q_{d,p}}{\partial \eta} \right)^T. \tag{B.2}$$

Assuming that the maximum likelihood estimator $\hat{q}_{d,p}$ follows a Normal distribution, the confidence intervals can be calculated from the variance $\mathrm{Var}(\hat{q_{d,p}})$. However, this assumption may be poor. To take into account the dependence of the maxima of different durations and stations in the estimation of uncertainties, the delta method can be adjusted accordingly (Van de Vyver, 2012). In this case, however, we did not apply this adjustment, since we verified the coverage only for confidence intervals estimated from independent maxima.

A common and simplistic method for estimating confidence intervals without the assumption of normality is bootstrapping. Here we apply the ordinary non-parametric bootstrap percentile method. Therefore, firstly a sample is created by drawing from the data with replacement, and then the model parameters are estimated. From those, we calculate the resulting return levels. By 1000 repetitions of this process, we obtain a distribution of return levels. We estimate the lower and upper bound of the confidence interval from the empirical 0.025 and 0.975 quantiles of the distribution of return levels. When we apply the bootstrap method to the observations in the study area, we sample the years by randomly drawing with replacement. Therefore all available maxima for a given year are used collectively in the bootstrap sample. We expect that in this way the dependency structure of the observations is taken into account.

To analyze the coverage of both described methods, we proceed as follows: We draw random values from a d-GEV distribution with the parameters $\tilde{\mu} = 3, \sigma_0 = 5, \xi = 0.06, \theta = 0.05, \eta = 0.7$ to get a sample of size n. The parameters are chosen to be comparable to those estimated for individual stations in the research area. However, in this sample the maxima are independent of each other, contrary to what we expect for the observations. For this sample, the 95%-confidence intervals are estimated for a given quantile $q_{d,p}$ using both methods. Then it is tested whether the actual value for $q_{d,p}$ is included in the estimated confidence intervals. This process is repeated 1000 times. Finally, the coverage of the confidence interval is estimated from the relative frequency of how often the actual value was within the confidence interval. The results are presented in Figure B.1. The coverage for the delta method intervals varies with duration and probability and deviates strongly from 95%. The bootstrapped confidence intervals, on the other hand, show a reasonable behavior.

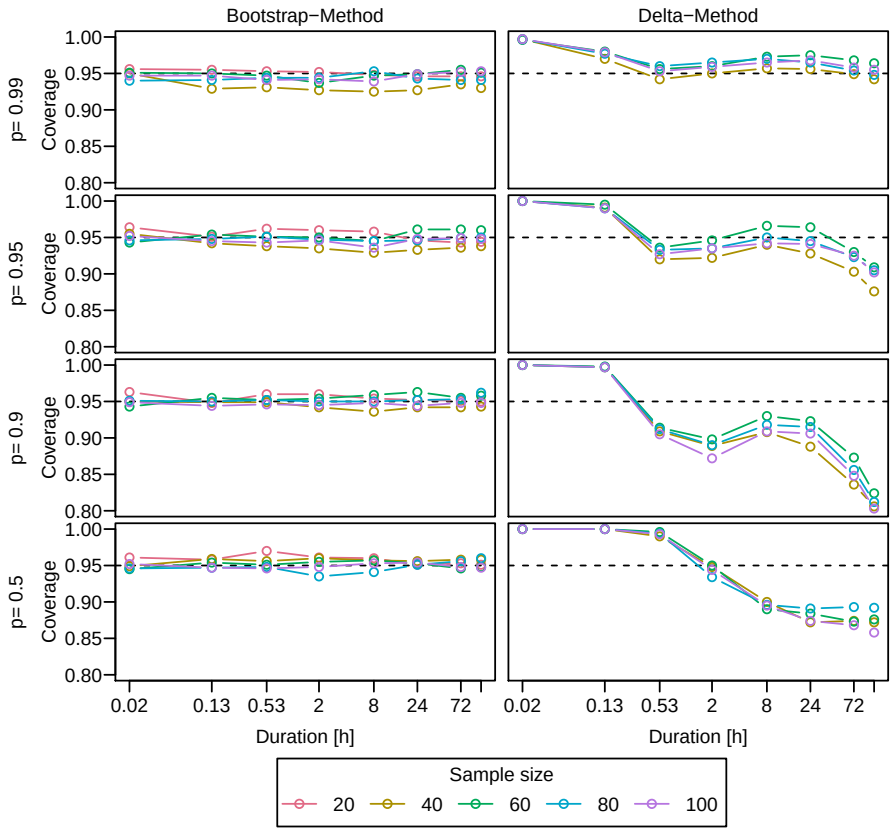


Figure B.1: Coverage of Confidence Intervals, obtained by the bootstrap method (**left column**) and the delta method (**right column**). The coverage was calculated by re-sampling from a known d-GEV distribution 1000 times. Different colors indicate different sample sizes, which correspond to the length of the time series in years in this context. Different rows represent different non-exceedance probabilities p, whereby the confidence intervals were examined for the corresponding quantile estimates.

Appendix C

Supporting Information for Study II (Chpt. 5)

C.1 Initial Values

The estimation of d-GEV parameters was conducted with the R base `optim()` function (R Core Team, 2020) and the Nelder-Mead method. The quality of the fitted model depends on the initial values passed to the function. Each optimization i was repeated m times with different initial values that are derived through different techniques. The sets of initial functions, i.e., $s_i \in \{\tilde{\mu}', \sigma_0', \xi', \theta', \eta', \eta_2', \tau'\}$, were collected as suggestions, and the individual parameters were named with version indices (v1, v2, etc.). All suggestions were subsequently used as initial values in the model, and the suggestion which led to the smallest negative log likelihood was selected. Table C.1 gives an overview of the combinations of initial values.

All initial-value techniques were based on the same first step. An individual GEV distribution was fitted to each duration d separately, with moment estimators as initial values (Coles, 2001), and the three GEV parameters of location μ, scale σ, and shape ξ were stored for each duration. In the next step, a function was fitted to each of the parameters with respect to the duration. Since we assumed no dependence of the shape parameter on the duration, we chose $\xi'_{\mathrm{v1}} = \mathrm{median}(\xi)$ for all suggestions. According to Eqs. (5.4) and (5.5), we fitted $\ln(\sigma(d))$ and $\ln(\mu(d))$ as a function of $\ln(d)$ in a linear regression with a simple slope and y intercept as follows:

$$\ln(\sigma(d)) \sim -(\eta + \eta_2)\ln(d) + \ln(\sigma_0) \tag{C.1}$$

$$\ln(\mu(d)) \sim -\eta \ln(d) + (\ln(\sigma_0) + \ln(\tilde{\mu})) \tag{C.2}$$

with given $\sigma(d)$, $\mu(d)$, and d. From this fit, we extracted $\sigma'_{0,\mathrm{v1}} = \exp(\ln(\sigma_0))$ (Eq. C.1), $\tilde{\mu}'_{\mathrm{v1}} = \tilde{\mu}$ (combine Eqs. C.1 and C.2), $\eta'_{\mathrm{v1}} = \eta$, and $\eta'_{2,\mathrm{v1}} = \eta_2$. For the most simple suggestion of initial parameters, we chose $\theta'_{\mathrm{v1}} = 0$ and $\tau'_{\mathrm{v1}} = 0$. In the next steps, we further elaborated the ways of finding good initial values. Version 2 of the duration exponents η'_{v2} and $\eta'_{2,\mathrm{v2}}$ were found, using only d$\geq$1 h, because the slope is mainly characterized by this duration regime. This is the second set of suggestions for initial values, together with version 1 of the other parameters.

Table C.1: Overview of initial value combinations (suggestions). The initial values for the parameters $\tilde{\mu}'_{\mathrm{v1}}$, $\sigma'_{0,\mathrm{v1}}$ and ξ'_{v1} are the same in all combinations.

Suggestion no.	**Version**			
	θ'	η'_1	η'_2	τ'
1	v1	v1	v1	v1
2	v1	v2	v2	v1
3	v2	v1	v1	v1
4	v2	v2	v2	v1
5	v1	v1	v1	v2
6	v1	v2	v2	v2
7	v2	v1	v1	v2
8	v2	v2	v2	v2

Another initial duration offset θ'_{v2} could be estimated by fitting a nonlinear squares regression (R function `nls`) as follows:

$$\ln(\sigma(d)) \sim -(\eta + \eta_2)\ln(d + \theta) + \ln(\sigma_0) \tag{C.3}$$

$$\ln(\mu(d)) \sim -\eta\ln(d + \theta) + \ln(\sigma_0\tilde{\mu}) \tag{C.4}$$

with given $\sigma(d)$, $\mu(d)$, and d. The mean of both estimates for θ was used for θ'_{v2}. These functions were less stable and provided worse initial values for σ'_0 and $\tilde{\mu}'$ than Eq. (C.1) and C.2. That is why we used them only for estimating initial θ'_{v2} in the `nls` function and not for estimating initial $\tilde{\mu}$, σ_0, η_1, or η_2. The new initial estimate θ'_{v2} was combined with η'_{v1} and $\eta'_{2,\mathrm{v1}}$ in one set (suggestion 3) and with η'_{v2} and $\eta'_{2,\mathrm{v2}}$ in another set (suggestion 4).

The same `nls` function was used for an estimation of the initial τ'_{v2}, taking only d≥1 h and no duration offset (curvature), as follows:

$$\ln(\sigma(d)) \sim \ln(\sigma_0 d^{-(\eta+\eta_2)} + \tau) \tag{C.5}$$

$$\ln(\mu(d)) \sim \ln(\sigma_0\tilde{\mu}d^{-\eta} + \tau) \tag{C.6}$$

with given $\sigma(d)$, $\mu(d)$, and d. Again, the mean of both estimates of τ was used as τ'_{v2}. To define suggestions 5-8, this second version of τ' was combined with θ'_{v1}, η'_{v1}, $\eta'_{2,\mathrm{v1}}$ or θ'_{v1}, η'_{v2}, $\eta'_{2,\mathrm{v2}}$ or θ'_{v2}, η'_{v1}, $\eta'_{2,\mathrm{v1}}$ or θ'_{v2}, η'_{v2}, $\eta'_{2,\mathrm{v2}}$. The different combinations of initial value versions are listed again in Table C.1.

C.2 Simulated Data

For the coverage analysis in Sect. 5.3.3 and Appendix C.3 about duration sample choice, we did not use the original data set but simulated data from a d-GEV distribution. The simulated data were drawn, according to Eqs. (5.3), (5.10), and (5.11), with a random number $0 < p < 1$ and the parameters $\tilde{\mu} = 3.2$, $\sigma_0 = 5.8$, $\xi = 0.21$, $\theta = 0.089$, $\eta = 0.78$,

$\eta_2 = 0.09$, and $\tau = 0.10$. These values were based on realistic parameter values from station data, fitted for the features of curvature, multiscaling, and flattening. When disabling one or more of the features, the values of all the parameters will change. For the coverage analysis, the true quantile intensity z could be calculated directly, using these parameters and Eqs. (5.3), (5.10), and (5.11).

C.3 Influence of Duration Sample Choice

We investigate how the choice of durations that are used to train the model influences the model performance. Since the number of training data points is much higher for long durations $d \geq 24$ (one day), there is a possibility that these duration regimes are overrepresented in the training phase, and thus, model performance is worse for short durations. To account for this effect, the model is trained twice, (1) with simulated maxima (Appendix C.2) and the set of aggregated durations that was used for analysis in this study (Eqs. 5.1) and (2) with simulated annual maxima of a different set of durations that focuses more on short durations (numerically in hours) as follows:

$$\begin{aligned} d_2 \in \{&1, 2, 3, ..., 15, 16, 18, 20, ..., 30, 32, 33, 36, 39, ..., 57\,\text{min}, \\ &1, 2, 3, ..., 6, 8, 10, 12, 15, 16, 18, 21\,\text{h}, 1, 3, 5\,\text{d}\}. \end{aligned} \tag{C.7}$$

In this way, there is more training data for short durations available, which might shift the model's performance focus to other duration regimes. However, it is important to note that this is only an artificial increase in available data, since the additional data points do not contain substantial new information.

The model with the new artificial training data set (Eq. C.7) is now verified against the same model with the previously used artificial data set (Eq. 5.1) with the following results: all QSIs are below 0.05 for all durations d and all quantiles $q \in \{0.5, 0.8, 0.9, 0.95, 0.98\}$ (not shown). Thus, the results do not indicate that the choice of accumulation duration significantly influences how well the model performs for certain duration regimes.

C.4 Model Diagnosis

In order to evaluate whether the GEV distribution is an appropriate choice for this analysis, we provide quantile-quantile (QQ) plots (Fig. C.1) for the stations of Bever and Buchenhofen, as chosen in Sect. 5.3.2. While for Buchenhofen all values follow the angle bisecting line, for Bever only a few outlying events, which all correspond to higher quantiles, leave the confidence intervals. However, their number is small compared to the number of shown data points. So, we conclude that the GEV distribution is a suitable assumption in our case.

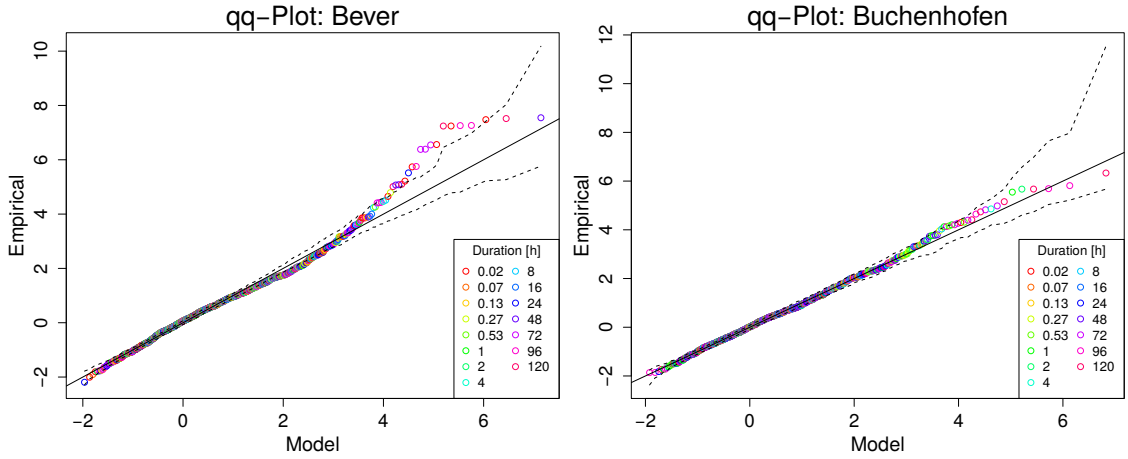


Figure C.1: QQ plots for selected stations. Confidence intervals were obtained by simulating transformed Fréchet distributed values from the model distribution and extracting a 95% interval.

C.5 Overview of Reference Models for Verification

In Sect. 5.3.1, the feature skill was evaluated by comparing models with a reference where the considered feature is disabled. For clarification, Table C.2 lists the models and reference models that were used in Figs. 5.3 and 5.4, together with the parameter restriction to zero, if applied. These specifications refer to Eqs. (5.10) and (5.11).

Table C.2: Overview of models and references for verification.

Column (title); row (number)	**Model features and parameters**	**Reference features and parameters**
Curvature		
1	$\mathrm{IDF_c}$	IDF
	$\theta \neq 0, \eta_2 = 0, \tau = 0$	$\theta = 0, \eta_2 = 0, \tau = 0$
2	$\mathrm{IDF_{cm}}$	$\mathrm{IDF_m}$
	$\theta \neq 0, \eta_2 \neq 0, \tau = 0$	$\theta = 0, \eta_2 \neq 0, \tau = 0$
3	$\mathrm{IDF_{cf}}$	$\mathrm{IDF_f}$
	$\theta \neq 0, \eta_2 = 0, \tau \neq 0$	$\theta = 0, \eta_2 = 0, \tau \neq 0$
4	$\mathrm{IDF_{cmf}}$	$\mathrm{IDF_{mf}}$
	$\theta \neq 0, \eta_2 \neq 0, \tau \neq 0$	$\theta = 0, \eta_2 \neq 0, \tau \neq 0$
Multiscaling		
1	$\mathrm{IDF_m}$	IDF
	$\theta = 0, \eta_2 \neq 0, \tau = 0$	$\theta = 0, \eta_2 = 0, \tau = 0$
2	$\mathrm{IDF_{cm}}$	$\mathrm{IDF_c}$
	$\theta \neq 0, \eta_2 \neq 0, \tau = 0$	$\theta \neq 0, \eta_2 = 0, \tau = 0$
3	$\mathrm{IDF_{mf}}$	$\mathrm{IDF_f}$
	$\theta = 0, \eta_2 \neq 0, \tau \neq 0$	$\theta = 0, \eta_2 = 0, \tau \neq 0$
4	$\mathrm{IDF_{cmf}}$	$\mathrm{IDF_{cf}}$
	$\theta \neq 0, \eta_2 \neq 0, \tau \neq 0$	$\theta \neq 0, \eta_2 = 0, \tau \neq 0$
Flattening		
1	$\mathrm{IDF_f}$	IDF
	$\theta = 0, \eta_2 = 0, \tau \neq 0$	$\theta = 0, \eta_2 = 0, \tau = 0$
2	$\mathrm{IDF_{cf}}$	$\mathrm{IDF_c}$
	$\theta \neq 0, \eta_2 = 0, \tau \neq 0$	$\theta \neq 0, \eta_2 = 0, \tau = 0$
3	$\mathrm{IDF_{mf}}$	$\mathrm{IDF_m}$
	$\theta = 0, \eta_2 \neq 0, \tau \neq 0$	$\theta = 0, \eta_2 \neq 0, \tau = 0$
4	$\mathrm{IDF_{cmf}}$	$\mathrm{IDF_{cm}}$
	$\theta \neq 0, \eta_2 \neq 0, \tau \neq 0$	$\theta \neq 0, \eta_2 \neq 0, \tau = 0$

Appendix D

Supporting Information for Study III (Chpt. 6)

D.1 Model Diagnosis for Station Bever-Talsperre

The distribution of monthly maxima for a range of durations is modeled using the d-GEV distribution. We visually inspect, whether the d-GEV distribution is a reasonable approximation for the distribution of monthly maxima at the considered stations using quantile-quantile (q-q) plots. As an example, we present the q-q plots for the station Bever-Talsperre with respect to each month in Fig. D.1. The different aggregation times are indicated by different colors. We find that the d-GEV distribution describes the monthly maxima sufficiently well for each month. There are few outliers that correspond to the limits of the duration range, i.e. very short or very long durations.

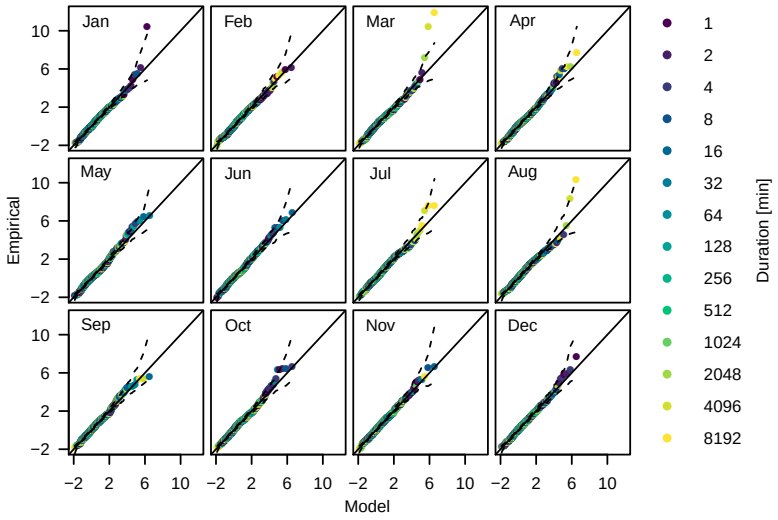


Figure D.1: Diagnostic q-q plots of the d-GEV model for each month at station Bever-Talsperre. The observations and the modeled quantiles are transformed to standard Gumbel $G(\mu = 0, \sigma = 1, \xi = 0)$ to remove the duration dependency. Dashed lines represent 95% confidence intervals.

D.2 Annual Maxima Station Bever-Talsperre

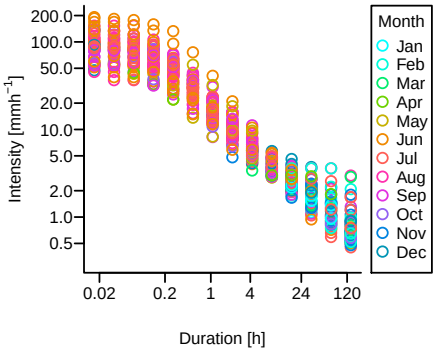


Figure D.2: Annual intensity maxima for different durations at station Bever-Talsperre. The months in which the respective maxima occurred are represented by colors.

The annual maxima of different durations for the station Bever-Talsperre are presented in Fig. D.2. The colors indicate the month in which the respective maxima occurred. It is evident that at this station the maxima of short durations occur mostly in summer, while from $d \gtrsim 8\,\mathrm{h}$ the maxima originate from different seasons, especially autumn and winter. This is in agreement with the results presented in Fig. 4 (top left). In addition, we find that the span of the data exhibits a minimum at $d \approx 8\,\mathrm{h}$. This is consistent with the minimum value of the shape parameter ξ at this duration observed in Fig. 7 (c).

D.3 Phase Difference in Seasonal Variations of Location and Scale Parameter

From the parameter estimates of $\tilde{\mu}$ in Fig. 2, it can not be clearly determined whether modeling variations of $\tilde{\mu}$ throughout the year is a reasonable choice. Since the modified location parameter is defined as $\tilde{\mu} = \mu(d)/\sigma(d)$, setting $\tilde{\mu}(\mathrm{doy}) = \mathrm{const.}$ would result in not allowing a phase difference in the annual cycle of the location parameter μ and the scale parameter σ for any fixed duration d. Maraun et al. (2009) investigated this relationship for daily precipitation sums in the UK and found that this assumption is not justified, because the annual cycles of these two parameters are slightly out of phase. We investigate this in an exploratory analysis by modeling the individual durations. For this purpose, we model the monthly maxima of each duration separately using a GEV distribution with monthly covariates. Fig. D.3 shows the resultig parameter estimates μ and σ for some durations at station Bever-Talsperre in the first two columns as lines. For comparison, the estimates resulting from separately modeling the maxima of each duration and month with a GEV distribution are shown as dots. The right column presents the resulting estimates for μ/σ for each duration. Since μ/σ shows a clear variation for each duration, we conclude that assuming no phase shift in the annual cycle of μ and σ for each duration may restrict the model too much.

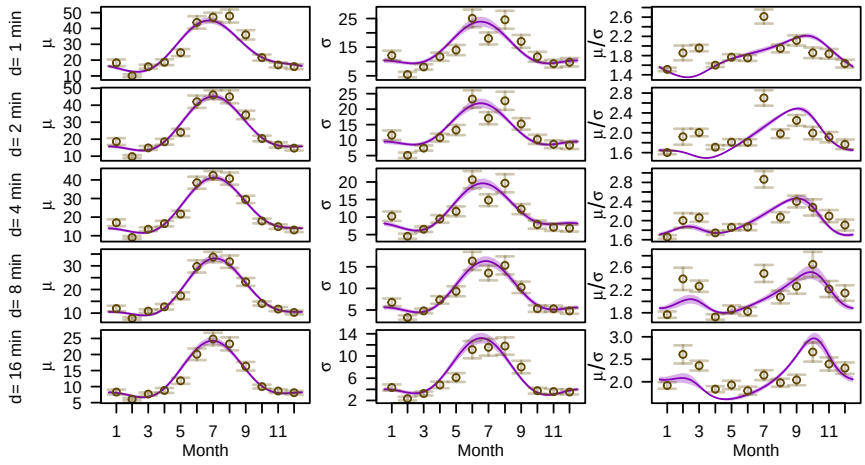


Figure D.3: Parameter estimates for fixed durations modeling monthly maxima using (1) a GEV with monthly covariates (purple lines) and (2) a separate GEV model for each month (dots). The error bars and shaded areas show the 95% confidence intervals obtained via the estimated Fisher information matrix.

Appendix E

Supporting Information for Chapter 7

E.1 Station Data Used to Estimate the IDF Relationship

To model the spatial variations of precipitation intensity in Berlin and the surrounding area, we use precipitation data from DWD stations in Berlin and Brandenburg. We only consider gauge stations where measurements with a minutely measurement frequency are available. In Berlin and Brandenburg this results in 50 stations, the locations of which are presented in Figure E.1. At each of these stations, we use all available data with the measurement frequencies: minutely, 10-minutely, hourly, and daily, and combine them into one time series per station. The length of the time series is different for different stations and measurement frequencies. Some time series for daily precipitation sums begin before the year 1900, while the measurements with minutely resolution are only available since approximately the mid-1990s.

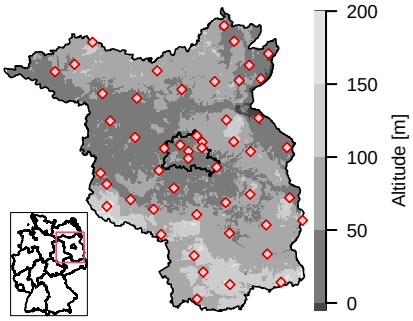


Figure E.1: Locations of all 50 stations within Berlin and Brandenburg used for the analysis. The gray scale indicates the altitude, obtained from `http://www.diva-gis.org/gdata`.

E.2 Radar-Based Data for the Analysis of Selected Events

In order to analyze a precipitation event, particularly with respect to its spatial extent, a gridded precipitation data set is required. We use the product RADKLIM provided by the DWD, which is derived from measurements of the German radar network [30]. The 5-minute precipitation rates estimated based on the radar data are adjusted to measurements from gauge stations and subjected to quality control by the DWD [29]. The data are available on a grid with a size of 1 km x 1 km in polar stereographic projection and cover the period 2001 to 2019. For the analysis of the events discussed here, the radar product for the periods 29.06. - 30.06.2017 and 02.08.2019 was projected onto a regular grid and cropped to the Berlin area.

E.3 Spatial d-GEV Model

To estimate IDF curves and return level maps, we apply the methods described in Sections 4.2.2-4.2.5 to the data set of annual maxima at stations in Berlin and Brandenburg described in Section E.1. However, we adopt a modification for the spatial covariates of the d-GEV distribution (Eq. 4.11). In addition to longitude and latitude, we use the altitude of the stations as covariates, as suggested by Fischer et al. (2019). This results in the following model for each d-GEV parameter $\phi \in \{\tilde{\mu}, \sigma_0, \xi, \theta, \eta\}$:

$$\begin{aligned}\phi = \phi_0 &+ \sum_{j=1}^{J} \alpha_j^{\phi} P_j(\text{lon}) + \sum_{k=1}^{K} \beta_k^{\phi} P_k(\text{lat}) + \sum_{l=1}^{L} \gamma_l^{\phi} P_l(\text{alt}) + \sum_{j=1}^{J}\sum_{k=1}^{K} \delta_{j,k}^{\phi} P_j(\text{lon}) P_k(\text{lat}) \\ &+ \sum_{j=1}^{J}\sum_{l=1}^{L} \epsilon_{j,l}^{\phi} P_j(\text{lon}) P_l(\text{alt}) + \sum_{k=1}^{K}\sum_{l=1}^{L} \zeta_{k,l}^{\phi} P_k(\text{lat}) P_l(\text{alt}),\end{aligned} \tag{E.1}$$

with the intercept ϕ_0 and regression coefficients $\alpha_j^{\phi}, \beta_k^{\phi}, \gamma_l^{\phi}, \delta_{j,k}^{\phi}, \epsilon_{j,l}^{\phi}, \zeta_{k,l}^{\phi}$ and orthogonal polynomials $P(\cdot)$ of the covariates with the maximum order of $J = K = L = 6$.

Based on the spatial d-GEV model for Berlin and Brandenburg, the annual exceedance probabilities of the selected events for different durations are estimated as follows: The event data, described in Section E.2, are aggregated at each grid point to time series for a range of durations between 5 min and 48 h with $d \in \{5, 10, 15, ..., 2880\}$ min. For each duration and grid point, the maximum precipitation intensity of the time series is considered. The annual non-exceedance probability $p = G(z, d)$ for intensity z follows from Eq. 4.8, where the annual exceedance probability is $1 - p$ and the average return period is $1/(1 - p)$.

Bibliography

Agilan, V. and Umamahesh, N. V.: What are the best covariates for developing non-stationary rainfall Intensity-Duration-Frequency relationship?, Adv. Water Resour., 101, 11–22, doi:10.1016/j.advwatres.2016.12.016, 2017.

Ariff, N., Jemain, A., Ibrahim, K., and Wan Zin, W.: IDF relationships using bivariate copula for storm events in Peninsular Malaysia, J. Hydrol., 470-471, 158–171, doi:10.1016/j.jhydrol.2012.08.045, 2012.

Arlot, S. and Celisse, A.: A survey of cross-validation procedures for model selection, Statist. Surv., 4, 40–79, doi:10.1214/09-SS054, 2010.

Asadi, P., Davison, A. C., and Engelke, S.: Extremes on river networks, Ann. Appl. Stat., 9, 2023 – 2050, doi:10.1214/15-AOAS863, 2015.

Barbero, R., Fowler, H. J., Lenderink, G., and Blenkinsop, S.: Is the intensification of precipitation extremes with global warming better detected at hourly than daily resolutions?, Geophys. Res. Lett., 44, 974–983, doi:10.1002/2016GL071917, 2017.

Barredo, J. I.: Normalised flood losses in Europe: 1970–2006, Nat. Hazards Earth Syst. Sci., 9, 97–104, doi:10.5194/nhess-9-97-2009, 2009.

Benestad, R. E., Lutz, J., Dyrrdal, A. V., Haugen, J. E., Parding, K. M., and Dobler, A.: Testing a simple formula for calculating approximate intensity-duration-frequency curves, Environ. Res. Lett., 16, 044 009, doi:10.1088/1748-9326/abd4ab, 2021.

Bentzien, S. and Friederichs, P.: Decomposition and graphical portrayal of the quantile score, Q. J. R. Meteorol. Soc., 140, 1924–1934, doi:10.1002/qj.2284, 2014.

Berg, P., Moseley, C., and Haerter, J.: Strong increase in convective precipitation in response to higher temperatures, Nat. Geosci., 6, 181–185, doi:10.1038/ngeo1731, 2013.

Berghäuser, L., Schoppa, L., Ulrich, J., Dillenardt, L., Jurado, O. E., Passow, C., Mohor, G. S., Seleem, O., Petrow, T., and Thieken, A. H.: Starkregen in Berlin: Meteorologische Ereignisrekonstruktion und Betroffenenbefragung, Universität Potsdam: Brandenburg, Germany, doi:10.25932/publishup-50056, 2021.

Bezak, N., Šraj, M., and Mikoš, M.: Copula-based IDF curves and empirical rainfall thresholds for flash floods and rainfall-induced landslides, J. Hydrol., 541, 272–284, doi:10.1016/j.jhydrol.2016.02.058, 2016.

Blanchet, J., Ceresetti, D., Molinié, G., and Creutin, J.-D.: A regional GEV scale-invariant framework for Intensity–Duration–Frequency analysis, J. Hydrol., 540, 82–95, doi:10.1016/j.jhydrol.2016.06.007, 2016.

Bougadis, J. and Adamowski, K.: Scaling model of a rainfall intensity-duration-frequency relationship, Hydrol. Process., 20, 3747–3757, doi:10.1002/hyp.6386, 2006.

Brémond, P., Grelot, F., and Agenais, A.-L.: Review Article: Economic evaluation of flood damage to agriculture – review and analysis of existing methods, Nat. Hazards Earth Syst. Sci., 13, 2493–2512, doi:10.5194/nhess-13-2493-2013, 2013.

Bronstert, A.: Floods and Climate Change: Interactions and Impacts, Risk Anal., 23, 545–557, doi:10.1111/1539-6924.00335, 2003.

Bronstert, A., Agarwal, A., Boessenkool, B., Crisologo, I., Fischer, M., Heistermann, M., Köhn-Reich, L., López-Tarazón, J. A., Moran, T., Ozturk, U., Reinhardt-Imjela, C., and Wendi, D.: Forensic hydro-meteorological analysis of an extreme flash flood: The 2016-05-29 event in Braunsbach, SW Germany, Sci. Total Environ., 630, 977–991, doi:10.1016/j.scitotenv.2018.02.241, 2018.

Bücher, A., Lilienthal, J., Kinsvater, P., and Fried, R.: Penalized quasi-maximum likelihood estimation for extreme value models with application to flood frequency analysis, Extremes, 24, 325–348, doi:10.1007/s10687-020-00379-y, 2021.

Burlando, P. and Rosso, R.: Scaling and muitiscaling models of depth-duration-frequency curves for storm precipitation, J. Hydrol., 187, 45 – 64, doi:10.1016/S0022-1694(96)03086-7, fractals, scaling and nonlinear variability in hydrology, 1996.

Cheng, L. and AghaKouchak, A.: Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate, Sci. Rep., 4, 1–6, doi:10.1038/srep07093, 2014.

Chow, V. T.: Frequency analysis of hydrologic data with special application to rainfall intensities, Univ. Ill. Bull., 50, 86, 1953.

Chow, V. T., Maidment, D. R., and Mays, L. W.: Applied Hydrology, McGraw-Hill series in water resources and environmental engineering, Tata McGraw-Hill Education, 1988.

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, Springer, doi:10.1198/tech.2002.s73, 2001.

Coles, S. G. and Dixon, M. J.: Likelihood-based inference for extreme value models, Extremes, 2, 5–23, doi:10.1023/A:1009905222644, 1999.

Courty, L. G., Wilby, R. L., Hillier, J. K., and Slater, L. J.: Intensity-duration-frequency curves at the global scale, Environ. Res. Lett., 14, 084 045, doi:10.1088/1748-9326/ab370a, 2019.

Davenport, F. V., Burke, M., and Diffenbaugh, N. S.: Contribution of historical precipitation change to US flood damages, P. Natl. Acad. Sci. USA, 118, doi:10.1073/pnas.2017524118, 2021.

Davini, P. and D'Andrea, F.: From CMIP3 to CMIP6: Northern Hemisphere Atmospheric Blocking Simulation in Present and Future Climate, J. Climate, 33, 10 021 – 10 038, doi:10.1175/JCLI-D-19-0862.1, 2020.

Davison, A. C. and Gholamrezaee, M. M.: Geostatistics of extremes, P. Roy. Soc. A, 468, 581–608, doi:10.1098/rspa.2011.0412, 2012.

Davison, A. C. and Hinkley, D. V.: Bootstrap methods and their application, 1, Cambridge university press, 1997.

Davison, A. C., Padoan, S. A., and Ribatet, M.: Statistical modeling of spatial extremes, Stat. Sci., 27, 161–186, doi:10.1214/11-STS376, 2012.

Detring, C., Müller, A., Schielicke, L., Névir, P., and Rust, H. W.: Occurrence and transition probabilities of omega and high-over-low blocking in the Euro-Atlantic region, Weather Clim. Dyn., 2, 927–952, doi:10.5194/wcd-2-927-2021, 2021.

Doswell, C. A., Brooks, H. E., and Maddox, R. A.: Flash Flood Forecasting: An Ingredients-Based Methodology, Weather Forecast., 11, 560 – 581, doi:10.1175/1520-0434(1996)011<0560:FFFAIB>2.0.CO;2, 1996.

Dotterweich, M.: The history of soil erosion and fluvial deposits in small catchments of central Europe: Deciphering the long-term interaction between humans and the environment — A review, in: The 39th Annual Binghamton Geomorphology Symposium: Fluvial Deposits and Environmental History: Geoarchaeology, Paleohydrology, and Adjustment to Environmental Change, Geomorphology, 101, 192–208, doi:10.1016/j.geomorph.2008.05.023, 2008.

Durrans, S. R.: Intensity-Duration-Frequency Curves, in: Rainfall: State of the Science, pp. 159–169, American Geophys. Union (AGU), doi:10.1029/2009GM000919, 2010.

Durrans, S. R., Eiffe, M. A., Thomas, W. O., and Goranflo, H. M.: Joint Seasonal/Annual Flood Frequency Analysis, J. Hydrol. Eng., 8, 181–189, doi:10.1061/(ASCE)1084-0699(2003)8:4(181), 2003.

DWA-Arbeitsgruppe HW-1.1: DWA-A 531: Starkregen in Abhängigkeit von Wiederkehrzeit und Dauer, Hennef, Germany, 2017.

DWD: KOSTRA-Atlas, available online: `https://www.dwd.de/DE/leistungen/kostra_dwd_rasterwerte/kostra_dwd_rasterwerte.html`, last access: 9 June 2021.

DWD Climate Data Center (CDC): Historical 1-minute station observations of precipitation for Germany, version v1, `https: //opendata.dwd.de/climate_environment/CDC/observations_ germany/climate/1_minute/precipitation/historical/`, last access: 5 March 2021, 2019.

DWD Climate Data Center (CDC): Historical daily precipitation observations for Germany, version v21.3, `https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/more_precip/historical/`, last access: 22 February 2022, 2021.

Dyrrdal, A. V., Lenkoski, A., Thorarinsdottir, T. L., and Stordal, F.: Bayesian hierarchical modeling of extreme hourly precipitation in Norway, Environmetrics, 26, 89–106, doi:10.1002/env.2301, 2015.

Erhardt, R. and Sisson, S. A.: Modelling Extremes Using Approximate Bayesian Computation, in: Extreme value modeling and risk analysis: Methods and applications, edited by Dey, D. K. and Yan, J., chap. 14, pp. 281–306, CRC Press, Boca Raton, USA, doi:10.1201/b19721, 2016.

Fauer, F. S., Ulrich, J., Jurado, O. E., and Rust, H. W.: Flexible and Consistent Quantile Estimation for Intensity-Duration-Frequency Curves, Hydrol. Earth Syst. Sci., 25, 6479–6494, doi:10.5194/hess-25-6479-2021, 2021a.

Fauer, F. S., Ulrich, J., Jurado, O. E., and Rust, H. W.: Annual Maxima of Station-based Rainfall Data over Different Accumulation Durations, doi:10.5281/zenodo.5012621, 2021b.

Fial, K.: Naturgefahrenreport 2021: Die Schaden-Chronik der deutschen Versicherer, Gesamtverband der Deutschen Versicherungswirtschaft e. V., `https://www.gdv.de/resource/blob/71294/ebaad3ff1563be2b92e0dd0ce4c0751c/download-naturgefahren-report-data.pdf`, last access: 10 February 2022, 2021.

Fischer, E. M., Sedláček, J., Hawkins, E., and Knutti, R.: Models agree on forced response pattern of precipitation and temperature extremes, Geophys. Res. Lett., 41, 8554–8562, doi:10.1002/2014GL062018, 2014.

Fischer, M., Rust, H. W., and Ulbrich, U.: Seasonal Cycle in German Daily Precipitation Extremes, Meteorol. Z., 27, 3–13, doi:10.1127/metz/2017/0845, 2018.

Fischer, M., Rust, H., and Ulbrich, U.: A spatial and seasonal climatology of extreme precipitation return-levels: A case study, Spat. Stat., 34, doi:10.1016/j.spasta.2017.11.007, 2019.

Friederichs, P.: Statistical downscaling of extreme precipitation events using extreme value theory, Extremes, 13, 109–132, doi:10.1007/s10687-010-0107-5, 2010.

Friederichs, P. and Thorarinsdottir, T. L.: Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction, Environmetrics, 23, 579–594, doi:10.1002/env.2176, 2012.

Fukutome, S., Schindler, A., and Capobianco, A.: MeteoSwiss extreme value analyses: User manual and documentation, Tech. Rep. , 255, 3rd Edition, 80 pp., Federal Office of

Meteorology and Climatology, MeteoSwiss, Zürich, Switzerland, URL `https://www.meteoswiss.admin.ch/content/dam/meteoswiss/de/service-und-publikationen/Publikationen/doc/nidex_technical_report_20181114.pdf`, 2018.

Ganguli, P. and Coulibaly, P.: Does nonstationarity in rainfall require nonstationary intensity–duration–frequency curves?, Hydrol. Earth Syst. Sci., 21, 6461–6483, doi:10.5194/hess-21-6461-2017, 2017.

García-Bartual, R. and Schneider, M.: Estimating maximum expected short-duration rainfall intensities from extreme convective storms, Phys. Chem. Earth Part B, 26, 675 – 681, doi:10.1016/S1464-1909(01)00068-5, 2001.

Gaur, A., Schardong, A., and Simonovic, S. P.: Gridded Extreme Precipitation Intensity–Duration–Frequency Estimates for the Canadian Landmass, J. Hydrol. Eng., 25, 05020 006, doi:10.1061/(ASCE)HE.1943-5584.0001924, 2020.

Gebauer, P.: Berliner Wetterkarte vom 29.06.2017, Berliner Wetterkarte e.V., 66 Nr. 124, 2017.

Gebauer, P., Myrcik, G., and Schenk, F.: Beiträge zur Berliner Wetterkarte: Berlin unter Wasser, Berliner Wetterkarte e.V., `https://berliner-wetterkarte.de/Beilagen/2017/BWK_Beitraege_20170714_Berlin_unter_Wasser.pdf`, last access: 11 January 2022, 2017.

Gneiting, T. and Ranjan, R.: Comparing Density Forecasts Using Threshold- and Quantile-Weighted Scoring Rules, J. Bus. Econ. Stat., 29, 411–422, doi:10.1198/jbes.2010.08110, 2011.

Goudenhoofdt, E., Delobbe, L., and Willems, P.: Regional frequency analysis of extreme rainfall in Belgium based on radar estimates, Hydrol. Earth Syst. Sci., 21, 5385–5399, doi:10.5194/hess-21-5385-2017, 2017.

Grieser, J., Staeger, T., and Schonwiese, C.-D.: Estimates and uncertainties of return periods of extreme daily precipitation in Germany, Meteorol. Z., 16, 553–564, doi:10.1127/0941-2948/2007/0235, 2007.

Grounds, M. A., LeClerc, J. E., and Joslyn, S.: Expressing Flood Likelihood: Return Period versus Probability, Weather Clim. Soc., 10, 5 – 17, doi:10.1175/WCAS-D-16-0107.1, 2018.

Guo, S.: A discussion on unbiased plotting positions for the general extreme value distribution, J. Hydrol., 121, 33–44, doi:10.1016/0022-1694(90)90223-K, 1990.

Gupta, V. K. and Waymire, E.: Multiscaling properties of spatial rainfall and river flow distributions, J. Geophys. Res. D, 95, 1999–2009, doi:10.1029/JD095iD03p01999, 1990.

Haberlandt, U. and Berndt, C.: The value of weather radar data for the estimation of design storms – an analysis for the Hannover region, P. Int. Ass. Hydrol. Sci., 373, 81–85, doi:10.5194/piahs-373-81-2016, 2016.

Hartmann, D., Klein Tank, A., Rusticucci, M., Alexander, L., Brönnimann, S., Charabi, Y., Dentener, F., Dlugokencky, E., Easterling, D., Kaplan, A., Soden, B., Thorne, P., Wild, M., and Zhai, P.: Observations: Atmosphere and Surface, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., book section 2, p. 159–254, Cambridge Univ. Press, Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/CBO9781107415324.008, 2013.

Hastie, T., Tibshirani, R., and Friedman, J.: Elements of Statistical Learning 2nd ed., Stanford University: Stanford, CA, USA, doi:10.1007/978-0-387-84858-7, 2009.

Hattermann, F. F., Kundzewicz, Z. W., Huang, S., Vetter, T., Gerstengarbe, F.-W., and Werner, P.: Climatological drivers of changes in flood hazard in Germany, Acta Geophysica, 61, 463–477, doi:10.2478/s11600-012-0070-4, 2013.

Hosking, J. R. M.: L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics, J. Roy. Stat. Soc. B Met., 52, 105–124, doi:10.1111/j.2517-6161.1990.tb01775.x, 1990.

Hosking, J. R. M. and Wallis, J. R.: Regional Frequency Analysis: An Approach Based on L-Moments, Cambridge University Press, 1997.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F.: Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments, Technometrics, 27, 251–261, doi:10.1080/00401706.1985.10488049, 1985.

Hu, G. and Franzke, C. L. E.: Evaluation of Daily Precipitation Extremes in Reanalysis and Gridded Observation-Based Data Sets Over Germany, Geophys. Res. Lett., 47, e2020GL089 624, doi:10.1029/2020GL089624, e2020GL089624 10.1029/2020GL089624, 2020.

Javelle, P., Ouarda, T. B., Lang, M., Bobée, B., Galéa, G., and Grésillon, J.-M.: Development of regional flood-duration–frequency curves based on the index-flood method, J. Hydrol., 258, 249–259, doi:10.1016/S0022-1694(01)00577-7, 2002.

Junghänel, T., Ertelund, H., and Deutschländer, T.: KOSTRA-DWD-2010R: Berichtzur Revisionderkoordinierten Starkregenregionalisierung und -auswertung des Deutschen Wetterdienstes in der Version 2010, Deutscher Wetterdienst, Abteilung Hydrometeorologie, Offenbach, Germany, URL `https://www.dwd.de/DE/leistungen/kostra_dwd_rasterwerte/download/bericht_revision_kostra_dwd_2010.pdf`, 2017.

Junghänel, T., Bissolli, P., Daßler, J., Fleckenstein, R., Imbery, F., Janssen, W., Kaspar, F., Lengfeld, K., Leppelt, T., Rauthea, M., Rauthe-Schöch, A., Rocek, M., Walawender, E., and Weigl, E.: Hydro-klimatologische Einordnung der Stark- und Dauerniederschläge in Teilen Deutschlands im Zusammenhang mit dem Tiefdruckgebiet „Bernd“ vom 12. bis 19. Juli 2021, DWD, `https://www.dwd.de/DE/leistungen/besondereereignisse`

/niederschlag/20210721_bericht_starkniederschlaege_tief_bernd.pdf?__blob=publicationFile&v=6, last access: 10 February 2022, 2021.

Jurado, O. E., Ulrich, J., Scheibel, M., and Rust, H. W.: Evaluating the performance of a max-stable process for estimating intensity-duration-frequency curves, Water, 12, 3314, doi:10.3390/w12123314, 2020.

Katz, R. W., Parlange, M. B., and Naveau, P.: Statistics of extremes in hydrology, Adv. Water Resour., 25, 1287–1304, doi:10.1016/S0309-1708(02)00056-8, 2002.

Kim, H., Kim, S., Shin, H., and Heo, J.-H.: Appropriate model selection methods for nonstationary generalized extreme value models, J. Hydrol., 547, 557–574, doi:10.1016/j.jhydrol.2017.02.005, 2017.

Knapp, A. K., Beier, C., Briske, D. D., Classen, A. T., Luo, Y., Reichstein, M., Smith, M. D., Smith, S. D., Bell, J. E., Fay, P. A., Heisler, J. L., Leavitt, S. W., Sherry, R., Smith, B., and Weng, E.: Consequences of More Extreme Precipitation Regimes for Terrestrial Ecosystems, BioScience, 58, 811–821, doi:10.1641/B580908, 2008.

Kochanek, K., Strupczewski, W. G., and Bogdanowicz, E.: On seasonal approach to flood frequency modelling. Part II: flood frequency analysis of Polish rivers, Hydrol. Process., 26, 717–730, doi:10.1002/hyp.8178, 2012.

Koenker, R. and Machado, J. A.: Goodness of Fit and Related Inference Processes for Quantile Regression, J. Am. Stat. Assoc., 94, 1296–1310, doi:10.1080/01621459.1999.10473882, 1999.

Koh, J.: Gradient boosting with extreme-value theory for wildfire prediction, https://arxiv.org/abs/2110.09497, doi:10.48550/arXiv.2110.09497, last access: 10 January 2022, 2021.

Koutsoyiannis, D., Kozonis, D., and Manetas, A.: A mathematical framework for studying rainfall intensity-duration-frequency relationships, J. Hydrol., 206, 118–135, doi:10.1016/S0022-1694(98)00097-3, 1998.

Kreibich, H., Bubeck, P., Kunz, M., Mahlke, H., Parolai, S., Khazai, B., Daniell, J., Lakes, T., and Schröter, K.: A review of multiple natural hazards and risks in Germany, Nat. Hazards, 74, 2279–2304, doi:10.1007/s11069-014-1265-6, 2014.

Kundzewicz, Z. W., Radziejewski, M., and Pinskwar, I.: Precipitation extremes in the changing climate of Europe, Clim. Res., 31, 51–58, doi:10.3354/cr031051, 2006.

Kunz, M., Mohr, S., and P.C., W.: Niederschlag, in: Klimawandel in Deutschland, edited by Brasseur, G., Jacob, D., and Schuck-Zöller, S., chap. 7, pp. 57–66, Springer Spektrum, Berlin, Heidelberg, doi:10.1007/978-3-662-50397-3_7, 2017.

Lazoglou, G., Anagnostopoulou, C., Tolika, K., and Kolyva-Machera, F.: A review of statistical methods to analyze extreme precipitation and temperature events in the Mediterranean region, Theor. Appl. Climatol., 136, 99–117, doi:10.1007/s00704-018-2467-8, 2019.

Lehmann, E., Phatak, A., Soltyk, S., Chia, J., Lau, R., and Palmer, M.: Bayesian hierarchical modelling of rainfall extremes, in: 20th International Congress on Modelling and Simulation, Adelaide, Australia, edited by Piantadosi, J and Anderssen, RS and Boland, J, pp. 2806–2812, 2013.

Lengfeld, K., Kirstetter, P.-E., Fowler, H. J., Yu, J., Becker, A., Flamig, Z., and Gourley, J.: Use of radar data for characterizing extreme precipitation at fine scales and short durations, Environ. Res. Lett., 15, 085 003, doi:10.1088/1748-9326/ab98b4, 2020.

Liew, S. C., Raghavan, S. V., and Liong, S.-Y.: How to construct future IDF curves, under changing climate, for sites with scarce rainfall records?, Hydrol. Process., 28, 3276–3287, doi:10.1002/hyp.9839, 2014.

Linnerooth-Bayer, J. and Amendola, A.: Introduction to Special Issue on Flood Risks in Europe, Risk Anal., 23, 537–543, doi:10.1111/1539-6924.00334, 2003.

Löffler, H.: Meteorologische Bodenmesstechnik (vormals: Instrumentenkunde), vol. 6, Leitfaden für die Ausbildung im Deutschen Wetterdienst, Selbstverlag des Deutschen Wetterdienstes, Offenbach am Main, `https://www.dwd.de/DE/leistungen/pbfb_verlag_leitfaeden/pdf_einzelbaende/leitfaden6_pdf.pdf?__blob=publicationFile&v=3`, last access: 8 January 2022, 2012.

Lutz, J., Grinde, L., and Dyrrdal, A. V.: Estimating Rainfall Design Values for the City of Oslo, Norway—Comparison of Methods and Quantification of Uncertainty, Water, 12, doi:10.3390/w12061735, 2020.

Mahdi, S. and Cenac, M.: Estimating Parameters of Gumbel Distribution using the Methods of Moments, probability weighted Moments and maximum likelihood, Rev. Mat., 12, 151–156, doi:10.15517/rmta.v12i1-2.259, 2005.

Mailhot, A., Lachance-Cloutier, S., Talbot, G., and Favre, A.-C.: Regional estimates of intense rainfall based on the Peak-Over-Threshold (POT) approach, J. Hydrol., 476, 188–199, doi:10.1016/j.jhydrol.2012.10.036, 2013.

Maraun, D., Rust, H. W., and Osborn, T. J.: The annual cycle of heavy precipitation across the United Kingdom: a model based on extreme value statistics, Int. J. Climatol., 29, 1731–1744, doi:10.1002/joc.1811, 2009.

Marra, F., Morin, E., Peleg, N., Mei, Y., and Anagnostou, E. N.: Intensity–duration–frequency curves from remote sensing rainfall estimates: comparing satellite and weather radar over the eastern Mediterranean, Hydrol. Earth Syst. Sci., 21, 2389–2404, doi:10.5194/hess-21-2389-2017, 2017.

Marsh, T. J. and Dale, M.: The UK Floods of 2000–2001: A Hydrometeorological Appraisal, Water Environ. J., 16, 180–188, doi:10.1111/j.1747-6593.2002.tb00392.x, 2002.

Mekis, E., Donaldson, N., Reid, J., Zucconi, A., Hoover, J., Li, Q., Nitu, R., and Melo, S.: An Overview of Surface-Based Precipitation Observations at Environment and Climate Change Canada, Atmos. Ocean, 56, 71–95, doi:10.1080/07055900.2018.1433627, 2018.

Mélèse, V., Blanchet, J., and Molinié, G.: Uncertainty estimation of Intensity–Duration–Frequency relationships: A regional analysis, J. Hydrol., 558, 579–591, doi:10.1016/j.jhydrol.2017.07.054, 2018.

Menabde, M., Seed, A., and Pegram, G.: A simple scaling model for extreme rainfall, Water Resour. Res., 35, 335–339, doi:10.1029/1998WR900012, 1999.

Meredith, E. P., Ulbrich, U., Rust, H. W., and Truhetz, H.: Present and future diurnal hourly precipitation in 0.11° EURO-CORDEX models and at convection-permitting resolution, Environ. Res. Comm., 3, 055 002, doi:10.1088/2515-7620/abf15e, 2021.

Messner, J. W., Mayr, G. J., and Zeileis, A.: Nonhomogeneous Boosting for Predictor Selection in Ensemble Postprocessing, Mon. Weather Rev., 145, 137 – 147, doi:10.1175/MWR-D-16-0088.1, 2017.

MeteoSwiss (Federal Office of Meteorology and Climatology): Maps of extreme precipitation, `https://www.meteoswiss.admin.ch/home/climate/swiss-climate-in-detail/extreme-value-analyses/maps-of-extreme-precipitation.html`, last access: 15 October 2020.

Moberg, A. and Jones, P. D.: Trends in indices for extremes in daily temperature and precipitation in central and western Europe, 1901–99, Int. J. Climatol., 25, 1149–1171, doi:10.1002/joc.1163, 2005.

Nguyen, V., Nguyen, T., and Wang, H.: Regional estimation of short duration rainfall extremes, in: Use of Historical Rainfall Series for Hydrological Modelling, Water Sci. Technol., 37, 15–19, doi:10.1016/S0273-1223(98)00311-4, 1998.

Niketta, D.: Berliner Wetterkarte vom 30.06.2017, Berliner Wetterkarte e.V., 66 Nr. 125, 2017.

Niketta, D.: Berliner Wetterkarte vom 03.08.2019, Berliner Wetterkarte e.V., 68 Nr. 146, 2019.

NOAA (National Oceanic and Atmospheric Administration): Precipitation Frequency Data Server, `https://hdsc.nws.noaa.gov/hdsc/pfds/`, last access: 15 October 2020.

Olsson, J., Södling, J., Berg, P., Wern, L., and Eronn, A.: Short-duration rainfall extremes in Sweden: a regional analysis, Hydrol. Res., 50, 945–960, doi:10.2166/nh.2019.073, 2019.

Ombadi, M., Nguyen, P., Sorooshian, S., and Hsu, K.-l.: Developing Intensity-Duration-Frequency (IDF) Curves From Satellite-Based Precipitation: Methodology and Evaluation, Water Resour. Res., 54, 7752–7766, doi:10.1029/2018WR022929, 2018.

Papalexiou, S. M. and Koutsoyiannis, D.: Battle of extreme value distributions: A global survey on extreme daily rainfall, Water Resour. Res., 49, 187–201, doi:10.1029/2012WR012557, 2013.

Paprotny, D., Sebastian, A., Morales-Nápoles, O., and Jonkman, S. N.: Trends in flood losses in Europe over the past 150 years, Nat. Commun., 9, doi:10.1038/s41467-018-04253-1, 2018.

Pasternack, A., Grieger, J., Rust, H. W., and Ulbrich, U.: Recalibrating decadal climate predictions – what is an adequate model for the drift?, Geoscientific Model Development, 14, 4335–4355, doi:10.5194/gmd-14-4335-2021, 2021.

Perica, S., Pavlovic, S., St. Laurent, M., Trypaluk, C., Unruh, D., and Wilhite, O.: NOAA Atlas 14: Precipitation-Frequency Atlas of the United States, Volume 11 Version 2.0, U.S. Department of Commerce National Oceanic and Atmospheric Administration National Weather Service, Silver Spring, Maryland, URL `https://www.weather.gov/media/owp/oh/hdsc/docs/Atlas14_Volume11.pdf`, last access: 10 January 2022, 2018.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, URL `https://www.R-project.org/`, 2020.

Rasmussen, P. F.: Generalized probability weighted moments: Application to the generalized Pareto Distribution, Water Resour. Res., 37, 1745–1751, doi:10.1029/2001WR900014, 2001.

Read, L. K. and Vogel, R. M.: Reliability, return periods, and risk under nonstationarity, Water Resour. Res., 51, 6381–6398, doi:10.1002/2015WR017089, 2015.

Ribatet, M., Dombry, C., and Oesting, M.: Spatial Extremes and Max-Stable Processes, in: Extreme value modeling and risk analysis: Methods and applications, edited by Dey, D. K. and Yan, J., chap. 9, pp. 257–280, CRC Press, Boca Raton, USA, doi:10.1201/b19721, 2016.

Ritschel, C., Ulbrich, U., Névir, P., and Rust, H. W.: Precipitation extremes on multiple timescales - Bartlett-Lewis rectangular pulse model and intensity-duration-frequency curves, Hydrol. Earth Syst. Sci., 21, 6501–6517, doi:10.5194/hess-21-6501-2017, 2017.

Roksvåg, T., Lutz, J., Grinde, L., Dyrrdal, A. V., and Thorarinsdottir, T. L.: Consistent intensity-duration-frequency curves by post-processing of estimated Bayesian posterior quantiles, J. Hydrol., 603, 127 000, doi:10.1016/j.jhydrol.2021.127000, 2021.

Rootzén, H. and Katz, R. W.: Design Life Level: Quantifying risk in a changing climate, Water Resour. Res., 49, 5964–5972, doi:10.1002/wrcr.20425, 2013.

Rottler, E., Francke, T., Bürger, G., and Bronstert, A.: Long-term changes in central European river discharge for 1869–2016: impact of changing snow covers, reservoir constructions and an intensified hydrological cycle, Hydrol. Earth Syst. Sci., 24, 1721–1740, doi:10.5194/hess-24-1721-2020, 2020.

Rust, H., Maraun, D., and Osborn, T.: Modelling seasonality in extreme precipitation, Eur. Phys. J. Spec. Top., 174, 99–111, doi:10.1140/epjst/e2009-01093-7, 2009.

Rust, H. W.: The effect of long-range dependence on modelling extremes with the generalised extreme value distribution, Eur. Phys. J. Spec. Top., 174, 91–97, doi:10.1140/epjst/e2009-01092-8, 2009.

Rust, H. W.: Statistische Klimatologie, University Lecture `http://users.met.fu-berlin.de/~HenningRust/teaching/StatKlim_Skript.pdf`, last access: 8 January 2022, 2021.

Scherrer, S. C., Fischer, E. M., Posselt, R., Liniger, M. A., Croci-Maspoli, M., and Knutti, R.: Emerging trends in heavy precipitation and hot temperature extremes in Switzerland, J. Geophys. Res. Atmos., 121, 2626–2637, doi:10.1002/2015JD024634, 2016.

Schneider, D. and Gebauer, P.: Beiträge zur Berliner Wetterkarte: Die Flutkatastrophe im Juli 2021 in Mitteleuropa aus meteorologischer Sicht, Berliner Wetterkarte e.V., `https://www.berliner-wetterkarte.de/Beilagen/2021/Die_Flutkatastrophe_2021.pdf`, last access: 10 February 2022, 2021.

Schweda, S.: Forschungsprojekt Starkregen: Summary November 2019, Gesamtverband der Deutschen Versicherungswirtschaft e. V., `https://www.gdv.de/resource/blob/52868/c6d7ffceab5d13fc0f7659496ced6421/forschungsprojekt-starkregen-summary-download-data.pdf`, last access: 10 February 2022, 2019.

Sebille, Q., Fougères, A.-L., and Mercadier, C.: Modeling extreme rainfall A comparative study of spatial extreme value models, Spat. Stat., 21, 187–208, doi:10.1016/j.spasta.2017.06.009, 2017.

Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., and et al.: Changes in climate extremes and their impacts on the natural physical environment, in: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation, edited by Field, CB and Barros, V and Stocker, TF and Dahe, Q and Dokken, DJ and Ebi, KL and Mastrandrea, MD and Mach, KJ and Plattner, GK and Allen, SK and Tignor, M and Midgley, PM, pp. 109–230, Cambridge University Press: Cambridge, UK, 2012.

Singh, V.: Elementary Hydrology, Prentice Hall, New Jersey, USA, 1992.

Soukissian, T. H. and Tsalis, C.: The effect of the generalized extreme value distribution parameter estimation methods in extreme wind speed prediction, Nat. Hazards, 78, 1777–1809, doi:10.1007/s11069-015-1800-0, 2015.

Steinheuer, J. and Friederichs, P.: Vertical profiles of wind gust statistics from a regional reanalysis using multivariate extreme value theory, Nonlinear Proc. Geoph., 27, 239–252, doi:10.5194/npg-27-239-2020, 2020.

Stephenson, A.: Bayesian inference for extreme value modelling, in: Extreme value modeling and risk analysis: Methods and applications, edited by Dey, D. K. and Yan, J., chap. 13, pp. 257–280, CRC Press, Boca Raton, USA, doi:10.1201/b19721, 2016.

Stephenson, A. G., Lehmann, E. A., and Phatak, A.: A max-stable process model for rainfall extremes at different accumulation durations, Weather Clim. Extrem., 13, 44–53, doi:10.1016/j.wace.2016.07.002, 2016.

Svensson, C. and Jones, D. A.: Review of rainfall frequency estimation methods, J. Flood Risk Manag., 3, 296–313, doi:10.1111/j.1753-318X.2010.01079.x, 2010.

Tank, A. M. G. K. and Können, G. P.: Trends in Indices of Daily Temperature and Precipitation Extremes in Europe, 1946–99, J. Clim., 16, 3665 – 3680, doi:10.1175/1520-0442(2003)016<3665:TIIODT>2.0.CO;2, 2003.

Thibaud, E., Aalto, J., Cooley, D. S., Davison, A. C., and Heikkinen, J.: Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures, Ann. Appl. Stat., 10, 2303 – 2324, doi:10.1214/16-AOAS980, 2016.

Trenberth, K. E., Dai, A., Rasmussen, R. M., and Parsons, D. B.: The Changing Character of Precipitation, B. Am. Meteorol. Soc., 84, 1205 – 1218, doi:10.1175/BAMS-84-9-1205, 2003.

Tyralis, H. and Langousis, A.: Estimation of intensity–duration–frequency curves using max-stable processes, Stoch. Environ. Res. Risk Assess., 33, 239–252, doi:10.1007/s00477-018-1577-2, 2019.

Ulrich, J. and Ritschel, C.: IDF: Estimation and Plotting of IDF Curves, URL `https://CRAN.R-project.org/package=IDF`, r package version 2.0.0, last access: 1 November 2020, 2019.

Ulrich, J. and Ritschel, C.: IDF: Estimation and Plotting of IDF Curves, URL `https://CRAN.R-project.org/package=IDF`, r package version 2.1.0, last access: 15 December 2021, 2021.

Ulrich, J., Jurado, O. E., Peter, M., Scheibel, M., and Rust, H. W.: Estimating IDF Curves Consistently over Durations with Spatial Covariates, Water, 12, 3119, doi:10.3390/w12113119, 2020.

Ulrich, J., Fauer, F. S., and Rust, H. W.: Monthly precipitation intensity maxima for 14 aggregation times at 132 stations in Germany, doi:10.5281/zenodo.5025657, 2021a.

Ulrich, J., Fauer, F. S., and Rust, H. W.: Modeling seasonal variations of extreme rainfall on different timescales in Germany, Hydrol. Earth Syst. Sci., 25, 6133–6149, doi:10.5194/hess-2021-336, 2021b.

Van de Vyver, H.: Spatial regression models for extreme precipitation in Belgium, Water Resour. Res., 48, doi:10.1029/2011WR011707, 2012.

Van de Vyver, H.: Bayesian estimation of rainfall intensity–duration–frequency relationships, J. Hydrol., 529, 1451–1463, doi:10.1016/j.jhydrol.2015.08.036, 2015.

Van de Vyver, H.: A multiscaling-based intensity–duration–frequency model for extreme precipitation, Hydrol. Process., 32, 1635–1647, doi:10.1002/hyp.11516, 2018.

Van de Vyver, H. and Demarée, G. R.: Construction of Intensity–Duration–Frequency (IDF) curves for precipitation at Lubumbashi, Congo, under the hypothesis of inadequate data, Hydrol. Sci. J. J. Sci. Hydrol., 55, 555–564, doi:10.1080/02626661003747390, 2010.

Vandeskog, S. M., Martino, S., Castro-Camilo, D., and Rue, H.: Modelling short-term precipitation extremes with the blended generalised extreme value distribution, `https://arxiv.org/abs/2105.09062`, doi:10.48550/arXiv.2105.09062, last access: 10 January 2022, 2021.

Veneziano, D. and Furcolo, P.: Multifractality of rainfall and scaling of intensity-duration-frequency curves, Water Resour. Res., 38, 42–1, doi:10.1029/2001WR000372, 2002.

Wake, B.: Flooding costs, Nat. Clim. Change, 3, 778, doi:10.1038/nclimate1997, 2013.

Westra, S., Alexander, L. V., and Zwiers, F. W.: Global Increasing Trends in Annual Maximum Daily Precipitation, J. Clim., 26, 3904 – 3918, doi:10.1175/JCLI-D-12-00502.1, 2013.

Wilks, D. S.: Statistical methods in the atmospheric sciences, vol. 100, Academic press, doi:10.1016/C2017-0-03921-6, 2011.

Willems, P.: Compound intensity/duration/frequency-relationships of extreme precipitation for two seasons and two storm types, J. Hydrol., 233, 189–205, doi:10.1016/S0022-1694(00)00233-X, 2000.

Yan, L., Xiong, L., Jiang, C., Zhang, M., Wang, D., and Xu, C.-Y.: Updating intensity–duration–frequency curves for urban infrastructure design under a changing environment, WIREs Water, 8, e1519, doi:10.1002/wat2.1519, 2021.

Yee, T. W. and Stephenson, A. G.: Vector generalized linear and additive extreme value models, Extremes, 10, 1–19, doi:10.1007/s10687-007-0032-4, 2007.

Zeder, J. and Fischer, E. M.: Observed extreme precipitation trends and scaling in Central Europe, Weather Clim. Extrem., 29, 100 266, doi:10.1016/j.wace.2020.100266, 2020.

Łupikasza, E. B.: Seasonal patterns and consistency of extreme precipitation trends in Europe, December 1950 to February 2008, Clim. Res., 72, 217–237, doi:10.3354/cr01467, 2017.

List of Abbreviations

BHM	Bayesian Hierarchical Model
d-GEV	Duration-Dependent GEV
DDF	Depth-Duration-Frequency (curve)
DWA	German Association for Water, Wastewater and Waste (*Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall*)
DWD	German Meteorological Service (*Deutscher Wetterdienst)*
EVT	Extreme Value Theory
GEV	Generalized Extreme Value (distribution)
IDF	Intensity-Duration-Frequency (curve)
MLE	Maximum Likelihood Estimation
NOAA	National Oceanic and Atmospheric Administration
PMLE	Penalized Maximum Likelihood Estimation
PWM	Probability Weighted Moments
QDF	Flood-Duration-Frequency (curve)
QS	Quantile Score
QSI	Quantile Skill Index
QSS	Quantile Skill Score
VGLM	Vector Generalized Additive Model
VGLM	Vector Generalized Linear Model

List of Figures

List of Tables

Acknowledgments

It took the support of many people in many different ways to complete this thesis, and I would like to express my sincere gratitude to all of them.
First of all, I would like to thank my supervisors Henning Rust and Uwe Ulbrich for giving me the opportunity to be part of your working groups and supporting me with guidance and advice, while also allowing me to work independently.
Henning, thank you for your constant encouragement, for always being available to address any of my questions and for creating a pleasant and friendly work environment.
The meteorological institute has always felt like a welcoming place to me, so I would like to thank all my co-workers and especially the members of the *StatMet* and *CliDia* working groups not only for your helpful discussions and comments, but also for providing a friendly and familiar working atmosphere.
I also gained a lot of valuable experience as a member of the graduate school *NatRiskChange* and I am especially grateful to the chairpersons Annegret Thieken and Axel Bronstert and the project manager Theresia Petrow for all your efforts. I don't take your commitment to help us PhD students maintain as much normality as possible in these challenging times for granted.
To my fellow PhD students, above all Oscar, Lisa, Christian, Felix, Madlen, Andreas, Lukas, Omar, Joscha, ... thank you for the countless valuable discussions (on- and off-topic) and the wonderful moments spent together at lunch, table tennis matches, bike rides, campfires, and so on. I wish we could have shared more of those experiences, but it is for this reason that I am so grateful for the time we did have.
Warm thanks goes to Oscar for all of your work related and personal support – you are the best office mate/ co-author/ friend anyone could ask for. Thank you for the countless times of proof-reading anything I ever wrote and for always helping me answer my questions.
And many thanks also to you, dear Lisa, without our shared writing sessions and your motivation, this work would only have half the amount of pages.
Lastly, I would like to express my gratitude to my friends and family for your incredible support. Thank you, Mama and Andrea, you are the best grandmas anyone could possibly want for their child, without your active assistance and valuable time I certainly would not have gotten this far.
Sebastian, I can't tell you how grateful I am to have you in my life. Thank you for just taking over all the day-to-day tasks without complaining, so that I could fully focus on writing, and for constantly encouraging me. Thank you for sticking together, in both bright and dark, crazy or stressful times.

Selbstständigkeitserklärung (Declaration)

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Sämtliche wissentlich verwendete Textausschnitte, Zitate oder Inhalte anderer Verfasser wurden ausdrücklich als solche gekennzeichnet. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Universität zur Prüfung vorgelegen.

Berlin, den

Jana Ulrich