technische universität
dortmund

Group-based regionalization in flood frequency analysis considering heterogeneity

by

Jona Lilienthal

Dissertation

in partial fulfilment of the requirements for the degree of
Doktor der Naturwissenschaften
presented to the

Department of Statistics, TU Dortmund University

Dortmund, 2019

Referees:
Prof. Dr. R. Fried
Dr. U. Ligges

Date of oral examination:
20 November 2019

Contents

1 Introduction

Throughout history people have always settled near rivers. Rivers provide access to fresh water and therefore good farming conditions and are often important waterways for trade. However, close habitation also entails the risk of destruction through flooding. Already in ancient Egypt the water level of the Nile was observed in so-called nilometers to understand and predict the occurrence of floods. In Germany the first gauging stations were installed on the Elbe and Rhine rivers around 1775.

In the last two decades, several floods in Germany and Europe demonstrated the risk that living near rivers inherently carries. In August of 2002 extreme rainfall throughout several days led to floods on Elbe and Donau affecting people in Germany, Czech Republic, and Austria. Alone in Germany 21 deaths and an economic damage of about 11 billion Euro were reported (Baumgarten et al., 2011). In May and June of 2013, again resulting from heavy rainfall, floods affected numerous river networks in several countries in Western and Central Europe and caused 14 deaths and a damage of around 6 billion Euro in Germany (DKKV, 2015).

Flood risks can be partially limited by measures like dams or reservoirs but there will always remain some risk. According to the Federal Water Act (WHG, Wasserhaushaltsgesetz), the governments of Germany's federal states define areas in which a recurrent flooding has to be expected. In those areas special regulations, like limitations on farming and a ban on new construction sites, are implemented. To determine such areas or the dimensions of flood protection measures, one has to estimate the recurrence time of flood events. A common quantity is the 100 year flood, i.e. the flood that is expected to be reached once in 100 years on average. Expressed statistically, such a flood appears in any given year with a chance of one percent.

In order to determine the 100 year flood (or floods even much more seldom) from the flood records, often only yearly maxima are used. This has different advantages: While daily or monthly observations are autocorrelated, yearly maxima are mostly stochastically independent and their distribution can be approximated by a generalized extreme value distribution. On the other hand, the restriction to yearly maxima leads to a shortage of data, typical to the analysis of floods. With only few exceptions, most gauging stations in Germany record data since 40–80 years, often interrupted by years without records due to modernisation or maintenance. Since we are mostly interested in rare events, the estimation from relatively short records can be very volatile. To improve this situation, information from different gauging

stations is used jointly. This is called regional flood frequency analysis and can be performed using different approaches. Often groups of gauging stations are built that share some distribution characteristic. For example, in the Index Flood model it is assumed that the site's distributions only vary by some site-specific scaling factor. In the popular approach of Hosking and Wallis (1997) this model is used and distribution parameters are estimated using L-moments, an alternative to product moments that are linear combinations of expectations of order statistics.

This dissertation is mainly based on three works dealing with different aspects of regional flood frequency analysis. Naturally each work itself contains parts in which basic notation and methodology are presented. To avoid repetition, these are given in a unified manner in Chapter 2 which gives an overview of the basics of flood data analysis, the modelling of statistical extremes and multivariate distributions, parameter estimation, and lastly of regional flood frequency analysis.

As mentioned before, regional flood frequency analysis is designed to transfer information from multiple sites to one specific target site. Therefore it is an important step to assure that this transfer is beneficial, i.e. that the assumption of a similar behaviour is appropriate. Chapter 3 deals with this subject in case of intersite dependence and highly skewed data, a scenario in which the most common L-moment-based procedure gives poor results. This problem is faced by modelling the intersite dependence with copulas and by using trimmed L-moments instead of L-moments. The proposed generalized procedure, analysed in a comprehensive simulation study, is able to overcome the known drawbacks.

An alternative to L-moments are probability-weighted moments that can be calculated from L-moments and vice versa. In Chapter 4 a limit theorem for sample probability-weighted moments of multiple sites is presented and extended to L-moments and trimmed L-moments as well as resulting parameter and quantile estimators of generalized extreme value distributions. Using this theory and a consistent estimator, an asymptotic homogeneity test is derived and examined. This test is an alternative to the procedure of the preceding chapter with advantages in computation speed, but it requires a higher number of observations to give good results.

The last work, presented in Chapter 5, applies penalized Maximum-Likelihood estimators to regional flood frequency analysis. The main goal is to provide a new information pooling scheme by transferring the Index Flood principles to penalization. Penalization techniques typically contain a parameter controlling the degree of penalization. A cross-validation procedure is presented that allows for a data-driven choice. As a result, the new estimator is able to give good results even in situations in which we are not fully confident that the Index Flood model is appropriate.

Chapter 6 summarises the most important results from all works. The following appendix contains further material regarding aspects from Chapter 4 and 5.

2 Basics

In this chapter the basic principles and statistical theories are given that are needed to understand the following chapters. First the basics of the analysis of floods are described. The data itself and its collection are of special interest here. Afterwards an excursion into extreme value analysis shows how the behaviour of univariate extreme events can be modelled. Two possibilities are presented: either we model the statistical distribution of block maxima or the distribution of exceedances above some fixed threshold. Copula theory is applied to model multivariate extremes. Its basics and the most important copula classes (regarding our application) are described in the following section. Multiple options to estimate distribution parameters are given subsequently. First probability-weighted moments, L-moments, as well as trimmed L-moments are defined and it is described how they can be used to estimate parameters. These statistics are alternative moments, which are especially popular in hydrology because of good small sample properties. As an alternative to L-moment based methods a penalized Maximum-Likelihood estimator is given that will be applied to regional flood frequency analysis in a later chapter. The last section in this chapter deals with the practice of regionalization in hydrology. The basic steps are described and a popular L-moment based procedure is given.

2.1 Basic principles of flood data analysis

When flood data is analysed the main goal is to formulate statements about the likelihood of specific events or likewise about what events have to be expected every, say, 100 years on average. Such assessments and their precision are important to identify regions with an increased vulnerability to flooding and to plan structures like dams or reservoirs, which are important measures to restrict danger to life or property.

To be able to define what a flood is and what properties of a flood can be measured, first we need to take a look at the collection of data. At measuring stations the water level and flow velocity are constantly measured. Together with knowledge about the stream's cross-sectional area from the flow velocity and water level the discharge or stream flow can be calculated which is measured in m^3/sec (Wittenberg, 2011). Plotting of the discharge over time gives us a hydrograph which can be used to determine floods and their properties.

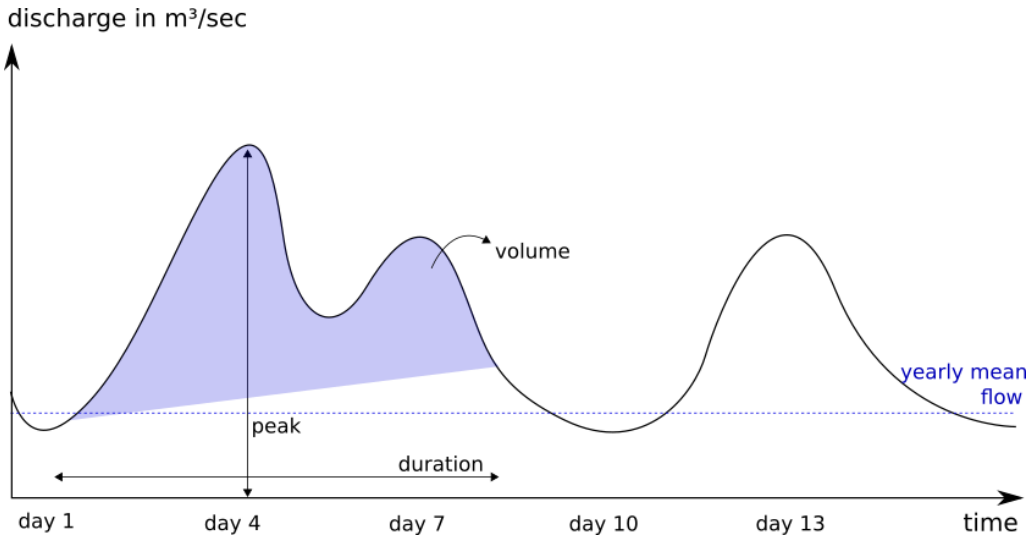

Figure 2.1: Example of a hydrograph of two separate events with peaks at day 4 and day 13.

First the different flood events have to be separated from the graph, see Figure 2.1 as an example. For this purpose multiple criteria are applied. Events are different, if the peaks of the flow rate are more than seven days apart or if the difference between the highest rate of an event and the lowest rate prior to that event is bigger than the difference between the lowest rate between two possible events and the yearly mean flow rate (LAWA, 1997). In the example the first two peaks belong to one event, because the peaks are only three days apart and the difference between the minimum between the peaks and the second peak is smaller than the difference between that minimum and the yearly mean flow. The third peak at day 13 is considered to be a new event, because it is more than seven days apart (and the second condition would also give the same decision).

The properties of one flood event are peak, duration, and volume (see again Figure 2.1). For more information how to assess these properties see, for example, Fohrer et al. (2016). Often only the peak is analysed, but recently more and more multivariate studies have been published that analyse multiple properties jointly (Chebana and Ouarda, 2009; Requena, Chebana, and Mediero, 2016).

However, in the statistical analysis of floods mostly not all separated flood events are used but only a subset of them. Two procedures are popular which use different subsets of data and which are motivated by different theorems leading to either a generalized extreme value distribution or a generalized Pareto distribution to approximate the distributional behaviour of floods. The following section will give a short introduction into the analysis of extreme values and how it is applied in hydrology.

In hydrological applications we often talk about the return period or recurrence

interval of events (even though being criticised for being prone to misuse, see Serinaldi, 2015). The return period describes the average time span, in which a flood exceeding a certain value is to be expected. The often used 100-year flood therefore indicates the height of a flood that is reached or exceeded one time on average every 100 years. Statistically this means that such a flood has a 1% probability of occurrence in each year (assuming the distribution of floods does not change over time). As a formula the return period T of an event exceeding the value x can be expressed as

$$T = \frac{\mu}{P(X > x)}, \tag{2.1}$$

with X denoting the random variable of the process of interest and μ denoting the average time between two realizations. In case of yearly maxima we have $\mu = 1$ [year] and the return period (in years) equals the reciprocal of the exceedance probability.

2.2 Statistical modelling of extreme events

The goal of extreme value analysis is to describe the behaviour and frequency of extreme events, i.e. unusually rare ones. This section gives a concise overview over this topic that is needed to understand the remainder of this dissertation. For a thorough introduction it is referred to Coles (2001) and de Haan and Ferreira (2006).

The two most important results in extreme value analysis are the Fisher-Tippett-Gnedenko theorem and the theorem of Pickands-Balkema-de Haan. They describe the asymptotic distributions of block maxima for increasing block length and of threshold exceedances for increasing threshold. They are of huge importance in the analysis of hydrological data (as well as in other disciplines in which extremal data is analysed) because they offer a theoretical justification for the choice of specific distribution families to model the distribution of events like extreme flows. In this dissertation the block maxima method is used to model yearly maximal flows. Because of their importance, though, both methods and the corresponding limiting distributions are described shortly.

2.2.1 Theorem of Fisher-Tippett-Gnedenko

First we want to turn our attention to the question how maxima of a set of random variables are distributed. Let $Y_1, \ldots, Y_n$ be independent and identically distributed random variables with distribution function F. Further let $M_n = \max(Y_1, \ldots, Y_n)$

be the maximum of this set. It is known that the distribution of M_n is given as:

$$\begin{aligned} F_{M_n}(x) = P(M_n \leq x) &= P(Y_1 \leq x, \ldots, Y_n \leq x) \\ &= P(Y_1 \leq x) \times \cdots \times P(Y_n \leq x) = F^n(x). \end{aligned} \tag{2.2}$$

Because $F(x) \in [0, 1]$ it follows that $F^n(x)$ converges to either zero or one for increasing n. For this reason we consider the distribution of a standardized statistic instead. Let $a_n > 0$ and $b_n \in \mathbb{R}$, so that

$$\frac{\max(Y_1, \ldots, Y_n) - b_n}{a_n} \tag{2.3}$$

converges to a non-degenerate distribution G for $n \to \infty$ (de Haan and Ferreira, 2006, Theorem 1.1.3; Leadbetter, 1974, Theorem 2.1). According to the theorem of Fisher-Tippett-Gnedenko (Fisher and Tippett, 1928; Gnedenko, 1943), this distribution G is a member of either the Fréchet, Gumbel, or Weibull distribution family, which can be summarized in the family of the generalized extreme value distribution.

2.2.2 Generalized extreme value distribution

The generalized extreme value distribution (GEV) is a distribution family with a location parameter $\mu \in \mathbb{R}$, a scale parameter $\sigma > 0$, and a shape parameter $\xi \in \mathbb{R}$ (Jenkinson, 1955).

The distribution function of the GEV is given as

$$F(x) = \begin{cases} \exp\left(-(1 + \xi \frac{x-\mu}{\sigma})^{-1/\xi}\right), & \xi \neq 0, \\ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right), & \xi = 0, \end{cases} \tag{2.4}$$

for $1 + \xi(x - \mu)/\sigma > 0$.

The quantile function F^{-1} can be stated explicitly as

$$F^{-1}(p) = \begin{cases} \mu + \frac{\sigma}{\xi}\left((-\log(p))^{-\xi} - 1\right), & \xi \neq 0, \\ \mu - \sigma \log(-\log(p)), & \xi = 0. \end{cases} \tag{2.5}$$

The GEV has the important property that it unites the three families possible as limit distribution for block maxima:

- Weibull family for negative shape parameters,
- Gumbel family for $\xi = 0$,
- Fréchet family for positive shape parameters.

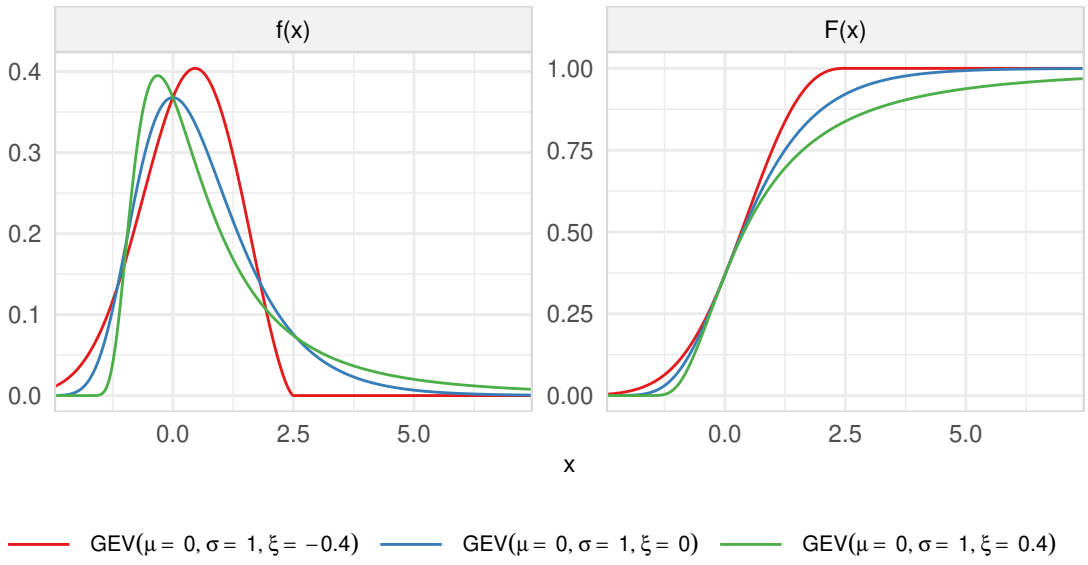

Figure 2.2: Density (left) and distribution function (right) of the three GEV subclasses: Weibull ($\xi < 0$), Gumbel ($\xi = 0$), and Fréchet ($\xi > 0$).

The Weibull family features a finite upper tail, the Gumbel family is unrestricted, and the Fréchet family has a finite lower tail and a heavy upper tail. A representation of these three families is given in Figure 2.2. Which of the three families is appropriate to model the asymptotic limit distribution of (2.3) depends on the distribution of the underlying variables $Y_1, \ldots, Y_n$.

Moments of GEV distributions exist depending on the shape parameter ξ. The mean and variance of a generalized extreme value distributed random variable are

$$E(X) = \begin{cases} \mu + \sigma\gamma, & \xi = 0, \\ \mu + \frac{\sigma}{\xi}(\Gamma(1-\xi) - 1), & \xi < 1, \\ \infty, & \xi \geq 1, \end{cases} \tag{2.6}$$

$$Var(X) = \begin{cases} \sigma^2 \frac{\pi^2}{6}, & \xi = 0, \\ \frac{\sigma^2}{\xi^2}(\Gamma(1-2\xi) - (\Gamma(1-\xi))^2), & \xi < 1/2, \\ \infty, & \xi \geq 1/2, \end{cases} \tag{2.7}$$

with γ denoting Euler's constant.

2.2.3 Theorem of Pickands-Balkema-de Haan

An alternative to the block maxima approach is the peaks-over-threshold (POT) method. Instead of modelling the distribution of maxima of a set of random variables,

the distribution of the exceedances over a threshold is modelled. Therefore consider a random variable Y with distribution function F. The distribution of the exceedance above some threshold u is given as

$$F_u(y) = P(Y - u \leq y \mid Y > u) = \frac{F(y+u) - F(u)}{1 - F(u)}, \quad y > 0. \tag{2.8}$$

According to the theorem of Pickands-Balkema-de Haan (Balkema and de Haan, 1974; Pickands, 1975) the asymptotic limit distribution of the excess F_u for $u \to \infty$ has to be a member of the generalized Pareto distribution family.

2.2.4 Generalized Pareto distribution

The generalized Pareto distribution (GPD) has two parameters: scale parameter $\sigma > 0$ and shape parameter $\xi \in \mathbb{R}$ (Coles, 2001; Hosking and Wallis, 1987). In some literature the GPD is given as a location-scale family with three parameters, adding a location parameter $\mu \in \mathbb{R}$ indicating the threshold or lower limit (e.g. Wang, 1991).

The distribution function of the GPD can be written as

$$F(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\sigma}\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(\frac{x}{\sigma}\right), & \xi = 0, \end{cases} \tag{2.9}$$

with $x > 0$ for $\xi \geq 0$ and $0 \leq x \leq -\sigma/\xi$ for $\xi < 0$.

The mean and variance of a generalized Pareto distributed random variable X are given as

$$E(X) = \begin{cases} \frac{\sigma}{1-\xi}, & \xi < 1, \\ \infty, & \xi \geq 1, \end{cases} \tag{2.10}$$

$$Var(X) = \begin{cases} \frac{\sigma^2}{(1-\xi)^2(1-2\xi)}, & \xi < 1/2, \\ \infty, & \xi \geq 1/2. \end{cases} \tag{2.11}$$

Figure 2.3 characterises density and distribution function of the GPD for different shapes.

2.2.5 Application in hydrology

These two limit theorems enable us to theoretically justify the statistical modelling of extreme floods. The practical application differs fundamentally in both approaches.

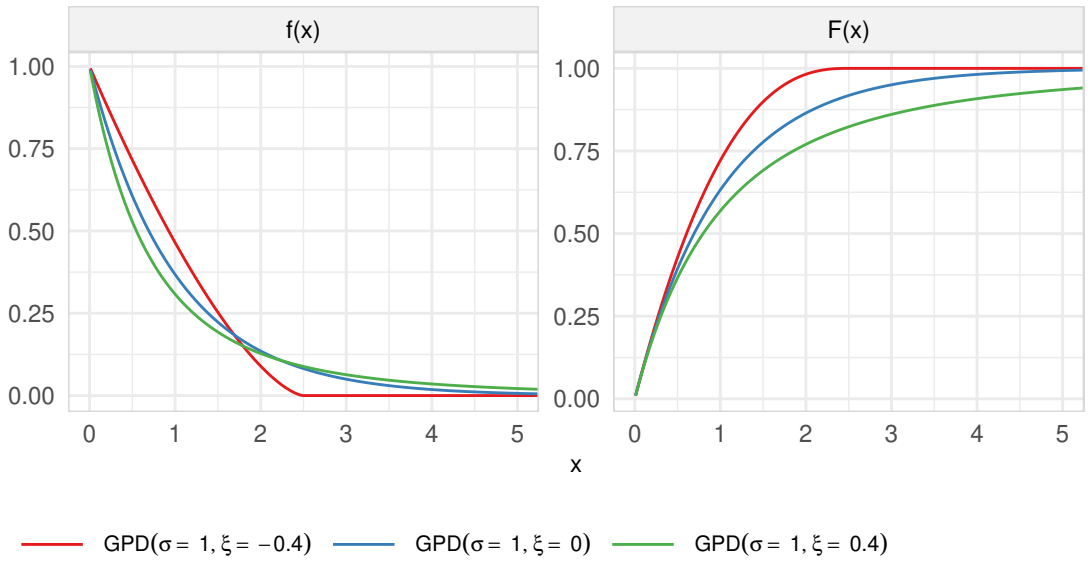

Figure 2.3: Density (left) and distribution function (right) of the GPD.

If using the block maxima approach, maxima of specified time periods are built. These periods normally span over one year, the so-called hydrological year. In Germany it is defined to begin on 1 November and end on 31 October the following year. The distribution of the maxima is modelled using the generalized extreme value distribution, and return periods can be derived directly from the fitted distribution.

With the peaks-over-threshold approach we use all observations that exceed a threshold that has to be specified prior to the analysis. The threshold should be chosen not to be too high, so that enough data can be analysed. According to Cunnane (1973) using the POT approach is beneficial in comparison to the annual maxima approach if at least 1.65 peaks per year are selected. On the other hand it should neither be too low, so that the approximation by using the limit distribution family is still good enough. The choice of the threshold has a significant influence on parameter and quantile estimates (Beguería, 2005) and should therefore be performed with caution. One option are graphical representations of the estimated GPD shape parameter depending on the threshold value (Coles, 2001). Using the outcome of a statistical goodness-of-fit test depending on the threshold can lead to an automated threshold selection (Choulakian and Stephens, 2001; Durocher et al., 2018). The generalized Pareto distribution then models the exceedances above the threshold. To calculate annual return periods, also the frequency of exceedances has to be taken into account which is commonly modelled using a Poisson process (e.g. Cunnane, 1973; Katz, Parlange, and Naveau, 2002; Rosbjerg, 1985).

Which approach is chosen depends also on the data available. The POT approach requires data of all individual flood events. Often only monthly or yearly maxima

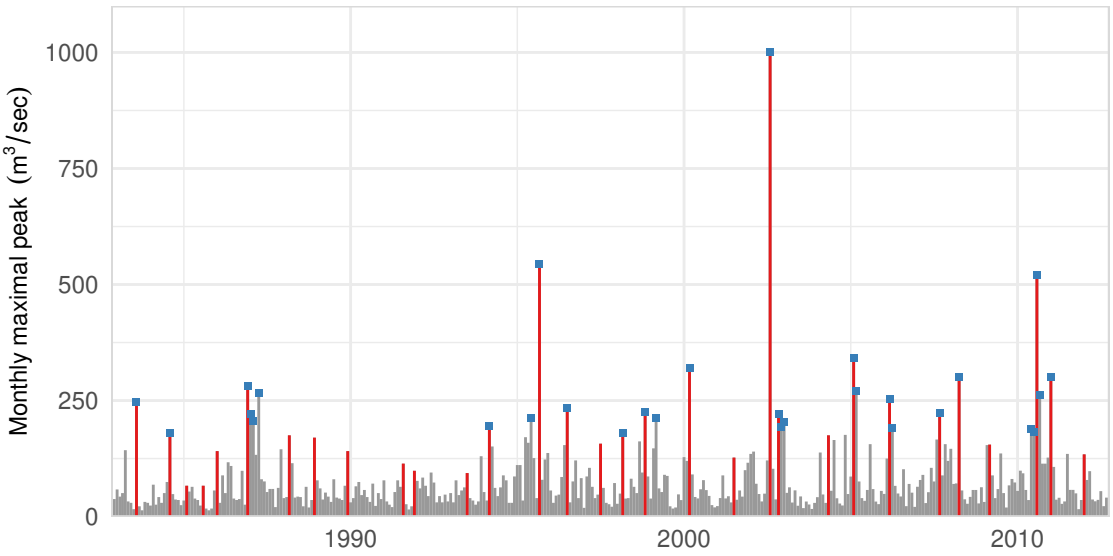

Figure 2.4: Comparison of block maxima and POT approach using an excerpt from data of Wechselburg. Red observations mark yearly maxima, blue squares mark exceedances over a threshold of $180\,m^3/sec$.

are available. While we can assume that most exceedances are present in sets of monthly maxima, we cannot assume this for yearly maximal data.

Figure 2.4 compares the data basis of the block maxima and POT approach using a real life example. Differences between the used data are clearly visible. From 1988 to 1993, for example, no yearly maximum was high enough to be used in a POT analysis. Reversely there were multiple events around the year 1987 that we would consider in the POT analysis but not in a block maxima approach.

The popularity of using the GEV and GPD to model the statistical behaviour of extreme events originates in the two limit theorems that state them as limiting distributions as well as in their flexibility of modelling distributions with a wide variety of shapes. However, it should be considered that these theoretical results are asymptotic ones. Because of the commonly used block length of one year as well as the fixed chosen threshold the limiting distributions can only be considered as approximations. The quality of the approximation depends on various factors and on the distribution of the underlying random variables and can in practise hardly be determined. Further, it should stay in mind that both limit theorems assume independent and identically distributed variables. In practice this assumption is often not fulfilled. If we consider flood events of a river, the distribution normally varies over the year based on different flood origins. During winter and spring, floods are mostly generated by snow melt, while during summer local heavy rain events are the reason for high water levels. For assessment of the independence of such data

points, the temporal resolution of observations is the essential factor. Weekly, daily, or even hourly observations are strongly dependent while with monthly maximum values we tend to assume independence.

However, applications of the GEV and GPD to hydrological flood data and comparisons to other distribution families have shown that these distributions mostly give a good approximation of the distribution of yearly maxima or of threshold exceedances. Still, it is recommended to assess each case separately and to not simply rely on the limit theorems.

2.3 Modelling multivariate distributions

Often modelling of the joint behaviour of multiple random variables or random vectors is of interest. These can be different characteristics of one object or one characteristic that is measured on multiple objects. In the context of flood analysis these are, for example, the flood peak and duration at one station or the measurement of the flood peak at several stations.

A possibility to model the statistical behaviour of random vectors are multivariate distributions, e.g. the multivariate normal distribution. The disadvantage of this approach is that the shape and strength of dependence between the dimensions are limited. Often dependence between variables manifests not over the whole support of the distribution but only in the tails, i.e. in extraordinary small events (like droughts) or high events (like snow melts). A measure to assess such dependence in the tails is the so-called tail dependence. In case of the multivariate normal distribution we can adjust the strength of the mean dependence, but not the tail dependence.

A much more flexible approach is to model multivariate events with copulas. Copulas are functions that connect marginal distributions to joint distributions and contain all information about the dependence structure. They allow for a separate modelling of marginal behaviour and dependence. Through a huge variety of different copula families with different properties, dependences can be modelled in a much more sophisticated and more complex manner.

This section contains the main definitions and the main theorem regarding copula theory as well as the most important copula families that we need to model and to describe flood events. For a more thorough and extensive introduction it is referred to Nelsen (2006). In this section we mostly deal with the joint distribution of random variables $X_1, \ldots, X_d$. For the sake of simplicity and to be closer to the usual notation, the variables are denoted as X and Y in the case of two dimensions.

2.3.1 Copulas

A function $C : [0,1]^d \mapsto [0,1]$ is called copula, if the following properties hold:

$$\text{i)} \quad C(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_d) = 0 \;\; \forall\, i = 1, \ldots, d, \tag{2.12}$$

$$\text{ii)} \quad C(1, \ldots, 1, \underbrace{u_i}_{i\text{-th position}}, 1, \ldots, 1) = u_i \;\; \forall\, i = 1, \ldots, d, \tag{2.13}$$

$$\text{iii)} \quad C \text{ is } d\text{-increasing.} \tag{2.14}$$

The last property means that the volume of each hyper rectangle under C is non-negative. In one dimension this simply means that C is increasing, i.e. that $C(u_2) - C(u_1) \geq 0$ holds for $u_1 \leq u_2$. In case of two dimensions this translates to the property that

$$C(u_{12}, u_{22}) - C(u_{12}, u_{21}) - C(u_{11}, u_{22}) + C(u_{11}, u_{21}) \geq 0. \tag{2.15}$$

holds for $u_{11}, u_{12}, u_{21}, u_{22} \in [0,1]$. Figure 2.5 illustrates this. The contour lines are representing the value of $C(\cdot, \cdot)$. The function C is d-increasing, if the volume of every hyper rectangle $[u_{11}, u_{12}] \times [u_{21}, u_{22}]$ is non-negative. A more precise definition for arbitrary dimensions can be found in Nelsen (2006, Definition 2.10.1 and 2.10.2).

2.3.2 Sklar's theorem

A copula function can be seen as multivariate distribution with uniformly on $[0,1]$ distributed marginals. Sklar's theorem enables us to combine arbitrary marginal distributions and copulas to multivariate distributions.

Let $X_1, \ldots, X_d$ be random variables with marginal distributions $F_j,\ j = 1, \ldots, d$, and joint distribution $F(x_1, \ldots, x_d) = P(X_1 \leq x_1, \ldots, X_d \leq x_d)$. According to Sklar's theorem (Sklar, 1959) a copula C exists so that the joint distribution F can be expressed as

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)) \;\forall\, x_1, \ldots, x_d \in \mathbb{R}. \tag{2.16}$$

All information about the dependence structure between $X_1, \ldots X_d$ is contained in C, all information about the marginal distributions in $F_1, \ldots, F_d$, respectively. If the marginals are continuous, C is unique.

Given a joint distribution function F and marginal distributions $F_1, \ldots, F_d$ and the corresponding inverse functions $F_1^{-1}, \ldots, F_d^{-1}$, the copula function C can be stated as

$$C(u_1, \ldots, u_d) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)) \;\forall\, u_1, \ldots, u_d \in [0,1]. \tag{2.17}$$

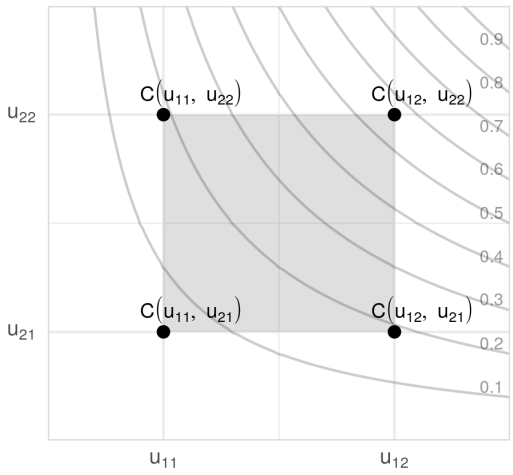

Figure 2.5: Graphical illustration of the third copula property for bivariate copulas. Contour lines give the value of $C(\cdot, \cdot)$. The C-volume of every hyper rectangle (grey) must not be negative.

To model the behaviour of multivariate events it can be used that the reverse of Sklar's theorem holds as well. Given marginal distributions $F_1, \ldots, F_d$ and a copula function C the combination $C(F_1(x_1), \ldots, F_d(x_d))$ is a multivariate distribution. Therefore, the modelling of multivariate events can be split into modelling of marginal distributions and modelling of the dependence structure separately. In the next sections the most important classes of copulas are presented.

2.3.3 Archimedean copulas

Archimedean copulas are a popular class of copulas and defined using a so-called generator ϕ. Let $\phi : [0, 1] \mapsto [0, \infty]$ be a continuous, strictly decreasing function with $\phi(1) = 0$. The pseudo-inverse of ϕ is defined as

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t), & 0 \leq t \leq \phi(0), \\ 0, & \phi(0) < t \leq \infty. \end{cases} \tag{2.18}$$

Then, a function C defined as

$$C(u_1, u_2) = \phi^{[-1]}(\phi(u_1) + \phi(u_2)) \tag{2.19}$$

is a (bivariate) copula if and only if ϕ is convex (Nelsen, 2006, Theorem 4.1.4). The function ϕ is called generator of C.

Independently from the used generator, Archimedean copulas share some properties. They are symmetrical, $C(u_1, u_2) = C(u_2, u_1)\ \forall\, u_1, u_2 \in [0,1]$, and associative, $C(C(u_1, u_2), u_3) = C(u_1, C(u_2, u_3))\ \forall\, u_1, u_2, u_3 \in [0,1]$.

The class of bivariate Archimedean copulas can be extended to arbitrary dimensions $d > 2$ using the associativity and a special property of the generator. Let ϕ be a generator that is completely monotonic on $[0, \infty)$, i.e. it satisfies

$$(-1)^k \frac{d^k}{dt^k}\phi(t) \geq 0\ \forall\, t \in (0, \infty). \tag{2.20}$$

for $k \in \mathbb{N}_0$. Then, according to Nelsen (2006, Theorem 4.6.2),

$$C(u_1, \ldots, u_d) = \phi^{[-1]}(\phi(u_1) + \ldots + \phi(u_d)) \tag{2.21}$$

is a d-dimensional copula.

2.3.4 Extreme value copulas

Extreme value copulas are another important class of copulas, especially when modelling extreme events (Gudendorf and Segers, 2010). The main application is the modelling of multivariate extreme value distributions that result from combining an extreme value copula with generalized extreme value distributed marginals (Durante and Salvadori, 2010). Their defining property is the maximum-stability. This means that a copula C is an extreme value copula if

$$C(u_1^t, \ldots, u_d^t) = C^t(u_1, \ldots, u_d) \tag{2.22}$$

holds for all $u_1, \ldots, u_d \in [0,1]$ and $t > 0$.

Every bivariate extreme value copula can be expressed as

$$C(u_1, u_2) = \exp\left(\log(u_1 u_2)\, A\left(\frac{\log u_2}{\log(u_1 u_2)}\right)\right), \tag{2.23}$$

using a convex function $A : [0,1] \mapsto [1/2, 1]$ called Pickands dependence function (Pickands, 1981). Pickands dependence function A can be written depending on the copula C as

$$A(t) = -\ln C(e^{-(1-t)}, e^{-t}), \tag{2.24}$$

fulfils

$$\max(t, 1-t) \leq A(t) \leq 1 \quad \forall\, t \in [0,1] \tag{2.25}$$

and can be used as a description of the dependence. The lower and upper limits of equation (2.25) characterise the special cases of complete dependence or independence, respectively. Figure 2.6 gives examples of different dependence functions. Because the dependence function describes the copula C uniquely, it is often used as a mean

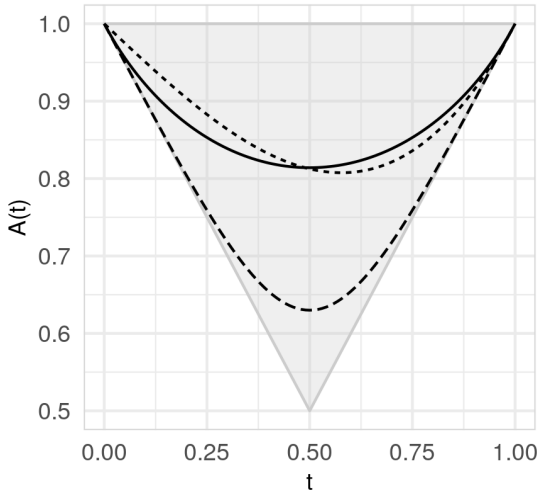

Figure 2.6: Different dependence functions: ——, medium symmetric dependence; - - - -, strong symmetric dependence, ·····, medium asymmetric dependence. Grey area marks lower and upper limits of equation (2.25).

to estimate copula parameters or to construct parameter tests (Genest et al., 2011; Gudendorf and Segers, 2012).

A generalization to d-dimensional extreme value copulas is possible and has the expression

$$C(u_1, \ldots, u_d) = \exp\left(\left(\sum_{j=1}^{d} \log(u_j)\right) A\left(\frac{\log(u_1)}{\sum_{j=1}^{d} \log(u_j)}, \ldots, \frac{\log(u_d)}{\sum_{j=1}^{d} \log(u_j)}\right)\right), \quad (2.26)$$

in which $A : \Delta_{d-1} \mapsto [1/d, 1]$ is convex with

$$\max(t_1, \ldots, t_d) \leq A(t_1, \ldots, t_d) \leq 1 \; \forall \, t_1, \ldots, t_d \in \Delta_{d-1} \quad (2.27)$$

and $\Delta_{d-1} = \{(t_1, \ldots, t_d) \in [0, 1]^d : \sum_{j=1}^{d} t_j = 1\}$.

Different estimators were proposed to estimate the (bivariate) Pickands dependence function from observed data. Genest and Segers (2009) compared them and recommended a rank based version of the estimator of Capéraà, Fougères, and Genest (1997). Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be bivariate observations that have the same distribution like (X, Y) and whose dependence can be described by bivariate copula C. Then pseudo observations are calculated as

$$U_i = \frac{1}{n+1} \sum_{i'=1}^{n} \mathbb{1}(X_{i'} \leq X_i), \quad V_i = \frac{1}{n+1} \sum_{i'=1}^{n} \mathbb{1}(Y_{i'} \leq Y_i). \quad (2.28)$$

The so-called CFG estimator of the logarithm of Pickands dependence function A is

$$\log A_n^{CFG}(t) = -\gamma - \frac{1}{n}\sum_{i=1}^{n} \log\left(\min\left(\frac{-\log U_i}{1-t},\ \frac{-\log V_i}{t}\right)\right), \tag{2.29}$$

with $\gamma \approx 0.577$ being Euler's constant. This estimator does not fulfil the condition $A(0) = A(1) = 1$ resulting from inequality (2.25). Therefore often a correction is applied:

$$\log A_{n,corr}^{CFG}(t) = \log A_n^{CFG}(t) - (1-t)\log A_n^{CFG}(0) - t \log A_n^{CFG}(1). \tag{2.30}$$

This corrected estimator is asymptotically normal and unbiased (Genest and Segers, 2009). In this dissertation only estimators based on bivariate Pickands dependence functions are applied. For estimators of copulas with more than two dimensions it is referred to Gudendorf and Segers (2012).

2.3.5 Mixed copulas

In popular statistical software only few extreme value copulas are implemented that can model more than two dimensions. A useful result from Durante and Salvadori (2010, Corollary 2) allows for combining or mixing of copulas. This way, a much more complex dependency structure can be created from simple copulas.

Let C_{θ_1} and C_{θ_2} be copulas and $a_j \in [0,1],\ j = 1,\ldots,d$, then

$$C(u_1,\ldots,u_d) = C_{\theta_1}(u_1^{a_1},\ldots,u_d^{a_d})\ C_{\theta_2}(u_1^{1-a_1},\ldots,u_d^{1-a_d}) \tag{2.31}$$

is also a copula. If C_{θ_1} and C_{θ_2} are extreme value copulas, C is also an extreme value copula. Because C is a (extreme value) copula again, through concatenation even more complex structures can be generated by the cost of additional parameters.

The introduced parameters $a_1,\ldots,a_d$ define the mixing proportions between the two copulas for each dimension separately and thus enable even asymmetrical dependence structures (using symmetrical copulas). This is especially helpful in hydrological applications because the river network inherently contains a directed structure. Because of river junctions and the direction of flow, stations more downstream are much more affected by upstream stations than otherwise.

2.3.6 Measures of dependence

Measures of dependence summarise the whole dependence structure to one value, but they are often useful to obtain a first impression of the degree of dependence or can be used to fit copulas to the data. The most popular measure is the (Bravais-Pearson) correlation, which measures the linear dependence of two random variables. Correlation is invariant against linear transformations but not generally against monotonous

ones. Since a copula is invariant against monotonous increasing transformations of the marginals (Genest and Favre, 2007), the correlation is not suitable as a measure of the dependence inherent in a copula.

In the following, two important measures of dependence that are invariant against monotonous transformations are given: the rank correlation of Spearman, that only considers the order of observations and not the observations itself, and the concordance measure of Kendall, that only takes pairwise comparisons into account. It is presented how both can be used to describe and fit copulas models. Further, the tail dependence is introduced which is a measure that assesses the dependence in the upper or lower tail of distributions.

Spearman's rho

The rank correlation of Spearman is calculated similarly to Pearson's correlation coefficient but uses observation ranks instead of the observations itself. This way the coefficient is invariant against monotonous transformations. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be random variables and let

$$R_i = \sum_{i'=1}^{n} \mathbb{1}(X_{i'} \leq X_i) \quad \text{and} \quad S_i = \sum_{i'=1}^{n} \mathbb{1}(Y_{i'} \leq Y_i) \tag{2.32}$$

be their ranks. Then the rank correlation is defined as

$$\rho_n = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2 \sum_{i=1}^{n}(S_i - \bar{S})^2}}, \tag{2.33}$$

with $\bar{R} = \bar{S} = \frac{n+1}{2}$ being the means of the ranks.

The corresponding rank correlation of a copula C can be derived directly as (Genest and Favre, 2007):

$$\rho = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3. \tag{2.34}$$

If C is an extreme value copula, the rank correlation can be calculated from Pickands dependence function as (Salvadori and De Michele, 2011):

$$\rho = 12 \int_0^1 \frac{1}{(1 + A(t))^2} dt - 3. \tag{2.35}$$

Rank correlation ρ_n of equation (2.33) is an asymptotically unbiased estimator of ρ in equations (2.34) and (2.35).

Kendall's tau

Kendall's tau, also called concordance measure, compares the amount of concordant observation pairs with the amount of discordant ones. A pair of observations, (X_1, Y_1) and (X_2, Y_2), is called concordant, if $X_1 < X_2$ and $Y_1 < Y_2$ or if $X_1 > X_2$ and $Y_1 > Y_2$ holds, i.e. if the order in each dimension is equal. Otherwise the pair is called discordant. Given random variables $(X_1, Y_1), \ldots, (X_n, Y_n)$ Kendall's tau can be calculated as

$$\tau_n = (C - D)/\binom{n}{2} = \frac{2}{n(n-1)} \sum_{i<j} \text{sgn}(X_i - X_j)\text{sgn}(Y_i - Y_j), \tag{2.36}$$

with C being the amount of concordant pairs, D the amount of discordant pairs, and sgn being the sign function.

If copula C describes the dependence between X and Y, Kendall's tau can be derived as (Genest and Favre, 2007):

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1. \tag{2.37}$$

If C is Archimedean with generator ϕ the following form can be used:

$$\tau = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt. \tag{2.38}$$

For an extreme value copula C we can also use Pickands dependence function A for calculation (Salvadori and De Michele, 2011):

$$\tau = \int_0^1 \frac{t(1-t)}{A(t)} dA'(t). \tag{2.39}$$

Again, Kendall's tau τ_n of equation (2.36) is an asymptotically unbiased estimator of τ in (2.37), (2.38), and (2.39).

Tail dependence

Tail dependence (Joe, 1997) are two measures that quantify the dependence within the lower and upper tails of two distributions. Therefore it is an important measure to summarise the dependence of a copula. Let X and Y be two random variables with cumulative distribution functions F_X and F_Y, respectively. Then upper tail dependence is defined as

$$\lambda_U = \lim_{t \to 1^-} P(Y > F_Y^{-1}(t) | X > F_X^{-1}(t)), \tag{2.40}$$

and lower tail dependence is defined as

$$\lambda_L = \lim_{t \to 0^+} P(Y \leq F_Y^{-1}(t) | X \leq F_X^{-1}(t)). \tag{2.41}$$

If those limits exist and a copula C describes the dependence between X and Y, according to Nelsen (Theorem 5.4.2, 2006) it follows that

$$\lambda_U = 2 - \lim_{t \to 1^-} \frac{1 - C(t,t)}{1-t} \quad \text{and} \quad \lambda_L = \lim_{t \to 0^+} \frac{C(t,t)}{t}. \tag{2.42}$$

For bivariate Archimedean copulas the following forms using generator ϕ are resulting from equation (2.42):

$$\begin{aligned} \lambda_U &= 2 - \lim_{t \to 1^-} \frac{1 - \phi^{[-1]}(2\phi(t))}{1-t} = 2 - \lim_{x \to 0^+} \frac{1 - \phi^{[-1]}(2x)}{1 - \phi^{[-1]}(x)}, \\ \lambda_L &= \lim_{t \to 0^+} \frac{\phi^{[-1]}(2\phi(t))}{t} = \lim_{x \to \infty} \frac{\phi^{[-1]}(2x)}{\phi^{[-1]}(x)}. \end{aligned} \tag{2.43}$$

For bivariate extreme value copulas with dependence function A we get

$$\begin{aligned} \lambda_U &= 2 - 2\,A(1/2), \\ \lambda_L &= \begin{cases} 0, & A(1/2) > 1/2, \\ 1, & A(1/2) = 1/2. \end{cases} \end{aligned} \tag{2.44}$$

2.3.7 Important copula families

After relevant classes of copulas as well as summary statistics have been presented, important copula families and their properties are given now.

First we consider simple copulas that are directly constructed from the joint cumulative distribution function using equation (2.17). Assuming independence between the dimensions the joint distribution function can be written as product of marginal distribution functions. The **independence copula** therefore is written as

$$\begin{aligned} \Pi(u_1, \ldots, u_d) := C(u_1, \ldots, u_d) &= F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)) \\ &= F_1(F_1^{-1}(u_1)) \cdot \ldots \cdot F_d(F_d^{-1}(u_d)) = \prod_{i=1}^{d} u_i. \end{aligned} \tag{2.45}$$

If the distribution function can be written as minimum of the corresponding marginal distribution functions, the dimensions feature a perfect positive dependence, they

are comonotonic. In this case the **copula of comonotonicity** is resulting:

$$\begin{aligned} M(u_1, \ldots, u_d) := C(u_1, \ldots, u_d) &= F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)) \\ &= \min_{i=1,\ldots,d} F_i(F_i^{-1}(u_i)) = \min_{i=1,\ldots,d} u_i. \end{aligned} \tag{2.46}$$

Of course, these two trivial copulas do mostly not allow for a realistic modelling of dependence, but they are important for the modelling of extreme cases and they are special cases in many copula families.

Another basic copula family are **Gaussian copulas**. Because of the popularity of the multivariate Gaussian distribution the corresponding copula is an important family even in this work, although this copula family is less appropriate in extreme value analysis. As well, the Gaussian copula is defined by equation (2.17):

$$C(u_1, \ldots, u_d) = \Phi_R\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right), \tag{2.47}$$

with Φ^{-1} being the inverse cumulative distribution function of a standard Gaussian distribution and Φ_R being the joint distribution function of a centred multivariate Gaussian distribution with correlation matrix R. Therefore the Gaussian copula describes the dependence structure that exists within multivariate Gaussian distributions. The parameter $R \in [-1, 1]^{d \times d}$ indicates the correlation structure.

As already stated, the Gaussian copula is rather inappropriate when modelling extreme events (its misuse is even said to have had an impact on the financial crisis of 2008, Salmon, 2012). This is due to the fact that no upper or lower tail dependence can be modelled but only dependencies in the centre of the distributions by modifying the correlation parameter. Extreme events, however, tend to have dependencies particularly in the tails whereas dependencies in the centre are often much lower.

A simple copula with upper tail dependence is the **Gumbel copula**. It is defined as

$$C_\theta(u_1, \ldots, u_d) = \exp\left(-\left(\sum_{i=1}^{d}(-\ln u_i)^\theta\right)^{1/\theta}\right), \tag{2.48}$$

with $\theta \in [1, \infty)$. The independence copula and the copula of comonotonicity are limit cases of this family and result from $\theta = 1$ and $\theta \to \infty$, respectively. The Gumbel copula family belongs to the group of Archimedean copulas as well as to the group of extreme value copulas with dependence function $A(t) = (t^\theta + (1-t)^\theta)^{1/\theta}$ in the bivariate case.

Because the Gumbel copula only describes relatively simple, symmetric dependencies, **mixed Gumbel copulas** are constructed according to Section 2.3.5. Using two Gumbel copulas C_{θ_1} and C_{θ_2} and mixing parameters $a_1, \ldots, a_d \in [0, 1]$ the dependence

Table 2.1: Overview of used copula families and some of their properties.

Copula	Parameter	Generator	ρ	τ	λ_L	λ_U
Independence	-	$\phi(t) = -\ln(t)$	0	0	0	0
Comonotonicity	-	-	1	1	1	1
Gaussian	$R \in [-1,1]^{d\times d}$	-			0	0
Clayton	$\theta \in (0,\infty)$	$\phi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$		$\theta/(\theta+2)$	$2^{-1/\theta}$	0
Gumbel	$\theta \in [1,\infty)$	$\phi(t) = (-\ln t)^{\theta}$		$1 - 1/\theta$	0	$2 - 2^{1/\theta}$

function in the bivariate case is written as

$$A(t) = ((a_1 t)^{\theta_1} + (a_2(1-t))^{\theta_1})^{1/\theta_1} + (((1-a_1)t)^{\theta_2} + ((1-a_2)(1-t))^{\theta_2})^{1/\theta_2}, \quad (2.49)$$

(Salvadori and De Michele, 2011) from which the measures of dependence can be calculated depending on θ_1 and θ_2 using equations (2.35), (2.39), and (2.44).

The **Clayton copula** is used as a strong contrast to the Gumbel copula and mixed Gumbel copula. It is defined as

$$C_\theta(u_1, \ldots, u_d) = \left(\sum_{i=1}^{d} u_i^{-\theta} - 1\right)^{-1/\theta}, \quad (2.50)$$

with $\theta > 0$. Like the Gumbel copula the Clayton copula family is Archimedean and contains the independence copula and copula of monotonicity as limiting cases for $\theta \to 0$ and $\theta \to \infty$, respectively. However, the Clayton copula is not an extreme value copula and describes no upper tail dependence but only lower tail dependence.

Table 2.1 contains the simple copula models used in this dissertation, their parameters and some important properties. Figure 2.7 illustrates the spectrum of different dependence structures that can be modelled with the copulas presented. In it $n = 10\,000$ random numbers that are distributed according to different copulas are displayed. Copula families vary along the rows, degree of dependency along the columns. In each column, the rank correlation is the same for each copula family.

The topmost line displays Gaussian dependence, that has no lower or upper tail dependence. The second and third line from the top give the Clayton and Gumbel copula with lower and upper tail dependence, respectively. The bottom line contains a mixed Gumbel copula featuring asymmetric dependence. Although the rank correlation is identical in each column, the dependence structures are quite different. This highlights the importance of attentive modelling of dependence.

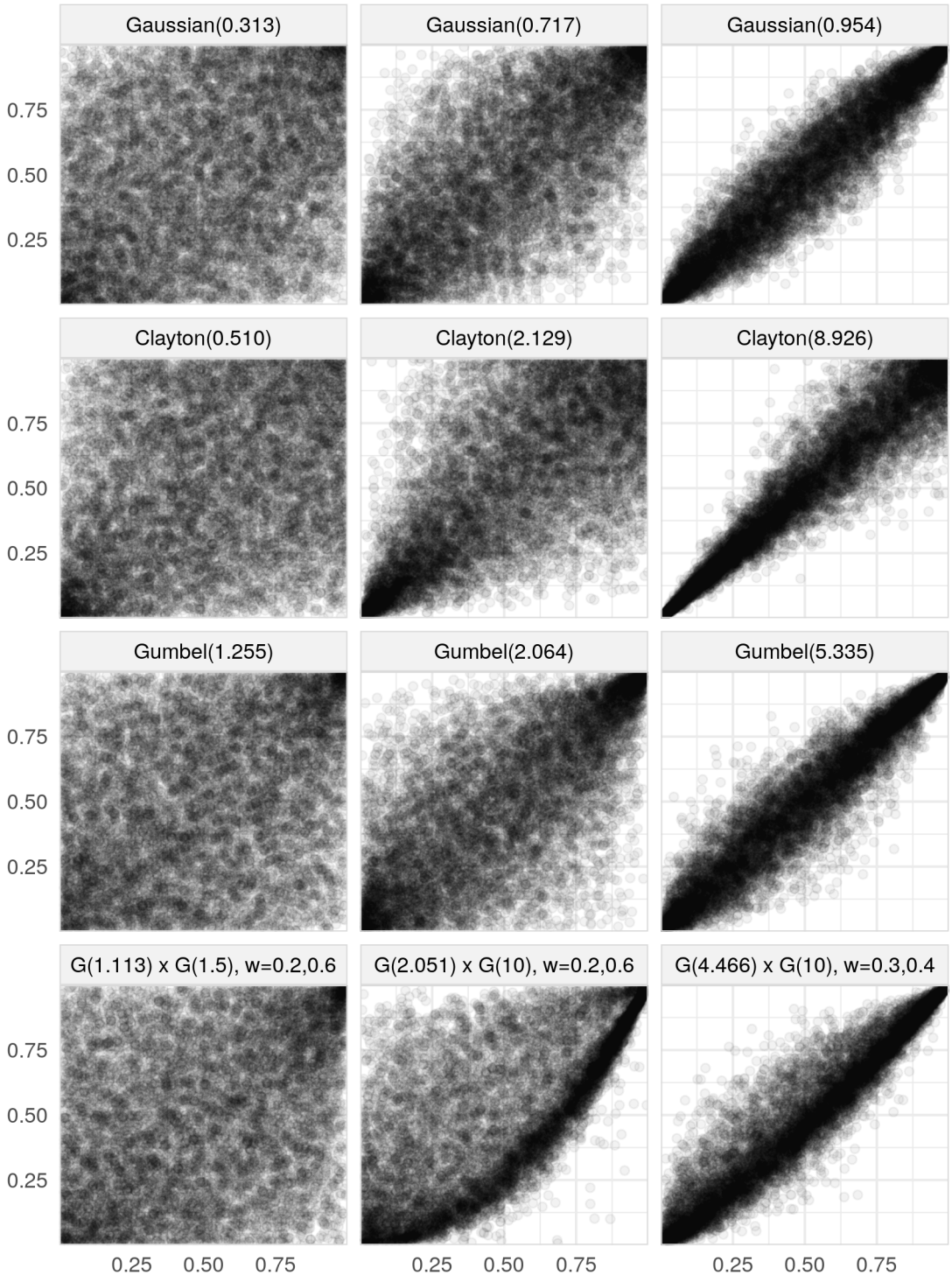

Figure 2.7: Comparison of random numbers generated using different copula models (rows) and different parameters (columns). Rank correlation ρ is identical in each column and amounts to $\rho = 0.3, 0.7, 0.95$. This illustrates differences between copula models with equally specified rank correlation.

2.3.8 Fitting of copulas

Often copulas are fitted to data using the Maximum-Pseudolikelihood approach, which maximizes the sum of logarithms of copula densities evaluated at pseudo-observations. Let $X_{1;j}, \ldots, X_{n;j}$ be variables distributed like X_j and let $F_{n;j}$ be the corresponding empirical distribution function, $j = 1, \ldots, d$. The Maximum-Pseudolikelihood estimator is the solution to

$$\arg\max_{\theta} \ell(\theta) = \arg\max_{\theta} \sum_{i=1}^{n} \log c_\theta(F_{n;1}(X_{i;1}), \ldots, F_{n;d}(X_{i;d})), \tag{2.51}$$

in which $c_\theta(u_1, \ldots, u_d) = \frac{\partial^d}{\partial u_1 \ldots \partial u_d} C_\theta(u_1, \ldots, u_d)$ indicates the density of a copula C_θ. This approach often needs numerical optimization but is generally applicable.

In many cases the relation between copula parameters and measures like Spearman's ρ or Kendall's τ can be used to derive the parameters similar to the method of moments in estimation of distribution parameters.

In Section 2.3.6 relations between bivariate copulas and measures of dependence were described. For example, for a Gumbel copula with parameter θ Kendall's tau is $\tau = 1 - 1/\theta$, or, conversely, the copula parameter dependent on Kendall's tau is $\theta = 1/(1 - \tau)$. If Kendall's tau is calculated from data as $\hat{\tau}$, the corresponding parameter of a Gumbel copula can be estimated as

$$\hat{\theta} = \frac{1}{1 - \hat{\tau}}. \tag{2.52}$$

This is only possible for bivariate copulas, because Kendall's tau is only defined for the bivariate case. However, if the assumption of equal dependence between all dimensions is justified, a mean of the pairwise correlation can be used to estimate the parameter. If the assumption is not justified, this leads to biased estimates. In these cases simple copula models are inappropriate and more complex models and estimations are needed.

For extreme value copulas Pickands dependence function can be used for parameter estimation. Salvadori and De Michele (2011) recommend this in the context of parameter estimation of mixed copulas. Their estimator minimizes the sum of pairwise quadratic differences between the theoretical and empirical Pickands dependence function at evaluation points $x_k = k/n, k = 1, \ldots, n-1$:

$$\hat{\theta} = \arg\min_{\theta} \sum_{j=1}^{d-1} \sum_{j'=i+1}^{d} \sum_{k=1}^{n-1} |A_{jj'}(x_k; \theta) - \hat{A}_{jj'}(x_k)|^2. \tag{2.53}$$

Here $A_{jj'}(x_k; \theta)$ is the value of Pickands dependence function between dimension j and j' of a copula C with parameter θ and $\hat{A}_{jj'}$ is the corresponding estimator (for example the corrected CFG estimator of equation (2.30)).

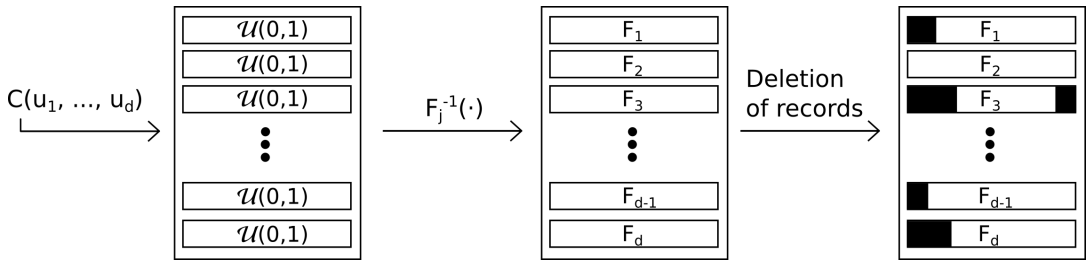

Figure 2.8: Schematic representation of generating multivariate data sets of different lengths with a copula C and marginal distributions F_j.

2.3.9 Random number generation

Being multivariate distributions, copulas can be used to generate synthetic data sets which feature specific dependence structures and marginal distributions. Figure 2.8 gives a schematic overview on the procedure to generate data sets of different record lengths given a copula C and marginal distributions $F_1, \ldots, F_d$.

First, uniformly distributed data with specified dependence structure is generated with copula C. Since a copula is a distribution with uniformly distributed margins, each component follows a uniform distribution on $[0, 1]$ at first. To get the desired marginal distributions, the inverse probability integral transform is applied at each margin as a next step. If the data should consist of different record lengths or missing values, this can be achieved by deletion of respective entries as final step. This is often necessary to mimic the structure of real data in hydrology that normally features different observation periods and gaps due to maintenance or reconstruction.

Depending on the copula, different algorithms can be used to do the first step. One general approach is to begin with one independent uniformly-distributed vector $(v_1, \ldots, v_d)'$ and recursively transform each component using the conditional distribution depending on the former components (Bouyé et al., 2000). To generate random numbers from Archimedean copulas, like Gumbel or Clayton, a more efficient approach is described in Marshall and Olkin (1988). Algorithms for generating random numbers for mixed copulas can be found in Durante and Salvadori (2010) and Salvadori and De Michele (2010).

2.4 Distribution parameter estimation

In this section different estimators are presented for fitting distribution parameters to data. In hydrology (in contrast to other disciplines) probability-weighted moments or L-moments are very popular because they offer better estimation properties for small samples in comparison to Maximum-Likelihood methods. Trimmed L-moments are a generalization that has advantages when analysing highly skewed data. As an

alternative the penalized Maximum-Likelihood estimator is presented. This is a procedure that makes it possible to restrict the parameter space or to penalize unrealistic estimations. This way, at the expense of some bias, the estimation variance can often be reduced which results in a better overall quadratic deviation in many situations, especially in small sample cases. In Chapter 5 the application of the penalized Maximum-Likelihood estimator will be further investigated and applied to regional flood frequency analysis.

2.4.1 Probability-weighted moments

Probability-weighted moments (PWMs) were described first by Greenwood et al. (1979) with the goal to estimate parameters for distributions that are only defined by their inverse distribution function. Similar to the method of moments, relations between distribution parameters and probability-weighted moments can be used to convert empirical PWMs to parameter estimations. In Hosking, Wallis, and Wood (1985) the PWMs of a generalized extreme value distribution and their corresponding PWM estimators are calculated. Superiority to Maximum-Likelihood estimation in small samples ($n < 50$) is described in this work. PWMs have been replaced by L-moments in hydrological applications (see the following section), but because many properties of PWMs can be transferred to L-Moments, they are still relevant.

Definition & estimation

Let F be the cumulative distribution function of a continuous random variable X. Further let the inverse distribution function F^{-1} be given in explicit form. Then probability-weighted moments are defined as

$$M_{p,r,s} = E(X^p F(X)^r (1 - F(X))^s) = \int_0^1 (F^{-1}(x))^p x^r (1-x)^s dx, \tag{2.54}$$

in which $p, r, s \in \mathbb{R}$. With $r = s = 0$ this leads to classical non-centred product moments.

For derivation of parameters, Greenwood et al. (1979) uses either $M_{1,0,s}, s = 0, 1, \ldots$, or $M_{1,r,0}, r = 0, 1, \ldots$, where the moments of one of these series can each be represented as a linear combination of the moments of the other series. The possible advantage of these series to describe a distribution is that the power of X is set to one and only the powers of $F(X)$ or $(1 - F(X))$ are varied. Therefore, extraordinary values do not influence the outcome that much as for ordinary product moments. Hosking, Wallis, and Wood (1985) exclusively use the series $M_{1,r,0}, r = 0, 1, \ldots$, which from now on we will call

$$\beta_r = M_{1,r,0} = E(XF(X)^r). \tag{2.55}$$

From an ordered sample $x_{1:n} < \ldots < x_{n:n}$ an unbiased estimation of β_r can be calculated by (Hosking, Wallis, and Wood, 1985)

$$\hat{\beta}_r = \frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)!(n-r-1)!}{(i-r-1)!(n-1)!}x_{i:n} = \frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)(i-2)\ldots(i-r)}{(n-1)(n-2)\ldots(n-r)}x_{i:n}. \quad (2.56)$$

The derivation of PWM-based parameter estimators is skipped here, because it is analogous to the derivation using L- or TL-moments given in the following sections. It is referred to Landwehr, Matalas, and Wallis (1979) and Hosking, Wallis, and Wood (1985) for more information on this.

2.4.2 L-moments

L-moments (Hosking, 1990) are highly popular in hydrology as measures to describe and identify distributions as well as to derive distribution parameters. In comparison to classic product moments they are less sensitive to outliers or measurement errors and they yield better properties than Maximum-Likelihood methods for parameter estimation in small sample cases. They are closely related to probability-weighted moments but are more descriptive and easier to interpret.

Definition

The most descriptive definition of L-moments (more possibilities to describe them follow) defines them as linear combinations of expectations of order statistics. Let X be a continuous random variable with cumulative distribution function F and further let $X_{1:n} \leq \ldots \leq X_{n:n}$ be increasing order statistics of a hypothetical sample of size n from the distribution of X. Then the first four L-moments are defined as

$$\lambda_1 = E(X_{1:1}), \quad (2.57)$$

$$\lambda_2 = \frac{1}{2}E(X_{2:2} - X_{1:2}), \quad (2.58)$$

$$\lambda_3 = \frac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3}), \quad (2.59)$$

$$\lambda_4 = \frac{1}{4}E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}). \quad (2.60)$$

Hence, the first L-moment is equal to the expectation of X and therefore a measure of the central location of distribution F. The second L-moment is calculated as the half of the expected range that results if a sample of size two is drawn. It is a measure of variability. The third L-moment measures skewness because, under the assumption of symmetry, the distance of the upper and lower order statistic, $X_{3:3}, X_{1:3}$, respectively, to the mean order statistic, $X_{2:3}$, would be equal and therefore

$X_{3:3} - 2X_{2:3} + X_{1:3} = (X_{3:3} - X_{2:3}) - (X_{2:3} - X_{1:3}) = 0$ would follow. Positive values indicate a positive skewness because the distance between the upper and middle order statistic is higher than between the lower and middle one. For negative values it is the other way round. The fourth L-moment describes the kurtosis of the distribution. This becomes clear because $X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4} = (X_{4:4} - X_{1:4}) - 3(X_{3:4} - X_{2:4})$ compares the distance between the extreme order statistics with the distance of the two central order statistics. High values therefore indicate leptokurtic distributions, low values platykurtic ones.

Generally, the L-moment of order r can be written as

$$\lambda_r = \frac{1}{r} \sum_{i=0}^{r-1} (-1)^i \binom{r-1}{i} E(X_{r-i:r}). \tag{2.61}$$

Alternatively, L-moments can be described by integrals over the inverse distribution function of F. The expectation of an order statistic can be written as

$$E(X_{i:n}) = \frac{n!}{(i-1)!(n-i)!} \int_0^1 F^{-1}(x) x^{i-1} (1-x)^{n-i} dx \tag{2.62}$$

(David and Nagaraja, 2004). Then, according to Hosking (1990) the r-th L-moment is given by

$$\lambda_r = \int_0^1 F^{-1}(x) P^*_{r-1}(x) dx, \tag{2.63}$$

with $P^*_r(x) = \sum_{i=0}^{r} p^*_{r,i} x^i$ being the so-called r-th shifted Legendre polynomial and $p^*_{r,i} = (-1)^{r-i} \binom{r}{i} \binom{r+i}{i}$.

L-moments can also be written as linear combination of probability-weighted moments (and vice versa). With $\beta_0, \ldots, \beta_{r-1}$ being the first r PWMs (see equation (2.55)), the r-th L-moment is given as

$$\lambda_r = \sum_{i=0}^{r-1} p^*_{r-1,i} \, \beta_i. \tag{2.64}$$

Analogous to product moments, L-moments of higher order are usually standardised to be independent of the unit of measurement. These L-moment ratios are calculated as

$$\tau_2 = \lambda_2 / \lambda_1, \quad \tau_r = \lambda_r / \lambda_2, \; r \in \{3, 4, \ldots\}. \tag{2.65}$$

The L-moment ratios τ_2, τ_3, τ_4 are called L-CV (coefficient of variation), L-skewness, and L-kurtosis, respectively.

L-moments of a distribution exist if the first moment, the expectation, exists. In this case, the set of L-moments uniquely describes the corresponding distribution.

Empirical L-moments

Empirical L-moments can be calculated by plugging in empirical PWMs (see equation (2.56)) into equation (2.64). Let $x_{1:n} \leq \ldots \leq x_{n:n}$ be an increasingly ordered sample of size n. The r-th empirical L-moment can be calculated as

$$\hat{\lambda}_r = \sum_{k=0}^{r-1} p^*_{r-1,k}\, \hat{\beta}_k = \sum_{k=0}^{r-1} \frac{1}{n} \sum_{i=1}^{n} \frac{(i-1)(i-2)\cdots(i-k)}{(n-1)(n-2)\cdots(n-k)} x_{i:n}. \tag{2.66}$$

Alternatively the property

$$E(X_{m:r}) = \binom{n}{r}^{-1} \sum_{k=m}^{n-r+m} \binom{k-1}{m-1}\binom{n-k}{r-m} E(X_{k:n}) \tag{2.67}$$

of representation (2.61) can be used, which leads to estimator (Hosking, 1990)

$$\hat{\lambda}_r = \frac{1}{r}\binom{n}{r}^{-1} \sum_{i=1}^{n}\sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j}\binom{i-1}{r-1-j}\binom{n-i}{j} x_{i:n}. \tag{2.68}$$

In Hosking and Balakrishnan (2015) another estimator is proposed that uses recursively calculated weights. In a study they compare the PWM-based estimator (2.66), the direct estimator (2.68), and the recursive estimator. The latter method leads to the smallest calculation errors (due to rounding errors) while the PWM-based calculation is the worst in this regard.

Empirical L-moment ratios are calculated as

$$\hat{\tau}_2 = \hat{\lambda}_2/\hat{\lambda}_1, \quad \hat{\tau}_r = \hat{\lambda}_r/\hat{\lambda}_2,\ r \in \{3, 4, \ldots\}. \tag{2.69}$$

Empirical L-moments are unbiased estimators of their population counterparts while empirical L-moment ratios are biased. According to Hosking and Wallis (1997) this bias is moderate for samples consisting of at least 20 observations.

L-moment based parameter estimation

Besides their application as descriptive measures, L-moments are mostly used to derive parameter estimations. Analogue to the regular method of moments, relations between distribution parameters and L-moments are used to derive parameter estimation formulas.

This section presents the L-moment parameter estimators for a generalized extreme value distribution. The estimators for other distributions can be found in Hosking (1990) or can be derived analogously.

The first three L-moments of a GEV(μ, σ, ξ) can be calculated by putting its inverse distribution function into equation (2.63), which leads to:

$$\lambda_1 = \mu - \frac{\sigma}{\xi}(1 - \Gamma(1-\xi)), \tag{2.70}$$

$$\lambda_2 = \frac{\sigma}{\xi}(2^\xi - 1)\Gamma(1-\xi), \tag{2.71}$$

$$\lambda_3 = \frac{\sigma}{\xi}(2 \cdot 3^\xi - 3 \cdot 2^\xi + 1)\Gamma(1-\xi). \tag{2.72}$$

Hence, the L-skewness is

$$\tau_3 = \frac{\lambda_3}{\lambda_2} = \frac{2 \cdot 3^\xi - 3 \cdot 2^\xi + 1}{2^\xi - 1}, \tag{2.73}$$

and only depends on the shape parameter ξ.

For parameter estimation the formulas (2.70), (2.71), and (2.73) are solved for the parameters and the empirical L-moments or L-moment ratios are plugged in. The value of $\hat{\xi}$ can be determined numerically from equation (2.73), or the following approximation can be used (Hosking, 1990):

$$\hat{\xi} = -7.859z - 2.9554z^2 \quad \text{with} \quad z = \frac{2}{3 + \hat{\lambda}_3/\hat{\lambda}_2} - \frac{\ln 2}{\ln 3}. \tag{2.74}$$

Estimators of $\hat{\sigma}$ and $\hat{\mu}$ then follow as

$$\hat{\sigma} = \hat{\lambda}_2\hat{\xi}(1 + 2^{\hat{\xi}})\Gamma(1-\hat{\xi}), \tag{2.75}$$

$$\hat{\mu} = \hat{\lambda}_1 + \frac{\hat{\sigma}}{\hat{\xi}}(1 - \Gamma(1-\hat{\xi})). \tag{2.76}$$

2.4.3 TL-moments

An alternative to L-moments is a generalization of them, the trimmed L-moments, or TL-moments. In comparison to L-moments two added parameters allow for trimming of upper and lower order statistics. This can lead to a more robust parameter estimation, which is an advantage especially when dealing with skewed data. TL-moments were introduced by Elamir and Seheult (2003) at first, Hosking (2007) and Hosking and Balakrishnan (2015) developed their theory further.

Definition

As before let us assume that $X_{1:n} \leq \ldots \leq X_{n:n}$ are order statistics from a hypothetical sample of size n following a continuous distribution F. The r-th TL(s,t)-moment

with trimming parameters $s, t \in \mathbb{N}_0$ is given by (Elamir and Seheult, 2003)

$$\lambda_r^{(s,t)} = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(X_{r+s-k:r+s+t}). \tag{2.77}$$

With trimming parameters set to zero, $s = t = 0$, the regular L-moments result as a special case. The trimming parameters therefore indicate how many lower or upper order statistics are skipped in the linear combination. For example, the first TL(1,1)-moment, $\lambda_1^{(1,1)} = E(X_{2:3})$, is the expectation of the central element of a sample of size three. The second TL(1,1)-moment, $\lambda_2^{(1,1)} = \frac{1}{2}E(X_{3:4} - E_{2:4})$, is half of the expected range of the inner elements from a sample of size four. In the formula of both moments the smallest and the highest order statistic are skipped corresponding to a trimming of $s = 1$ and $t = 1$.

Analogously to L-moments, TL-moments of a distribution F can be expressed using an integral over the inverse distribution function:

$$\begin{aligned} \lambda_r^{(s,t)} = \frac{1}{r} \sum_{k=0}^{r-1} \binom{r-1}{k} \frac{(r+s+t)!}{(r+s-k-1)!(t+k)!} \times \\ \int_0^1 F^{-1}(x) x^{r+s-k-1} (1-x)^{t+k} dx. \end{aligned} \tag{2.78}$$

TL-moments can also be stated as linear combinations of PWMs (Hosking, 2007):

$$\lambda_r^{(s,t)} = \sum_{i=s}^{r+s+t-1} z_{r-1,i}^{(s,t)} \beta_i, \tag{2.79}$$

with

$$z_{r,i}^{(s,t)} = \frac{r!(r+s+t+1)!}{(r+1)(r+s)!(r+t)!} (-1)^{s+r+i} \binom{r+t}{s+i} \binom{r+i}{r}. \tag{2.80}$$

TL-moment ratios are given as

$$\tau_2^{(s,t)} = \lambda_2^{(s,t)} / \lambda_1^{(s,t)}, \quad \tau_r^{(s,t)} = \lambda_r^{(s,t)} / \lambda_2, \quad r \in \{3, 4, \ldots\}, \tag{2.81}$$

with $\tau_2^{(s,t)}, \tau_3^{(s,t)}$, and $\tau_4^{(s,t)}$ being called TL(s,t)-CV, TL(s,t)-skewness, and TL(s,t)-kurtosis, respectively.

All TL(s,t)-moments of a distribution exist if the expectations $E(\max(-X, 0)^{1/(s+1)})$ and $E(\max(X, 0)^{1/(t+1)})$ exist (Hosking, 2007). This is an advantage over L-moments because this condition is weaker than that for L-moments. For example, L-moments of a Cauchy distribution do not exist (because the first moment does not exist), but TL(1,1)-moments can be calculated and used for parameter estimation. For other distribution families the L-moments do not exist for all parameter combinations (L-moments of a GEV distribution only exist for $\xi < 1$, for example). In these cases

TL-moments can be useful for parameter estimators that are not restricted to that space or, at least, to a larger parameter space.

One property of TL-moments useful for calculating empirical TL-moments is that TL-moments, if they exist, can be expressed as a linear combination of TL-moments with lower trimming parameters (Hosking, 2007):

$$(2r+s+t-1)\lambda_r^{(s,t)} = (r+s+t)\lambda_r^{(s,t-1)} - \frac{1}{r}(r+1)(r+s)\lambda_{r+1}^{(s,t-1)}, \tag{2.82}$$

$$(2r+s+t-1)\lambda_r^{(s,t)} = (r+s+t)\lambda_r^{(s-1,t)} + \frac{1}{r}(r+1)(r+t)\lambda_{r+1}^{(s-1,t)}. \tag{2.83}$$

By chaining this so-called recurrence property, TL-moments of arbitrary trimming can be reduced to a combination of simple L-moments, for example:

$$\lambda_r^{(0,2)} = \frac{(r+1)(r+2)}{2r(2r+1)}\lambda_r - \frac{r+2}{2r}\lambda_{r+1} + \frac{r+2}{2(2r+1)}\lambda_{r+2}. \tag{2.84}$$

Empirical TL-moments

From an ordered sample $x_{1:n} \leq \ldots \leq x_{n:n}$ the empirical r-th TL(s,t)-moment can be calculated as (Elamir and Seheult, 2003; Hosking, 2007)

$$\begin{aligned}\hat{\lambda}_r^{(s,t)} &= \sum_{j=s+1}^{n-t} \frac{1}{r\binom{n}{r+s+t}} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k}\binom{j-1}{r+s-k-1}\binom{n-j}{t+k} x_{j:n} \\ &= \sum_{j=s+1}^{n-t} w_j\, x_{j:n}.\end{aligned} \tag{2.85}$$

This estimator can be deduced by plugging in the PMW estimator from equation (2.56) into equation (2.79). Alternatively, an estimator can be constructed using L-moment estimators and the recurrence properties (2.82) and (2.83). In Hosking and Balakrishnan (2015) different implementations are compared regarding computation time and numerical stability. Their result is that for medium degrees of trimming ($s+t<3$) calculation using the recurrence properties is superior to other methods.

Corresponding empirical TL-moment ratios are calculated as

$$\hat{\tau}_2^{(s,t)} = \hat{\lambda}_2^{(s,t)}/\hat{\lambda}_1^{(s,t)} \text{ and } \hat{\tau}_r^{(s,t)} = \hat{\lambda}_r^{(s,t)}/\hat{\lambda}_2^{(s,t)},\ \ r \in \{3,4,\ldots\}. \tag{2.86}$$

The differences between trimming parameters regarding calculation of empirical TL-moments can be best compared by means of the weights w_j of formula (2.85). Figure 2.9 illustrates the weights for the first three TL-moments for increasing asymmetrical or symmetrical trimmings. For the first TL-moments it is easily noticeable that the classic L-moment (which is the arithmetic mean) weighs all

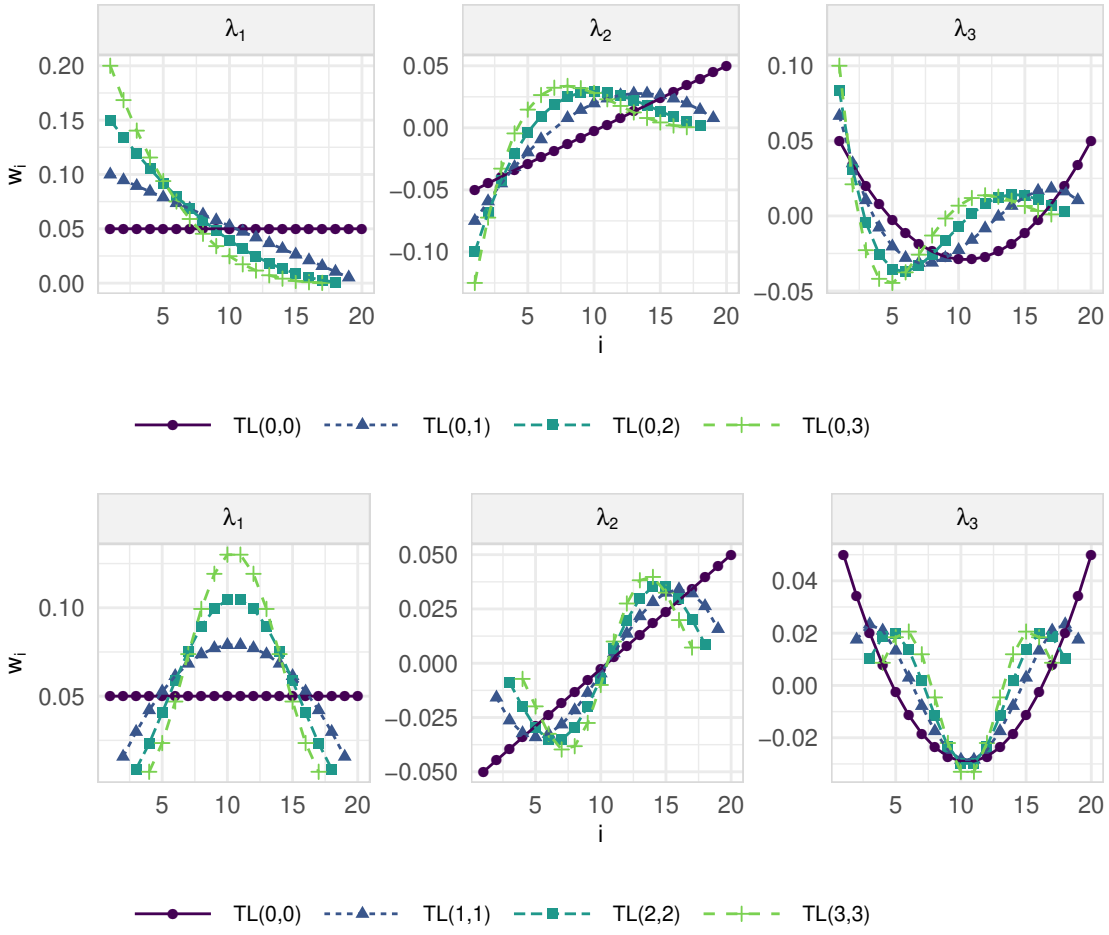

Figure 2.9: Weights of ordered observations for empirical asymmetrically (top) or symmetrically trimmed (bottom) TL-moments with different trimming parameters.

observations equally. Asymmetrically trimmed TL(0,t)-moments decreasingly weigh larger observations and increasingly weigh the lower ones, while symmetrical trimming puts emphasise on the central observations. Higher moments are more difficult to interpret, but it is visible that different trimmings emphasize different sections of the ordered sample. For TL-moments with positive trimming the smallest and/or highest observations are neglected in the calculation (see equation (2.85)), so they do not get any weight at all.

Estimation of distribution parameters

Parameter estimation using TL-moments works analogously to that for L-moments. By formula (2.78) theoretical TL-moments dependent of distribution parameters

can be calculated using the inverse distribution function. By plugging in empirical TL-moments or TL-moment ratios into those equations, parameter estimations can be deduced.

Elamir and Seheult (2003) give calculations of TL(1,1)-moments of the Gaussian, Cauchy, logistic, and exponential distribution. Formulas for TL(0,1)-, TL(0,2)-, and TL(1,1)-moments of the generalized extreme value distribution can be found in Lilienthal (2014).

2.4.4 Penalized Maximum-Likelihood estimator

The penalized or regularized Maximum-Likelihood estimator (PMLE) is an adoption of the regular Maximum-Likelihood estimator (MLE). It adds a penalization term that can be used to push the model in a specified direction in order to avoid unrealistic estimations or to add further information to the estimation. Popular examples of PMLE are the Ridge and the Lasso estimator in regression analysis. Here, high (quadratic or absolute) values of the coefficient vector are penalized which leads to improved estimation properties and an in-built variable selection in case of the Lasso estimator.

Let f_θ be the density function of distribution F_θ with parameter vector $\theta \in \Theta$ and let $X_1, \ldots, X_n$ be i.i.d. observation variables following distribution F_θ. The regular MLE chooses the parameter vector that optimizes the product of density values or likewise the sum of logarithms of density values:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \prod_{i=1}^{n} f_\theta(X_i) = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \log f_\theta(X_i). \tag{2.87}$$

More information on this estimator for the special case that F_θ is a GEV distribution can be found in Prescott and Walden (1980). Since the support of the GEV distribution depends on the parameters, classical Maximum-Likelihood asymptotics are not trivial. Smith (1985) stated that asymptotic normality holds for $\xi > -0.5$, which was completely proven in Bücher and Segers (2017).

Penalized estimation adds a penalization term to that maximization that is depending on θ. The PMLE is given as

$$\hat{\theta}_\lambda = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \log f_\theta(X_i) - \lambda' \Omega(\theta), \tag{2.88}$$

with hyperparameter $\lambda \in \mathbb{R}^p$ and penalization term $\Omega : \Theta \to \mathbb{R}^p$.

The penalization term is used to influence the parameter estimation. It can be used to avoid unrealistic parameter estimations or to include expert knowledge or further sources of information, like from nearby or similar stations in the hydrological context.

This new possibilities come with the need to choose the degree of penalization by modifying the hyperparameter. This often leads to a trade-off between variance and bias. For $\lambda = 0$ the ML estimator is calculated that is unbiased but often yields high variance. With increasing λ usually the bias raises and the variance decreases. By skilfully selecting the hyperparameter often an improvement in terms of MSE can be achieved.

In Chapter 5 this estimator is applied to hydrological issues and especially to regional flood frequency analysis.

2.5 Regional flood frequency analysis

We now come back to analysing the distributional behaviour of flood peaks. Since most measuring stations were built during the last century, nowadays often only the data of some decades is available. Using the block maxima approach, i.e. only using the maxima of each year, often leaves us with relatively small sample records. This is critical because often the interest of the analysis is the estimation of a rather high quantile, like the 99.9%-quantile corresponding to a return period of 1000 years. The extrapolation to such extreme values hugely amplifies the uncertainty of parameter estimation based on small samples.

A tool to improve this situation is regionalization. The idea is that the data of a station is not analysed on its own, but that further information from similar measuring stations is somehow included to improve estimation uncertainty. A special case of regionalization is the situation in which a quantile should be calculated at a site with no prior measurements at all, i.e. at a site with a record length of zero. In this case we have to completely rely on regional information to obtain some assessment.

The term “regionalization” does not imply a specific method of analysis, but it stands generally for procedures in which regional information is used to improve a local estimator. Applied procedures vary vastly in general conception and in statistical methodology.

One possibility is to use regression techniques to establish relations between flood characteristics like the 100-year flood and site characteristics like the mean height or the mean slope. Regression models constructed using observations from a set of stations and their characteristics can be applied to sites with only few observations or to sites with no observations at all. A comparison of the performance of regression models and parameter estimation techniques for estimating flood quantiles at ungauged sites can be found in Pandey and Nguyen (1999). Another approach are interpolation methods like topological kriging (Skøien, Merz, and Blöschl, 2006) or a kriging technique applied in a physiographical space that is built by applying canonical correlation analysis or principal component analysis on site characteristics

(Chokmani and Ouarda, 2004). A very popular approach nowadays is based on the Index Flood model and uses regionalized L-moments to build a common distributional form. It is now given in more detail.

The core concept of the so-called Index Flood model (Dalrymple, 1960) is the assumption that the distributions of annual peaks at several stations in a region is the same if normalized by some scaling factor called index flood. Expressed as a formula this means that for $X_1, \dots, X_d$ being the random variables of d sites with cumulative distribution functions F_j, its inverse F_j^{-1}, and with s_j being some scaling factor:

$$\mathcal{H}_{0,\mathrm{IF}} : \; F_j^{-1}(p) = s_j \cdot q(p) \; \forall \; j = 1, \dots, d. \tag{2.89}$$

The p-quantile of site j therefore is split into the site-specific factor s_j, called index flood, and the dimensionless common quantile function q, often called regional growth curve. The index flood is mostly selected as sample mean of annual peaks or as locally calculated location parameter (Sveinsson, Salas, and Boes, 2003) if sample data is available. Calculation of the growth curve is often done using L-moments of normalized observations.

This model represents quite strong assumptions to the distributions F_j. Other models have weaker assumptions, e.g. only an identical tail behaviour is assumed that is measured using the extreme value index. See Kinsvater, Fried, and Lilienthal (2016) for regional estimation and homogeneity testing in such a model.

Model (2.89) is important for all steps of the process because it defines what sites are regarded as "homogeneous". The whole process of regionalization can be separated in three general steps:

1. Building of groups of similar stations,
2. Calculating the regional growth curve,
3. Deriving local quantiles.

Now the three steps are described in more detail.

Building homogeneous groups

The first step is very important because the benefit of regional flood frequency analysis is highly dependent on how good the model (2.89) describes the data. If the model is inappropriate for the group, this leads to biased estimations of the regional growth curve and therefore to inaccurate quantile estimates (Lettenmaier, Wallis, and Wood, 1987). The size of the group is some kind of trade off because the homogeneity assumption is often more questionable the more sites are included in the group. On the other hand, the less sites are included the less information can be regionalized.

Basically there are three different kinds of groups: Geographically contiguous regions, fixed partitions of the available set of sites, and groups individually selected for one target site. Contiguous regions have some advantages as well as disadvantages. New sites can easily be included to an analysis based on their geographical location. On the downside, nearby stations often have high intersite dependencies because they are influenced by the same weather, which reduces the amount of information gained by applying a regional analysis.

Fixed partitions do not need to be regions in the literal sense, i.e. the sites do not necessarily need to be geographically close. Statistical methods like cluster analysis (Burn, 1989), classification and regression trees (Walther et al., 2012), canonical correlation analysis (Ouarda et al., 2001), or others are used to delineate groups out of the set of available sites. The data basis of such procedures are mostly not the flood records itself but site characteristics like catchment size, mean height, or river length in the catchment. The reason for this is that it enables us to assign new sites with no prior measurements to groups by only collecting those characteristics.

Individual groups (called regions of influence, Burn, 1990) are built for each target site individually. They can form contiguous regions around the target site or can be built by using statistical methods to find similar sites.

After a group is built, its homogeneity is assessed using discordance measures and tests. Depending on the results, adjustments can be made to improve the homogeneity. Because of its importance, homogeneity testing takes an important role in this dissertation. Chapter 3 analyses drawbacks of the very popular Hosking-Wallis homogeneity measure (Hosking and Wallis, 1997) and generalizes it to be applicable in more general situations. In Chapter 4 a new homogeneity test based on a limit theorem of PWM estimators is constructed.

Calculation of the regional growth curve

After a set of stations has been selected for which the regionalization seems to be beneficial, the dimensionless regional growth curve is calculated. First, a parametric family for this curve has to be chosen. A reasonable choice based on asymptotic extreme value theory are the GEV and GPD distributions depending on whether the block maxima or POT approach is used (see 2.2.2 and 2.2.4). Since they only appear as limiting distributions, in hydrology often other families like the log normal distribution (Stedinger, 1980), generalized logistic distribution, Pearson type 3 (Bobée, 1973), and log Pearson type 3 distribution (Bobée, 1975) are used.

To select a distribution, a popular graphical approach is the use of L-moment ratio diagrams (Vogel and Fennessey, 1993) that display the empirical L-skewness, empirical L-kurtosis, and the corresponding theoretical ratios of different distribution families. Figure 2.10 gives an example using 25 stations consisting of $n = 100$ samples generated by a GEV(10, 5, 0.2). The observations (filled points) spread around the

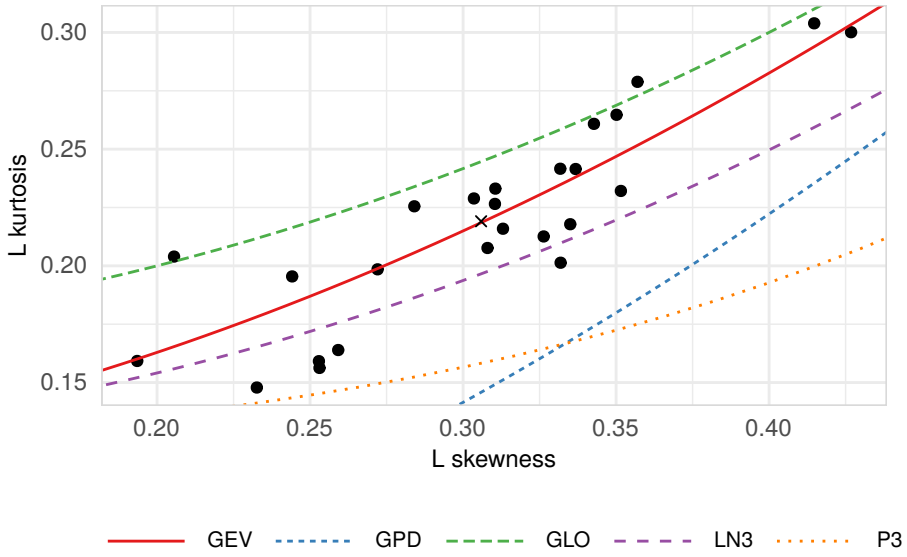

Figure 2.10: Example of a L-moment ratio diagram with random data from a GEV(10, 5, 0.2) with record lengths $n = 100$.

red GEV curve, which indicates all combinations of L-skewness and L-kurtosis in the GEV family. This example illustrates how difficult it is to choose an appropriate distribution family from empirical data because a graphical representation often does not give clear decisions. As an alternative to graphical methods, Hosking and Wallis (1997) gave a goodness-of-fit measure based on the L-kurtosis.

When a parametric distribution family is selected, the growth curve is calculated from the pooled information. To do this, remember that the Index Flood model (2.89) assumes that the flood peak distribution of the sites of interest is equal up to a site-specific factor. If $X_1, \ldots, X_d$ denote random variables following the flood peak distribution of d sites and with $s_1, \ldots, s_d$ denoting the site-specific factors, we get that $X_1/s_1, \ldots, X_d/s_d$ are following the same distribution (according to the model). In the approach of Hosking and Wallis (1997) regionalized L-moment ratios are used to calculate the parameters of the growth curve. They are calculated as weighted means of local L-moment ratios:

$$\tau_{r;R} = \frac{\sum_{j=1}^{d} n_j \tau_{r;j}}{\sum_{j=1}^{d} n_j}, \quad r \in \{2, 3, \ldots\}, \tag{2.90}$$

with $\tau_{r;j}$ being the r-th L-moment ratio of the scaled sample of site j, $x_{1,j}/s_j, \ldots, x_{n_j,j}/s_j$, and n_j being the record length of that site. With those regionalized L-moment ratios a distribution is then fit as described in Section 2.4.2 and the quantile function of that distribution is used as a growth curve estimate.

Deriving local quantiles

The last step in regional flood frequency analysis is to calculate a quantile estimate at a given site. Two quantities are needed: the site-specific index flood and the regional growth curve.

If observations are available at the given site, the index flood is calculated based on them (using the same statistic as in the previous data pooling scheme). If no observations are available, one has to estimate the index flood by using regional information and/or site characteristics and applying regression or interpolation techniques.

The other part, the growth curve, comes from the group's regional information and was calculated in the preceding step. If multiple groups were built and the site of interest is not already assigned to one group, this has to be done. This step is dependent on the statistical method that was used to build the groups.

The p-quantile of the site of interest can then be calculated as

$$\hat{F}_j^{-1}(p) = \hat{s}_j \times \hat{q}(p), \tag{2.91}$$

with $\hat{s}_j$ denoting the index flood and $\hat{q}$ the regional growth curve.

2.6 Software

All implementations, experiments, simulations studies, tables, and most of the figures in this dissertation have been calculated or made using the statistical software environment `R` (R Core Team, 2019). The most important packages used are `evd` (Stephenson, 2002) for providing the extreme value distribution, `copula` (Hofert et al., 2018) for modelling copulas, and `ggplot2` (Wickham, 2016) for graphical plotting. In Chapter 3 a new homogeneity test is compared to existing procedures that are taken from the package `homtest` (Viglione, 2012).

An important concept in this dissertation are TL-moments. Because existing implementations were either slow or limited in options (like trimming parameters), the package `TLMoments` (Lilienthal, 2019) has been built and made available on `CRAN`. The main focus of this package is the implementation of fast algorithms to calculate TL-moments and parameter estimations which makes it feasible to conduct comprehensive simulations studies using TL-moments. This has been done by taking advantage of the recurrence properties described in equations (2.82) and (2.83) (see also Hosking and Balakrishnan, 2015) and by implementing core parts in `C++`. Covariance estimators of the sample TL-moments and resulting parameter and quantile estimators described in Chapter 4 were also implemented in this package.

3 Homogeneity test for skewed and cross-correlated data

In Section 2.5 the procedure of regional flood frequency analysis was described with special attention to the Index Flood procedure. In all regionalization techniques information resulting or transferred from a set of stations is used for quantile estimation at the site of interest. This can only be of benefit if the sites share some common flood frequency characteristic. In the Index Flood model the flood distribution of scaled variables is assumed to be the same for all stations. Test procedures are used to assess if this assumption is appropriate or if modifications to the group are to be recommended. The following chapter deals with this problem in the case of skewed and cross-correlated data which is typical for flood observations in Germany.

It is based on the work "Homogeneity testing for skewed and cross-correlated data in regional flood frequency analysis" published in the Journal of Hydrology (Lilienthal, Fried, and Schumann, 2018).

3.1 Introduction

As described previously, homogeneity of a set of stations is crucial for the benefit of regionalization. A group of stations is called homogeneous, if the Index Flood model (2.89) is an appropriate description of their statistical behaviour. Dalrymple (1960), who also introduced the model itself, proposed a homogeneity test, but due to its lack of power other procedures have been suggested (Fill and Stedinger, 1995; Lu and Stedinger, 1992; Wiltshire, 1986). Hosking and Wallis (1993) introduced the nowadays most common Hosking-Wallis (HW) procedure based on the variability of L-moment ratios (Hosking, 1990). The former univariate procedure was later on extended to multivariate situations (Chebana and Ouarda, 2007; Chebana and Ouarda, 2009).

Although being widely used, the HW procedure is not free of drawbacks. Viglione, Laio, and Claps (2007) compared it to two rank based tests and concluded that the L-moment based HW procedure is better for little skewed distributions while rank based tests are better for higher skewness. Another problem lies in cross-correlation between the stations. Hosking and Wallis (1988) examined the effect of intersite

dependence to regional flood frequency analysis in general and Castellarin, Burn, and Brath (2008) to the HW procedure in particular. Their result is that cross-correlation reduces the power of the test, which means that heterogeneous cross-correlated groups misleadingly tend to look homogeneous. Their proposed solution is an empirical corrector that is added to the test statistic. In a recent study Masselot, Chebana, and Ouarda (2017) argued that the original rejection threshold is not well justified and enhanced the test by replacing parametric assumptions with a nonparametric procedure.

This study takes up the drawbacks regarding skewed distributions and intersite dependence. Trimmed L-moments (Elamir and Seheult, 2003) are used to replace regular L-moments and the intersite dependence is modelled with copulas.

In Section 3.2 the original Hosking-Wallis procedure is explained. The new approach is motivated and the differences to the original Hosking-Wallis procedure are shown in Section 3.3. In Section 3.4 simulation studies are performed to compare the new method to existing approaches and to assess the quality of the modifications. Results and comparisons to other approaches are discussed in Section 3.5. A case study follows in Section 3.6, in which the need for modifications of the original procedure is revealed. Thereafter the main results are summarized.

3.2 Hosking-Wallis heterogeneity measure

The heterogeneity measure of Hosking and Wallis (1993) is based on the comparison of the observed sample variability of L-moment ratios (Hosking, 1990) and the expected variability under the assumption of homogeneity. It can be divided into three parts: the calculation of a statistic, the calculation of coefficients to normalize this statistic, and finally the decision about homogeneity.

First, L-moment ratios have to be calculated at each site as well as regionalized averages of them. Let there be d samples of length $n_1, \ldots, n_d$ and let $\hat{\tau}_{2;j}$, $\hat{\tau}_{3;j}$, and $\hat{\tau}_{4;j}$ be the empirical L coefficient of variation (L-CV), L-skewness, and L-kurtosis of sample $j = 1, \ldots, d$, respectively. The regional averaged L-moment ratios are defined by

$$\hat{\tau}_{2;R} = \frac{\sum_{j=1}^{d} n_j \hat{\tau}_{2;j}}{\sum_{j=1}^{d} n_j}, \quad \hat{\tau}_{3;R} = \frac{\sum_{j=1}^{d} n_j \hat{\tau}_{3;j}}{\sum_{j=1}^{d} n_j}, \quad \hat{\tau}_{4;R} = \frac{\sum_{j=1}^{d} n_j \hat{\tau}_{4;j}}{\sum_{j=1}^{d} n_j}. \tag{3.1}$$

Hosking and Wallis (1993) defined three different measures of dispersion using

L-moment ratios. The two most often applied ones are

$$V_1 = \sum_{j=1}^{d} n_j (\hat{\tau}_{2;j} - \hat{\tau}_{2;R})^2 / \sum_{j=1}^{d} n_j, \tag{3.2}$$

$$V_2 = \sum_{j=1}^{d} n_j ((\hat{\tau}_{2;j} - \hat{\tau}_{2;R})^2 + (\hat{\tau}_{3;j} - \hat{\tau}_{3;R})^2)^{1/2} / \sum_{j=1}^{d} n_j. \tag{3.3}$$

These statistics have to be normalized afterwards. This means that

$$H_i = \frac{V_i - \mu_i}{\sigma_i}, \quad i = 1, 2, \tag{3.4}$$

is calculated with appropriate values of μ_i and σ_i. To select these, a parametric bootstrap is performed. Using the regional averages of L-CV, L-skewness, and L-kurtosis a kappa distribution (Hosking, 1994) is fitted. A large number of datasets is then drawn from this kappa distribution, each of them consisting of d samples with corresponding sample lengths $n_1, \ldots, n_d$. For each bootstrap dataset the value of V_i is calculated and the mean and standard deviation over all of these values are inserted for μ_i and σ_i, respectively.

According to Hosking and Wallis (1993) the set of samples is called "acceptably homogeneous" if $H_i < 1$, "possibly heterogeneous" if $1 \leq H_i < 2$ and "definitely heterogeneous" if $H_i \geq 2$. Note that Hosking and Wallis (1993) did not formulate this as a test, but rather as a recommendation. Therefore no specific significance level is controlled.

3.3 Construction of a generalized Hosking-Wallis procedure

In this section we motivate our new approach and describe the differences to the original Hosking-Wallis procedure. For this we first review the papers of Castellarin, Burn, and Brath (2008) and Viglione, Laio, and Claps (2007), which dealt with two different drawbacks of the original procedure, and we explain the differences to our approach. The modified procedure is summarised afterwards.

3.3.1 Consideration of intersite dependence

The original Hosking-Wallis homogeneity measure assumes intersite independence, meaning that the observations of each station are independent of the other stations' observations for the same year. In practice this is a strict assumption, which is often not fulfilled. Stations in the same river network feature a natural dependence

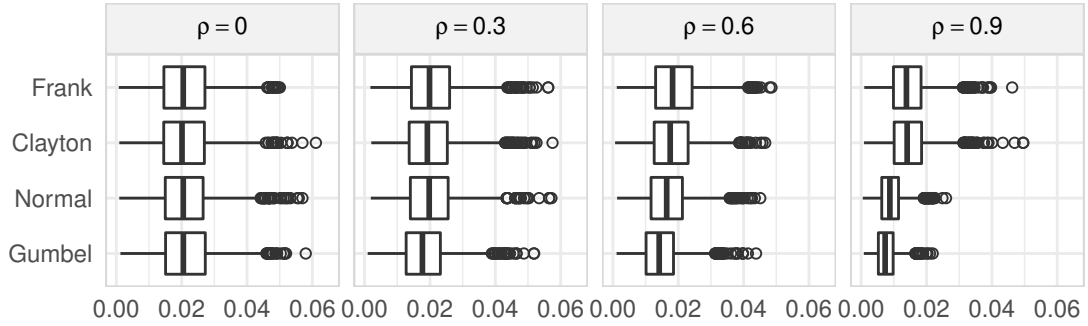

Figure 3.1: Sampling distribution of V_1 when normal, Gumbel, Frank, or Clayton copulas generate the data, grouped by increasing rank correlation.

because floods at upstream stations affect floods downstream. Additionally and more generally, all nearby stations are simultaneously influenced by events like snow melt or synoptic rainfalls.

Castellarin, Burn, and Brath (2008) investigated this problem. Their result is that cross-correlation reduces the power of the test. This means that heterogeneity of a group of stations is detected less often and therefore heterogeneous cross-correlated groups misleadingly tend to appear as homogeneous. Their proposed solution is an empirical corrector. They first calculate the test statistic H_1 (see formula (3.4)) under the assumption of intersite independence. After this, they calculate an adjusted test statistic

$$H_{1,\text{adj}} = H_1 + C \times \bar{\rho^2}(d-1), \tag{3.5}$$

with the average squared correlation $\bar{\rho^2}$, the number of stations d, and a constant correction coefficient C, which has to be derived by simulations. Castellarin, Burn, and Brath (2008) computed it as $C = 0.122$, but they noted that "the coefficient C is inevitably associated with the Monte Carlo simulation experiments performed in this study".

A drawback of this approach is that the dependence structure is only taken into account through the correlation ρ. Figure 3.1 shows the sampling distribution of the dispersion of L-skewness (V_1 given in formula (3.2)) when generating data with four common copulas and given rank correlations $\rho = 0, 0.3, 0.6, 0.9$. One can observe that the distribution changes more with increasing correlation and thus would require different corrections. Hence, following the above approach, we would need to calculate a specific correlation coefficient C for each dependence structure.

Our solution to the problem of cross-correlated data differs from the above approach. Instead of ignoring the assumption of independence and correcting for this afterwards, we dispose the assumption by allowing cross-correlated data. Therefore the procedure for generating bootstrap data to calculate the test statistic needs to be modified.

We use a d-dimensional copula to describe the intersite dependence of a group of d stations. This copula is taken to generate datasets which feature the same dependence structure like the observed data. This way the cross-correlation between the stations is directly included in the procedure.

3.3.2 Robustification against high skewness

Another weakness of the original procedure arises when analysing skewed data. Viglione, Laio, and Claps (2007) compared the Hosking-Wallis procedure to two rank-based test statistics, a generalization of the Anderson-Darling goodness of fit test and the Durbin and Knott test (Durbin and Knott, 1972). Viglione, Laio, and Claps (2007) concluded that L-moment based tests are better for little skewed distributions while rank-based tests are better for higher skewness. Their final recommendation is to use the Hosking-Wallis procedure if the regionalized L-skewness is lower than 0.23 and the Anderson-Darling test otherwise.

Our approach to face this issue is to improve the Hosking-Wallis procedure by using trimmed L-moments (see Section 2.4.3) instead of regular L-moments. This is done by substituting the L-moment ratios $\hat{\tau}_{2;j}, \hat{\tau}_{2;R}, \hat{\tau}_{3;j}, \hat{\tau}_{3;R}$ in formula (3.2) and (3.3) by corresponding TL-moment ratios $\hat{\tau}_{2;j}^{(s,t)}, \hat{\tau}_{2;R}^{(s,t)}, \hat{\tau}_{3;j}^{(s,t)}, \hat{\tau}_{3;R}^{(s,t)}$ with trimming parameters $s, t \in \mathbb{N}_0$, which need to be specified. A simulation study in Section 3.4 will analyse the influence of these parameters to the test performance.

3.3.3 Generalized Hosking-Wallis procedure

We now briefly describe our proposed generalized Hosking-Wallis procedure. Afterwards we provide further notes to the different steps.

To calculate the generalized Hosking-Wallis procedure the following steps are performed:

1. Analyse the data to identify suitable trimming parameters (s, t) and copula model C.
2. Calculate a TL-moment based statistic:

$$V_1 = \sum_{j=1}^{d} n_j (\hat{\tau}_{2;j}^{(s,t)} - \hat{\tau}_{2;R}^{(s,t)})^2 / \sum_{j=1}^{d} n_j, \tag{3.6}$$

$$V_2 = \frac{\sum_{j=1}^{d} n_j \sqrt{(\hat{\tau}_{2;j}^{(s,t)} - \hat{\tau}_{2;R}^{(s,t)})^2 + (\hat{\tau}_{3;j}^{(s,t)} - \hat{\tau}_{3;R}^{(s,t)})^2}}{\sum_{j=1}^{d} n_j}. \tag{3.7}$$

3. Fit the copula $\hat{C}$ and the marginal kappa distribution $\hat{K}$ to the data.

4. Generate N_{Sim} datasets of the same structure as the real data using the copula approach with copula $\hat{C}$ and equal kappa margins $\hat{K}$ and calculate (3.6) and/or (3.7) in each of them (called $V_{1,1}, \ldots, V_{1,N_{Sim}}$ and/or $V_{2,1}, \ldots, V_{2,N_{Sim}}$, respectively).

5. Use either a) or b) to derive a decision:

   a) Calculate $\mu_i = 1/N_{Sim} \sum_{j=1}^{N_{Sim}} V_{i,j}$ and $\sigma_i = 1/(N_{Sim}+1) \sum_{j=1}^{N_{Sim}} (V_{i,j} - \mu_i)^2$ to get
   $$H_i = \frac{V_i - \mu_i}{\sigma_i}, \quad i = 1, 2, \tag{3.8}$$
   and reject the homogeneity according to a critical value or use the original suggestion-type decision rule: Classify the group as "acceptably homogeneous" if $H_i < 1$, "possibly heterogeneous" if $1 \leq H_i < 2$, and "definitely heterogeneous" if $H_i \geq 2$.

   b) Use the position of the observed V_i inside the ordered bootstrap results to determine the unlikeliness of the observed data. Reject the null hypothesis at the $\alpha\%$ significance level if $V_i > \tilde{c}_\alpha$ with $\tilde{c}_\alpha$ being the empirical $(1-\alpha)$-quantile of $V_{i,1}, \ldots, V_{i,N_{Sim}}$.

A note should be made regarding the last step that contains the decision making. We decided to mention two options, the classic suggestion-type decision of the original procedure and a more statistical way that determines a critical value based on the bootstrap results, similar to the suggestion of Masselot, Chebana, and Ouarda (2017). In Section 3.4.2 we will compare these options and explain our choices in the simulation study.

The original procedure can be obtained by choosing L-moments (setting $s = t = 0$) and the independence copula $C(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i$ (see Section 2.3.7). Hence, this new procedure truly generalizes the original one.

The choice of trimming parameters and copula model needs further attention and will be analysed in the following simulation studies. The appropriate trimming seems to depend on the skewness of the analysed data. In situations with positively skewed data upper-trimmed TL-moments turn out to be preferable. In situations with negative or no skewness other choices might be better. A suitable copula model could be chosen with the help of Goodness-of-fit procedures (e.g. Genest, Rémillard, and Beaudoin, 2009). Small misspecifications are negligible as the simulation study will show.

A difficulty that arises when analysing real data is that the stations usually feature different record lengths and missing data. In Step 3 this has to be considered in the fitting procedure. Simple Gumbel or Clayton copulas can be fitted by calculating a weighted mean of the pairwise Spearman's ρ and using the inversion of ρ method (as described in Section 2.3.8). The given procedure for fitting the mixed Gumbel model uses only pairwise calculations of Pickands dependence function anyway, so

this is easily adopted by only considering the common records of each two stations. How data sets of unequal record lengths are simulated using a copula and marginal distributions is described in Section 2.3.9.

3.4 Simulation studies

In this section simulation studies are carried out to investigate the advantages and drawbacks of our proposed new procedure. Comparisons to other approaches will be given in Section 3.5.

3.4.1 Data generation

In the subsequent studies a scenario with 20 stations, each with a maximum record length of 60 years, is considered. In practice observation periods typically start at varying points in time. To reflect this, the beginning of each station's simulated observations is randomised for each data set. Four stations start in the "first" year, six 10 years later, another six 20 years later, and the last four stations start 30 years after the first observation. The group's mean sample length is $\bar{n} = 45$. An example how the data scheme looks like is illustrated in the left panel of Figure 3.2.

The simulated network of stations can contain intersite dependence and can either be homogeneous or heterogeneous. The joint distribution of the sites is constructed using the copula approach, which means that the marginal distributions and the copula are specified separately.

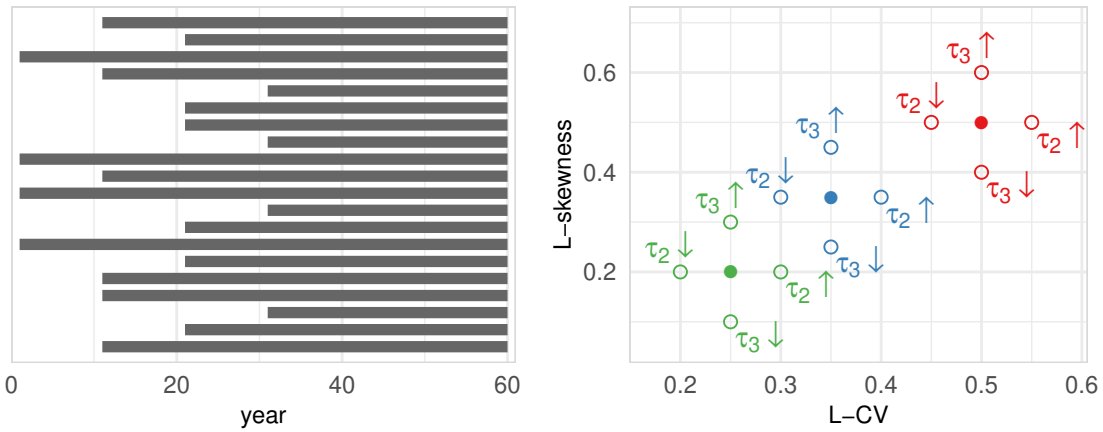

Figure 3.2: Left: Schematic example of simulated observation lengths. Right: Distributions employed in the simulation studies characterized by their L-CV and L-skewness.

Table 3.1: Overview of the configuration of the six studies: Data generation consists of different marginal distributions and copulas. Operated test procedure depends on trimming parameters and copula model.

	Data generation		Test configuration	
No.	Marginals	Copulas	TL	Copulas
1	little skewed medium skewed highly skewed	Gumbel(1.5)	$(0,0),(0,1),(0,2)$ $(0,3),(0,4),(1,1)$ $(1,2),(2,2)$	Gumbel
2	little skewed medium skewed highly skewed	Independence MixedGumbel*	$(0,1)$	Independence Gumbel MixedGumbel
3	medium skewed	Clayton(2)	$(0,0),(0,1),(0,2)$	Gumbel Clayton
4	medium skewed**	Gumbel(1.5)	$(0,0),(0,1),(0,2)$	Gumbel
5	medium skewed***	Gumbel(1.5)	$(0,0),(0,1),(0,2)$	Gumbel
6	medium skewed****	Gumbel(1.5)	$(0,0),(0,1),(0,2)$	Gumbel

* as in formula (2.31): $C_1 = \text{Gumbel}(3), C_2 = \text{Gumbel}(1.25), a = (.2, .4, .6, .8, .2, .4, \ldots, .8)$

** discordant sites vary on a free grid, *** artifical outliers added,

**** different mean record lengths and number of stations

The generalized extreme value distribution family is selected for every marginal distribution. 16 stations always have the same GEV parameters, while the parameters of four other stations are allowed to vary. Similarly to Viglione, Laio, and Claps (2007) we perform these modifications in the space of L-moment ratios, which means that we either modify the L-CV τ_2 or the L-skewness τ_3 of the discordant stations. These modifications are indicated as $\tau_2 \downarrow$, $\tau_2 \uparrow$, $\tau_3 \downarrow$, and $\tau_3 \uparrow$. The right panel of Figure 3.2 shows the $\tau_2 - \tau_3$-combinations chosen. We consider three parameter combinations as a base point (filled circles; parameters of station 1–16) and vary them at the τ_2- and τ_3-scale (unfilled circles; possibly modified parameters of stations 17–20 in the main study). These three combinations represent common distributions we found in real data and feature high, medium, or little skewness. To ease readability, we will use the terms CV and skewness without their L-prefix from now on.

To model the intersite dependence of these stations the independent, Gumbel, and mixed Gumbel copula are used. The latter is constructed using two Gumbel copulas and the result of formula (2.31). We chose these copulas because it allows us to examine copulas of different complexity within one framework (the independence copula is a special case of the Gumbel copula, which in turn is a special case of the mixed Gumbel copula). Additionally, all these models are extreme value copulas and therefore a logical choice to model the dependence of extremes. The Clayton copula

is neither an extreme value copula nor does it capture upper tail dependence. It is included to simulate a severe case of copula misspecification.

Each of the settings is replicated $B = 5000$ times. On the left hand side of Table 3.1 the different configurations of our studies are summarised.

3.4.2 Test application

Our proposed procedure is calculated with different configurations, i.e. different selections of trimming parameters or copula models. The right hand side of Table 3.1 summarises these options for the different studies. We use mainly TL-moments with upper-trimming. The reason for this lies in the fact that we deal with distributions with heavy upper tails. The influence of extraordinary high observations should therefore be restricted.

Both test statistics, H_1 and H_2, are calculated. Because the results of H_1 are superior to those of H_2 in almost every respect (Viglione, Laio, and Claps, 2007, experienced similar results), the test results based on H_2 are neglected in favour of a simpler presentation.

In Section 3.3.3 we provided two options to determine the test's outcome. The method of Hosking and Wallis is formally not a test and is meant more as a recommendation. Because of this there is no critical value whose exceedance defines the rejection of the null hypothesis. Some authors (Fill and Stedinger, 1995; Hosking and Wallis, 1997) suggest using $H > 1.64$ as a rejection criterion assuming V to be normally distributed. In fact, the original V statistic using L-moments is not normally distributed and the critical value c_α for which the rejection criterion $H > c_\alpha$ leads to a significance level of $(\alpha \times 100)\%$ depends strongly on the sizes and record lengths of the groups and especially on the skewness of the marginal distributions as we will see in the following.

Table 3.2 illustrates this by giving the empirical critical values $\hat{c}_{5\%}$ for using the V_1 statistic with L-moments based on a simulation of homogeneous regions consisting of 20 sites with different skewnesses and record lengths (5000 replications each). Other factors like the intersite dependence are not considered here, but have impact, too. In the second and third line the estimated rejection rates indicate that both fixed critical values fail to ensure a constant significance level. The rejection criterion $V_i > \tilde{c}_\alpha$ with $\tilde{c}_\alpha$ being set as the empirical $(1-\alpha)$-quantile of the simulated V (given in the fourth line) works well for little skewed data, but also fails when the data features higher skewness. The lower part of Table 3.2 gives the same information but using the TL(0,1)-moments to calculate V_1. It is visible that the empirical critical values at the $\alpha = 5\%$ level are more stable over the different situations. The decision rule $H_1 > 1.64$ leads to more stable rejection rates about 6-8%. The assumption of

Table 3.2: Empirical critical values that would lead to a 5% significance level ($\hat{c}_{5\%}$) and the empirical rejection rates for different decision rules (with $\tilde{c}_{5\%}$ being the 95%-quantile of the simulated V). Results are based on 5000 simulated homogeneous regions consisting of 20 sites. The values show the strong connection between the critical values and rejection rates and the data's skewness and record lengths when using L-moments.

		little skewed			medium skewed			highly skewed		
	$n =$	30	50	100	30	50	100	30	50	100
L	$\hat{c}_{5\%}$	1.84	1.76	1.75	2.29	2.07	2.00	2.77	2.59	2.31
	$\hat{P}(H_1 > 2)$	0.04	0.03	0.03	0.08	0.06	0.05	0.14	0.11	0.07
	$\hat{P}(H_1 > 1.64)$	0.07	0.06	0.06	0.13	0.10	0.09	0.22	0.17	0.12
	$\hat{P}(V_1 > \tilde{c}_{5\%})$	0.07	0.06	0.05	0.11	0.09	0.08	0.19	0.14	0.09
TL	$\hat{c}_{5\%}$	1.80	1.73	1.79	1.82	1.80	1.70	1.93	1.88	1.84
	$\hat{P}(H_1 > 2)$	0.04	0.03	0.03	0.04	0.04	0.03	0.04	0.04	0.03
	$\hat{P}(H_1 > 1.64)$	0.07	0.06	0.06	0.07	0.06	0.06	0.08	0.07	0.07
	$\hat{P}(V_1 > \tilde{c}_{5\%})$	0.06	0.05	0.06	0.06	0.06	0.05	0.08	0.07	0.07

normality seems to work better when using TL-moments in the V-statistic. There are no substantial differences between using the criterion $H_1 > 1.64$ and $V_1 > \tilde{c}_{5\%}$.

We choose $H_1 > 1.64$ as rejection criterion in this study. We do not see substantial improvements with using the empirical simulated quantile as a critical value since all methods fail to hold the significance level for skewed data. When using TL-moments, the resulting differences between these options are small anyway.

3.4.3 Test assessment

To assess the test's performance we need to distinguish between two cases. If the analysed group of stations is truly homogeneous we want the rate of rejections to be low. The proportion of rejections of the null hypothesis estimate the probability of committing a type-I-error and is subsequently called size of the test. If the group is truly heterogeneous a high rate of rejections is desirable and the proportion of rejections describes the power of the test in this case. Given a simulated group of stations, denoted by $\mathcal{G}$, and a simulation of B replications, we therefore calculate both measures as

$$\frac{1}{B}\sum_{b=1}^{B} \mathbb{1}(H_{1,b} \geq 1.64) = \begin{cases} \text{size}, & \text{if } \mathcal{G} \text{ is homogeneous,} \\ \text{power}, & \text{if } \mathcal{G} \text{ is heterogeneous.} \end{cases} \tag{3.9}$$

The indicator function $\mathbb{1}(\cdot)$ takes the value 1 if $H_{1,b} \geq 1.64$, and is 0 otherwise, with $H_{1,b}$ indicating the observed test statistic in replication b.

Because different tests can have different sizes a direct comparison regarding the power is often difficult or even not feasible. For a fair comparison we adjust the empirical power by choosing the critical value c_α as the value that leads to a size of $(\alpha \times 100)\%$ in the simulations (see also Zhang and Boos, 1994):

$$\begin{aligned} \text{size adjusted power} &= \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}(H_{1,b} \geq c_\alpha) \quad (\mathcal{G} \text{ is heterogeneous}), \\ \text{with } \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}(H_{1,b} \geq c_\alpha) &= \alpha \quad (\mathcal{G} \text{ is homogeneous}). \end{aligned} \tag{3.10}$$

In this study we present the size adjusted power for $\alpha = 5\%$.

3.4.4 Influence of the trimming

This simulation study assesses which trimming parameters are appropriate in the presence of intersite dependence as it is described by a Gumbel copula with parameter $\vartheta = 1.5$. The generalized Hosking-Wallis procedure using a Gumbel copula is calculated for several TL-moments. The results for L-, TL(0,1)-, TL(0,2)-, TL(0,3)-, and TL(0,4)- moments are given below. Other trimmings (see Table 3.1) lead to inferior results and are not reported here.

The results are depicted in Figure 3.3. The left panel contains the size and the right one displays the size adjusted power of the test for the different modifications. Both graphics are grouped row-wise by the three degrees of skewness. The different trimming parameters are indicated by different colours.

The size using L-moments (TL(0,0)) exceeds 5% noticeably, especially in the medium and highly skewed setting. For TL(0,1)-moments the size is generally the lowest, but still above the 5% level. Higher trimmings lead to increased sizes.

Looking at the size adjusted power of the tests, the first finding is that there are substantial differences between the different modifications. In the little skewed setting changes of CV are more likely to be detected than changes of skewness, while in the highly skewed setting the detection rates are more similar or even the other way round. In case of changes of CV the (0,1)-trimming generally is the most preferable one, other trimmings mostly lead to inferior rates or only minor improvements (in case of highly skewed data). For changes of skewness the size adjusted power improves for higher trimmings up to the trimming of (0,2) or (0,3) and then declines.

As a result we suggest the use of TL(0,1)-moments because their size adjusted power is always an improvement over L-moments while they also lead to an substantial

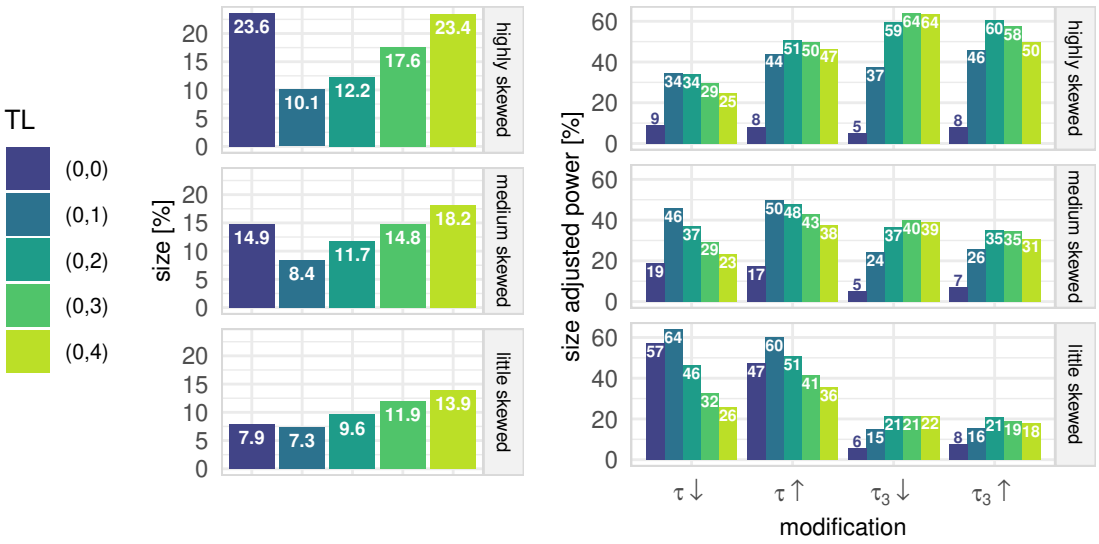

Figure 3.3: Size and size adjusted power of the generalized Hosking-Wallis procedure with different TL-moments and Gumbel copula. Synthetic data is generated with Gumbel(1.5) dependence structure.

decrease in size when standard decision rules are applied (i.e. $H > 2$ or $H > 1.64$). Higher trimmings can result in a more powerful test for some problems but more considerations would be necessary on how to control the significance level.

In the remaining studies we will choose TL(0,1)-moments to calculate the modified Hosking-Wallis statistics, unless stated otherwise.

3.4.5 Influence of the copula model choice

This section deals with the influence of the chosen copula model. The independence copula or the mixed Gumbel model were used to generate data sets. The generalized Hosking-Wallis procedure is applied to each data set using either the independence copula, the Gumbel copula, or the mixed Gumbel copula. Parameters of the latter two copula models have to be estimated. As stated before, TL(0,1)-moments are chosen to calculate H_1.

First we have a look at the results when the data is generated without intersite dependence, depicted in Figure 3.4. Fitting copulas to the independently generated data lead to increases in size. These are minor if simple Gumbel copulas are used and higher for the more complex mixed copula model. The size adjusted power rates are practically the same for all tests. The reason for this is that the fitted copula model is only used in the bootstrap procedure, which only affects the coefficients that standardize the V-statistic. This only changes the size of the test but not the size adjusted power rate.

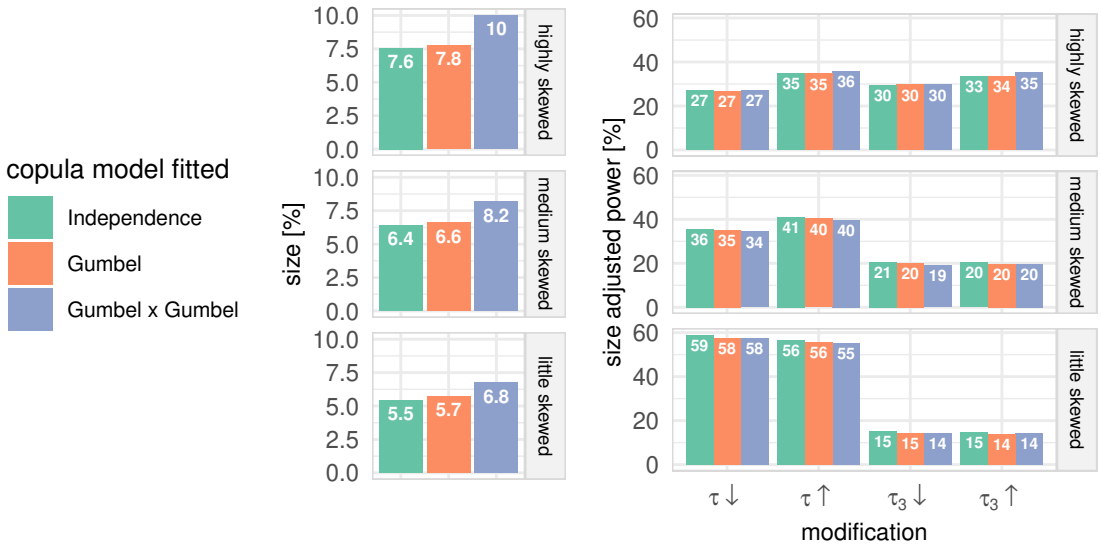

Figure 3.4: Size and size adjusted power of the generalized Hosking-Wallis procedure with different copula models and TL(0,1)-moments. Synthetic data is generated without dependence structure.

This demonstrates that there is generally little harm in the assumption of a simple dependence structure when in fact there is no intersite dependence present. The size is not drastically increased and the size adjusted power is not affected. Of course, one reason for this is that both the Gumbel copula and the mixed Gumbel copula include the independence copula as a special case. In fact, the mean fitted Gumbel parameter over all replications of the displayed settings here is $\bar{\theta} = 1.05$ (corresponding to a correlation of $\bar{\rho} = 0.05$) and therefore indicates almost the independence copula.

Now we have a look at the opposite case, meaning that the real dependence structure is more complex than the fitted one (Figure 3.5). The size using the independence copula is very low. The unconsidered dependence structure leads to a wrongly adjusted bootstrap, which corrupts the test. The differences between the Gumbel and the mixed Gumbel model are smaller here. The size for the simpler model is generally lower than that for the true model. The estimation of a simple copula seems more beneficial than the estimation of the true model with many parameters. As in the previous study, the size adjusted power rates are similar for all tests.

To investigate the case of complete misspecification of the copula model, data was generated in the medium-skewed situation using a Clayton(2) copula. The Clayton copula describes dependencies with a lower tail dependence and is therefore very different to the Gumbel model, which features upper tail dependence.

Figure 3.6 shows the results of this situation, in which the H_1 statistic was calculated under the assumption of Clayton or Gumbel copulas. To analyse the role of the trimmed L-moments in this setting, L, TL(0,1)-, and TL(0,2)-moments were chosen.

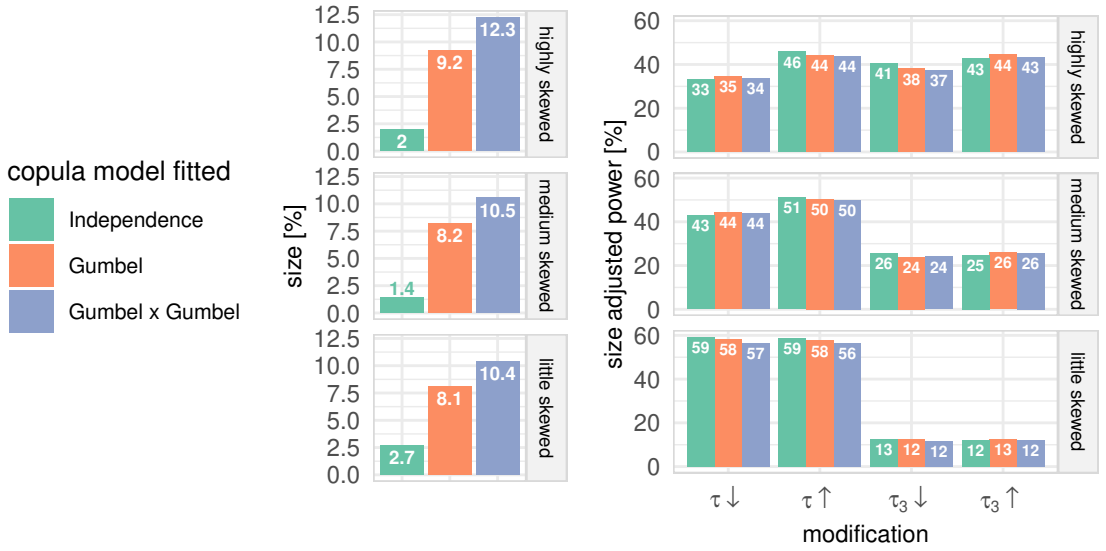

Figure 3.5: Size and size adjusted power of the generalized Hosking-Wallis procedure with different copula models and TL(0,1)-moments. Synthetic data is generated with mixed Gumbel model.

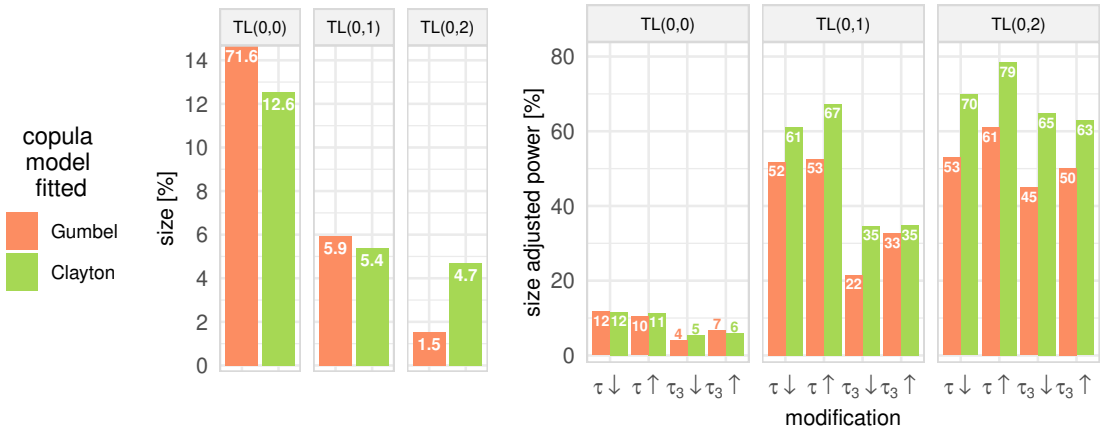

Figure 3.6: Size and size adjusted power of the generalized Hosking-Wallis procedure in the case of copula misspecification. Real copula is Clayton(2); Gumbel and Clayton copulas are fitted and different TL-moments are used.

The Gumbel copula is not to be recommended in the case of L-moments (TL(0,0)) because the resulting size in case of the copula misspecification considered here is by far too large. Fitting a Clayton copula leads to an excessive size, too. The results change with TL(0,1)-moments in the test statistic. The size is below 6% for both copula models. The size adjusted power is clearly better using the Clayton model, but reasonable results are achieved using the Gumbel model. Larger trimming leads to more conservative tests in this situation. The size adjusted power is even better

than for TL(0,1)-moment, but in practice this is irrelevant when the exact critical values are unknown.

These findings indicate that the application of trimmed L-moments leads to some robustness against misspecification of the copula model. Even in this highly constructed case, in which the copulas generating the data are very different from those fitted, the test using trimmed L-moments maintains some power while not exceeding the nominal level too much.

3.4.6 Sensitivity Analysis

We have seen that the power of the test is affected differently by the different types of modification to CV or skewness. It is important to investigate this further to be able to decide if an assumed heterogeneity can be discovered by the procedure.

For this, the medium skewed situation is regarded as base distribution ($\tau_2 = \tau_3 = 0.35$). Now the modified stations are varied to all possible combinations on the grid $(\tau_2, \tau_3) \in \{0.25, 0.26, \ldots, 0.45\} \times \{0.15, 0.17, \ldots, 0.55\}$. A Gumbel(1.5) copula describes the intersite dependence.

Figure 3.7 contains the size adjusted power of the test depending on CV and skewness of the modified station. Differences in the structure of the rejection rates between L- and TL-moments become obvious. With L-moments, the power raises mainly when the CV varies. The test statistic only incorporates the empirical CV, so this behaviour is to be expected. With TL-moments a variation of lower CV and higher skewness or inversely can easily be detected, but when CV and skewness changes in the same direction nearly no detection is possible. The reason for this is that

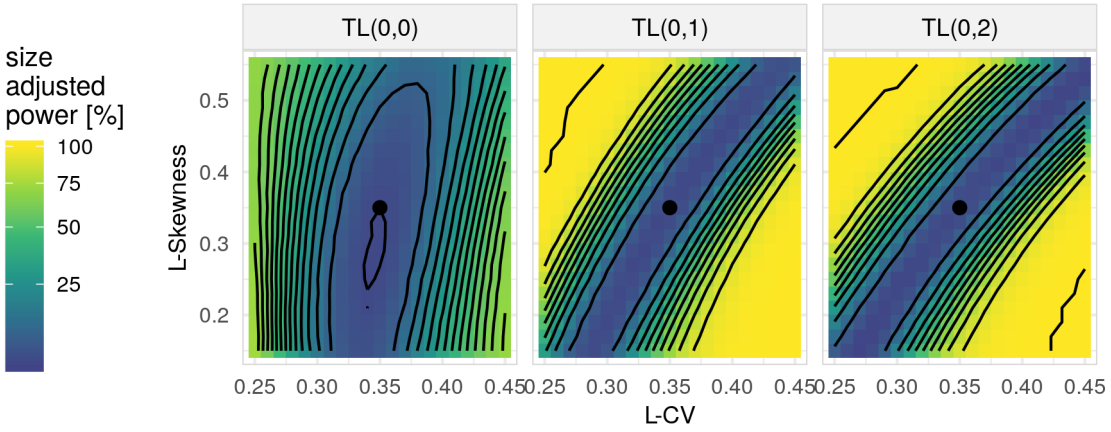

Figure 3.7: Size adjusted power rates of the generalized Hosking-Wallis procedure depending on the position of discordant site grouped by TL-moments.

the modified statistic now incorporates TL-moments and therefore the test detects variations in the TL-moment space.

3.4.7 Robustness against outliers

This analysis concerns the robustness against outliers. An outlier can occur because of a measurement error but also due to a rare event. The first type could be eliminated by simply removing it, but the second one is a valid observation and removing it would change the data. Due to the typically short observation periods, already a few very high measurements can have a big impact on the analysis of homogeneity. It is desirable that homogeneous groups do not appear heterogeneous due to some rare single events. By investigating the test size under the presence of a varying number of outliers we can check for this quality.

A homogeneous group of medium skewness and with a Gumbel(1.5) copula describing their intersite dependence is simulated. After generating data, up to five observations at one single station are changed to the theoretical 99%, 99.9% or 99.99%-quantile of the respective distribution. This reflects the situation in which an extraordinarily large, but still plausible, observation occurs.

The size was calculated applying L-, TL(0,1)-, and TL(0,2)-moments. The results, displayed in Figure 3.8, show that the robustness against extreme observations improves with the degree of the trimming. While the differences in size are quite small for outliers equalling the 99%-quantile, they increase for outliers of the size 99.9% or 99.99% (corresponding to return periods of 1000 or 10 000 years, respectively).

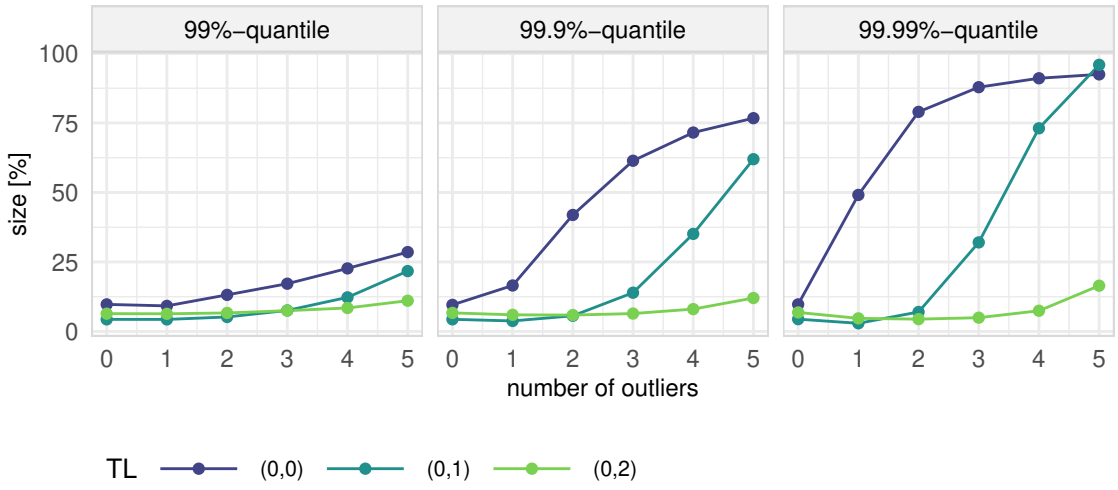

Figure 3.8: Size of the generalized Hosking-Wallis procedure depending on the number of artificial outliers added to one single site, grouped by outlier height.

3.4.8 Influence of record lengths and number of stations

Finally we examine the influence of record lengths and the number of stations on our findings. Figure 3.9 provides size and power for different group sizes and three mean records lengths (the length ratios of the stations are chosen as in the other simulations but adapted to the varied mean record lengths). The test data is generated from the medium skewness scenario with dependence structure Gumbel(1.5) and test statistics are calculated using TL(0,0)-, TL(0,1)-, and TL(0,2)-moments. To illustrate the case of heterogeneous groups (i.e. calculation of power rate) we choose the case of increased skewness at one out of each five sites.

Naturally we observe an increase in size adjusted power with increasing number of stations and mean record length. The size increases with the number of stations when using L-moments, but it is rather constant with trimmed L-moments. Increased mean record length can cause a decrease in size. For all group sizes and record lengths the choice of TL(0,1)-moments seems preferable. A higher trimming does not lead to an improved size adjusted power but to an increase in size, independent of record length and number of stations.

These findings suggest that our results apply also for other group sizes and record lengths than those examined before. It also indicates that the optimal trimming is not affected strongly by these parameters.

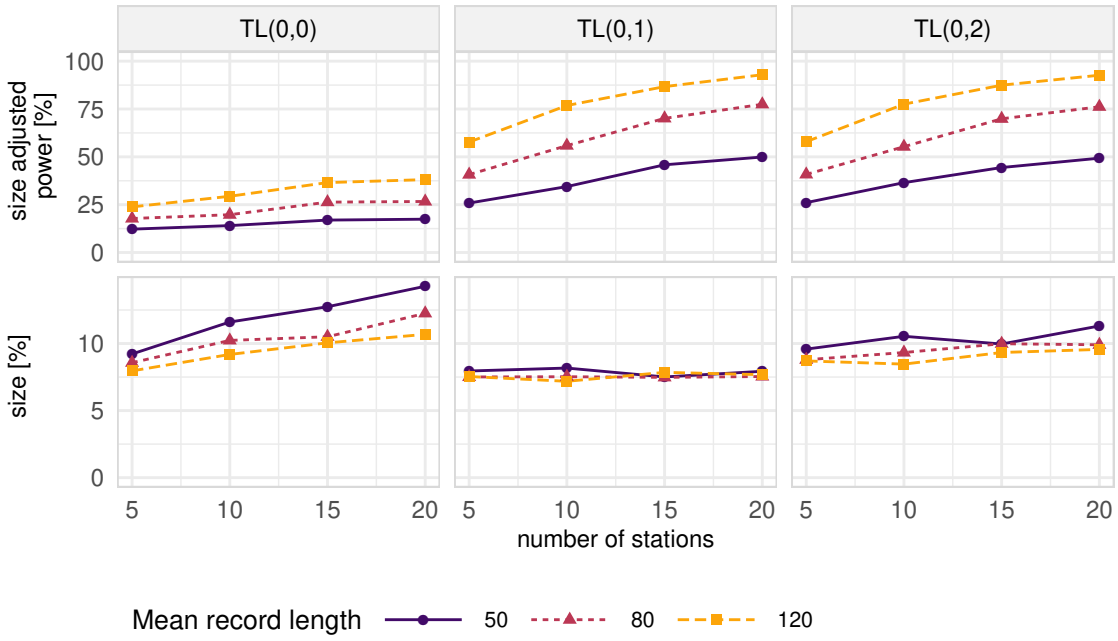

Figure 3.9: Influence of different group sizes and mean record lengths dependent on the degree of trimming.

3.5 Discussion

The simulation studies of the previous section show that our proposed procedure improves the results of heterogeneity tests in many situations. Due to the use of copulas, intersite dependence within the monitored group can be taken into account. The usage of trimmed L-moments can have several positive impacts. It shrinks the size of the test when skewed and cross-correlated regions are examined and simultaneously increases the power. Additionally they lead to a higher robustness against misspecification of the copula model as well as against outliers occurring at single sites. Lastly, our results hold even when varying the number of stations and the mean record lengths.

These benefits come with the need of specifying an appropriate trimming and a copula model. Our simulations suggest that the use of TL(0,1)-moments is preferable to the use of L-moments. Besides improved size and power rates, TL(0,1)-moments lead to an increased robustness against extraordinary high observations. However, it is likely that other trimmings could be more useful in other situations. If the marginal distributions feature a heavy lower tail, for example, lower-trimming appears promising. The specification of the copula model seems challenging because many different models have been proposed in the literature. Our finding, that the chosen copula model does not have to fit perfectly, eases this problem. Using a reasonable choice (like copulas featuring upper tail dependence) will be preferable over neglecting the dependence in any case.

A general problem of the HW procedure is its inability to hold the significance level for skewed data. The use of trimmed L-moments eases this problem but still increased sizes are too be expected.

Masselot, Chebana, and Ouarda (2017) introduce a nonparametric procedure to replace the parametric bootstrap in the HW test. Their suggestion is to use a permutation-based procedure that results in an increased power and decreased computational costs. However, as the bootstrap samples are drawn from the pooled set of all stations, it does not account for intersite dependence. The construction of a nonparametric procedure that keeps the dependence intact could be an interesting enhancement.

Multivariate versions of the HW procedure based on multivariate L-moments (Serfling and Xiao, 2007) have gained popularity (Chebana and Ouarda, 2007). Our proposed procedure could be adapted by defining multivariate TL-moments analogously. Difficulties could be the definition of an appropriate copula for the bootstrap, which must model the dependence of all variates at all sites simultaneously. A promising approach could be the use of nested Archimedean copulas (e.g. Hofert, 2010) that combine d copulas describing the dependence between multiple variates at d sites with one copula describing the intersite dependence.

In the following two subsections we compare our approach to the ones of Castellarin, Burn, and Brath (2008) and Viglione, Laio, and Claps (2007) by replication of their simulation studies. In each comparison only one of the two problems, heterogeneity or skewed data, is considered. In order to stay close to these studies, we use the rejection criterion $H_1 > 2$ here and present size and power rates. We calculated the size adjusted power as well and mention its behaviour.

3.5.1 Comparison to Castellarin

Castellarin, Burn, and Brath (2008) proposed an empirical corrector applied afterwards to the Hosking-Wallis test statistic to adjust for cross-correlation. We replicated their simulation with 20 stations of which 19 follow a GEV$(1, 0.4, 0)$ and one station follows a GEV$(1, 0.7, 0)$ as marginal distribution. Each station consists of $n = 25$ years of measurements. Because these authors used a multivariate normal distribution to generate cross-correlated data and then transformed the margins to the above-mentioned distributions, we utilised a Gaussian copula to generate similar dependence structures. Rank correlations of $\rho = 0, 0.1, \ldots, 0.8$ were considered. To simulate the case of homogeneity, data sets in which each margin follows a GEV$(1, 0.4, 0)$ distribution are considered as well.

For each setting 25 000 data sets are simulated. We apply the original Hosking-Wallis procedure, the corrected version (with correction coefficient $C = 0.122$ calculated by

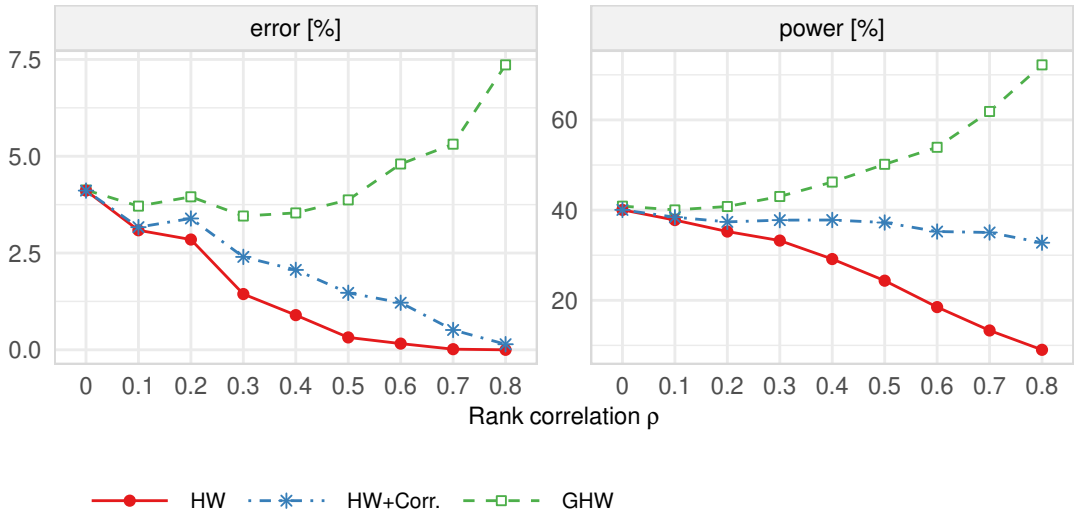

Figure 3.10: Comparison of size and power between original procedure (`HW`), corrected procedure (`HW+Corr`), and generalized Hosking-Wallis (`GHW`) procedure for different correlations.

Castellarin, Burn, and Brath, 2008), and the generalized Hosking-Wallis procedure using L-moments and the Gaussian copula model.

Figure 3.10 contains the size and power of the three procedures depending on the rank correlation coefficient. The size and power rates are similar in the independence case ($\rho = 0$), but differ with increasing correlation. The size of the original and corrected version shrinks with higher correlations while the size of the generalized version stays higher and exceeds 5% at the highest correlation. The power of the original procedure decreases, but the corrected version is able to compensate and leads to a stable curve. The generalized procedure, however, is even able to increase the power with increasing correlation.

If the size adjusted power is considered, all three tests yield similar rates. This is because the introduction of a copula only prevents the size from dropping due to intersite dependence and does not improve the heterogeneity detection itself.

This shows that in the settings considered here the generalized version can be more suited to face cross-correlation because it not only compensates the power reduction but instead can incorporate the dependence to increase the power. It has to be noted that in this simulation we specified the copula model accurately, which is difficult in practice.

3.5.2 Comparison to Viglione

Viglione, Laio, and Claps (2007) compared the Hosking-Wallis procedure to two rank-based procedures, the Anderson-Darling test and the Durbin-Knott test. Their final recommendation is to choose the Hosking-Wallis procedure if the L-skewness coefficient is below 0.23 and the Anderson-Darling-test otherwise. We redid most of their simulation study including our new approach. For this, several combinations of L-CV and L-skewness are considered as the mean of a group of stations. Sets of 11 stations are built with varying either L-CV, L-skewness, or none (which corresponds to the homogeneous case). Each station consisting of $n = 30$ measurements is simulated using the generalized extreme value distribution with parameters corresponding to the specific L-CV and L-skewness.

Besides the original Hosking-Wallis procedure and the Anderson-Darling test, we included the generalized Hosking-Wallis procedure using TL(0,1)-moments and an independence copula to the simulation.

Figure 3.11 (top panel) gives the size when analysing homogeneously built data sets. Comparing the new procedure (GHW) to the others, it can be observed that the size is comparable or smaller and that there are no regions in which the size of the new procedure exceeds 5% substantially. In the bottom panel of Figure 3.11 the power rates are given for the case that the station's L-CV τ_2 varies equidistantly around the group's mean L-CV $\bar{\tau}_2$. The spread of these variations, $\Delta\tau_2$, is set to

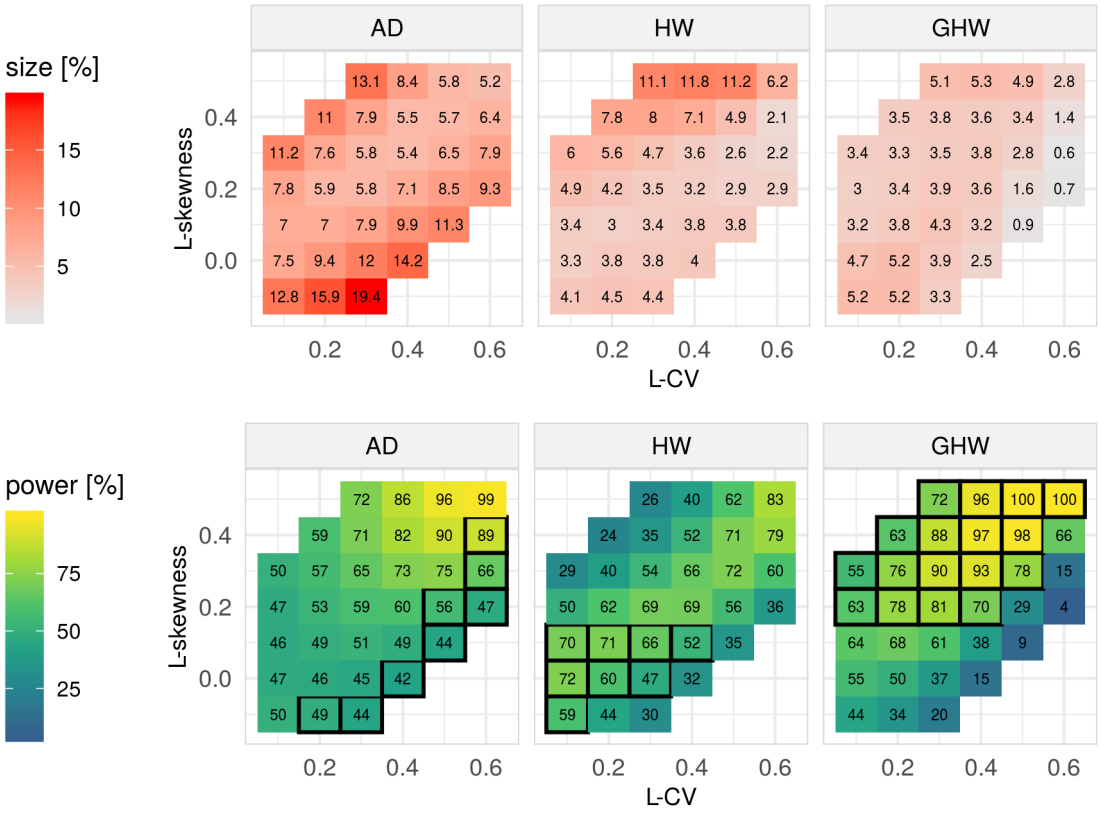

Figure 3.11: Size and power of the Anderson-Darling (AD), Hosking-Wallis (HW), and generalized Hosking-Wallis (GHW) procedure in the $\tau_2 - \tau_3$-space. Position on the grid determines the group's mean L-CV and L-skewness. Heterogeneity is constructed varying L-CV. Bordered tiles indicate the procedure with the highest power at each position in the bottom panel.

$\Delta\tau_2 = 0.5\bar{\tau}_2$, so with a mean L-CV of 0.4 we vary the station's L-CV between 0.3 and 0.5. The generalized Hosking-Wallis procedure is the best among these tests when L-skewness is roughly larger than or equal to 0.2. The original Hosking-Wallis procedure outperforms the other procedures when L-skewness is lower than 0.2. The Anderson-Darling test provides the highest power rates when the groups centre lies on the upper edge of L-CV, but in this region the size (see upper panel) is increased simultaneously. If the power was adjusted to an equal size, this became even more clear. In that case the best test is either the HW or the GHW test, with the GHW being best at the same positions like for the regular power.

Besides variation in L-CV, Viglione, Laio, and Claps (2007) examined variation in L-skewness τ_3. With $\Delta\tau_3$ denoting the group's spread in L-skewness, Figure 3.12 gives the power rates at four specific points in the $\tau_2 - \tau_3$-grid depending on the relative spread. As we can see, the new procedure can compete with the Anderson-Darling test, which was superior to the original Hosking-Wallis procedure in this setting.

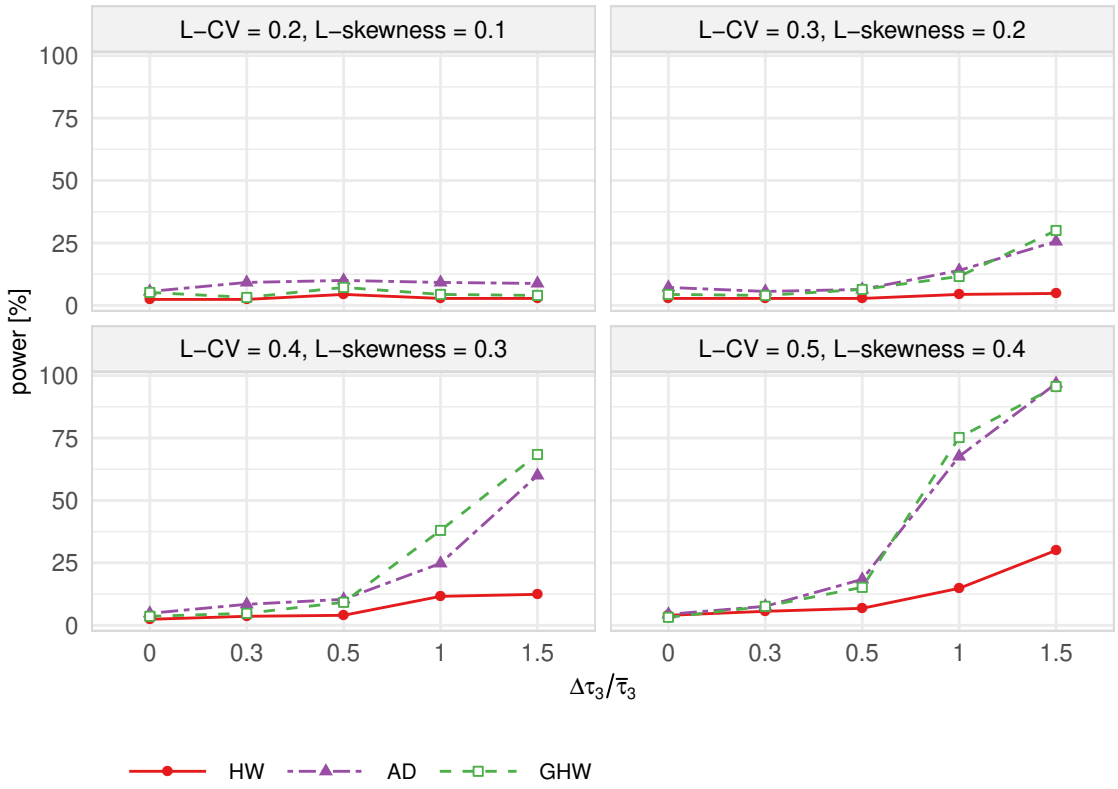

Figure 3.12: Rejection of the null hypothesis using the Anderson-Darling (AD), Hosking-Wallis (HW), and generalized Hosking-Wallis (GHW) procedure at specific points of L-CV and L-skewness. Heterogeneity is constructed varying L-skewness.

In summary, the new procedure outperforms the Anderson-Darling test when analysing highly skewed data sets. When the skewness is low and variations at L-CV scale are expected, the original Hosking-Wallis procedure can still be recommended.

3.6 Case study

A real data example is presented to illustrate the advantage of the modified procedure. For this purpose we use discharge data from 10 gauges in the Mulde river basin in the South-Eastern part of Germany with observation lengths of more than 75 years. This basin almost completely drains the north side of a low mountain range, the Ore Mountains, and is part of the Elbe river basin, which has suffered two extreme floods in the last 20 years (2002 and 2013; for more information see Fischer, Schumann, and Schulte, 2016). The Mulde basin consists of the watersheds of three main tributaries

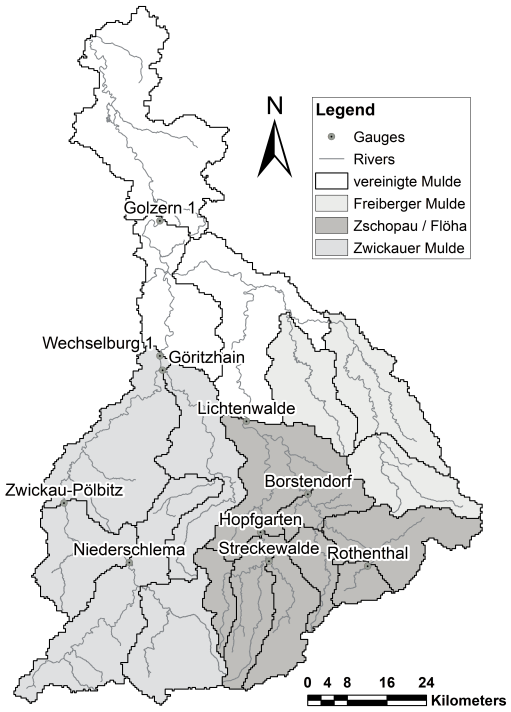

Figure 3.13: Map of the stations and river network.

– the Zwickauer Mulde, Zschopau and Freiberger Mulde rivers – and the catchment of the Vereinigte (united) Mulde River downstream. An overview of the basin, the river network, and the runoff gauges is given in Figure 3.13. The catchment covers an area of approximately 7400 km^2. The three main tributaries rise close to the mountain ridge at elevations between 760 and 1125 m above sea level. The outlet of the basin is located at an elevation of 81 m a.s.l.

The climatic seasonality of the runoff and flood regimes differ. The seasonal distribution of runoff has its maximum in late spring and its minimum in autumn. The flood peak averages are also high in spring (resulting from snowmelt and wet soil conditions), but there is often also a maximum in summer (July/August) caused by extreme rainfall events. With regard to higher spatial variability of summer flood events, the series of annual maxima of the summer period were used here.

Table 3.3 contains basic information about the gauges. Besides mean and standard deviation (sd) of the summerly maximum flood peak, the coefficient of variation (cv) and the discordancy measure (D_j, Hosking and Wallis, 1997) is given. The cv varies around a value of 1 at most of the stations but is clearly lower at "Goeritzhain". The discordancy measure, which is based on L-moment ratios, is relatively high for "Goeritzhain" and "Hopfgarten", although both values do not exceed the critical value of 2.491 (given in Hosking and Wallis, 1997). Both measures can be an indication of heterogeneity.

Table 3.3: Summary of maximum flood peaks at stations used in the case study.

Name of station	catchment area [km^2]	range of records	n	mean	sd	cv	D_j
Borstendorf	644	1929-2013	78	69.6	71.7	1.03	0.58
Goeritzhain	532	1910-2013	101	57.8	45.0	0.78	1.95
Golzern	5433	1912-2013	96	376.7	394.0	1.05	0.58
Hopfgarten	529	1911-2013	97	56.7	57.2	1.01	2.17
Lichtenwalde	1572	1910-2013	98	150.5	164.9	1.10	0.69
Niederschlema	754	1929-2013	78	88.8	87.7	0.99	0.35
Rothenthal	75	1929-2013	82	12.0	11.4	0.95	0.97
Streckewalde	206	1921-2013	88	22.9	21.2	0.93	0.75
Wechselburg	2099	1910-2013	101	174.6	171.8	0.98	1.28
Zwickau-Poelbitz	1021	1928-2013	79	105.6	112.0	1.06	0.70

Our task is therefore to assess the homogeneity of the given group of stations, e.g. in order to check if a regional flood frequency analysis is reasonable. A difficulty that arises is the high intersite dependence. Spearman's rank correlation coefficient was calculated for each pair of stations. Coefficients between 0.5 and 0.96 (between "Zwickau-Poelbitz" and "Niederschlema") verify the existence of medium to strong dependencies.

Figure 3.14 depicts a L-moment ratio diagram, which is commonly used to assess which distribution is suitable to model the data (Vogel and Fennessey, 1993). The group mean is near the GEV distribution line and inside the area covered by the four-parameter kappa distribution, which contains all pictured combinations below the GLO curve (see Hosking, 1994; Hosking and Wallis, 1993). Usage of the kappa distribution within the bootstrap procedure seems therefore reasonable. Additionally the goodness-of-fit measure of Hosking and Wallis (1997) was calculated for all distribution families which confirmed the findings ($Z^{GLO} = 0.24$, $Z^{GEV} = -0.25$, $Z^{GNO} = -1.51$, $Z^{P3} = -3.67$, $Z^{GPA} = -2.15$ with $|Z| > 1.64$ indicating a rejection of the assumption that the distribution fits the data). Because all stations lie between the medium skewed and highly skewed setting of the simulation study (see Figure 3.14), the recommendation derived in this study is to choose asymmetrically trimmed L-moments.

The group is tested with the original Hosking-Wallis procedure (i.e. with independence copula and L-moments), the corrected version of Castellarin, Burn, and Brath (2008), the Anderson-Darling procedure, and with the modified procedure using a Gumbel copula and TL(0,1)-moments. The results are given in Table 3.4. Our proposed procedure is the only one able to detect heterogeneity. The Hosking-Wallis procedure gives a negative value which is an indication of correlation between the sites according to Hosking and Wallis (1997). With this in mind, this result cannot be used to decide

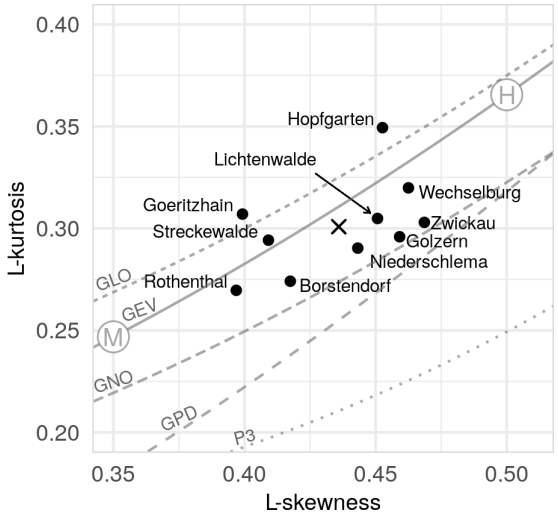

Figure 3.14: L-moment ratio diagram of stations and their group mean (cross). Curves give distribution families of generalized logistic (`GLO`), generalized extreme-value (`GEV`), generalized normal (`GNO`), generalized Pareto (`GPD`), and Pearson type 3 distribution (`P3`). Big circles give distributions employed in the simulation study (`M` and `H` indicating medium and high skewness, respectively).

about homogeneity. Based on the empirical coefficient of variation, we suspected the station "Goeritzhain" to be discordant to the others. All procedures yield homogeneity if the station "Goeritzhain" is excluded from the group. This result can be explained by different hydrological conditions. The catchment of the gauge Goeritzhain has a significant urbanised part (220 of 533 km^2), which is caused by the City of Chemnitz. In summertime this part of the catchment reacts faster and the flood peaks differ from other gauges.

Table 3.4: Results of homogeneity tests applied to our case study data (with and without the discordant site of Goeritzhain) for the classic and corrected HW procedure, the AD procedure, and our proposed approach.

Test	complete data	without Goeritzhain
HW	$H_1 = -0.88$	$H_1 = -1.61$
HW + Corr	$H_1 = -0.14$	$H_1 = -0.90$
GHW (TL(0,1), Gumbel)	$H_1 = 2.74$	$H_1 = 0.75$
AD*	$p = 0.52$	$p = 0.89$

* reject to α-level if $p < \alpha$.

3.7 Conclusions

We have proposed a generalization of the Hosking-Wallis procedure that uses trimmed L-moments and copula models to overcome the known disadvantages when highly-skewed or cross-correlated data occurs.

In simulation studies we investigated the choice of the degree of trimming of TL-moments and the selection of copula models as well as the robustness to extreme values occurring unusually often. A comparison to former studies showed that the new procedure is capable of improving drawbacks of the original procedure. A case study illustrated that the classic procedure can fail to detect heterogeneity due to intersite dependence and medium to high skewed distributions. The improved procedure, however, is able to detect heterogeneity in this application.

Overall the generalized procedure offers an improvement to the original procedure in many cases. The drawback, in exchange, is the need to specify a copula model and to select trimming parameters of L-moments. The most important observations/recommendations are as follows:

1. There is a disparity in the test's power rate depending on the direction in which the discordant site varies. It is important to keep in mind that some variations are not detectable and that the detectable region differs depending on whether L- or TL-moments are used.

2. Asymmetrical trimmed L-moments are beneficial in our settings. We recommend the application of TL(0,1)-moments when analysing positively skewed data. TL(0,2)-moments could be useful when analysing highly skewed data, but tend to increase the type-I-error.

3. The usage of simple copula models does not harm when analysing independent data if the independence copula is a special case of the used copula model.

4. When analysing dependent data, simple copula models seem to be sufficient to calculate adequate test statistics, even when the dependence structure is more complex. Trimmed L-moments lead to a robustification against copula misspecification.

5. Application of trimmed L-moments also leads to a more robust behaviour when a station experiences very high values unusually often. The degree of robustness increases with the degree of trimming, which needs to be chosen prior to the analysis.

4 Limit theorems for sample PWMs and TL-moments

In Section 2.4.1 and 2.4.3 the basics of probability-weighted moments and TL-moments have been provided. To build procedures like homogeneity tests based on such moments, it is useful to know the distribution of their sample versions. Hosking, Wallis, and Wood (1985) state that sample PWMs are asymptotically normal and give expressions to calculate the limiting covariance matrix for GEV distributed data. The exact covariance structure of sample TL-moments is analysed and a nonparametric estimation thereof is given in Elamir and Seheult (2003) and more generally in Hosking (2007). However, those works only consider the distributions of sample PWMs or TL-moments from a single distribution and not in a regional setting. Therefore, in this chapter a limit theorem for sample PWMs is constructed that is valid for local as well as for regional situations, i.e. the asymptotic joint distribution of PWMs at different stations is analysed. Limit laws for sample TL-moments and resulting GEV parameter and even quantile estimations can then be derived by application of the delta method. These results are used to construct an asymptotic test of homogeneity.

This chapter is based on the work "On the method of probability-weighted moments in regional frequency analysis" published as Discussion Paper (Lilienthal, Kinsvater, and Fried, 2016).

4.1 Limit theorem for sample PWMs

Let $\boldsymbol{X} = (X_1, \ldots, X_d)'$ be a d-dimensional random vector whose continuous marginal distribution functions are denoted by $F_j(x) = P(X_j \leq x)$, $j = 1, \ldots, d$. Note that we do not assume the components to be independent. Let $R \in \mathbb{N}$ be fixed. The first R PWMs of F_j are denoted by $\boldsymbol{\beta}_j = (\beta_{0;j}, \beta_{1;j}, \ldots, \beta_{R-1;j})' \in \mathbb{R}^R$, where

$$\beta_{r;j} = E(X_j F_j(X_j)^r) = \int_{\mathbb{R}} x F_j^r(x) dF_j(x), \; r = 0, \ldots, R-1, \; j = 1, \ldots, d. \tag{4.1}$$

All these local PWM vectors are summarized in $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_d') \in \mathbb{R}^{dR}$.

Suppose that $\boldsymbol{X}_i = (X_{i;1}, \ldots, X_{i;d})'$, $i = 1, \ldots, n$, denote independent copies of $\boldsymbol{X}$, where i can be interpreted as a time index and with $\{1, \ldots, n\}$ covering the full observation period. Keep in mind that in the application to the analysis of river flows it is unlikely that the observation period is the same for all d sites. Instead we want to assume that sites are built at different points in time but the observation ends at the same time. Let $n = n_1 \geq n_2 \geq \ldots \geq n_d$ denote the local sample lengths, which are rearranged by length for ease of presentation. We observe the scheme

$$\begin{array}{rrrrrrr} X_{1;1}, & X_{2;1}, & X_{3;1}, & X_{4;1}, & X_{5;1}, & \ldots, & X_{n;1}, \\ X_{a_2+1;2}, & X_{a_2+2;2}, & X_{a_2+3;2}, & \ldots, & X_{n;2}, & & \\ & \ddots & & \vdots & & & \\ X_{a_d+1;d}, & X_{a_d+2;d}, & \ldots, & X_{n;d}, & & & \end{array} \tag{4.2}$$

with $a_j = n - n_j$ and where each row contains only observations from the same station. It is important to account for the structure of this scheme to be able to properly capture the dependence between local estimates of probability-weighted moments. For the asymptotic results we let $n \to \infty$ and assume that $n_j/n \to c_j \in (0, 1)$ in order to account for possibly very different local sample lengths, i.e. we set $n_j = \lfloor nc_j \rfloor$.

The sample version of $\beta_{r;j}$ computed from $X_{a_j+1;j}, \ldots, X_{n;j}$ is given by

$$\hat{\beta}_{r;j} = \int_{\mathbb{R}} x \cdot F^r_{a_j+1:n;j}(x)\, dF_{a_j+1:n;j}(x) = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{a_j+i;j} \cdot F^r_{a_j+1:n;j}(X_{a_j+i;j}), \tag{4.3}$$

where $F_{\ell:m;j}$ is the empirical distribution function of $X_{\ell;j}, X_{\ell+1;j}, \ldots, X_{m;j}$. Sample counterparts of $\boldsymbol{\beta}_j \in \mathbb{R}^R$ and $\boldsymbol{\beta} \in \mathbb{R}^{dR}$ are denoted by

$$\hat{\boldsymbol{\beta}}_j = \left(\hat{\beta}_{0;j}, \ldots, \hat{\beta}_{R-1;j}\right)' \quad \text{and} \quad \hat{\boldsymbol{\beta}} = \left(\hat{\boldsymbol{\beta}}_1', \ldots, \hat{\boldsymbol{\beta}}_d'\right)', \tag{4.4}$$

respectively.

Theorem 1 *Suppose that $\boldsymbol{X}_i$, $i \geq 1$, is a sequence of independent copies of $\boldsymbol{X} = (X_1, \ldots, X_d)'$, whose PWMs are summarized in the vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_d')' \in \mathbb{R}^{dR}$ and with*

$$E\left[X_j F_j^k(X_j)\, X_\ell F_\ell^m(X_\ell)\right] < \infty \quad \textit{for all} \quad 1 \leq j, \ell \leq d \quad \textit{and} \quad 0 \leq k, m < R.$$

Suppose further that $\sup_{x \in \mathbb{R}} |x\{F_j(x)(1 - F_j(x))\}^w| < \infty$ for all $j = 1, \ldots, d$ and some $w \in [0, 1/2)$. Then, for fixed $\boldsymbol{r} \in (0, 1)^d$ and $n \to \infty$, we have that

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{D} \mathcal{N}\left(0,\ \Sigma\right),$$

where the limiting covariance matrix $\Sigma \in \mathbb{R}^{dR \times dR}$ is provided in Section 4.3.

This theorem has been derived and proven by Paul Kinsvater (see Lilienthal, Kinsvater, and Fried, 2016). Theorem 1 and a consistent estimator $\hat{\Sigma}$ of Σ allow us to develop asymptotically consistent methods for regional frequency analysis.

4.2 Limit theorems for sample TL-moments and resulting estimators

Instead of PWMs, in hydrological application nowadays mainly L-moments or more generally trimmed L-moments are used. As described in Section 2.4.2 and 2.4.3, they are interpretable analogously to summary statistics based on classical product moments, for instance, with λ_1, λ_2, and $\tau_3 = \lambda_3/\lambda_2$ representing location, dispersion, and skewness of the analysed distribution F, respectively. As shown in equations (2.64) and (2.79) every L-moment or TL-moment can be represented as a linear combination of a finite number of PWMs, provided F has finite mean. This fact, by referring to Theorem 1, allows us to derive limit theorems for sample TL-moments (and L-moments as a special case) and related methods.

4.2.1 At-site statistics

First we want to describe how limit theorems for at-site statistics, i.e. statistics of a single distribution F_j, can be deduced from Theorem 1. Therefore we assume that a set of local PWMs, $\boldsymbol{\beta}_j = (\beta_{0;j}, \ldots, \beta_{R-1;j})$, is given. According to Theorem 1 it follows that for some Σ_j

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\right) \xrightarrow{D} \mathcal{N}\left(0,\ \Sigma_j\right). \tag{4.5}$$

The TL-moment $\lambda_{m;j}^{(s,t)}$ of F_j of order $m \in \mathbb{N}$ with trimming $s, t \in \mathbb{N}_0$ is known to satisfy

$$\lambda_{m;j}^{(s,t)} = \sum_{i=0}^{m+s+t-1} z_{m-1,i}^{(s,t)}\, \beta_{i;j} = (\boldsymbol{z}_{m-1}^{(s,t)})' \boldsymbol{\beta}_j, \tag{4.6}$$

with $\boldsymbol{\beta}_j = (\beta_{0;j}, \ldots, \beta_{m+s+t-1;j})'$ being the vector of the first $m+s+t$ PWMs of F_j and $\boldsymbol{z}_{m-1}^{(s,t)} = (z_{m-1,0}^{(s,t)}, \ldots, z_{m-1,r+s+t-1}^{(s,t)})'$ being a coefficient vector with components

$$z_{m,i}^{(s,t)} = \frac{m!(m+s+t+1)!}{(m+1)(m+s)!(m+t)!}(-1)^{s+m+i}\binom{m+t}{i+s}\binom{m+i}{m}. \tag{4.7}$$

Now let $\Delta_{\beta\to\lambda}^{(s,t)} = (\boldsymbol{z}_0^{(s,t)}, \ldots, \boldsymbol{z}_{M-1}^{(s,t)})'$ denote the linear mapping such that $\boldsymbol{\lambda}_j^{(s,t)} = (\lambda_{1;j}^{(s,t)}, \ldots, \lambda_{M;j}^{(s,t)})' = \Delta_{\beta\to\lambda}^{(s,t)}\boldsymbol{\beta}_j$. Then the first M sample TL(s,t)-moments $\hat{\boldsymbol{\lambda}}_j^{(s,t)} =$

$(\hat{\lambda}_{1;j}^{(s,t)}, \ldots, \hat{\lambda}_{M;j}^{(s,t)})'$ are given by $\hat{\boldsymbol{\lambda}}_j^{(s,t)} = \Delta_{\beta\to\lambda}^{(s,t)} \hat{\boldsymbol{\beta}}_j$.

By using the affine property of multivariate normal distributions and the continuous mapping theorem, from equation (4.5) it follows that

$$\sqrt{n}\left(\hat{\boldsymbol{\lambda}}_j^{(s,t)} - \boldsymbol{\lambda}_j^{(s,t)}\right) \xrightarrow{D} \mathcal{N}\left(0,\ \Delta_{\beta\to\lambda}^{(s,t)} \Sigma_j (\Delta_{\beta\to\lambda}^{(s,t)})'\right). \tag{4.8}$$

So far we have introduced TL-moments as summary statistics of distributions without restricting to any parametric family. In practice, however, one usually assumes that $F_j = F_{\boldsymbol{\vartheta}_j}$ for some unknown parameter vector $\boldsymbol{\vartheta}_j \in \Theta \subset \mathbb{R}^p$. Relationships between TL-moments and the distribution parameters are employed which allows us to estimate these parameters by plugging in sample TL-moments into the formulas.

More specifically, let $g^{(s,t)} : \mathbb{R}^m \mapsto \mathbb{R}^p$ be a totally differentiable function that maps the first M TL(s,t)-moments of $F_{\boldsymbol{\vartheta}_j}$ onto its parameter vector $\boldsymbol{\vartheta}_j$. From the delta method, for $\hat{\boldsymbol{\vartheta}}_j = g^{(s,t)}(\hat{\boldsymbol{\lambda}}_j^{(s,t)})$ and $n \to \infty$, we obtain that

$$\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}_j - \boldsymbol{\vartheta}_j\right) \xrightarrow{D} \mathcal{N}\left(0,\ \Delta_{\boldsymbol{\lambda}_j\to\vartheta}^{(s,t)} \Delta_{\beta\to\lambda}^{(s,t)} \Sigma_j \left(\Delta_{\boldsymbol{\lambda}_j\to\vartheta}^{(s,t)} \Delta_{\beta\to\lambda}^{(s,t)}\right)'\right), \tag{4.9}$$

where $\Delta_{\boldsymbol{\lambda}_j\to\vartheta}^{(s,t)} = \frac{\partial}{\partial \boldsymbol{\lambda}} g^{(s,t)}(\boldsymbol{\lambda}_j) \in \mathbb{R}^{p\times m}$ denotes the Jacobi matrix of $g^{(s,t)}$ evaluated at $\boldsymbol{\lambda}_j \in \mathbb{R}^m$. Note that, according to Rao (1973, p. 388), the rank of the multivariate normal distribution is equal to the rank of $\Delta_{\boldsymbol{\lambda}_j\to\vartheta}^{(s,t)} \Delta_{\beta\to\lambda}^{(s,t)} \Sigma_j \left(\Delta_{\boldsymbol{\lambda}_j\to\vartheta}^{(s,t)} \Delta_{\beta\to\lambda}^{(s,t)}\right)'$. The matrices $\Delta_{\boldsymbol{\lambda}_j\to\vartheta}^{(s,t)}$ corresponding to TL(0,0)- and TL(0,1)-moments are summarized in Appendix A.

In flood frequency analysis we are usually interested in the estimation of some quantile $\hat{q}$. Suppose that $h : \mathbb{R}^p \to \mathbb{R}$ with $h(\boldsymbol{\vartheta}) = F_{\boldsymbol{\vartheta}}^{-1}(q)$ is differentiable and let $\Delta_{\boldsymbol{\vartheta}_j\to q} \in \mathbb{R}^{1\times p}$ denote the corresponding Jacobi matrix evaluated at $\boldsymbol{\vartheta}_j$. Again, from the delta method we obtain for $n \to \infty$

$$\sqrt{n}\left(\hat{\boldsymbol{q}} - \boldsymbol{q}\right) \xrightarrow{D} \mathcal{N}\left(0,\ \Delta_{\boldsymbol{\vartheta}_j\to q} \Delta_{\boldsymbol{\lambda}_j\to\vartheta}^{(s,t)} \Delta_{\beta\to\lambda}^{(s,t)} \Sigma_j \left(\Delta_{\boldsymbol{\vartheta}_j\to q} \Delta_{\boldsymbol{\lambda}_j\to\vartheta}^{(s,t)} \Delta_{\beta\to\lambda}^{(s,t)}\right)'\right). \tag{4.10}$$

Considering the GEV with parameter vector $\boldsymbol{\vartheta} = (\mu, \sigma, \xi)'$ and quantile function as in equation (2.5), the Jacobi matrix of h is given by

$$\begin{aligned} \Delta_{\boldsymbol{\vartheta}_j\to q} &= \frac{\partial}{\partial \boldsymbol{\vartheta}} h(\boldsymbol{\vartheta}) \\ &= \left(1, \frac{(-\log(q))^{-\xi} - 1}{\xi}, \frac{\sigma(\xi^{-1} - (-\log(q))^{-\xi}(\log(-\log(q)) + \xi^{-1}))}{\xi}\right). \end{aligned} \tag{4.11}$$

4.2.2 Joint estimation at multiple stations

We now switch to a regional scale by considering multivariate observations as given in scheme (4.2). Recall that $\hat{\boldsymbol{\beta}}$ from equation (4.4) contains sample PWMs of all d marginal distributions F_j involved. In analogy to equation (4.4), the vector of all sample TL(s,t)-moments is denoted by

$$\hat{\boldsymbol{\lambda}}^{(s,t)} = \left((\hat{\boldsymbol{\lambda}}_1^{(s,t)})', \ldots, (\hat{\boldsymbol{\lambda}}_d^{(s,t)})' \right)', \tag{4.12}$$

with population counterpart $\boldsymbol{\lambda}^{(s,t)} = \left((\boldsymbol{\lambda}_1^{(s,t)})', \ldots, (\boldsymbol{\lambda}_d^{(s,t)})' \right)' \in \mathbb{R}^{Md}$. By Theorem 1 and the delta method we obtain that for $n \to \infty$

$$\sqrt{n} \left(\hat{\boldsymbol{\lambda}}^{(s,t)} - \boldsymbol{\lambda}^{(s,t)} \right) \xrightarrow{D} \mathcal{N} \left(0,\ \tilde{\Delta}_{\beta \to \lambda}^{(s,t)} \, \Sigma \, (\tilde{\Delta}_{\beta \to \lambda}^{(s,t)})' \right), \tag{4.13}$$

with Σ being the limiting covariance matrix of Theorem 1 and with block-diagonal matrix

$$\tilde{\Delta}_{\beta \to \lambda}^{(s,t)} = \operatorname{diag}(\Delta_{\beta \to \lambda}^{(s,t)}, \ldots, \Delta_{\beta \to \lambda}^{(s,t)}) = \begin{pmatrix} \Delta_{\beta \to \lambda}^{(s,t)} & 0 & \ldots & 0 \\ 0 & \Delta_{\beta \to \lambda}^{(s,t)} & & \vdots \\ \vdots & & \ddots & \\ 0 & \ldots & & \Delta_{\beta \to \lambda}^{(s,t)} \end{pmatrix}. \tag{4.14}$$

Similarly, under the assumption that we have parametric margins $F_j = F_{\boldsymbol{\vartheta}_j}$ for all $j = 1, \ldots, d$ with block-diagonal matrices $\tilde{\Delta}_{\boldsymbol{\lambda} \to \vartheta}^{(s,t)} = \operatorname{diag}(\Delta_{\boldsymbol{\lambda}_1 \to \vartheta}^{(s,t)}, \ldots, \Delta_{\boldsymbol{\lambda}_d \to \vartheta}^{(s,t)})$ and $\tilde{\Delta}_{\boldsymbol{\vartheta} \to q} = \operatorname{diag}(\Delta_{\boldsymbol{\vartheta}_1 \to q}, \ldots, \Delta_{\boldsymbol{\vartheta}_d \to q})$ taken into account, one can easily obtain the joint limiting distribution of parameter and quantile estimators for all d stations.

4.3 Estimation of the limiting covariance matrix

The limiting covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{1,1} & \cdots & \Sigma_{1,d} \\ \vdots & \ddots & \vdots \\ \Sigma_{d,1} & \cdots & \Sigma_{d,d} \end{pmatrix} = \lim_{n \to \infty} \operatorname{Var} \left(\sqrt{n} \left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} \right) \right) \tag{4.15}$$

from Theorem 1 is defined block-wise by

$$\begin{aligned} \Sigma_{j,l} &= \lim_{n \to \infty} \operatorname{Cov} \left(\sqrt{n} \left(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j \right),\ \sqrt{n} \left(\hat{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}_\ell \right) \right) \\ &= \frac{\min(c_j, c_\ell)}{c_j \cdot c_\ell} \operatorname{Cov}(\boldsymbol{Z}_j,\ \boldsymbol{Z}_\ell) \in \mathbb{R}^{R \times R} \end{aligned} \tag{4.16}$$

and where $\boldsymbol{Z}_j = (Z_{0;j}, Z_{1;j}, \ldots, Z_{R-1;j})'$, $j = 1, \ldots, d$, are random vectors defined through

$$Z_{r;j} = X_j \cdot F_j^r(X_j) + \int_{\mathbb{R}} x \cdot r \cdot F_j^{r-1}(x) \cdot \mathbb{1}(X_j \leq x)\, dF_j(x),\ r = 0, \ldots, R-1. \quad (4.17)$$

According to Theorem 1, empirical probability-weighted moments are asymptotically jointly normal with limiting covariance matrix obtained from that of the variables defined in equation (4.17) (see the proof in Lilienthal, Kinsvater, and Fried, 2016, for more details).

4.3.1 Empirical estimator

Suppose that we have collected an observation scheme as given in (4.2). In practice the covariance matrices $\mathrm{Cov}(\boldsymbol{Z}_j, \boldsymbol{Z}_\ell)$ can be consistently estimated by their sample analogues: Let

$$\hat{Z}_{i,r;j} = X_{i;j} \cdot F^r_{a_j+1:n;j}(X_{i;j}) + \frac{1}{n_j} \sum_{i'=a_j+1}^{n} X_{i';j} \cdot r \cdot F^{r-1}_{a_j+1:n;j}(X_{i';j}) \cdot \mathbb{1}(X_{i;j} \leq X_{i';j}) \quad (4.18)$$

and $\hat{\boldsymbol{Z}}_{i;j} = (Z_{i,0;j}, Z_{i,1;j}, \ldots, Z_{i,R-1;j})'$, $i = a_j + 1, \ldots, n$. For $1 \leq j, \ell \leq d$, the covariance matrix $\mathrm{Cov}(\boldsymbol{Z}_j, \boldsymbol{Z}_\ell)$ is estimated by the empirical covariance matrix of the sample

$$\left\{ \left(\hat{\boldsymbol{Z}}_{\max(a_j,a_\ell)+1;j}, \hat{\boldsymbol{Z}}_{\max(a_j,a_\ell)+1;\ell} \right), \ldots, \left(\hat{\boldsymbol{Z}}_{n;j}, \hat{\boldsymbol{Z}}_{n;\ell} \right) \right\}. \quad (4.19)$$

The resulting estimator $\hat{\Sigma} = \left(\hat{\Sigma}_{j,\ell} \right)_{j,\ell=1,\ldots,d}$ with

$$\hat{\Sigma}_{j,\ell} = \frac{\min(c_j, c_\ell)}{c_j \cdot c_\ell} \widehat{\mathrm{Cov}}(\boldsymbol{Z}_j, \boldsymbol{Z}_\ell) \quad (4.20)$$

is called empirical estimator or nonparametric estimator.

Using $\hat{\Sigma}$ (which gives us an estimate of the limiting covariance matrix of $\sqrt{n}(\hat{\beta} - \beta)$) the limiting variance matrices of TL-Moments, parameter estimations, and quantile estimations can be estimated using the matrices $\tilde{\Delta}^{(s,t)}_{\beta \to \lambda}$, $\tilde{\Delta}^{(s,t)}_{\boldsymbol{\lambda} \to \vartheta}$, and $\tilde{\Delta}_{\boldsymbol{\vartheta} \to q}$, described in Section 4.2.

4.3.2 Parametric modification on the block diagonal

Hosking, Wallis, and Wood (1985) derived a parametric expression for the limiting covariance matrix of empirical PWMs of GEV distributed variables. In applications we often assume that the margins are GEV and known up to some finite dimensional parameters $\boldsymbol{\vartheta}_j$, i.e. $F_j = G_{\boldsymbol{\vartheta}_j}$. Therefore these expressions can be used to replace the

local parts of $\hat{\Sigma}$, i.e. $\hat{\Sigma}_{1,1}, \ldots, \hat{\Sigma}_{d,d}$, by their parametric estimates $\Sigma_{j,j}(\hat{\boldsymbol{\vartheta}}_j)$, where $\hat{\boldsymbol{\vartheta}}_j$ are consistent estimates of $\boldsymbol{\vartheta}_j$, e.g. TL-moment estimators of GEV parameters. The non-diagonal elements of $\hat{\Sigma}$, i.e. the parts describing covariances between PWMs of different distributions, are still estimated using the empirical estimator. The modified estimator of Σ is denoted by $\hat{\Sigma}_m$ and can also be used to derive estimations of the limiting variance of TL-moments, parameter estimates, and quantile estimates. The modified estimator $\hat{\Sigma}_m$ is not necessarily a valid covariance matrix. The mixture of nonparametric and parametric parts involved produces negative eigenvalues in some cases. In the following section this problem will be analysed.

4.3.3 Empirical analysis

We now want to empirically examine the consistency of the estimators $\hat{\Sigma}$ and $\hat{\Sigma}_m$ as well as the derived covariance estimators for TL-moments and distribution parameter estimates.

Therefore data from a two-dimensional distribution is generated with margins being defined as $F_j = GEV(10, 5, \xi_j), j \in \{1, 2\}$ and a Gumbel(θ) copula describing their dependence. The shape $\xi_j \in \{0, 0.05, 0.25, 0.45, 0.5\}$, the degree of dependence $\theta \in \{1.1, 1.4, 2.4\}$, and the number of observations $n \in \{25, 50, 100, 250, 500, 1000, 2500, 5000\}$ are varied. Each situation is replicated $B = 5000$ times.

To assess the performance of the respective estimator, relative distances to the true covariance matrix are calculated using the Frobenius norm:

$$d(\hat{\Sigma}, \Sigma) = \frac{||\hat{\Sigma}_b - \Sigma||_F}{||\Sigma||_F} = \frac{\sqrt{\sum_{j=1}^{d} \sum_{j'=1}^{d} (\hat{s}_{jj'} - s_{jj'})^2}}{\sqrt{\sum_{j=1}^{d} \sum_{j'=1}^{d} s_{jj'}^2}}, \tag{4.21}$$

with $\hat{\Sigma}_b = (\hat{s}_{jj'})_{jj'}$ being the estimator of $\Sigma = (s_{jj'})_{jj'}$ in one specific replication. Since the true covariance matrix Σ is unknown, we approximate it as the empirical covariance matrix in the corresponding situation built using 50 000 replications. The distance (4.21) is calculated for each of the $B = 5000$ estimations and the mean relative Frobenius distance $\frac{1}{B} \sum_{b=1}^{B} d(\hat{\Sigma}_b, \Sigma)$ is used to measure the quality of estimation. Note that, for the sake of simplicity, Σ and $\hat{\Sigma}$ here denote the limiting covariance matrix of either PWMs, TL-moments, or parameter estimators and the corresponding estimation thereof (empirical estimator or with parametric modification).

Results of this analysis for the PWM estimator are given in Figure 4.1. The panels do not differ much within each row, so the degree of dependence seems to have little influence on the distance to the true covariance matrix. For little and moderately skewed situations (upper two rows) both estimators tend to zero for increasing sample length, and for samples longer than $n = 250$ the differences between both estimators

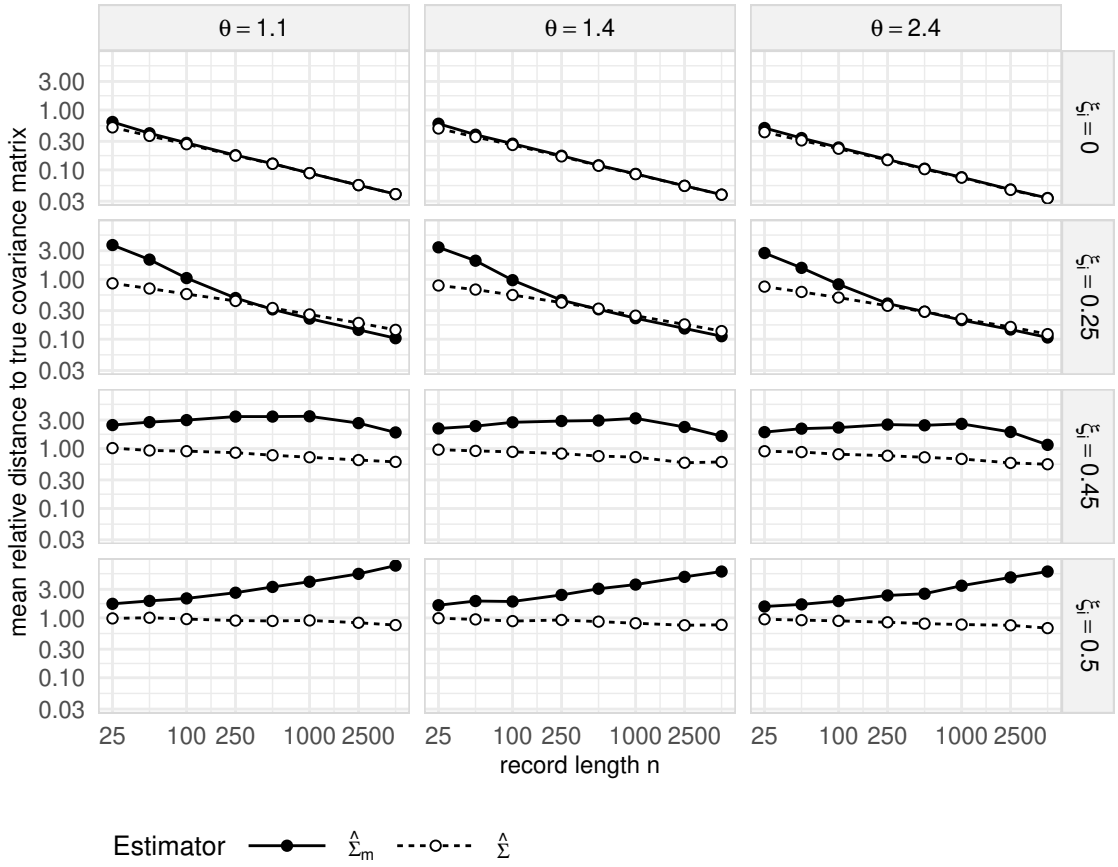

Figure 4.1: Mean relative Frobenius distances to true limiting covariance matrix of PWM estimator for increasing sample length n in different situations.

become very small. In settings with higher skewness the decrease of distance is very slow or even negative. One reason is that the parametric formulas used in the modified estimator are only valid for $\xi < 0.5$ (Hosking, Wallis, and Wood, 1985).

Situations in which both shape parameters differ have been analysed as well but are not reported in full detail for the sake of brevity. The results are similar to those given before with the larger shape parameter determining the behaviour.

We now have a look at the consistency of the covariance estimation of TL-moment estimators and parameter estimators. Because of the previous results and for the sake of simplicity, we restrict ourselves to the medium-dependence case of $\theta = 1.4$. Figure 4.2 displays the results. Again, only the situations with equal shapes are given, differing shapes lead to similar plots with the upper shape defining the behaviour. In scenarios of low to medium skewness, the distances between all covariance matrix estimators to the true covariance matrix tend to zero for increasing sample length. In situations of higher skewness this fails for TL(0,0)-moments and parameter estimators

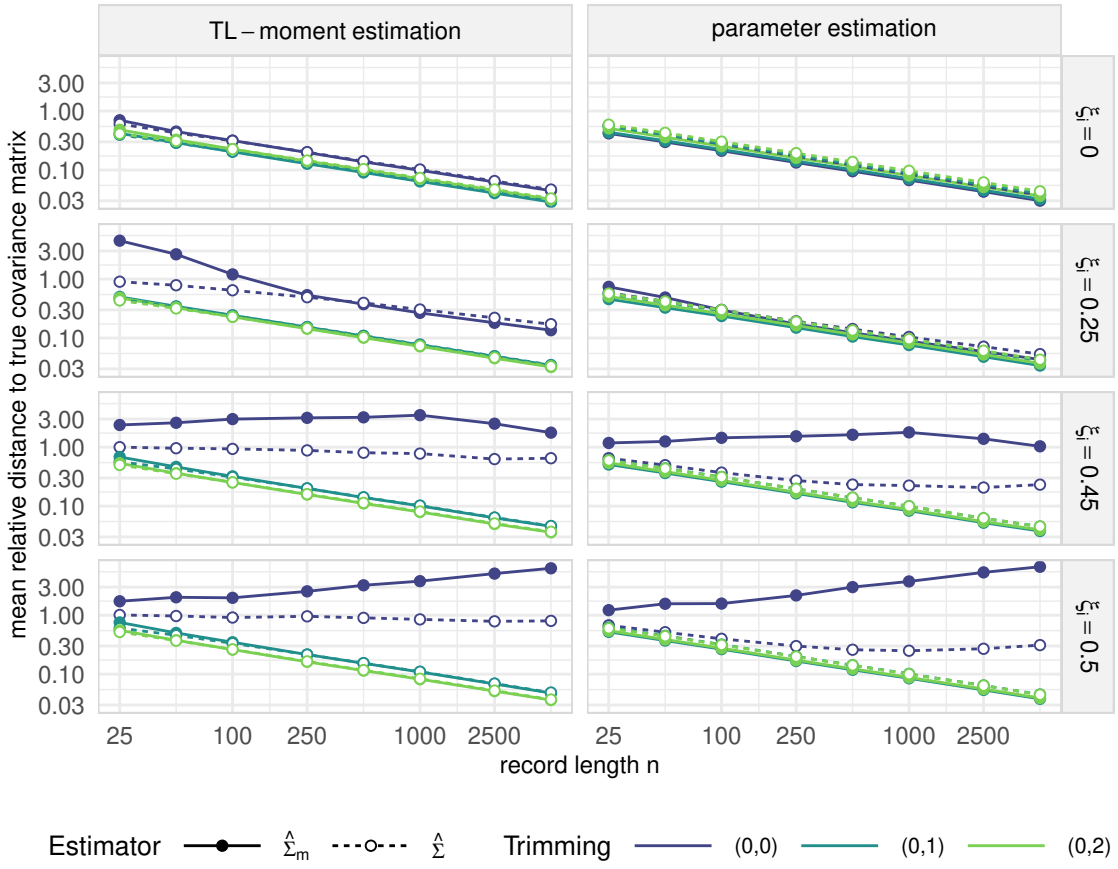

Figure 4.2: Mean relative Frobenius distances to true limiting covariance matrix of TL-moment estimators and parameter estimators for increasing sample length n in different situations ($\theta = 1.4$).

based on TL(0,0)-moments. Asymmetrically trimmed TL(0,1)- or TL(0,2)-moments and parameter estimators based on them are still consistent in these situations.

As mentioned earlier, the modified estimator can lead to matrices that are not positive semi-definite. Figure 4.3 shows the proportion of positive semi-definite covariance matrices given by the modified estimator. If no trimming is used there is a substantial proportion of matrices in most settings (except for situations with low skewness and low dependence) that are not positive semi-definite. For trimmed moments the proportion of positive semi-definite matrices is much higher and the proportion is only remarkably different from one for short record lengths (less than 100 observations) and high dependence.

The empirical analysis has shown that the limiting covariance matrix of PWM estimators can be estimated consistently in low to medium skewness situations and that these estimations can also be extended to estimators of TL-moments

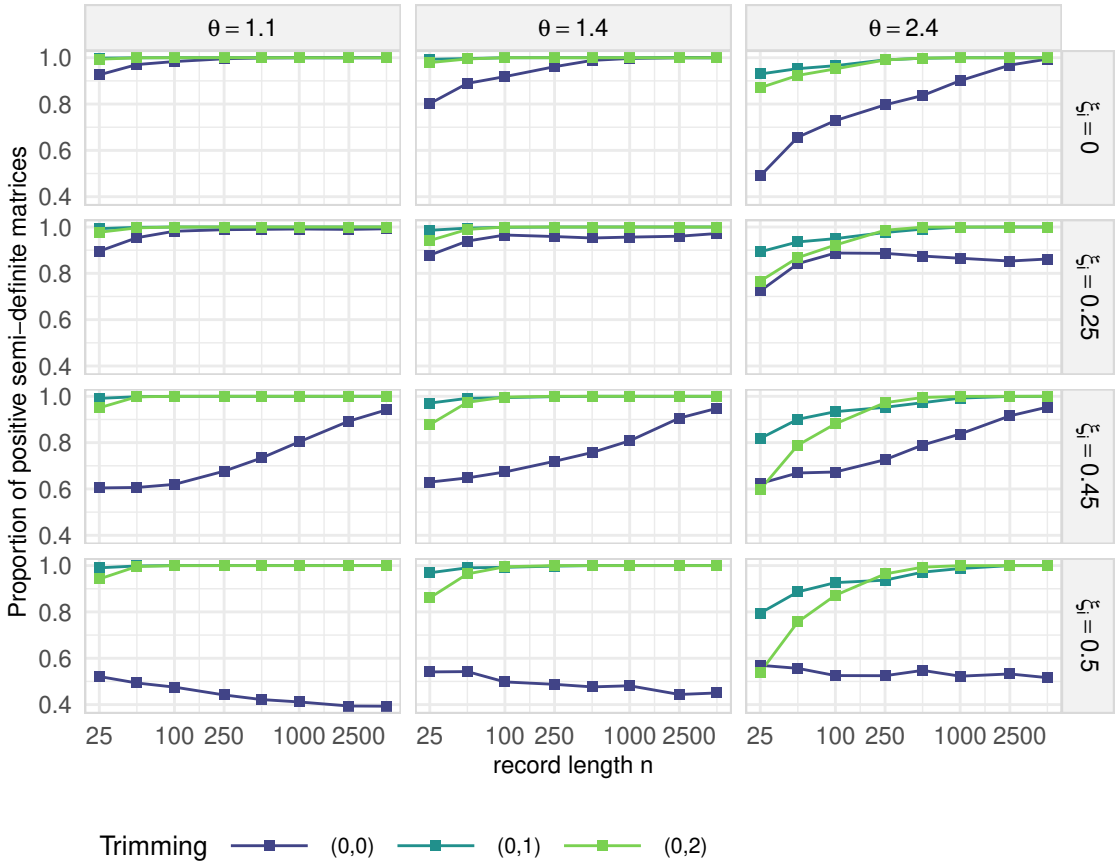

Figure 4.3: Proportion of positive semi-definite matrices yield by the modified estimator depending on record length.

and resulting parameter estimations. In the case of high skewness, the use of a non-negative trimming still leads to consistent estimators.

4.4 Test of regional homogeneity

When considering observations from multiple stations, e.g. scheme (4.2), in flood frequency analysis mostly the Index Flood assumption $\mathcal{H}_{0,IF}$ stated in (2.89) is applied in order to decrease the variability of estimation. However, while a moderate amount of heterogeneity of the group may still lead to an overall improvement compared to local estimation (Lettenmaier, Wallis, and Wood, 1987), strong heterogeneity typically leads to a severe bias which again increases the overall estimation error. It is thus important to be able to identify serious sources of heterogeneity. We are going to introduce a statistical test that proves to be advantageous in several aspects to competitive procedures from the literature.

4.4.1 Test construction

Suppose that we have observed scheme (4.2) with site-specific distribution functions $F_j = F_{\boldsymbol{\vartheta}_j}, j = 1, \ldots, d$, and that $F_{\boldsymbol{\vartheta}_j} = GEV(\mu_j, \sigma_j, \xi_j)$ is the GEV distribution with parameters $\boldsymbol{\vartheta}_j = (\mu_j, \sigma_j, \xi_j)'$. In this case the Index Flood model $\mathcal{H}_{0,IF}$ is equivalent to

$$\frac{\mu_i}{\sigma_i} = \ldots = \frac{\mu_d}{\sigma_d} \quad \text{and} \quad \xi_1 = \ldots = \xi_d. \tag{4.22}$$

Let $\hat{\boldsymbol{\vartheta}} = (\hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1, \ldots, \hat{\mu}_d, \hat{\sigma}_d, \hat{\xi}_d)'$ denote the TL(s,t)-moment estimator of local parameters from Section 4.2 obtained from scheme (4.2). Let $g : \mathbb{R}^{3d} \mapsto \mathbb{R}^{2d}$ denote the mapping $g(\boldsymbol{\vartheta}) = (\mu_1/\sigma_1, \xi_1, \ldots, \mu_d/\sigma_d, \xi_d)'$ and $\Delta_{\boldsymbol{\vartheta}\to\tilde{\vartheta}}$ the corresponding Jacobi-matrix $\frac{\partial}{\partial\boldsymbol{\vartheta}} g(\boldsymbol{\vartheta})$ evaluated at $\boldsymbol{\vartheta}$. Again, from the delta method, we obtain that

$$\sqrt{n}\left(g(\hat{\boldsymbol{\vartheta}}) - g(\boldsymbol{\vartheta})\right) \xrightarrow{D} \mathcal{N}\left(0,\ \Gamma\right), \tag{4.23}$$

with $\Gamma = \ \Delta_{\boldsymbol{\vartheta}\to\tilde{\vartheta}} \tilde{\Delta}^{(s,t)}_{\boldsymbol{\lambda}\to\vartheta} \tilde{\Delta}^{(s,t)}_{\beta\to\lambda} \, \Sigma \, (\Delta_{\boldsymbol{\vartheta}\to\tilde{\vartheta}} \tilde{\Delta}^{(s,t)}_{\boldsymbol{\lambda}\to\vartheta} \tilde{\Delta}^{(s,t)}_{\beta\to\lambda})'$ as $n \to \infty$.

Using

$$C = \begin{pmatrix} 1 & 0 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & & & \ddots & & & & \vdots \\ 0 & \cdots & & & & 1 & 0 & -1 \end{pmatrix} \in \mathbb{R}^{2(d-1),2d}, \tag{4.24}$$

hypothesis (4.22) can be expressed as

$$C \cdot g(\boldsymbol{\vartheta}) = \begin{pmatrix} \delta_1 - \delta_2 \\ \xi_1 - \xi_2 \\ \vdots \\ \delta_{d-1} - \delta_d \\ \xi_{d-1} - \xi_d \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}. \tag{4.25}$$

To check if assumption (4.22) holds, we use a Wald-type test statistic

$$T_n = n \, (C \; g(\hat{\boldsymbol{\vartheta}}))' (C \; \hat{\Gamma} \; C')^{-1} (C \; g(\hat{\boldsymbol{\vartheta}})). \tag{4.26}$$

Under $\mathcal{H}_{0,IF}$, for $n \longrightarrow \infty$ and under the assumptions of Theorem 1, we have that $T_n \xrightarrow{D} \chi^2_{2(d-1)}$ (which can be proved easily by application of the continuous mapping theorem and the theorem of Slutsky to the result (4.23)). Therefore, the hypothesis that $\mathcal{H}_{0,IF}$ holds is rejected at a significance level of α, if the observed value of T_n exceeds the $(1-\alpha)$-quantile of a $\chi^2_{2(d-1)}$.

4.4.2 Simulation study

To check the capability of the proposed homogeneity test, a simulation study is conducted at a nominal level of $\alpha = 5\%$. The data is generated from $d = 6$ stations with different local sample lengths, with margins $F_j = GEV(\mu_j, \sigma_j, \xi_j)$ and, for simplicity, with a Gumbel copula C_θ describing their intersite dependence.

Five of the six sites are selected concordant with $\mu_j = 10 \times j$, $\sigma_j = 5 \times j$, $j = 1, \ldots, 5$ and the shape parameter as either $\xi_j = 0.05$ (little skewed), $\xi_j = 0.25$ (medium skewed), or $\xi_j = 0.45$ (highly skewed), for $j = 1, \ldots, 5$. The sixth site can be discordant from the first five stations with $\mu_6 = 10 \times 6$, $\sigma_6 = \tilde{\sigma} \times 6$ and $\xi_6 = \tilde{\xi}$ varying on a grid around the parameters of the concordant sites. The dependence parameter controlling the intersite dependence is varied as $\theta \in \{1.1, 1.4, 2.4\}$.

We replicate the homogeneous situations, i.e. the sixth site is concordant to the first five sites, 10 000 times for $n_{max} \in \{25, 50, 100, 250, 500, 1000, 2500, 5000\}$. The heterogeneous situations are (due to their high number) replicated 5000 times for $n_{max} \in \{100, 500\}$. To reflect reality with different sample lengths, they are set as $n_1 = n_6 = n_{max}$, $n_2 = n_5 = \lfloor 0.85\, n_{max} \rfloor$, and $n_3 = n_4 = \lfloor 0.7\, n_{max} \rfloor$. Not completely observed sites (number 2 to 5) are truncated in the beginning (we refer to scheme (4.2)).

The test statistic (4.26) is calculated with $\hat{\Gamma}$ being constructed plugging in either the nonparametric $\hat{\Sigma}$ or the modified $\hat{\Sigma}_m$ (see Section 4.3.1 and 4.3.2) and by using either L-, TL(0,1)-, or TL(0,2)-moments to derive parameter estimations. If, in case of the modified estimator, the covariance estimate is not positive definite, the test is considered to not reject the null hypothesis. The proportion of test rejections, i.e. the proportion of p-values smaller than $\alpha = 5\%$, is interpreted as empirical error in the homogeneous case and as empirical power in the heterogeneous case.

Figure 4.4 gives the error rates (i.e. the proportion of test rejections in homogeneous situations) for increasing samples lengths and different estimators. The first and most important observation is that although most tests approach the nominal level of 5% for increasing n_{max} (with exception of the L-moment based estimation in the highly skewed setting), the tests are very liberal for small or medium sample lengths ($n_{max} \leq 100$). A test based on the nonparametric covariance estimator needs very high sample lengths ($n_{max} \geq 1000$) to have an error of reasonable size. For tests using the modified covariance estimator the results are better but still at least 100 observations per station are required to yield acceptable error rates below 10%. The TL(0,0)-based estimator seems not to work well in skewed situations. Comparing TL(0,1)- and TL(0,2)-based procedures, the former seems to be favourable for small to medium sized sample lengths in nearly all situations.

These results are a bit unexpected since we found the modified estimator to have larger distances to the true covariance matrix in the empirical analysis in Section 4.3.3. However, this is a different situation using the inverse of a non-linear transformation

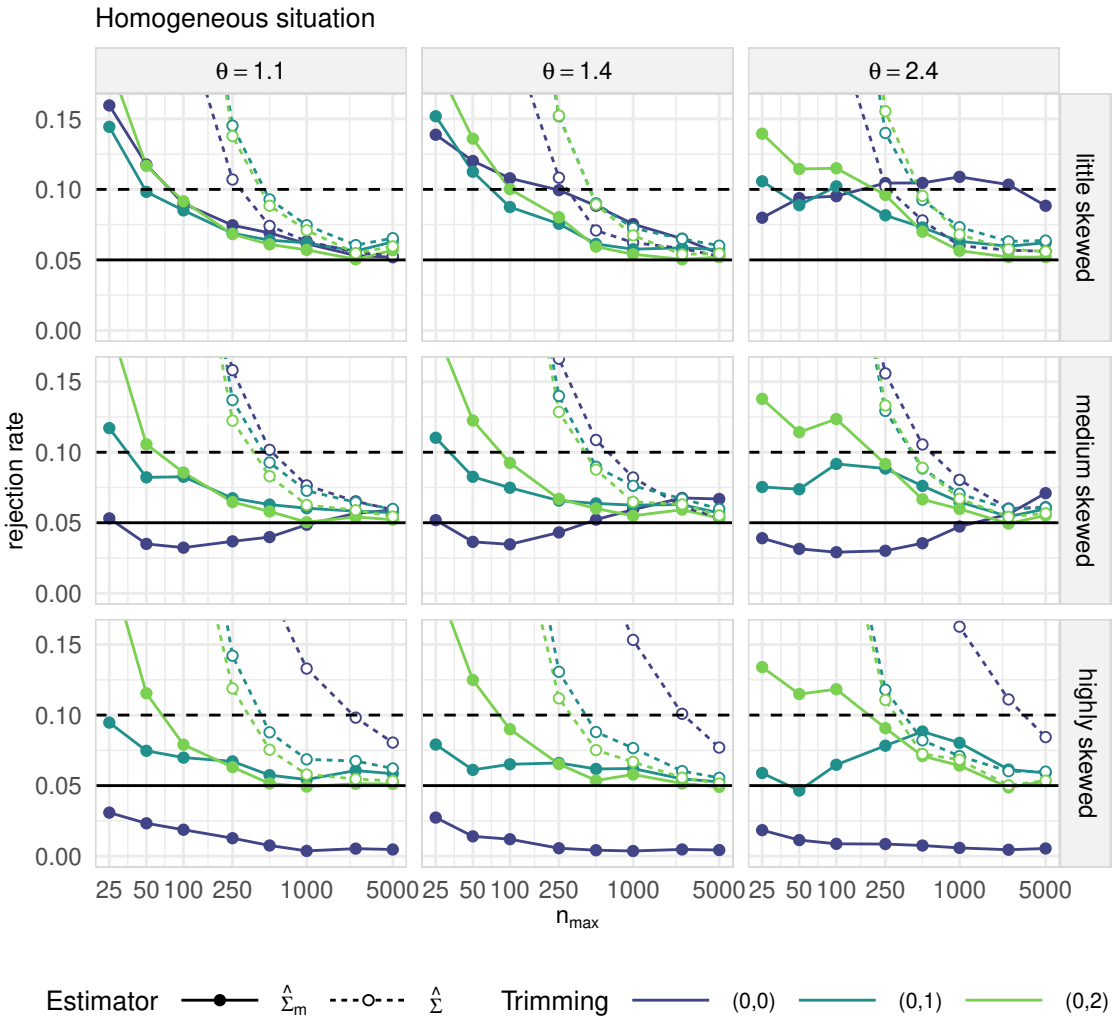

Figure 4.4: Error rates of the proposed test for increasing sample length.

of that matrix which results in better test properties for the modified estimator in the examined settings.

Considering these results, the analysis of the power of the test is presented for the procedure using the modified covariance estimator and TL(0,1)-based estimation. Figure 4.5 presents the rejection rates, depending on outlier shape $\tilde{\xi}$ and outlier scale $\tilde{\sigma}$, in different situations for $n_{max} = 100$. In the centre of each panel, i.e. in situations in which the varying station is concordant or almost concordant to the remaining stations, the rejection rates are relatively low, between 6.1% and around 12%. The further away the varying station is from the centre, the higher the proportion of test rejections.

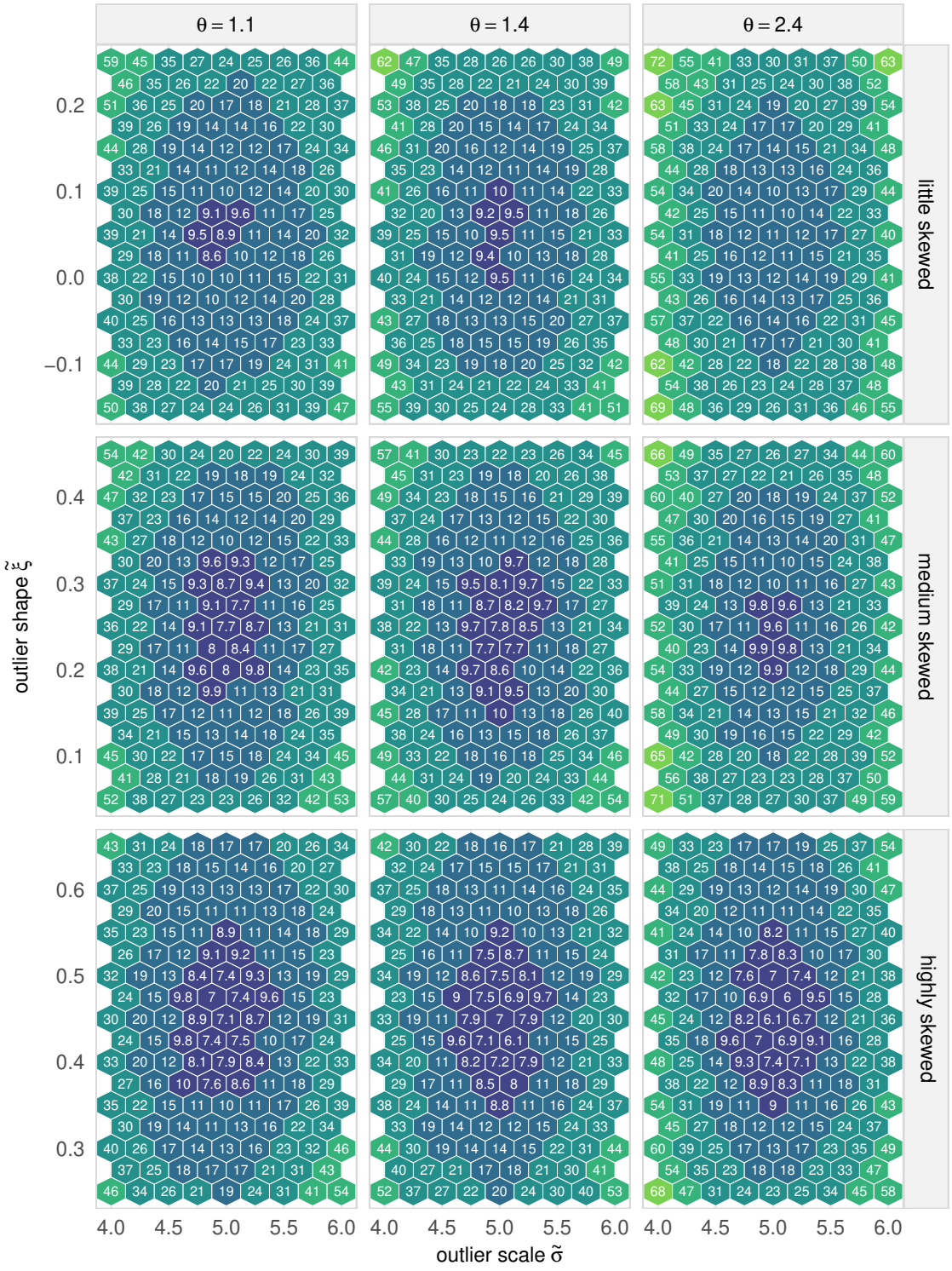

Figure 4.5: Rejection rates of the proposed test using the modified covariance estimator and TL(0,1)-moments, $n_{max} = 100$.

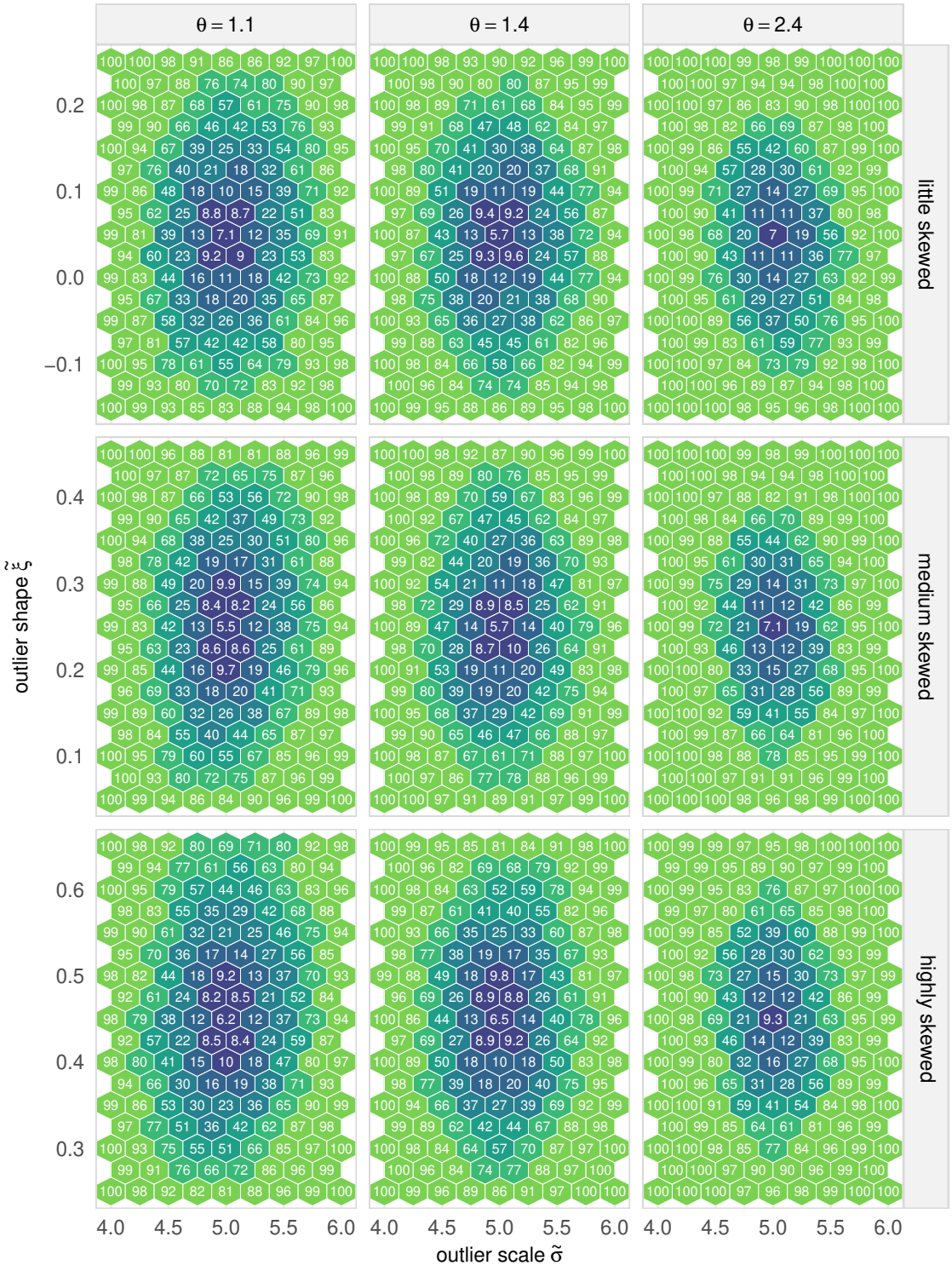

Figure 4.6: Rejection rates of the proposed test using the modified covariance estimator and TL(0,1)-moments, $n_{max} = 500$.

Figure 4.6 presents the same plot for $n_{max} = 500$. The type-I-error rates are at a lower level between 5.5% and 9.3% and heterogeneous situations are detected much earlier. These figures show that, despite being too liberal especially for small sample lengths, the test works as intended by diagnosing deviations from the common centre of the group.

4.5 Conclusions

In this chapter a multivariate limit theorem for PWM estimators in a regional setting is introduced. By application of the delta rule, the theorem regarding PWM estimators can be extended to describe the asymptotic limit distributions of TL-moment estimators and derived parameter and quantile estimators. In contrast to other approaches, our theory can be applied to situations with multiple sites described by different distributions and with intersite dependence. This allows for a wide variety of possible applications in regional analysis like the presented procedure for testing homogeneity.

An empirical estimator for the limiting variance matrix and, as an alternative, a modification that plugs in local parametric expressions, are given. Both estimators are analysed afterwards in a simulation study consisting of different record lengths and GEV shapes. This indicates that both options are consistent estimators of the PWM estimator's limiting variance for small to moderately large shapes. For higher shapes ($\xi \geq 0.5$) both estimators are not consistent anymore. These studies were also conducted for the estimation of the asymptotic covariance matrix of TL-moments and GEV parameter estimates using different asymmetrical trimming options. This showed that estimation of the limiting variance of trimmed TL-moments or parameter estimates based on them is consistent even for high shapes.

Based on the presented theorem, a homogeneity test was constructed that analyses whether the Index Flood model is appropriate for a specific group of observations under the assumption that we have GEV marginals. Therefore the Index Flood model was transferred to constraints regarding GEV parameters and those are tested using a Wald-type test statistic. The procedure was examined in a simulation study which indicates that the nominal significance level is achieved asymptotically. However, a very high record length is needed to ensure an error probability of first kind close to the nominal level if the empirical estimator is used. For the modified estimator and using TL(0,1)-based statistics a record length of $n = 100$ or higher mostly leads to acceptable error rates. In this configuration the test can detect violations of the Index Flood model.

In conclusion, these simulations indicate that the presented theorem and methods can be an option if record lengths are large enough. In many applications in regional flood frequency analysis this may be problematic, since it is mainly used in short data

situations. Hence, for testing the assumption of homogeneity in such applications the generalized Hosking-Wallis procedure presented in Chapter 3 is recommended. The presented theory, however, could be useful in more data-rich situations, e.g. when analysing monthly rainfall or when modelling financial risks. In these cases, an advantage over the bootstrap-based Hosking-Wallis procedure would be the much faster computation time.

5 Penalized Maximum-Likelihood estimation

The main focus of the following chapter is the estimation of distribution parameters in order to model the flood distribution at one or multiple sites. In hydrology and in particular in the Index Flood procedure, the most common method to estimate parameters is the L-moment method, described in Section 2.4.2. One reason is that Maximum-Likelihood estimators often give very volatile estimates in small sample scenarios that are common when analysing annual maxima of flood events. As an alternative to both, in this chapter the penalized Maximum-Likelihood estimator (see Section 2.4.4) is applied to flood frequency analysis. Its usefulness in two different aspects of information expansion (expert knowledge, seasonal information) and especially in regional flood frequency analysis is illustrated. For the latter the Index Flood model is adapted and used in a regularization framework.

The chapter is based on the work "Penalized Quasi-Maximum-Likelihood Estimation for Extreme Value Models with Application to Flood Frequency Analysis" (Bücher, Lilienthal, Kinsvater, and Fried, 2019) which is submitted but not yet published.

5.1 Introduction

As described in Section 2.2.5, a widely accepted framework for the analysis of annual maxima, or more generally of block maxima, relies on the assumption that the cumulative distribution function (c.d.f.) F belongs to the family of generalized extreme value distributions with parameters $\vartheta = (\mu, \sigma, \xi)' \in \Theta = \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$ called location, scale, and shape, respectively. Throughout this chapter, we will denote the GEV c.d.f. as G_ϑ and the corresponding density as g_ϑ. Being particularly interested in high quantiles (i.e. the right tail), note that the GEV family can handle a wide variety of right tail behaviour, with limited right tails for $\xi < 0$, exponential tails for $\xi = 0$, and arbitrarily heavy tails for $\xi > 0$. The drawback of this flexibility shows up in the estimation of the parameter vector ϑ, particularly by a volatile estimation of the shape ξ resulting in a volatile quantile estimate. Different attempts have been made to reduce the estimation uncertainty for such estimation problems, in statistics for extremes in general and particularly in flood frequency analysis.

For instance, probability-weighted moments or L-moments (see Sections 2.4.1 and 2.4.2) have been proposed as alternatives to moment or Maximum-Likelihood estimators. Indeed, the former show a superior performance in typical small sample cases (Hosking, Wallis, and Wood, 1985), which has been mainly attributed to their restricted parameter space (Coles and Dixon, 1999).

Alternative approaches are based on reducing the model complexity, for instance, by restricting oneself to the two-parametric sub-family with a predefined shape like $\xi = 0$, resulting in the location-scale Gumbel model (Lu and Stedinger, 1992). The shortcoming of this approach is that only tails of one specific form (exponential if $\xi = 0$) are taken into account, which is not appropriate for many practical applications, in particular those that are primarily interested in the tails.

Finally, several attempts have been made to include additional sources of information into flood analyses. Regional methods are the main focus of this dissertation, but in this chapter we will also address seasonal methods. While regional methods try to reduce the variability of a quantile estimator at a specific site by taking observations from other sites into account, seasonal methods do not only use time series on an annual scale but consider, say, monthly maximal flows, allowing for seasonal variability (Baratti et al., 2012; Buishand and Demaré, 1990; Waylen and Woo, 1982).

The two last-mentioned approaches, the reduction of local model complexity and the homogenization of a collection of stations, can be considered in the framework of regularization. Let F_n denote the empirical c.d.f. of the data $X_1, \ldots, X_n$ and suppose that one aims at minimizing some risk measure $R(\vartheta;\ F)$ with respect to a model parameter $\vartheta \in \Theta$, where the c.d.f. F of the data is unknown. As for instance demonstrated in Vapnik (2000) by a simple regression example, minimizing the empirical counterpart $R(\vartheta;\ F_n)$ over the whole parameter space Θ is typically not the best strategy in finite samples. A more sophisticated and often preferable strategy (reducing possible overfitting) takes an additional penalty term $\Omega(\vartheta) \geq 0$ into account, which can be interpreted as measuring model complexity or representing a-priori expert knowledge:

$$\hat{\vartheta}_\Omega = \operatorname*{arg\,min}_{\vartheta \in \Theta} R(\vartheta;\ F_n) + \Omega(\vartheta). \tag{5.1}$$

The idea of accounting for model complexity in the estimation of GEV parameters is not new. In fact, using the so-called cross-entropy risk $R(\vartheta;\ F) = -E[\log g_\vartheta(X)]$, where g_ϑ is the density of a GEV with parameter ϑ and X is a generic random variable with $F(x) = P(X \leq x)$, then minimizing the empirical cross-entropy $R(\vartheta;\ F_n) = -n^{-1} \sum_{i=1}^n \log g_\vartheta(X_i)$ with respect to ϑ is equivalent to Maximum-Likelihood (ML) estimation. When including a non-zero penalty, the resulting estimators are therefore called penalized Maximum-Likelihood (PML) estimators. Coles and Dixon (1999) and Martins and Stedinger (2000) propose two slightly

different estimators of GEV parameters of this particular form (5.1), with a regularizer $\Omega(\vartheta)$ depending only on the shape ξ, thus aiming at ruling out unusual values of the shape parameter. However, no asymptotic theory is provided and it is even unknown whether (and under what conditions) the estimators are consistent. The same is true for related approaches in extremes for hydrology, e.g. the PML estimators in Song et al. (2018) proposed for non-stationary Pearson-type 3 distributions. It is worthwhile to mention that, due to the fact that the support of the GEV distribution depends on the parameter, even the asymptotic analysis of the classical ML estimator is actually quite complicated, and has just recently been fully worked out in Bücher and Segers (2017) and Dombry and Ferreira (2017).

The main focus of this chapter is the issue of choosing a suitable penalizing function Ω for some non-trivial problems with the prime example being regional flood frequency analysis based on the Index Flood assumption. Moreover, a data-adaptive approach to select a tuning parameter that controls the level of penalization in finite samples is provided. We illustrate that the proposed method performs very well compared to existing standard methods in an extensive simulation study and that it yields easily interpretable results in a case study. In Bücher, Lilienthal, Kinsvater, and Fried (2019) also asymptotic results in a quite general multivariate setting are presented that were proven by Axel Bücher. These are given in this dissertation in short.

The remainder of this chapter is organized as follows: Section 5.2 provides illustrations of possible applications of penalized (quasi) ML estimators in flood frequency analysis. Section 5.3 presents theoretical properties of such estimators in a general multivariate framework with GEV marginals. The degree of penalization is controlled by a hyperparameter, and the problem of its selection is treated in Section 5.4. An extensive simulation study in Section 5.5 compares the Index Flood penalization to estimators common in hydrology. A case study in Section 5.6 illustrates the applicability to hydrological data.

5.2 Regularization in flood frequency analysis

Within this section, we illustrate the broad applicability of penalized (quasi) maximum likelihood techniques in flood frequency analysis. For illustrative purposes, we start with a simple approach based on penalizing unusual GEV shape parameters in a univariate setting. Then we discuss two possibilities to include additional data by joint estimation of a set of stations using an Index Flood like penalization (adding regional information) and by using monthly instead of annual maxima (adding seasonal information).

5.2.1 Simple shape parameter penalization

Let $X_1, \ldots, X_n$ represent the data, consisting of independent and identically distributed observations with unknown distribution function $F(x) = P(X_i \leq x)$. We are interested in the estimation of a high quantile $q = F^{-1}(p)$ from a rather small sample length n. Often enough, flood frequency analysts need to deal with $p \geq 0.99$ and $n \leq 50$.

Restriction to a 2-parametric sub-family of the GEV-model, like the Gumbel or a GEV distribution with a fixed shape parameter ξ_c, reduces the variance of a respective quantile estimator but possibly leads to a large bias. As a first application, we use penalization as an alternative to such a strict reduction of model complexity. More precisely, suppose that an expert claims that the true shape parameter ξ_0 is close to $\xi_c = 0.2$. This knowledge may be reflected by choosing a penalty term of the form $\Omega_\lambda(\vartheta) = \lambda(\xi - \xi_c)^2$ with hyperparameter $\lambda \geq 0$ reflecting our confidence in this prior belief and by considering the PML estimator

$$\hat{\vartheta}_\lambda \in \arg\max_{\vartheta \in \Theta} \sum_{i=1}^{n} \log g_\vartheta(X_i) - \lambda \, (\xi - \xi_c)^2. \tag{5.2}$$

If the expert was perfectly sure that actually $\xi_0 = \xi_c$ holds, we should choose $\lambda = \infty$ and thus enforce an estimate of ϑ with third component $\hat{\xi} = \xi_c$ (using the convention

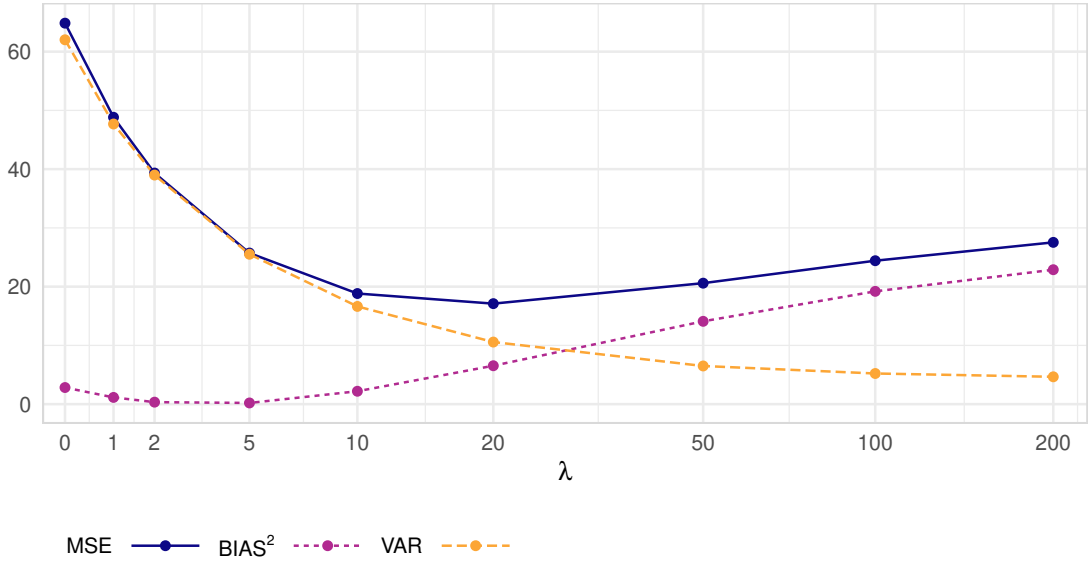

Figure 5.1: Empirical MSE, squared bias, and variance of a 99%-quantile estimate from PML parameter estimation with penalized deviations of the shape parameter to a ξ_c. Even though $\xi_c = 0.2$ does not correspond to the true $\xi_0 = 0.4$, regularization is still beneficial ($n = 50$).

that $\infty \cdot 0 = 0$). Alternatively, we can select any value $0 \leq \lambda < \infty$ reflecting the uncertainty in the expert's prior information, with $\lambda = 0$ leading to the ordinary ML estimator.

For further insight, we present the outcome of a small simulation experiment. Figure 5.1 depicts common empirical performance measures of estimators $\hat{q}_\lambda = G^{-1}_{\hat{\vartheta}_\lambda}(0.99)$ with $\hat{\vartheta}_\lambda$ from equation (5.2), $\xi_c = 0.2$, and increasing values of λ. The measures are computed from 10 000 independent samples of size $n = 50$, each with true parameter $\vartheta_0 = (\mu_0, \sigma_0, \xi_0)' = (2,\ 1,\ 0.4)'$. Note that our prior information reflected by $\Omega_\lambda(\vartheta) = \lambda(\xi - 0.2)^2$ is not centred around the true value of $\xi_0 = 0.4$. The (almost) unbiasedness of the ML estimator (for $\lambda = 0$) is outweighed by larger variability. Increasing the value of λ can be interpreted as trading variance for bias. In this example, the estimator $\hat{q}_\lambda$ with $\lambda = 20$ performs best in terms of empirical mean squared error. In fact, every value $\lambda > 0$ leads to better performance than $\lambda = 0$ (the empirical mean squared error for $\lambda = \infty$ is 31.16) although $\xi_0 = 0.4$ is not close to our a-priori guess $\xi_c = 0.2$. This can be explained by the strong imbalance between a small sample length and a comparably high quantile. Also note that neither $\lambda = 0$ nor $\lambda = \infty$ are optimal. Selecting a good value of λ is the most critical task in application of this method and will be discussed later on.

5.2.2 Penalization inspired by the Index Flood model

Practitioners typically have data from several stations in the same region available. The local record lengths often vary substantially over the stations, typically with different start times (times of gauge installations) and a common end time. The data scheme can hence be written as

$$\left.\begin{array}{rrrrrrr} X_{1;1}, & X_{2;1}, & X_{3;1}, & X_{4;1}, & X_{5;1}, & \ldots, & X_{n;1} \sim F_1 \\ X_{a_2+1;2}, & X_{a_2+2;2}, & X_{a_2+3;2}, & \ldots, & X_{n;2} & \sim F_2 \\ & & \ddots & & \vdots & & \\ X_{a_d+1;d}, & X_{a_d+2;d}, & \ldots, & X_{n;d} & \sim F_d \end{array}\right\} \begin{array}{l} \text{observations} \\ \text{from } d \text{ sites} \end{array} \tag{5.3}$$

with n denoting the end time, $a_j + 1$, $0 \leq a_j \leq n$, the start times, and $n_j = n - a_j$ the record lengths. For ease of presentation, we arranged the samples in (5.3) such that the first station corresponds to that with the full sample length of $n_1 = n$.

We assume that the random vectors $\boldsymbol{X}_i = (X_{i;1}, \ldots, X_{i;d})'$, consisting of possibly partially observed values for the different years, are independent and identically distributed with GEV margins $F_j \in \{G_{\vartheta_j} :\ \vartheta_j = (\mu_j, \sigma_j, \xi_j)' \in \Theta\}$ for all $j = 1, \ldots, d$. Note that we neither assume the d components $X_{i;j}$ for the same time point i to be independent nor that we impose a specific model for the spatial dependence.

Recall the Index Flood assumption from equation (2.89). If we additionally assume that the common distribution is a member of the GEV family, i.e. $H = G_{\vartheta_0}$ for

certain parameters $\vartheta_0 = (\mu_0, \sigma_0, \xi_0)' \in \Theta$, the hypothesis $\mathcal{H}_{0,\mathrm{IF}}$ is equivalent to $\vartheta_j = (\mu_j, \sigma_j, \xi_j)'$ satisfying

$$\frac{\mu_1}{\sigma_1} = \ldots = \frac{\mu_d}{\sigma_d} = \delta_0 \quad \text{and} \quad \xi_1 = \ldots = \xi_d = \xi_0 \quad \text{for some } \delta_0, \xi_0 \in \mathbb{R}. \tag{5.4}$$

A straightforward combination of the Index Flood principle and penalization techniques suggests to penalize deviations between $\delta_j = \mu_j/\sigma_j$ and δ_0 and between ξ_j and ξ_0. Because δ_0 and ξ_0 are not known, we replace them by approximations δ_c and ξ_c, which can be chosen as weighted means, $\delta_c = \sum_{j=1}^d w_j \delta_j$ and $\xi_c = \sum_{j=1}^d w_j \xi_j$ with weights $w_j = n_j / \sum_{j'=1}^d n_{j'}$, or using a-priori knowledge. A suitable penalization is given by

$$\Omega_{\boldsymbol{\lambda}}(\boldsymbol{\vartheta}) = ((\delta_1 - \delta_c)^2, \ldots, (\delta_d - \delta_c)^2, (\xi_1 - \xi_c)^2, \ldots, (\xi_d - \xi_c)^2)\boldsymbol{\lambda}, \tag{5.5}$$

with hyperparameter $\boldsymbol{\lambda} = (\lambda_{11}, \ldots, \lambda_{1d}, \lambda_{21}, \ldots, \lambda_{2d})' \in [0, \infty]^{2d}$. This results in the penalized quasi ML estimator

$$\hat{\boldsymbol{\vartheta}}_{\boldsymbol{\lambda}} \in \operatorname*{arg\,max}_{\boldsymbol{\vartheta} \in \Theta^d} \sum_{j=1}^{d} \sum_{i=a_j+1}^{n} \log g_{\vartheta_j}(X_{i;j}) - \sum_{j=1}^{d} \left\{ \lambda_{1j} \, (\delta_j - \delta_c)^2 + \lambda_{2j} \, (\xi_j - \xi_c)^2 \right\}. \tag{5.6}$$

The term quasi refers to the fact that the likelihood is derived under the additional assumption of spatially independent observations which is actually not necessary for consistency of the estimator, see Section 5.3.

In this application, increasing the hyperparameters $\boldsymbol{\lambda}$ reflects stronger belief in $\xi_j \approx \xi_c$ and $\delta_j \approx \delta_c$ for all $j = 1, \ldots, d$ or weaker certainty about the quality of the local estimator. In fact, both options of regular flood frequency analysis, calculation of local or regional estimates, are included as special cases when choosing $\lambda = 0$ or $\lambda \to \infty$, respectively. The elegance of this approach lies in the fact that strange local estimates are effectively ruled out without relying completely on the restrictive application of the Index Flood model or an arbitrary initial guess. The performance of this estimator in finite samples will be analysed in detail by simulations in Section 5.5, and by a real-data application in Section 5.6.

5.2.3 Penalization inspired by seasonal smoothness assumptions

An analysis that considers seasonal or monthly instead of annual maxima allows to expand the available information and can improve the estimation of very high quantiles because, due to different flood origins (like snow melt or heavy rainfall), stochastic characteristics vary over the course of a year. At a particular station, the observed monthly maximal flows are denoted by $M_1^{(m)}, \ldots, M_n^{(m)} \sim F^{(m)}$, $m =$

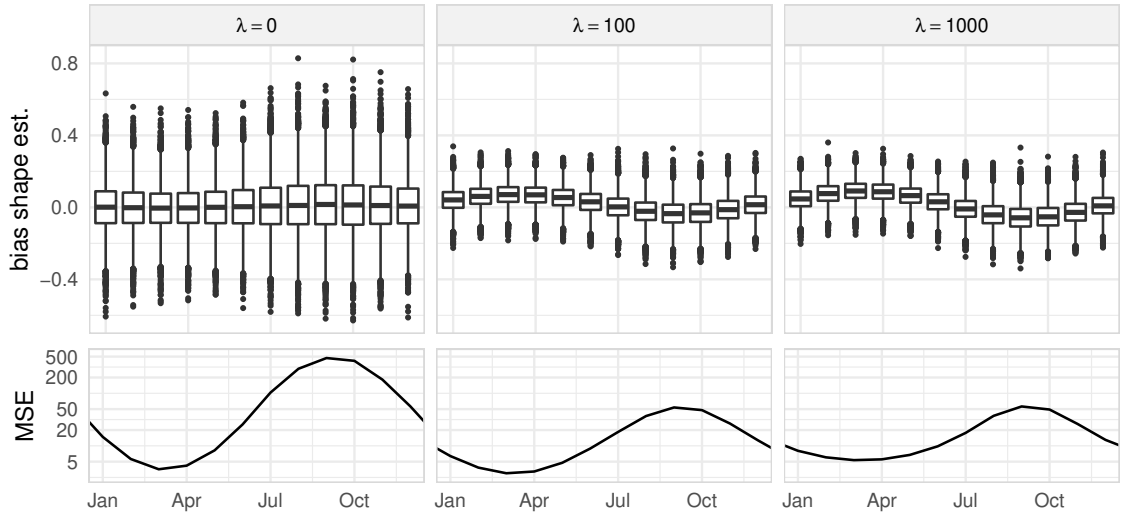

Figure 5.2: Error of the shape parameter estimation (top) and its empirical MSE (bottom). $\lambda = 0$ leads to the smallest bias but $\lambda = 100$ to the smallest MSE.

$1, \ldots, 12$. Under the assumption of independence of the monthly maxima, quantiles of the annual maximal flows $X_i = \max\left\{M_i^{(1)}, \ldots, M_i^{(12)}\right\}$ are given by

$$F^{-1}(p) = \left(F^{(1)} \cdot \ldots \cdot F^{(12)}\right)^{-1}(p). \tag{5.7}$$

It is reasonable to assume that the distributions of the monthly maxima $F^{(m)}$ are given by GEV distributions G_{ϑ_m} with parameters $\vartheta_m = (\mu_m, \sigma_m, \xi_m)'$, $m = 1, \ldots, 12$. We estimate the vector of unknown model parameters $\boldsymbol{\vartheta} = (\vartheta_1', \ldots, \vartheta_{12}')'$ by

$$\hat{\boldsymbol{\vartheta}}_\lambda \in \operatorname*{arg\,max}_{\boldsymbol{\vartheta} \in \Theta^{12}} \sum_{m=1}^{12} \sum_{i=1}^{n} \log g_{\vartheta_m}\left(M_i^{(m)}\right) - \Omega_\lambda(\boldsymbol{\vartheta}), \tag{5.8}$$

using a penalty Ω_λ that prefers gradually changing shape parameters $\xi_1, \ldots, \xi_{12}$ over the year. More specifically, we set

$$\Omega_\lambda(\boldsymbol{\vartheta}) = \Omega_\lambda(\xi_1, \ldots, \xi_{12}) = \lambda\left\{\sum_{m=1}^{11} (\xi_m - \xi_{m+1})^2 + (\xi_{12} - \xi_1)^2\right\}, \tag{5.9}$$

which implies a natural periodicity of one year. Note that we could have also incorporated similar penalties for location and scale parameters.

Figure 5.2 shows the outcome of a simulation experiment based on 10 000 independent samples of $n = 50$ independent GEV observations per month with $\mu_0 = 2$, $\sigma_0 = 1$ and shapes following a sinus curve $\xi_0^{(m)} = 0.35 + 0.25 \sin(m\pi/6 + 3)$, $m = 1, \ldots, 12$, with a period of one year. The boxplots illustrate the distribution of the bias of

the shape estimates for each month and different penalties $\lambda \in \{0, 100, 1000\}$. The corresponding empirical MSE are depicted below. The regular ML estimate ($\lambda = 0$) leads to the lowest bias, but trading some variance for bias, a much smaller MSE can be achieved by $\lambda = 100$. This choice also leads to the smallest mean square error of the yearly 99%-quantile estimate, calculated using equation (5.7) among the considered penalties (MSE of 226 compared to 1728 for $\lambda = 0$ and 232 for $\lambda = 1000$). An approach using only yearly maxima would have resulted in a MSE of 839, so the seasonal model yields a substantial gain in this situation.

5.2.4 Further extensions

The examples presented before assumed stationary distributions (over the years), but, due to known or unknown causes like river regulations or climate change, the assumption of stationarity is often hardly justified. PML estimators can be applied in such scenarios. An intuitive way to model a time-dependent distribution $F_t = G(\vartheta_t)$ is splitting the time span $\{1, \dots, n\}$ into b blocks for which we assume stationarity, i.e.

$$\vartheta_t = \kappa(t) = \begin{cases} \vartheta_1 = (\mu_1, \sigma_1, \xi_1)', & t \in [i_0, i_1), \\ \vdots & \vdots \\ \vartheta_b = (\mu_b, \sigma_b, \xi_b)', & t \in [i_{b-1}, i_b], \end{cases} \tag{5.10}$$

for given $1 = i_0 < i_1 < \cdots < i_{b-1} < i_b = n$. It is reasonable to penalize differences between parameters of consecutive blocks, for example $\Omega_\lambda(\boldsymbol{\vartheta}) = \lambda \sum_{j=1}^{b} (\xi_j - \xi_{j-1})^2$, possibly in addition to other penalizations. The main focus of this paper is to analyse PML estimators in the context of regionalization, and we restrict to stationarity in the following sections.

In the previous three subsections we have focused on squared distances in the penalization term. As an alternative, one could use absolute differences as in LASSO regression (Tibshirani, 1996), which leads to a built-in variable selection in regression problems by automatically setting coefficients to zero. In our applications, however, there is no particular advantage in setting individual parameters to an exact pre-described value. Throughout our simulation study described in Section 5.5, we have checked the performance of absolute differences (similar to a LASSO approach) and of a combination of absolute and squared differences (similar to the so-called elastic net) in different settings, but these choices lead to inferior empirical MSEs as compared to quadratic differences and also to higher computation times. We concentrate on quadratic differences in this work and use the BFGS algorithm, a quasi-Newton method, for the optimization of the objective function (5.2), (5.6), or (5.8), respectively, to be given in a general form in (5.11) in the next section.

5.3 Theoretical results

This section contains the theoretical results derived and proven by Axel Bücher (see Bücher, Lilienthal, Kinsvater, and Fried, 2019).

We show that the penalized quasi ML estimator exists (i.e. the maximization problem has a solution) and is consistent under fairly general conditions on the penalty. We also provide a result about the rate of consistency, which turns out to depend explicitly on the strength of penalization.

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ with $\boldsymbol{X}_i = (X_{i;1}, \ldots, X_{i;d})'$ denote an i.i.d. sequence from $\boldsymbol{X} = (X_1, \ldots, X_d)'$, a d-dimensional random vector with marginal cumulative distribution functions denoted by $F_1, \ldots, F_d$. We assume that the marginal laws are from the GEV-family, that is, there exists $\vartheta_{0j} = (\mu_{0j}, \sigma_{0j}, \xi_{0j})' \in \Theta_{-1} = \mathbb{R} \times (0, \infty) \times (-1, \infty)$ such that $F_j = G_{\vartheta_{0j}}$, for $j = 1, \ldots, d$. Note that the parameter set Θ_{-1} is restricted to $\xi_{0j} > -1$ since otherwise the classical ML estimator for the GEV parameters is not consistent (Dombry, 2015). The dependence between the coordinates of $\boldsymbol{X}$ is left unspecified.

Note that the setting of Section 5.2.2 fits into this framework, with d denoting the number of sites, as long as $a_j = 0$ for $j = 2, \ldots, d$ (the results can however be easily extended to situations with $n' = n - a_d \to \infty$). The setting of Section 5.2.3 is accomplished with $d = 12$; additionally, the coordinates of $\boldsymbol{X}$ are assumed to be stochastically independent then.

Let $\boldsymbol{\vartheta}_0 = (\vartheta_{01}', \ldots, \vartheta_{0d}')' \in \Theta_{-1}^d$ denote the stacked vector of true marginal parameters. A generic vector of marginal parameters will be denoted by $\boldsymbol{\vartheta} = (\vartheta_1', \ldots, \vartheta_d')'$ with $\vartheta_j = (\mu_j, \sigma_j, \xi_j)'$. Let $\hat{\boldsymbol{\vartheta}}$ denote any local maximum of the function

$$Q_n(\boldsymbol{\vartheta}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d} \log g_{\vartheta_j}(X_{i;j}) - \frac{1}{n} \boldsymbol{\lambda}_n' \Omega(\boldsymbol{\vartheta}) \equiv \frac{1}{n} \sum_{i=1}^{n} \ell_{\boldsymbol{\vartheta}}(\boldsymbol{X}_i) - \frac{1}{n} \boldsymbol{\lambda}_n' \Omega(\boldsymbol{\vartheta}), \tag{5.11}$$

where $\Omega : \Theta_{-1}^d \to [0, \infty)^m$ denotes an arbitrary penalty function.

The following main result shows that there always exists a strongly consistent local maximizer as soon as the smoothing parameter is of smaller order than n. Similar results have been obtained for Lasso-type estimators in a linear regression model in Knight and Fu (2000), although their results are easier to obtain due to the convexity of their criterion function.

Proposition 2 (Strong consistency) *Let K denote an arbitrarily large compact subset of Θ_{-1}^d, containing $\boldsymbol{\vartheta}_0$ in its interior. Suppose that the penalty Ω is continuous. Then, provided $\boldsymbol{\lambda} = \boldsymbol{\lambda}_n$ satisfies $\|\boldsymbol{\lambda}_n\| = o(n)$ as $n \to \infty$, any estimator $\hat{\boldsymbol{\vartheta}}_n$ such that*

$$Q_n(\hat{\boldsymbol{\vartheta}}_n) = \sup_{\boldsymbol{\vartheta} \in K} Q_n(\boldsymbol{\vartheta}), \tag{5.12}$$

such maximizers always existing, is strongly consistent for $\boldsymbol{\vartheta}_0$*, as* $n \to \infty$*.*

While the estimator is strongly consistent for any smoothing parameter of the order $o(n)$, it turns out that the rate of convergence of $\|\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\|$ to zero in fact depends on the precise order of the smoothing parameter. The following second main results shows that we obtain the usual parametric rate for $\|\boldsymbol{\lambda}_n\| = O(\sqrt{n})$ and smaller rates for $\|\boldsymbol{\lambda}_n\|$ between $n^{1/2}$ and n, asymptotically. Similar results have been obtained for Lasso-type estimators in simple linear regression models in Pötscher and Leeb (2009, Section 4). For technical reasons, we restrict ourselves to the reduced parameter set $\Theta_{-1/2} = \mathbb{R} \times (0, \infty) \times (-1/2, \infty)$, as this is the set where the GEV family is differentiable in quadratic mean and the usual ML estimator is $\sqrt{n}$-consistent and asymptotically normal, see Bücher and Segers (2017).

Proposition 3 (Rate of convergence) *Suppose that the conditions of Proposition 2 are met, with* K *denoting a compact subset of* $\Theta^d_{-1/2}$ *containing* $\boldsymbol{\vartheta}_0$ *in its interior. Additionally, let* Q *be Lipschitz-continuous on* K*. Then, as* $n \to \infty$*,*

$$\|\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\| = \begin{cases} O_P(n^{-1/2}) & \text{if } \|\boldsymbol{\lambda}_n\| = O(\sqrt{n}), \\ O_P(n^{-1/2+\kappa}) & \text{if } \|\boldsymbol{\lambda}_n\| = O(n^{1/2+\kappa}) \text{ for some } \kappa \in [0, 1/2). \end{cases} \tag{5.13}$$

An empirical illustration of the consistency statements with rate can be found in Section B in the appendix.

5.4 Hyperparameter selection

In this section strategies to select appropriate values of $\boldsymbol{\lambda}$ are discussed. We restrict attention to estimator (5.6) inspired by the Index Flood model, but similar approaches are applicable to the seasonal smoothing estimator (5.8) or the general estimator (5.12).

The first strategy is a cross-validation procedure in which λ can be selected to be identical for all sites or individually for every site. A combination of the individually and globally selected λ is given as an extension. Cross-validation strategies rely only on the observed measurements and can be unreliable if available data is scarce. Often additional data is availabe that contains information about the fit to the selected group. A possibility to use such information is presented as a last option.

5.4.1 Cross-validation

We propose a cross-validation procedure based on the empirical cross-entropy. The set of observed years, $I = \{1, \ldots, n\}$, is partitioned evenly into K subsets, $I_1, \ldots, I_K \subset I$, that do not necessarily consist of consecutive years and are chosen randomly. Let $F_n^{(k)}$ be the empirical c.d.f. of the k-th subset and let $\hat{\boldsymbol{\vartheta}}_{\boldsymbol{\lambda}}^{(-k)} = ((\hat{\vartheta}_{\boldsymbol{\lambda};1}^{(-k)})', \ldots, (\hat{\vartheta}_{\boldsymbol{\lambda};d}^{(-k)})')'$ be the estimator of $\boldsymbol{\vartheta}_0$ calculated without the data of the k-th group. Select the parameter $\boldsymbol{\lambda} \in [0, \infty]^m$ that minimizes the sum of empirical cross-entropies $R(\hat{\boldsymbol{\vartheta}}_{\boldsymbol{\lambda}}^{(-k)}, F_n^{(k)})$ over all groups, i.e.

$$\boldsymbol{\lambda}^{CV} = \underset{\boldsymbol{\lambda} \in [0,\infty]^m}{\arg\min} \sum_{k=1}^{K} R(\hat{\boldsymbol{\vartheta}}_{\boldsymbol{\lambda}}^{(-k)}, F_n^{(k)}) = \underset{\boldsymbol{\lambda} \in [0,\infty]^m}{\arg\max} \sum_{k=1}^{K} \sum_{i \in I_k} \sum_{\{j:a_j<i\}} \log g_{\hat{\boldsymbol{\vartheta}}_{\boldsymbol{\lambda};j}^{(-k)}}(X_{i;j}). \quad (5.14)$$

The much higher computational costs of a Leave-one-out cross validation using $K = n$ did not lead to a better quality of the selected hyperparameter in our experiments. Therefore, we choose $K = 10$ groups in our simulations and application.

If $\boldsymbol{\lambda}$ is high dimensional, the optimization of equation (5.14) can become very complex or even not feasible. In this case, constraints on $\boldsymbol{\lambda}$ can simplify calculations. More precisely, for some $m' \leq m$, let $\tau : [0, \infty]^{m'} \to [0, \infty]^m$ be a given fixed function. The resulting constrained estimator associated with τ is written as $\boldsymbol{\lambda}^{CV} = \tau(\boldsymbol{\lambda}_{cons}^{CV})$ with

$$\boldsymbol{\lambda}_{cons}^{CV} = \underset{\boldsymbol{\lambda} \in [0,\infty]^{m'}}{\arg\max} \sum_{k=1}^{K} \sum_{i \in I_k} \sum_{\{j:a_j<i\}} g_{\hat{\boldsymbol{\vartheta}}_{\tau(\boldsymbol{\lambda});j}^{(-k)}}(X_{i;j}). \quad (5.15)$$

The most simple constraint is equality of all hyperparameters, i.e. $\lambda_{1j} = \lambda_{2j} = \lambda$ for all $j = 1, \ldots, d$, which is achieved using $\tau(\lambda) = (\lambda, \ldots, \lambda)', \ \lambda \in [0, \infty]$. We refer to hyperparameters derived using this τ as $\boldsymbol{\lambda}_{global}^{CV}$.

Note that equality of all hyperparameters does not imply that the penalization effect is the same for sites with different record lengths. Indeed, the log-Likelihood part of equation (5.6) consists of different numbers of observations while the penalization term is independent of the observation length. Hence, the ratio between those two parts is different according to the length of records, penalizing sites with few records (relatively) more than sites with many records.

Alternatively, to have stronger differences in the penalization effect but still a feasible dimension of $\boldsymbol{\lambda}$, the constraint $\lambda_{1j} = \lambda_{2j} = \lambda_j$ for all $j = 1, \ldots, d$ can be used which is achieved by $\tau(\lambda_1, \ldots, \lambda_d) = (\lambda_1, \ldots, \lambda_d, \lambda_1, \ldots, \lambda_d)'$. We denote this selection as $\boldsymbol{\lambda}_{local}^{CV}$.

As we will see in the results of the simulation study (Section 5.5), globally selected hyperparameters tend to have high bias and low variance while individually selected hyperparameters tend to have low bias and high variance. To investigate whether

combinations of the local and global $\boldsymbol{\lambda}$ result in a better estimation, we also consider

$$\boldsymbol{\lambda}^{CV}_{comb,p} = p\boldsymbol{\lambda}^{CV}_{local} + (1-p)\boldsymbol{\lambda}^{CV}_{global},\ p \in [0,1]. \tag{5.16}$$

5.4.2 Surrogate variables

A disadvantage of cross-validation procedures is the dependence on recorded data. Short record lengths complicate the determination of appropriate regularisation factors, so it could be useful to resort to additional variables containing information on how good a site fits to a group. Because of their use as an surrogate, we will call them surrogate variables.

The idea is to adjust between the local (ordinary ML estimation) and regional estimation based on variables other than the observations that are used to derive the estimations. Those variables can be meteorological data or physiographical characteristics like mean height, mean slope, or river length, that are often already used in regional flood frequency analysis to construct groups of stations.

Such procedures are often based on distances between sites, for example in space of physiographical variables (cluster analysis procedures) or in space of canonical variables (Ouarda et al., 2001). Distances between sites and centroids of those groups can be used as surrogate variables.

To integrate surrogate variables into the PML estimate, we investigate a three-step procedure:

1. Transformation/Adjustment of the surrogate variable.
2. Identification of relevant λ-space.
3. Construction of a function that maps surrogate variables to λ values.

In the following more details are given on the different steps.

To be able to construct a reasonable mapping into the space of λ values, we assume that small absolute values of the surrogate indicate a good fit to the group while bigger absolute values argue against it. In many cases this assumption may already be fulfilled, for example when using distance-based measures. In other cases a transformation like centring or normalization may be necessary.

The second step is to identify a space $[\lambda_{lower}, \lambda_{upper}]$ that is relevant for the PML estimator, i.e. the space in which changes of λ lead to different estimations. Therefore we measure the distance from the estimated tuples $(\hat{\delta}_{\boldsymbol{\lambda},j}, \hat{\xi}_{\boldsymbol{\lambda},j})'$ using the hyperparameter $\boldsymbol{\lambda} = (\lambda, \ldots, \lambda)'$ to the regional values (δ_0, ξ_0):

$$w(\lambda) = \sum_{j=1}^{d} (\hat{\delta}_{\boldsymbol{\lambda},j} - \delta_0)^2 + (\hat{\xi}_{\boldsymbol{\lambda},j} - \xi_0)^2. \tag{5.17}$$

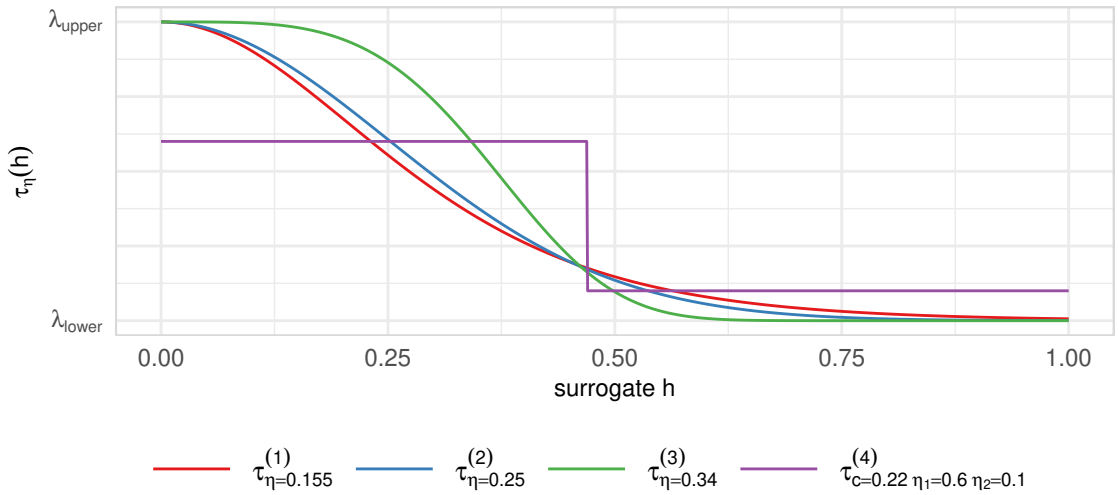

Figure 5.3: Different τ functions, $[\lambda_{lower}, \lambda_{upper}] = [0, 1]$.

The distance function w is decreasing in λ with $w(\lambda) \to 0$ for $\lambda \to \infty$. We normalize w by using $\tilde{w}(\lambda) = w(\lambda)/w(0)$ with $w(0)$ being the distance of the regular ML estimates. The boundaries λ_{lower} and λ_{upper} are chosen so that $\tilde{w}(\lambda_{lower}) = 0.01$ and $\tilde{w}(\lambda_{upper}) = 0.99$.

After we have identified the relevant space, we have to select a function $\tau : [0, \infty] \mapsto [\lambda_{lower}, \lambda_{upper}]$. Considering the assumption that small absolute values indicate a good fit to the group, the following functions are possible choices:

$$\tau_\eta^{(1)}(h) = \lambda_{lower} + (\lambda_{upper} - \lambda_{lower}) \cdot \frac{4\exp(-h/\eta)}{(1+\exp(-h/\eta))^2}, \ \eta \in (0, \infty), \tag{5.18}$$

$$\tau_\eta^{(2)}(h) = \lambda_{lower} + (\lambda_{upper} - \lambda_{lower}) \cdot \exp\left(-\frac{1}{2}\left(\frac{h}{\eta}\right)^2\right), \ \eta \in (0, \infty), \tag{5.19}$$

$$\tau_\eta^{(3)}(h) = \lambda_{lower} + (\lambda_{upper} - \lambda_{lower}) \cdot \exp\left(-\frac{1}{2}\left(\frac{h}{\eta}\right)^4\right), \ \eta \in (0, \infty), \tag{5.20}$$

$$\tau_\eta^{(4)}(h) = \eta_1 \cdot \mathbb{1}_{[0,c]}(h^2) + \eta_2 \cdot \mathbb{1}_{(c,\infty)}(h^2), \ c \in [0, \infty), \eta_1, \eta_2 \in [\lambda_{lower}, \lambda_{upper}]. \tag{5.21}$$

These functions are inspired by the densities of the logistic and Gaussian distribution and by a step function. The shape of these functions is illustrated in Figure 5.3.

Appropriate values of the parameter vector η can be chosen by a cross-validation procedure using the PML estimation with $\boldsymbol{\lambda} = \tau_\eta(\boldsymbol{h})$:

$$\eta = \arg\max_\eta \sum_{k=1}^{K} \sum_{i \in I_k} \sum_{\{j:a_j<i\}} g_{\hat{\boldsymbol{\vartheta}}^{(-k)}_{\tau_\eta(\boldsymbol{h});j}}(X_{i;j}). \tag{5.22}$$

The case study in Section 5.6 will illustrate the application of this technique to regional flood frequency analysis.

5.5 Simulation study

In this section we compare the performance of the penalized quasi ML estimator for regional flood quantile estimation with standard methods in this field.

5.5.1 Scenarios

We generate several synthetic data sets of different types and different lengths. We consider four types of heterogeneity: (I) a setting in which the sites are divided into two groups (called "groups"), (II) sites with linearly varying parameters ("linear"), (III) a setting with single sites that vary from the rest ("single"), and (IV) a setting with parameters that are arranged in a spherical fashion ("spherical"). All sites follow $\mathrm{GEV}(\mu_j, \sigma_j, \xi_j)$ distributions with the location parameter of station $j = 1, \ldots, d$, set to $\mu_j = 5 \times j$. The location-scale ratio $\delta_j = \mu_j/\sigma_j$ (and hence the scale parameter) and the shape parameter ξ_j of station j are selected using the following formulas in the four settings (I)-(IV):

$$\delta_j = 1.8 + r \times \Delta_1(j,d), \qquad \xi_j = 0.2 + 2r \times \Delta_2(j,d), \tag{5.23}$$

$$\text{with} \begin{cases} \Delta_1(j,d) = \Delta_2(j,d) = \mathrm{sign}(\frac{j-1}{d-1} - \frac{1}{2}), & \text{(I)} \\ \Delta_1(j,d) = \Delta_2(j,d) = \frac{j-1}{d-1} - \frac{1}{2}, & \text{(II)} \\ \Delta_1(j,d) = \mathbb{1}_{\{1,2\}}(j) - \mathbb{1}_{\{3,4\}}(j), \Delta_2(j,d) = \mathbb{1}_{\{1,3\}}(j) - \mathbb{1}_{\{2,4\}}(j), & \text{(III)} \\ \Delta_1(j,d) = \cos(\frac{j}{d}2\pi), \Delta_2(j,d) = \sin(\frac{j}{d}2\pi), & \text{(IV)} \end{cases}$$

and with parameter $r \in \mathbb{R}_+$ controlling the degree of heterogeneity, $\mathbb{1}$ denoting the indicator function and sign the sign function. Figure 5.4 illustrates the four settings. The central coordinate (1.8, 0.2) was chosen because it is an average coordinate in the case study presented in Section 5.6. We select record lengths between 20 and 100 observations and $d = 12$ stations. Quantile estimates of different heights are calculated from $B = 5000$ replications of each scenario using the methods described in the following section.

For the ease of a clear presentation, we only present results in spatially independent settings. Alternative simulation scenarios based on dependent data (with dependence described by a Gumbel copula) did not exhibit any fundamental qualitative differences, aside from increased variances of the estimators for all methods.

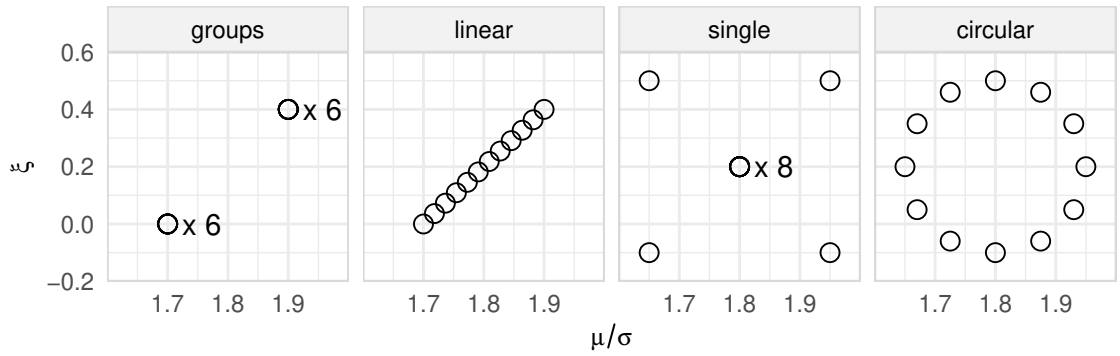

Figure 5.4: Representation of the four data settings. Sites differ group-wise, linearly, with single outliers, or in a circular fashion in terms of loc-scale-ratio and shape parameter.

5.5.2 Methods

We compare local and regional methods that are based either on Maximum-Likelihood (including our proposed penalized estimator) or L-moments.

The local L-moment method, denoted as `l-local`, calculates L-moments for each site individually and converts them to GEV parameters $\hat{\vartheta}_j^L = (\hat{\mu}_j^L, \hat{\sigma}_j^L, \hat{\xi}_j^L)', j = 1, \dots, d$. The regional L-moment method, `l-regional`, uses the famous regional flood frequency approach of Hosking and Wallis (1997), which is based on the Index Flood model given in equation (2.89). L-moments are calculated from the normalized series $X_{ij}/s_j, i = 1, \dots, n_j$, with individual index floods $s_j, j = 1, \dots, d$, being calculated as local arithmetic means. Regional L-moments are built as weighted means of these, with weights equal to the record lengths. Regional GEV parameters $\hat{\vartheta}_R = (\hat{\mu}_R, \hat{\sigma}_R, \hat{\xi}_R)'$ are calculated by converting the regional L-moments to GEV parameters. Local parameter estimates are then given through $\hat{\vartheta}_j^{LReg} = (\hat{\mu}_R\, s_j, \hat{\sigma}_R\, s_j, \hat{\xi}_R)', j = 1, \dots, d$. Note that Hosking and Wallis (1997) describe a much more comprehensive procedure, beginning with data screenings, identifications of homogeneous regions, and tests to check assumptions. We only concentrate on the data information pooling scheme in this study.

The local Maximum-Likelihood approach, denoted as `ml-local`, calculates ML estimates at each site individually by optimizing

$$\hat{\vartheta}_j^{ML} = \arg\max_{\vartheta \in \Theta} \sum_{i=a_j+1}^{n} \log g_\vartheta(X_{i;j}), \quad j = 1, \dots, d. \tag{5.24}$$

Starting values for the numerical optimization are chosen from L-moments.

Our proposed method is the penalized (quasi) ML estimator described in equation (5.6). Throughout the optimization, we fix δ_c and ξ_c using weighted means of

local L-estimates $\delta_c = n^{-1} \sum_{j=1}^{d} n_j \hat{\mu}_j^L / \hat{\sigma}_j^L$ and $\xi_c = n^{-1} \sum_{j=1}^{d} n_j \hat{\xi}_j^L$. This reduces the optimization problem to an individual maximization at each site:

$$\begin{aligned}\hat{\boldsymbol{\vartheta}}_{\boldsymbol{\lambda}}^{MLReg} &= \underset{\boldsymbol{\vartheta} \in \Theta^d}{\arg\max} \sum_{j=1}^{d} \sum_{i=a_j+1}^{n} \log g_{\vartheta_j}(X_{i;j}) - \sum_{j=1}^{d} \left(\lambda_{1j} \left(\delta_j - \delta_c\right)^2 + \lambda_{2j} \left(\xi_j - \xi_c\right)^2 \right) \\ &= \begin{pmatrix} \underset{\vartheta_1 \in \Theta}{\arg\max} \sum_{i=a_1+1}^{n} \log g_{\vartheta_1}(X_{i;1}) - \lambda_{11} \left(\delta_1 - \delta_c\right)^2 - \lambda_{21} \left(\xi_1 - \xi_c\right)^2 \\ \vdots \\ \underset{\vartheta_d \in \Theta}{\arg\max} \sum_{i=a_d+1}^{n} \log g_{\vartheta_d}(X_{i;d}) - \lambda_{1d} \left(\delta_d - \delta_c\right)^2 - \lambda_{2d} \left(\xi_d - \xi_c\right)^2 \end{pmatrix}. \end{aligned} \tag{5.25}$$

To determine appropriate hyperparameters $\boldsymbol{\lambda}$ we use cross-validation as described in Section 5.4.1 with $K = 10$ subsets. We use and compare the constrained hyperparameters $\boldsymbol{\lambda}_{global}^{CV}$, $\boldsymbol{\lambda}_{local}^{CV}$, as well as combinations $\boldsymbol{\lambda}_{comb,p}^{CV}$ with $p \in \{0.25, 0.5, 0.75\}$. This method will be denoted by `pml-gl`, `pml-ll`, `pml-cl-0.25`, `pml-cl-0.5`, or `pml-cl-0.75`, respectively.

The parameter estimates $\hat{\vartheta}$ of all methods are converted to quantile estimates by $\hat{q} = F_{\hat{\vartheta}}^{-1}(p)$, $p \in (0, 1)$.

5.5.3 Performance measures

We use common performance measures to assess the quality of the methods. Let $q_j = q_j(F_j)$ be a specific quantile of a distribution F_j and $\hat{q}_{b,j} = \hat{q}_{b,j}(\hat{\boldsymbol{\vartheta}}_{\boldsymbol{\lambda},b})$ the corresponding estimation in sample $b = 1, \ldots, B$. For each method we calculate the average empirical relative mean squared error as

$$\text{relMSE} = d^{-1} \sum_{j=1}^{d} B^{-1} \sum_{b=1}^{B} \frac{(\hat{q}_{b,j} - q_j)^2}{q_j^2}. \tag{5.26}$$

We also examine the composition of this measure by calculating the mean empirical relative squared bias and mean empirical relative variance as

$$\text{relSqBias} = d^{-1} \sum_{j=1}^{d} \left(B^{-1} \sum_{b=1}^{B} \frac{\hat{q}_{b,j} - q_j}{q_j} \right)^2, \tag{5.27}$$

$$\text{relVar} = d^{-1} \sum_{j=1}^{d} B^{-1} \sum_{b=1}^{B} \left(\frac{\hat{q}_{b,j} - B^{-1} \sum_{b'=1}^{B} \hat{q}_{b',j}}{q_j} \right)^2. \tag{5.28}$$

5.5.4 Results

Figure 5.5 displays the relative MSE of the 99%-quantile estimation for the penalized ML methods with different hyperparameters in the linear and the single setting. The two settings not displayed are qualitatively comparable to the linear one. The global $\boldsymbol{\lambda}$-selection, which selects the same hyperparameter for all sites, is the best choice in most of these situations. The relative MSE tends to get worse if a higher proportion of the local selection is used, with the only exception being the single setting with a high degree of heterogeneity. We therefore stick with $\boldsymbol{\lambda}_{global}$ for PML estimation in the following.

Figure 5.6 depicts the relative MSE of the estimates for the 99%-quantile for record lengths of $n = 80$ and two settings. These illustrations are representative also for other quantiles, record lengths (as we will see later), and the other two settings. Both L-moment based methods perform well for their intended application, the regional one for homogeneous groups (small r) and the local one for heterogeneous groups (large r), but they lack quality if they are applied to the contrary situation. The PML estimator overcomes this problem by allowing to gradually choose between local and regional estimation. Using the globally selected hyperparameter $\boldsymbol{\lambda}$, it performs best or close to the best in all these situations, independently of the degree of heterogeneity r. The local L-moment based estimation outperforms the local ML based one in all settings considered here. As already discussed in Hosking, Wallis, and Wood (1985), this is likely due to the short record length.

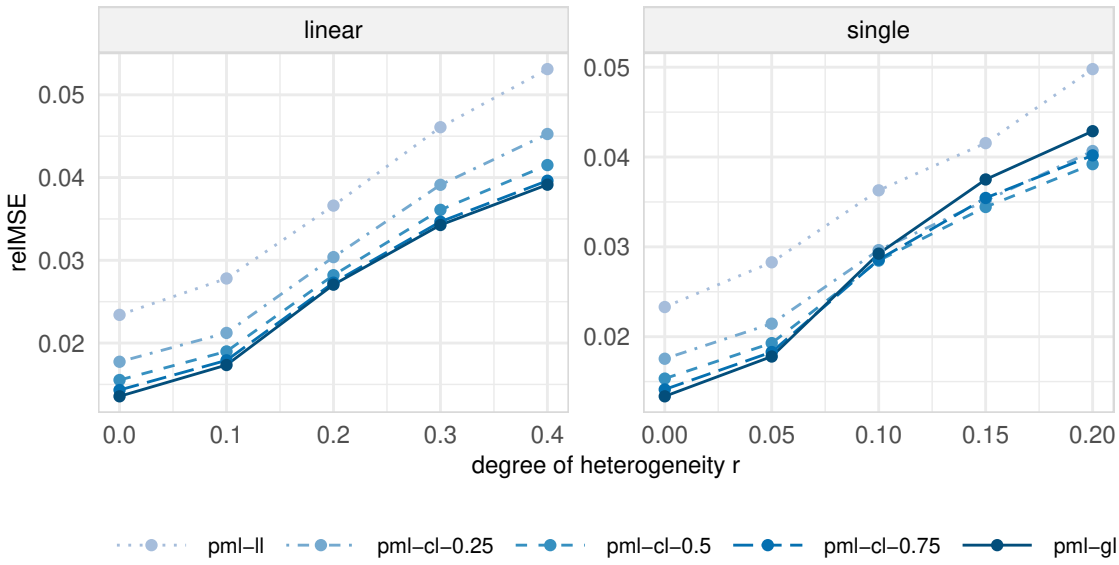

Figure 5.5: Relative MSE depending on the heterogeneity r for different $\boldsymbol{\lambda}$-selections ($n = 80$). The two settings not displayed are qualitatively comparable to the linear case.

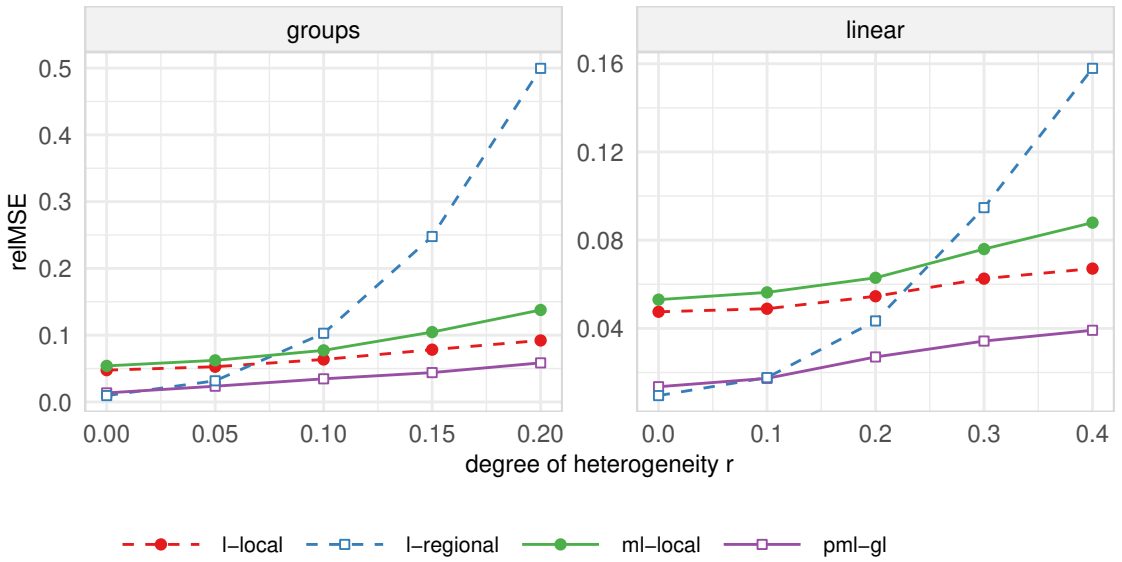

Figure 5.6: Relative MSE of the 99%-quantile estimators depending on the distance r for two settings and records of length $n = 80$.

The top panels of Figure 5.7 show the influence of the record length on the relative MSE in the linear setting for three degrees of heterogeneity. The local ML method fails for record lengths smaller than, say, $n = 40$, but it catches up with increasing record length while the L-moment estimations are not that much influenced by small record lengths. The PML estimator gives good results for record lengths larger than $n = 30$ and is nearly as good as the regional L-moment estimator in homogeneous groups ($r = 0$) and surpasses all other methods in groups of higher heterogeneity.

The bottom panels of Figure 5.7 show the MSE of the estimation of different quantiles in the linear setting for $n = 80$. The methods show stable relative performances for all quantiles and each degree of heterogeneity. For homogeneous groups, the local methods show much larger MSE than the regional ones. As opposed to the regional L-moment estimator, the PML estimator remains the best choice among these methods as the heterogeneity increases.

Figure 5.8 finally splits the MSE into the squared bias and the variance. The squared bias increases rapidly with increasing heterogeneity for the regional L-moment method, while for the other methods it is rather small as compared to the variance. The variance is substantially smaller for the regional estimators than for the local ones, with a small advantage for the regional L-moment estimator in this respect.

Overall, the PML estimator combines a small squared bias with a low variance, which results in a good relative MSE. The proposed cross-validation procedure is able to provide hyperparameters that adapt to local or regional solutions depending on the

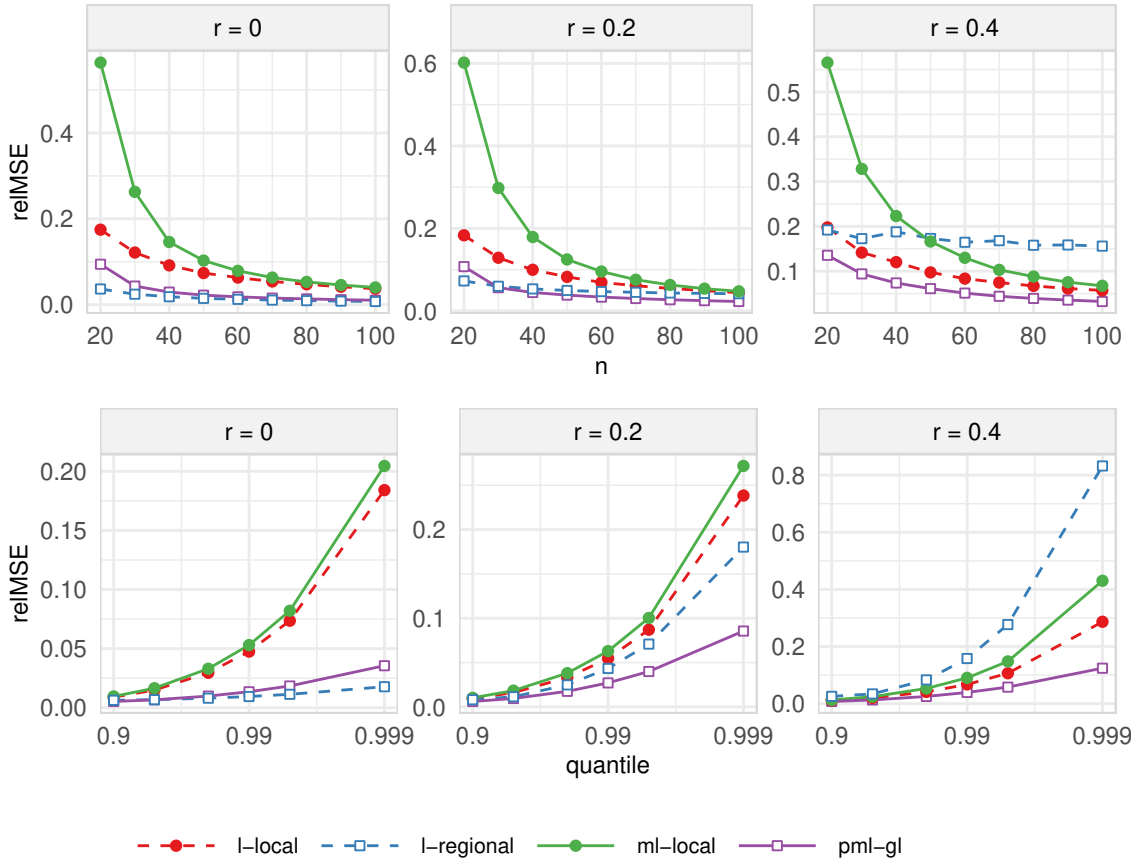

Figure 5.7: Results for the linear setting. Top: Relative MSEs for the 99%-quantile and different record lengths n. Bottom: Relative MSE as a function of the estimated quantile.

data situation and can reduce the relative mean square error substantially in this way.

5.6 Case study

We illustrate the application of our PML estimator with a case study. The data set consists of annual flood peaks at 26 stations in the Elbe river basin in Saxony, Germany, located in the north side of the Ore Mountains (with a mountaintop of 1244 m a.s.l.) and its foothills. The sites differ in mean height (from 168 m to 754 m a.s.l.) and catchment area (from around 36 km^2 to 5433 km^2) and consist of record lengths between 64 and 103 years.

Section 5.5 showed that the proposed PML estimator yields comparably good results both in homogeneous and heterogeneous situations. Because regional flood frequency

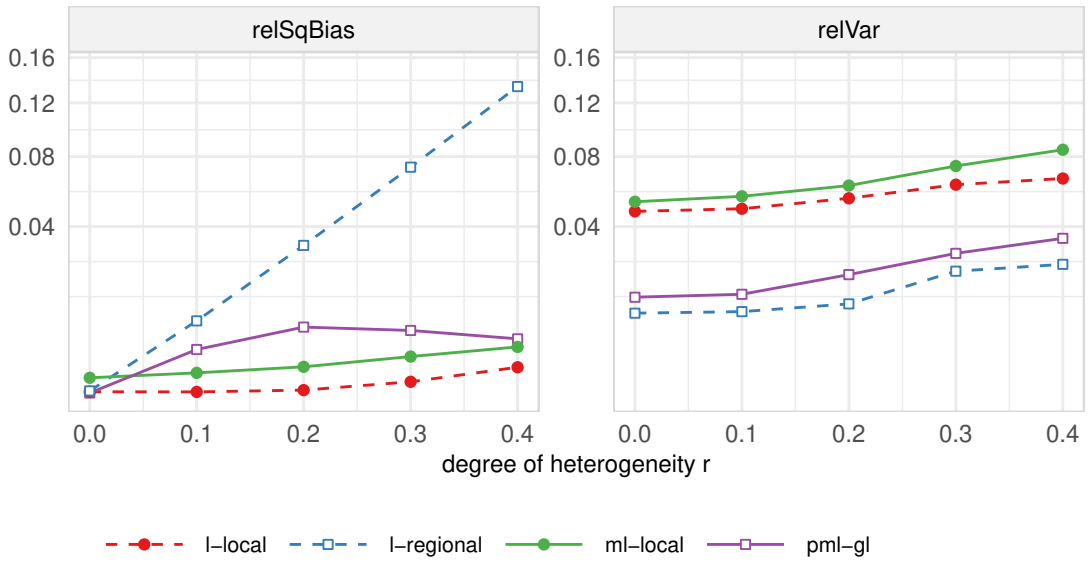

Figure 5.8: Relative squared bias and relative variance of 99%-quantile estimators depending on the heterogeneity r in the linear setting, $n = 80$.

analysis is most advantageous if the group is homogeneous, site characteristics like catchment area or mean height are used to construct two groups by application of a k-means clustering on standardized site characteristics. One group (mostly) contains sites with small catchment areas located in higher areas of the Ore Mountains, while the other group includes sites with bigger catchment areas further downstream. Smaller catchments are more strongly affected by single events and therefore often feature larger shape parameters in a GEV model. To analyse the influence of the group-building process, the estimates are calculated once with a single group containing all sites and once after division into these two groups.

The PML estimator of equation (5.25) is calculated for each group with a global $\boldsymbol{\lambda}$ (i.e. $\lambda_{1j} = \lambda_{2j} = \lambda \; \forall j = 1, \ldots, d$). Selecting δ_c and ξ_c as weighted means of the corresponding local values resulted in unstable results over varying hyperparameters. Therefore, as in the simulation study, we fix the group centres at pre-specified weighted means of local L-moment estimations throughout the optimization.

Figure 5.9 shows the estimates without and with group-building. In both plots the lines indicate all estimates obtained by the PML estimator using $\lambda \in [0, \infty)$, with the local ML estimate (i.e. $\lambda = 0$) being the most outward point of the line. The symbols indicate the estimates chosen by the crossvalidated $\boldsymbol{\lambda}^{CV}_{global}$. Without grouping, the estimates vary moderately around the centre, clearly less than ordinary ML estimates would do. With two groups, there are clear differences: the first group (filled circles) has a medium level of regionalization, resulting in estimates in the middle of the path. Regionalization is much stronger for the other group, with all estimates being

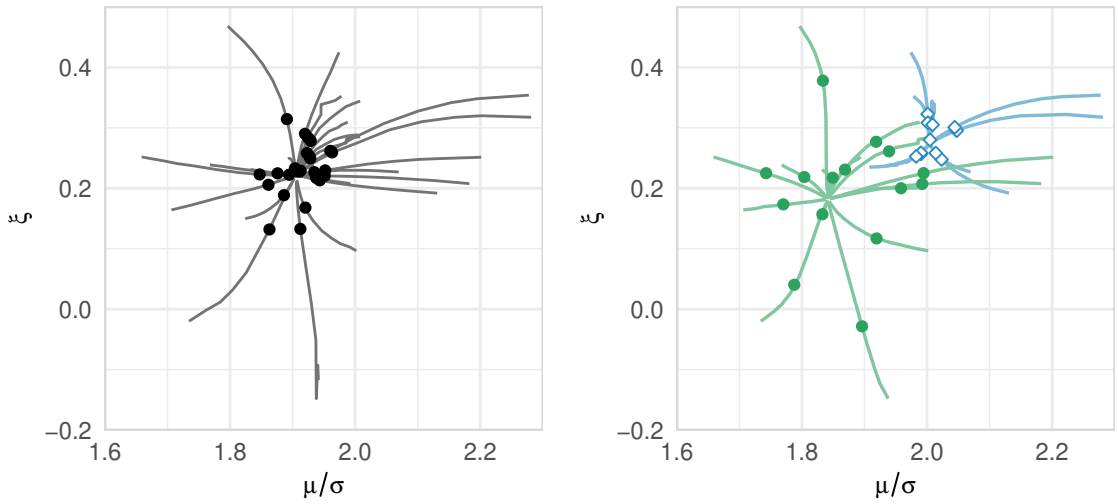

Figure 5.9: PML estimates without (left) and with group-building (right). Lines mark all possible estimates for different hyperparameters, symbols indicate the crossvalidated solution.

closer to the centre of the group.

We now use the PML estimation for both groups with a surrogate variable as described in Section 5.4.2. Because the groups were built using distances in the space of site characteristics, we use the distances to the group centres as surrogate and standardize them by their standard deviation. In both groups the PML estimator of equation (5.6) is calculated with δ_c and ξ_c being the respective group centres (i.e. the weighted means of local L-moment estimates of the respective group). The surrogate variables are mapped to the λ-space using the mapping functions of equations (5.18)–(5.21) with parameters of those functions determined by cross-validation procedures (see equation (5.22)).

We found mapping $\tau_\eta^{(2)}$ leading to the lowest empirical cross-entropy. The left panel of Figure 5.10 shows the resulting $\boldsymbol{\lambda}$-mappings for both groups dependent of the surrogate as well as the positions of the sites of both groups. The hyperparameters of the sites of one group (unfilled squares) are rather high, which results in a high degree of regionalization. The other group has both high and low parameters and therefore both high and low degrees of regionalization. The right panel of Figure 5.10 depicts the chosen estimates as before.

This use of surrogate variables to determine the degree of regionalization results in quite similar estimations as in the observation-driven cross-validation used before. We therefore believe that the use of further information to generate hyperparameters is a good alternative to the observation-based cross-validation, especially in cases of record lengths shorter than in this case study.

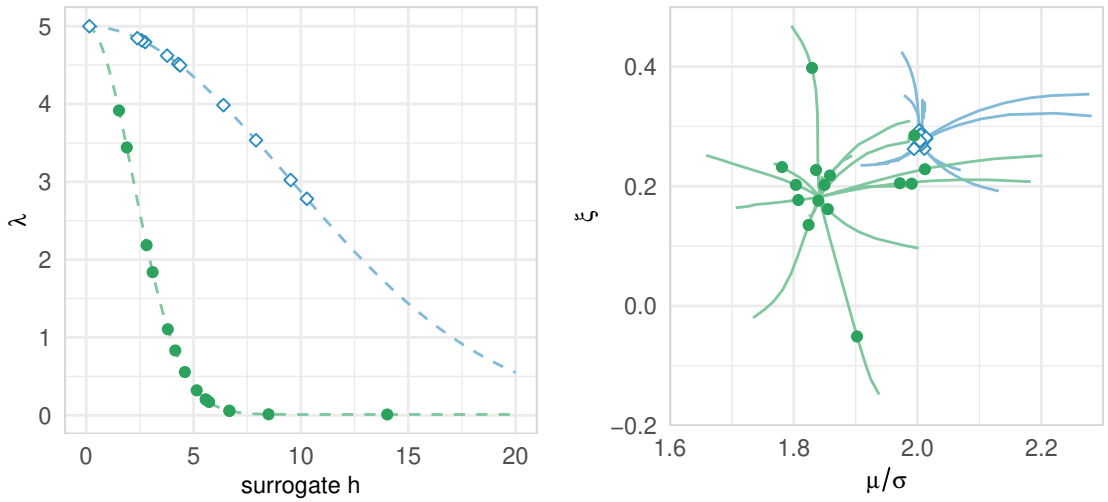

Figure 5.10: Left: Mapping between distances to centre of the groups and the crossvalidated hyperparameter. Right: Estimates using those hyperparameters.

Another promising possibility to use the PML estimator in such a setting is to modify the penalization term to penalize to all group centres simultaneously with the strength determined by the distance to the respective group. This way sites that have nearly equal fits to both groups are affected by both and not only by one of them. With $(\delta_{0g}, \xi_{0g})', g = 1, 2$, being the centre of the g-th group we change the penalizing term of equation (5.6) to

$$\begin{aligned}\boldsymbol{\lambda}'\Omega(\boldsymbol{\theta}) = \sum_{j=1}^{d}\Big(&\lambda_{1j}\left((\delta_j - \delta_{01})^2 + (\xi_j - \xi_{01})^2\right) \\ &+ \lambda_{2j}\left((\delta_j - \delta_{02})^2 + (\xi_j - \xi_{02})^2\right)\Big).\end{aligned} \tag{5.29}$$

To obtain hyperparameters, the distances to both group centres are calculated for each site and the set of distances is standardized by the standard deviation over all distances. Then the mapping functions are used to convert those distances to hyperparameters. As before, the parameters of the mapping function are deduced from a cross-validation procedure.

The left panel of Figure 5.11 depicts the combinations of the resulting hyperparameters for all 26 sites. The distribution of those parameters is quite diverse, but most sites have a high hyperparameter for one of the groups and a low one for the other. Few sites are influenced by both groups and few by neither.

In the right panel of Figure 5.11 the GEV estimates are given. The crosses indicate the centres of both groups. Nearly all estimates are spreading in the area between those two centre points now.

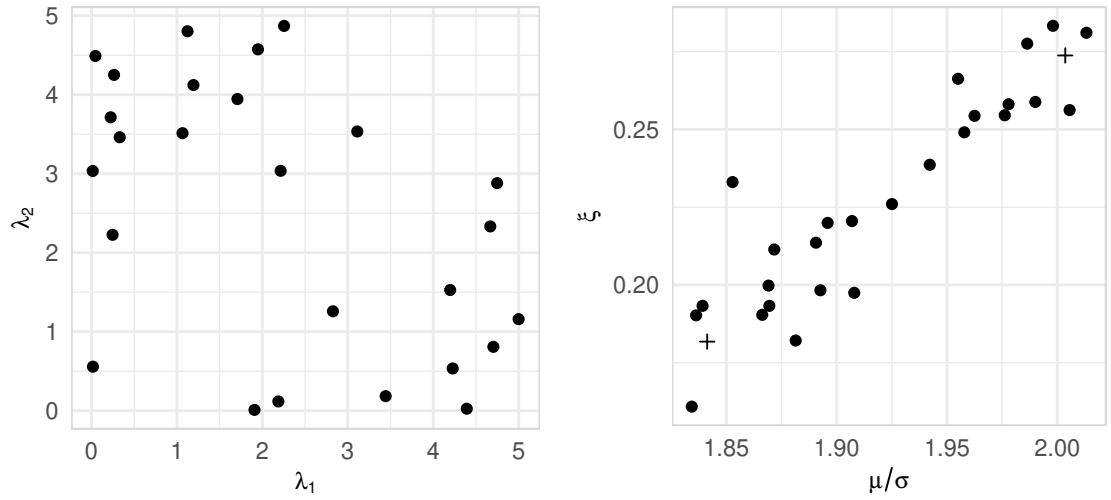

Figure 5.11: Left: Resulting combinations of hyperparameters for each site. Right: Regularized ML estimates of all sites. Crosses indicate both group centres.

In comparison to the former calculation, the sites' estimates vary about a wider margin between the two centres. This allows for individual estimates that are still stabilized by the penalization.

5.7 Conclusions

In this chapter applications of penalized quasi ML estimators in flood frequency analysis were presented. Possible applications cover simple constraints on the shape parameter, seasonal constraints, and an Index Flood like regularization for regional flood frequency analysis. The last application is of particular interest and analysed using synthetic data in a simulation study and real data in a case study. The penalization term is chosen to represent the popular Index Flood model by penalizing deviations from local parameter estimates to regional ones.

The crucial point in regularization techniques is the choice of the hyperparameters. In case of Index Flood regularization, the hyperparameters control the balance between local and regional estimates. As opposed to other methods, this enables us to adjust the degree of regionalization for the group or individually for each site. We have compared cross-validation procedures that calculate the hyperparameters with different constraints. Simulations indicate that the reduction of variance in case of short record lengths outperforms the increase of the squared bias if the same hyperparameters are chosen for all sites. An alternative to observation-based cross-validation procedures is the use of additional information that is often available in regional flood frequency analysis.

The main result of the simulation study is that the penalized ML estimator applied to regional flood frequency analysis generally provides good quantile estimates and, in particular, better estimates than other methods when there are uncertainties about the homogeneity of the group of sites. While the local and the regional L-moment estimators offer good results in situations they are actually made for, they lack quality if the situation is unclear or misspecified. The penalized estimator overcomes this problem by allowing to gradually choose between local and regional estimation.

The real-world applicability has been demonstrated for a set of 26 gauges in Germany which was divided into two groups. Even though the penalized estimator works for heterogeneous groups, we suggest to perform such a group-building process because the benefit of regional analysis is larger if the group is rather homogeneous. Using this example, we have shown how surrogate information like the distance of the stations to the centre of the group in the space of site characteristics can be used to derive hyperparameters. This might be a promising alternative to observation-based cross-validation in situations of short record lengths. We have also shown how the estimates of sites can be influenced by multiple groups at once. This could reflect reality better in which there often is no native membership to one group but different degrees of membership to multiple groups.

6 Summary

This dissertation deals with the problem of estimating the recurrence time of rare flood events, especially in the case of short data records. In such scenarios regional flood frequency analysis is used, a multi-step procedure with the goal of improving quantile estimates by pooling information across different gauging stations. Different aspects of regional flood frequency analysis using the Index Flood model are analysed, and improvements for parts of the procedure are proposed.

In group-based regional flood frequency analysis, sets of stations are built from which a similar flood distribution is assumed. In the Index Flood model, this means that the flood distributions of all stations are the same except for a site-specific scaling factor. In other words, the quantile function can be split into a site-specific factor and a common curve that is the same for all sites in the group. Because the validity of this assumption is of crucial importance for the benefits of regionalization, it is commonly checked using homogeneity tests. After possible reassignments of stations to the groups, the information of records within a group is pooled. The most popular approach for this is the calculation of regionalized L-moments that result as a weighted mean of local L-moments. Using these regionalized moments, distribution parameters can be deduced that define the common curve. In combination with the site-specific factor, quantile estimates can then be calculated.

Each of the main chapters of this dissertation focuses attention on specific steps of this procedure. The first work investigates the commonly used homogeneity testing procedure of Hosking and Wallis (1997) based on L-moments. This procedure has known drawbacks if the distribution of the observations is highly skewed or if the sites are spatially correlated. A new generalized procedure is proposed that uses copulas to model the intersite dependence and trimmed L-moments as a more robust replacement of L-moments. The influence of both changes is analysed in a comprehensive simulation study. The main outcomes are that TL(0,1)-moments work very well in the scenario of medium to high skewness and that the use of simple copula models clearly improves the results of the test, even if the true dependence is of a more complex nature. Another benefit from using asymmetrically trimmed L-moments is an increased robustness against outliers or extreme events.

The second main chapter is more technical. The asymptotic distribution of sample probability-weighted moments is described in a setting of multiple sites of different record lengths. This theory is then extended to sample TL-moments and GEV parameter and quantile estimators based on them. An estimator for the limiting

covariance matrix is given and analysed. The applicability of the theory is illustrated by the construction of a homogeneity test. This test works well if TL(0,1)-moments are used to derive GEV parameters but needs a record length of at least 100 observations at each site to give acceptable error rates.

The last main chapter deals with penalized Maximum-Likelihood estimation in flood frequency analysis. The main part provides an alternative to the information pooling scheme based on L-moments. Under the assumption of generalized extreme value distributed data, the Index Flood model is translated to restrictions on the parameter space. The penalty term of the optimization problem is then chosen to reflect those restrictions. The influence of the penalty term can be controlled by a hyperparameter, which in this case gives us the possibility to control the degree of regionalization. Using a crossvalidated hyperparameter, this leads to a procedure that automatically finds a compromise between local and regional estimation based on the data. This is especially useful in situations in which homogeneity is unclear. A simulation study indicates that this approach works nearly as good as pure regional methods if the homogeneity assumption is completely true and better than its competitors if the assumption does not hold.

Overall, this dissertation presents different approaches and improvements to steps of a group-based regionalization procedure. A special interest is the assessment of the homogeneity of a given group that is analysed with two different approaches. However, due to short record lengths or limitations in the homogeneity testing procedures, heterogeneous groups are often still hard to detect. In such situations the presented penalized Maximum-Likelihood estimator can be applied that gives comparatively good results both in homogeneous and heterogeneous scenarios. It should be stressed that application of this estimator does not supersede the group building steps. The benefit of regionalization is highest if the homogeneity assumption is fulfilled.

This dissertation only covers a small part of regionalization, so there is much potential for further research and development. The first step of group-based regionalization, the group building itself, has not been discussed in detail. Research could be invested in how to optimally delineate groups. The difficulty lies in the problem of using only site characteristics for this task. This is necessary in order to being able to dispatch unrecorded sites to groups.

In the data pooling step the penalized Maximum-Likelihood estimator offers a wide range of further extensions by adjusting the penalization term. One possibility for further research is an analysis of how strong deviations of the location-scale ratio and deviations of the shape from their respective group centres should be penalized and if that should be dependent on the quantile of interest. In the case study of Section 5.6 an approach is presented that penalizes the estimates to not only one group but to multiple groups simultaneously. This approach, similar to a fuzzy clustering procedure, seems promising and worth pursuing further.

This dissertation mainly used the assumption that the distribution of the data can be approximated by a generalized extreme value distribution and does not vary over time. The presented methodology could be adapted to other approaches including different distribution families or a time-dependent modelling allowing for trends or seasonal effects. Section 5.2.3 and 5.2.4 contain ideas on how to achieve this in the data pooling step using the penalized Maximum-Likelihood estimator. It could be worthwhile to investigate this further. Since we have focused on group-based procedures using the Index Flood model, also adaptation to other regionalization techniques seems interesting.

Bibliography

Balkema, A. A., & de Haan, L. (1974). Residual Life Time at Great Age. *The Annals of Probability*, *2*(5), 792–804. doi:10.1214/aop/1176996548

Baratti, E., Montanari, A., Castellarin, A., Salinas, J. L., Viglione, A., & Bezzi, A. (2012). Estimating the flood frequency distribution at seasonal and annual time scales. *Hydrology and Earth System Sciences*, *16*(12), 4651–4660. doi:10.5194/hess-16-4651-2012

Baumgarten, C., Christiansen, E., Naumann, S., Penn-Bressel, G., Rechenberg, J., & Walter, A.-B. (2011). *Hochwasser: Verstehen, Erkennen, Handeln!* Dessau-Roßlau: Umweltbundesamt.

Beguería, S. (2005). Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value. *Journal of Hydrology*, *303*(1–4), 215–230. doi:10.1016/J.JHYDROL.2004.07.015

Bobée, B. (1973). Sample error of T-year events commuted by fitting a Pearson type 3 distribution. *Water Resources Research*, *9*(5), 1264–1270. doi:10.1029/WR009i005p01264

Bobée, B. (1975). The Log Pearson type 3 distribution and its application in hydrology. *Water Resources Research*, *11*(5), 681–689. doi:10.1029/WR011i005p00681

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., & Roncalli, T. (2000). Copulas for Finance - A Reading Guide and Some Applications. *SSRN Electronic Journal*. doi:10.2139/ssrn.1032533

Bücher, A., Lilienthal, J., Kinsvater, P., & Fried, R. (2019). Penalized Quasi-Maximum-Likelihood Estimation for Extreme Value Models with Application to Flood Frequency Analysis. unpublished work.

Bücher, A., & Segers, J. (2017). On the maximum likelihood estimator for the Generalized Extreme-Value distribution. *Extremes*, *20*(4), 1–34. doi:10.1007/s10687-017-0292-6

Buishand, T. A., & Demaré, G. R. (1990). Estimation of the annual maximum distribution from samples of maxima in separate seasons. *Stochastic Hydrology and Hydraulics*, *4*(2), 89–103. doi:10.1007/BF01543284

Burn, D. H. (1989). Cluster Analysis as Applied to Regional Flood Frequency. *Journal of Water Resources Planning and Management*, *115*(5), 567–582. doi:10.1061/(ASCE)0733-9496(1989)115:5(567)

Burn, D. H. (1990). Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, *26*(10), 2257–2265. doi:10.1029/WR026i010p02257

Capéraà, P., Fougères, A. L., & Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, *84*(3), 567–577. doi:10.1093/biomet/84.3.567

Castellarin, A., Burn, D. H., & Brath, A. (2008). Homogeneity testing: How homogeneous do heterogeneous cross-correlated regions seem? *Journal of Hydrology*, *360*(1–4), 67–76. doi:10.1016/j.jhydrol.2008.07.014

Chebana, F., & Ouarda, T. B. M. J. (2007). Multivariate L-moment homogeneity test. *Water Resources Research*, *43*(8). doi:10.1029/2006WR005639

Chebana, F., & Ouarda, T. B. M. J. (2009). Index flood-based multivariate regional frequency analysis. *Water Resources Research*, *45*(10), 1–15. doi:10.1029/2008WR007490

Chokmani, K., & Ouarda, T. B. M. J. (2004). Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. *Water Resources Research*, *40*(12). doi:10.1029/2003WR002983

Choulakian, V., & Stephens, M. A. (2001). Goodness-of-fit tests for the generalized pareto distribution. *Technometrics*, *43*(4), 478–484. doi:10.1198/00401700152672573

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. London: Springer-Verlag.

Coles, S. G., & Dixon, M. J. (1999). Likelihood-Based Inference for Extreme Value Models. *Extremes*, *2*(1), 5–23. doi:10.1023/A:1009905222644

Cunnane, C. (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology*, *18*(3–4), 257–271. doi:10.1016/0022-1694(73)90051-6

Dalrymple, T. (1960). *Flood-frequency analyses, manual of hydrology: Part 3*. USGPO. doi:10.3133/wsp1543A

David, H. A., & Nagaraja, H. N. (2004). *Order statistics* (3rd ed.). Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons.

de Haan, L., & Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. New York, NY: Springer-Verlag.

DKKV. (2015). Das Hochwasser im Juni 2013: Bewährungsprobe für das Hochwasserrisikomanagement in Deutschland. *DKKV-Schriftenreihe Nr. 53*.

Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli*, *21*(1), 420–436. doi:10.3150/13-BEJ573

Dombry, C., & Ferreira, A. (2017). Maximum likelihood estimators based on the block maxima method.

Durante, F., & Salvadori, G. (2010). On the construction of multivariate extreme value models via copulas. *Environmetrics*, *21*(2), 143–161. doi:10.1002/env.988

Durbin, J., & Knott, M. (1972). Components of Cramer-von Mises statistics. I. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 290–307. doi:10.1111/j.2517-6161.1972.tb00908.x

Durocher, M., Mostofi Zadeh, S., Burn, D. H., & Ashkar, F. (2018). Comparison of automatic procedures for selecting flood peaks over threshold based on goodness-of-fit tests. *Hydrological Processes*, *32*(18), 2874–2887. doi:10.1002/hyp.13223

Elamir, E. A., & Seheult, A. H. (2003). Trimmed L-moments. *Computational Statistics & Data Analysis*, *43*(3), 299–314. doi:10.1016/S0167-9473(02)00250-5

Fill, H. D., & Stedinger, J. R. (1995). Homogeneity tests based upon Gumbel distribution and a critical appraisal of Dalrymple's test. *Journal of Hydrology*, *166*(1-2), 81–105. doi:10.1016/0022-1694(94)02599-7

Fischer, S., Schumann, A., & Schulte, M. (2016). Characterisation of seasonal flood types according to timescales in mixed probability distributions. *Journal of Hydrology*, *539*, 38–56. doi:10.1016/j.jhydrol.2016.05.005

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*(02), 180–190. doi:10.1017/S0305004100015681

Fohrer, N., Bormann, H., Miegel, K., Casper, M., Bronstert, A., Schumann, A., & Weiler, M. (2016). *Hydrologie* (1. Auflage). Bern: Haupt Verlag.

Genest, C., & Favre, A.-C. (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering*, *12*(4), 347–368. doi:10.1061/(ASCE)1084-0699(2007)12:4(347)

Genest, C., Kojadinovic, I., Nešlehová, J., & Yan, J. (2011). A goodness-of-fit test for bivariate extreme-value copulas. *Bernoulli*, *17*(1), 253–275. doi:10.3150/10-BEJ279

Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, *44*(2), 199–213. doi:10.1016/j.insmatheco.2007.10.005

Genest, C., & Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, *37*(5B), 2990–3022. doi:10.1214/08-AOS672

Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *The Annals of Mathematics*, *44*(3), 423–453. doi:10.2307/1968974

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, *15*(5), 1049–1054. doi:10.1029/WR015i005p01049

Gudendorf, G., & Segers, J. (2010). Extreme-Value Copulas. In P. Jaworski, F. Durante, H. W., & R. T. (Eds.), *Copula Theory and Its Applications. Lecture Notes in Statistics* (Vol. 198, pp. 127–145). doi:10.1007/978-3-642-12465-5_6

Gudendorf, G., & Segers, J. (2012). Nonparametric estimation of multivariate extreme-value copulas. *Journal of Statistical Planning and Inference*, *142*(12), 3073–3085. doi:10.1016/j.jspi.2012.05.007

Hofert, M. (2010). Construction and Sampling of Nested Archimedean Copulas. In J. P., D. F., H. W., & R. T. (Eds.), *Copula Theory and Its Applications. Lecture Notes in Statistics* (Vol. 198, pp. 147–160). doi:10.1007/978-3-642-12465-5_7

Hofert, M., Kojadinovic, I., Maechler, M., & Yan, J. (2018). *copula: Multivariate Dependence with Copulas*. R package version 0.999-19.1.

Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*(1), 105–124. doi:10.1111/j.2517-6161.1990.tb01775.x

Hosking, J. R. M. (1994). The four-parameter kappa distribution. *IBM Journal of Research and Development*, *38*(3), 251–258. doi:10.1147/rd.383.0251

Hosking, J. R. M. (2007). Some theory and practical uses of trimmed L-moments. *Journal of Statistical Planning and Inference*, *137*(9), 3024–3039. doi:10.1016/j.jspi.2006.12.002

Hosking, J. R. M., & Balakrishnan, N. (2015). A uniqueness result for L-estimators, with applications to L-moments. *Statistical Methodology*, *24*, 69–80. doi:10.1016/j.stamet.2014.08.002

Hosking, J. R. M., & Wallis, J. R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, *29*(3), 339–349. doi:10.1080/00401706.1987.10488243

Hosking, J. R. M., & Wallis, J. R. (1988). The Effect of Intersite Dependence on Regional Flood Frequency Analysis. *Water Resources Research*, *24*(4), 588–600.

Hosking, J. R. M., & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, *29*(2), 271–281. doi:10.1029/92WR01980

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge: Cambridge University Press.

Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments. *Technometrics*, *27*(3), 251–261. doi:10.1080/00401706.1985.10488049

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, *81*(348), 158–171. doi:10.1002/qj.49708134804

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. London, New York: Chapman & Hall/CRC.

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, *25*(8), 1287–1304. doi:https://doi.org/10.1016/S0309-1708(02)00056-8

Kinsvater, P., Fried, R., & Lilienthal, J. (2016). Regional extreme value index estimation and a test of tail homogeneity. *Environmetrics*, *27*(2), 103–115. doi:10.1002/env.2376

Knight, K., & Fu, W. (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics*, *28*(5), 1356–1378. doi:10.1214/aos/1015957397

Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel Parameters and quantiles. *Water Resources Research*, *15*(5), 1055–1064. doi:10.1029/WR015i005p01055

LAWA. (1997). *Pegelvorschrift - Stammtext* (4. überarbeitete Auflage). Berlin/Bonn: Kulturbuchverlag.

Leadbetter, M. R. (1974). On extreme values in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *28*, 289–303. doi:10.1007/BF00532947

Lettenmaier, D. P., Wallis, J. R., & Wood, E. F. (1987). Effect of regional heterogeneity on flood frequency estimation. *Water Resources Research*, *23*(2), 313–323. doi:10.1029/WR023i002p00313

Lilienthal, J. (2014). *Regionalisierung in der Hochwasserstatistik* (Masterarbeit, Fakultät Statistik, Technische Universität Dortmund).

Lilienthal, J. (2019). *TLMoments: Calculate TL-Moments and Convert Them to Distribution Parameters*. R package version 0.7.4.3.

Lilienthal, J., Fried, R., & Schumann, A. (2018). Homogeneity testing for skewed and cross-correlated data in regional flood frequency analysis. *Journal of Hydrology*, *556*, 557–571. doi:10.1016/j.jhydrol.2017.10.056

Lilienthal, J., Kinsvater, P., & Fried, R. (2016). On the method of probability weighted moments in regional frequency analysis. *SFB 823 Discussion*, *63/2016*.

Lu, L. H., & Stedinger, J. R. (1992). Variance of two- and three-parameter GEV/PWM quantile estimators: formulae, confidence intervals, and a comparison. *Journal of Hydrology*, *138*(1–2), 247–267. doi:10.1016/0022-1694(92)90167-T

Marshall, A. W., & Olkin, I. (1988). Families of Multivariate Distributions. *Journal of the American Statistical Association*, *83*(403), 834–841. doi:10.1080/01621459.1988.10478671

Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, *36*(3), 737–744. doi:10.1029/1999WR900330

Masselot, P., Chebana, F., & Ouarda, T. B. M. J. (2017). Fast and direct nonparametric procedures in the L-moment homogeneity test. *Stochastic Environmental Research and Risk Assessment*, *31*(2), 509–522. doi:10.1007/s00477-016-1248-0

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer Series in Statistics. New York, NY: Springer-Verlag.

Ouarda, T. B. M. J., Girard, C., Cavadias, G. S., & Bobée, B. (2001). Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology*, *254*, 157–173. doi:10.1016/S0022-1694(01)00488-7

Pandey, G. R., & Nguyen, V.-T.-V. (1999). A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology*, *225*(1–2), 92–101. doi:10.1016/S0022-1694(99)00135-3

Pickands, J. (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd session of the international statistical institute* (Vol. 49, pp. 859–878, 894–902).

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, *3*(1), 119–131. doi:10.1214/aos/1176343003

Pötscher, B. M., & Leeb, H. (2009). On the distribution of penalized maximum likelihood estimators: The LASSO, SCAD, and thresholding. *Journal of Multivariate Analysis*, *100*(9), 2065–2082. doi:10.1016/j.jmva.2009.06.010

Prescott, P., & Walden, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, *67*(3), 723–724. doi:10.1093/biomet/67.3.723

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications* (2nd ed.). Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons, Inc.

Requena, A. I., Chebana, F., & Mediero, L. (2016). A complete procedure for multivariate index-flood model application. *Journal of Hydrology*, *535*, 559–580. doi:10.1016/j.jhydrol.2016.02.004

Rosbjerg, D. (1985). Estimation in partial duration series with independent and dependent peak values. *Journal of Hydrology*, *76*(1–2), 183–195. doi:10.1016/0022-1694(85)90098-8

Salmon, F. (2012). The formula that killed Wall Street. *Significance*, *9*(1), 16–20. Reprinted from *Wired*, February 2009. doi:10.1111/j.1740-9713.2012.00538.x

Salvadori, G., & De Michele, C. (2010). Multivariate multiparameter extreme value models and return periods: A copula approach. *Water Resources Research*, *46*(10). doi:10.1029/2009WR009040

Salvadori, G., & De Michele, C. (2011). Estimating strategies for multiparameter multivariate extreme value copulas. *Hydrology and Earth System Sciences*, *15*(1), 141–150. doi:10.5194/hess-15-141-2011

Serfling, R., & Xiao, P. (2007). A contribution to multivariate L-moments: L-comoment matrices. *Journal of Multivariate Analysis*, *98*(9), 1765–1781. doi:10.1016/j.jmva.2007.01.008

Serinaldi, F. (2015). Dismissing return periods! *Stochastic Environmental Research and Risk Assessment*, *29*(4), 1179–1189. doi:10.1007/s00477-014-0916-1

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

Skøien, J. O., Merz, R., & Blöschl, G. (2006). Top-kriging - geostatistics on stream networks. *Hydrology and Earth System Sciences*, *10*(2), 277–287. doi:10.5194/hess-10-277-2006

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, *72*(1), 67–90. doi:10.1093/biomet/72.1.67

Song, X., Lu, F., Wang, H., Xiao, W., & Zhu, K. (2018). Penalized maximum likelihood estimators for the nonstationary Pearson type 3 distribution. *Journal of Hydrology*, *567*, 579–589. doi:10.1016/J.JHYDROL.2018.10.035

Stedinger, J. R. (1980). Fitting log normal distributions to hydrologic data. *Water Resources Research*, *16*(3), 481–490. doi:10.1029/WR016i003p00481

Stephenson, A. G. (2002). evd: Extreme Value Distributions. *R News*, *2*(2).

Sveinsson, O. G., Salas, J. D., & Boes, D. C. (2003). Uncertainty of quantile estimators using the population index flood method. *Water Resources Research*, *39*(8). doi:10.1029/2002WR001594

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x

Vapnik, V. N. (2000). *The Nature of Statistical Learning*. Statistics for Engineering and Information Science. New York, NY: Springer-Verlag.

Viglione, A. (2012). *homtest: Homogeneity tests for Regional Frequency Analysis*. R package version 1.0-5.

Viglione, A., Laio, F., & Claps, P. (2007). A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research*, *43*(3), 1–10. doi:10.1029/2006WR005095

Vogel, R. M., & Fennessey, N. M. (1993). L moment diagrams should replace product moment diagrams. *Water Resources Research*, *29*(6), 1745–1752. doi:10.1029/93WR00341

Walther, J., Fischer, B., Horn, S., Merz, R., Illarenas Salina, J. L., & Laaha, G. (2012). Flächenhafte Bestimmung von Hochwasserspenden. *Schriftenreihe des LfULG*, *3*.

Wang, Q. J. (1991). The POT model described by the generalized Pareto distribution with Poisson arrival rate. *Journal of Hydrology*, *129*(1-4), 263–280. doi:10.1016/0022-1694(91)90054-L

Waylen, P., & Woo, M. K. (1982). Prediction of annual floods generated by mixed processes. *Water Resources Research*, *18*(4), 1283–1286. doi:10.1029/WR018i004p01283

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.

Wiltshire, S. E. (1986). Regional flood frequency analysis I: Homogeneity statistics. *Hydrological Sciences Journal*, *31*(3), 321–333. doi:10.1080/02626668609491051

Wittenberg, H. (2011). *Praktische Hydrologie* (1. Auflage). Studium. Wiesbaden: Vieweg+Teubner.

Zhang, J., & Boos, D. D. (1994). Adjusted power estimates in Monte Carlo experiments. *Communications in Statistics - Simulation and Computation*, *23*(1), 165–173. doi:10.1080/03610919408813162

A Appendix to Chapter 4

This section presents the corresponding Jacobi matrices involved in formulas (4.9) and (4.10) for TL(0,0)- and TL(0,1)-moments.

TL(0,0)

Let $\boldsymbol{\vartheta} = (\mu, \sigma, \xi)'$ with $\xi < 1$ and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)'$ denote parameters and untrimmed L-moments of a GEV distribution, respectively, and let $g^{(0,0)}$ be the function that maps L-moments to GEV parameters. The Jacobi matrix $\Delta^{(0,0)}_{\boldsymbol{\lambda}_j \to \vartheta} = \frac{\partial}{\partial \boldsymbol{\lambda}} g^{(0,0)}(\boldsymbol{\lambda})$ involved in the asymptotic distribution of L-moment estimators is approximated by that of the explicit solution. For the latter we obtain

$$A = \begin{pmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{pmatrix} \tag{A.1}$$

with

$$\begin{aligned}
a_{12} =& \frac{\log(2)\ \lambda_2\ (\Gamma(1-\pi)-1)\ 2^{\pi}\,\rho\,\theta}{\Gamma(1-\pi)\ (1-2^{\pi})^2} + \frac{\lambda_2\,\psi_0(1-\pi)\ (\Gamma(1-\pi)-1)\ \rho\,\theta}{\Gamma(1-\pi)\ (1-2^{\pi})} - \\
& \frac{\lambda_2\,\psi_0(1-\pi)\ \rho\,\theta}{1-2^{\pi}} + \frac{\Gamma(1-\pi)-1}{\Gamma(1-\pi)\ (1-2^{\pi})} \\
a_{13} =& -\frac{\log(2)\ {\lambda_2}^2\ (\Gamma(1-\pi)-1)\ 2^{\pi+1}\,\rho\,\zeta^2}{\Gamma(1-\pi)\ (1-2^{\pi})^2} - \\
& \frac{2\,{\lambda_2}^2\,\psi_0(1-\pi)\ (\Gamma(1-\pi)-1)\ \rho\,\zeta^2}{\Gamma(1-\pi)\ (1-2^{\pi})} + \frac{2\,{\lambda_2}^2\,\psi_0(1-\pi)\ \rho\,\zeta^2}{1-2^{\pi}} \\
a_{22} =& -\frac{\log(2)\ \lambda_2\,\pi\,2^{\pi}\,\rho\,\theta}{\Gamma(1-\pi)\ (1-2^{\pi})^2} - \frac{\lambda_2\,\rho\,\theta\ (\psi_0(1-\pi)\ \pi+1)}{\Gamma(1-\pi)\ (1-2^{\pi})} - \frac{\pi}{\Gamma(1-\pi)\ (1-2^{\pi})} \\
a_{23} =& \frac{\log(2)\ {\lambda_2}^2\,\pi\,2^{\pi+1}\,\rho\,\zeta^2}{\Gamma(1-\pi)\ (1-2^{\pi})^2} + \frac{2\,{\lambda_2}^2\,\rho\,\zeta^2\ (\psi_0(1-\pi)\ \pi+1)}{\Gamma(1-\pi)\ (1-2^{\pi})} \\
a_{32} =& -2\,\lambda_3\ (2\,b\,\kappa\,\lambda_3 - a\,\lambda_3 + 6\,b\,\kappa\,\lambda_2 - 4\,b\,\lambda_2 - 3\,a\,\lambda_2)\ \zeta^3 \\
a_{33} =& 2\,\lambda_2\ (2\,b\,\kappa\,\lambda_3 - a\,\lambda_3 + 6\,b\,\kappa\,\lambda_2 - 4\,b\,\lambda_2 - 3\,a\,\lambda_2)\ \zeta^3
\end{aligned}$$

and with $a = -7.859,\ \ b = -2.9554,\ \ \kappa = \dfrac{\log 2}{\log 3}$,

$$\begin{aligned}
&\eta = (2\lambda_2\zeta - \kappa),\\
&\pi = b\eta^2 + a\eta,\\
&\zeta = 1/(\lambda_3 + 3\lambda_2),\\
&\theta = (2\zeta - 6\lambda_2\zeta^2),\\
&\rho = 2b\eta + a,\\
&\psi_0(x) = \Gamma'(x)/\Gamma(x).
\end{aligned}$$

TL(0,1)

Considering now trimmed L-moments $\boldsymbol{\lambda}^{(0,1)}$ of a GEV distribution with parameters $\boldsymbol{\vartheta}$ and mapping function $g^{(0,1)}$. The Jacobi matrix $\Delta^{(0,1)}_{\boldsymbol{\lambda}_j^{(0,1)}\to\vartheta} = \frac{\partial}{\partial\boldsymbol{\lambda}^{(0,1)}} g^{(0,1)}(\boldsymbol{\lambda}^{(0,1)})$ is approximated by

$$A = \begin{pmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{pmatrix}, \tag{A.2}$$

where

$$\begin{aligned}
a_{12} =& -\frac{2\,\lambda_2^{(0,1)}\,\psi_0\left(-\pi\right)\,\left(-2\,b\,\left(\zeta-\eta\right)\,\left(\lambda_2^{(0,1)}\,\zeta-\kappa\right)-a\,\left(\zeta-\eta\right)\right)+2}{3\,\pi\,\left(-2^{\pi+1}+3^{\pi}+1\right)\,\gamma\left(-\pi\right)}\\
&-\frac{2\,\lambda_2^{(0,1)}\,\left(\log\left(3\right)\,\iota\,3^{\pi}-\log\left(2\right)\,\iota\,2^{\pi+1}\right)\left(1-\left(2^{\pi}-2\right)\,\pi\,\gamma(-\pi)\right)}{3\,\pi\,\left(-2^{\pi+1}+3^{\pi}+1\right)^2\,\gamma\left(-\pi\right)}\\
&-\frac{2\,\iota\,\lambda_2^{(0,1)}}{3\,\pi^2\,\left(-2^{\pi+1}+3^{\pi}+1\right)\,\gamma\left(-\pi\right)}-\frac{\log\left(2\right)\,\iota\,\lambda_2^{(0,1)}\,2^{\pi+1}+2\,\left(2^{\pi}-2\right)}{3\,\left(-2^{\pi+1}+3^{\pi}+1\right)}\\
a_{13} =& -\frac{2\,\lambda_2^{(0,1)}\,\left(\log\left(2\right)\,2^{\pi+1}\,\rho-\log\left(3\right)\,3^{\pi}\,\rho\right)\,\left(1-\left(2^{\pi}-2\right)\,\pi\,\gamma(-\pi)\right)}{3\,\pi\,\left(-2^{\pi+1}+3^{\pi}+1\right)^2\,\gamma\left(-\pi\right)}\\
&-\frac{2\,\lambda_2^{(0,1)}\,\rho\,\left(\psi_0(-\pi)\,\pi-2^{\pi}\log(2)-1\right)}{3\,\pi^2\,\left(-2^{\pi+1}+3^{\pi}+1\right)\,\gamma\left(-\pi\right)}\\
a_{22} =& -\frac{2\,\lambda_2^{(0,1)}\,\psi_0\left(-\pi\right)\,\left(-2\,b\,\left(\zeta-\eta\right)\,\left(\lambda_2^{(0,1)}\,\zeta-\kappa\right)-a\,\left(\zeta-\eta\right)\right)+2}{3\,\left(-2^{\pi+1}+3^{\pi}+1\right)\,\gamma\left(-\pi\right)}\\
&-\frac{2\,\lambda_2^{(0,1)}\,\left(\log\left(3\right)\,\iota\,3^{\pi}-\log\left(2\right)\,\iota\,2^{\pi+1}\right)}{3\,\left(-2^{\pi+1}+3^{\pi}+1\right)^2\,\gamma\left(-\pi\right)}\\
a_{23} =& -\frac{2\,\lambda_2^{(0,1)}\,\left(\log\left(2\right)\,2^{\pi+1}\,\rho-\log\left(3\right)\,3^{\pi}\,\rho\right)}{3\,\left(-2^{\pi+1}+3^{\pi}+1\right)^2\,\gamma\left(-\pi\right)}-\frac{2\,\lambda_2^{(0,1)}\,\psi_0\left(-\pi\right)\,\rho}{3\,\left(-2^{\pi+1}+3^{\pi}+1\right)\,\gamma\left(-\pi\right)}\\
a_{32} =& \iota
\end{aligned}$$

$$a_{33} = -\rho$$

and with $a = -8.5674$, $b = 0.6760$, $\kappa = \dfrac{2\log 2 - \log 3}{3\log 3 - 2\log 4}$,

$$\begin{aligned}
\theta &= 3(\lambda_3^{(0,1)} + 2\lambda_2^{(0,1)}),\\
\zeta &= 10/(3\theta),\\
\eta &= 20\lambda_2^{(0,1)}/\theta^2,\\
\pi &= b(\lambda_2^{(0,1)}\zeta - \kappa)^2 + a(\lambda_2^{(0,1)}\zeta - \kappa)\\
\rho &= -b\eta(\lambda_2^{(0,1)}\zeta - \kappa),\\
\iota &= 2b(\zeta - \eta)(\lambda_2^{(0,1)}\zeta - \kappa) + a(\zeta - \eta)\\
\psi_0(x) &= \Gamma'(x)/\Gamma(x)
\end{aligned}$$

B Appendix to Chapter 5

In this section we examine the validity of the results derived in Section 5.3 using simulation studies. Therefore we analyse the consistency of the estimator (see Proposition 2 on page 91) and the rate of convergence (see Proposition 3 on page 92).

Study design

We consider a sample of observations $X_1 \ldots, X_n$ following a GEV distribution:

$$X_i \sim \mathrm{GEV}(\mu_0, \sigma_0, \xi_0), \;\; i = 1, \ldots, n. \tag{B.1}$$

While we chose the location and scale parameter to be $\mu_0 = 10$ and $\sigma_0 = 10/1.8$, we investigate three different settings regarding the true shape parameter ξ_0: a light-tailed scenario ($\xi_0 = -0.3$), a heavy-tailed scenario ($\xi_0 = 0.3$), and the Gumbel case in which $\xi_0 = 0$.

A realisation of $N = 100\,000$ observations is drawn for each setting and the regularized ML estimator $\boldsymbol{\vartheta}_n$ that maximizes

$$Q_n(\boldsymbol{\vartheta}) = \frac{1}{n} \sum_{i=1}^{n} \log g_{\boldsymbol{\vartheta}}(X_i) - \frac{1}{n} \boldsymbol{\lambda}_n' \, \Omega(\boldsymbol{\vartheta}) \tag{B.2}$$

is calculated for $n = 50, 100, 200, 500, 1000, 2000, 5000, 10\,000, 50\,000, 100\,000$ (using the first n observations and neglecting the remaining). The regularization term is selected as

$$\Omega(\boldsymbol{\vartheta}) = \Big((\mu_j/\sigma_j - \delta_c)^2, (\xi_j - \xi_c)^2 \Big)' \tag{B.3}$$

with $\delta_c = 2$ and $\xi_c = 0$.

The smoothing parameter is varied as $\boldsymbol{\lambda}_n \in \{(n^p, n^p)' : p \in \{0,\ 1/4,\ 1/2,\ 3/4,\ 1\}\}$,

that is, $\boldsymbol{\lambda}_n$ satisfies

$$||\boldsymbol{\lambda}_n|| = \begin{cases} o(1), & \text{if } \boldsymbol{\lambda}_n = (1,1)', \\ O(n^{1/4}), & \text{if } \boldsymbol{\lambda}_n = (n^{1/4}, n^{1/4})', \\ O(\sqrt{n}), & \text{if } \boldsymbol{\lambda}_n = (\sqrt{n}, \sqrt{n})', \\ O(n^{3/4}), & \text{if } \boldsymbol{\lambda}_n = (n^{3/4}, n^{3/4})', \\ O(n), & \text{if } \boldsymbol{\lambda}_n = (n,n)'. \end{cases} \tag{B.4}$$

Each scenario is replicated $B = 10\,000$ times and the estimator of the b-th replication using n observations and smoothing parameter $\boldsymbol{\lambda}_n$ is denoted as $\hat{\boldsymbol{\vartheta}}^{(b)}_{n,\boldsymbol{\lambda}},\ b = 1, \ldots, B$.

Strong consistency

To check for strong consistency we examine the behaviour of $\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}}$. Figure B.1, B.2, and B.3 display the paths of the entries of $\hat{\boldsymbol{\vartheta}}^{(i)}_{n,\boldsymbol{\lambda}}$ for increasing n and different smoothing parameters in the light-tailed, Gumbel, and heavy-tailed scenario, respectively. For $\boldsymbol{\lambda}_n \in \{(0,0)', (\sqrt{n}, \sqrt{n})'\}$ all estimates seemingly tend towards the true parameter, but for $\boldsymbol{\lambda}_n = (n,n)'$ they tend to wrong values. This corroborates Proposition 1 that states consistency as long as $||\boldsymbol{\lambda}_n|| = o(n)$.

MSE-consistency

Although not being addressed theoretically, we also investigate the MSE-consistency of the estimator. For $\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}}$ to be a MSE consistent estimator for $\boldsymbol{\vartheta}_0$ the mean squared error $MSE(\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}}) = E((\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}} - \boldsymbol{\vartheta}_0)^2) = Var(\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}}) + Bias(\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}}, \boldsymbol{\vartheta}_0)^2$ needs to tend to zero for increasing n. We check this by investigating the empirical mean squared error $\widehat{MSE}(\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}}) = B^{-1} \sum_{i=1}^{B} (\hat{\boldsymbol{\vartheta}}^{(i)}_{n,\boldsymbol{\lambda}} - \boldsymbol{\vartheta}_0)^2$ for increasing n.

In Figure B.4 the components of the empirical MSE vector for a true shape $\xi_0 = 0.3$ are given. Except for $\boldsymbol{\lambda}_n = (n,n)'$ all choices of $\boldsymbol{\lambda}_n$ lead to an empirical MSE tending towards zero. For $\boldsymbol{\lambda}_n = (n,n)'$ the empirical MSE stabilizes for increasing n, which suggests that the resulting estimator is not MSE-consistent.

Rate of convergence

Proposition 2 provides the rate of convergence, saying that $||\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}} - \boldsymbol{\vartheta}_0|| = O_{\mathbb{P}}(n^{-1/2})$ if $||\boldsymbol{\lambda}_n|| = O(\sqrt{n})$. This means that in this case $\sqrt{n}||\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}} - \boldsymbol{\vartheta}_0|| = O_{\mathbb{P}}(1)$ (is limited in probability).

In Figure B.5 the empirical non-exceedance probabilities for different fixed limits c are given for a true shape $\xi_0 = 0.3$. The figure affirms the results of Proposition 2, showing stable probabilities $P(\sqrt{n}||\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}} - \boldsymbol{\vartheta}_0|| < c)$ for fixed c and increasing n if $||\boldsymbol{\lambda}_n|| = O(\sqrt{n})$ (upper panel). If $||\boldsymbol{\lambda}_n|| = O(n^{3/4})$, such probabilities decrease to 0, but note that $P(n^{1/4}||\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}} - \boldsymbol{\vartheta}_0|| < c)$ then even increases towards 1 (lower panel).

The results of Proposition 2 can further be affirmed by analysing the density functions of $\sqrt{n}||\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}} - \boldsymbol{\vartheta}_0||$ and $n^{1/4}||\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}} - \boldsymbol{\vartheta}_0||$ (Figure B.6).

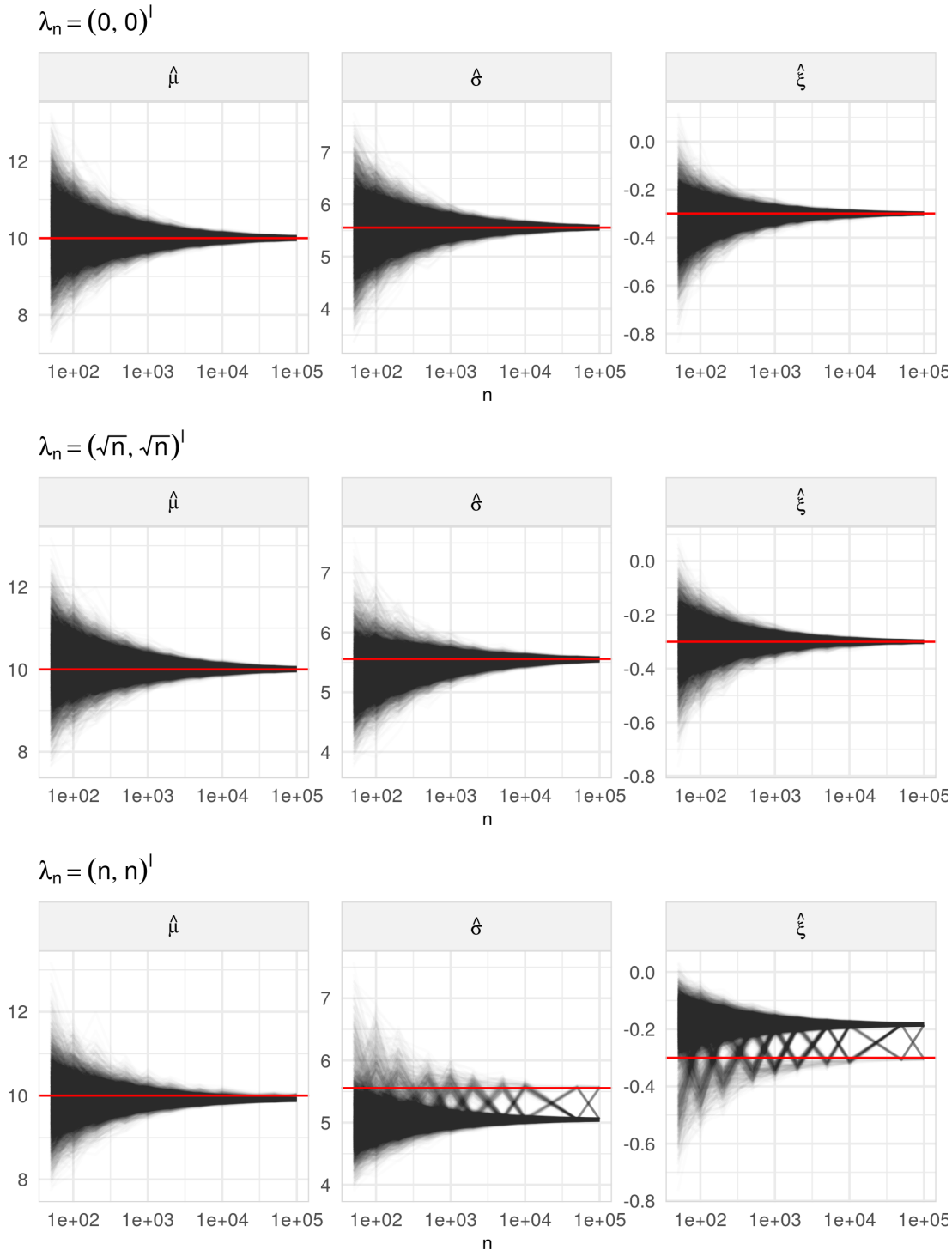

Figure B.1: Behavior of 10 000 estimations $\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}}$ for increasing n and different $\boldsymbol{\lambda}_n$ for $\xi_0 = -0.3$. Red line marks the true parameter.

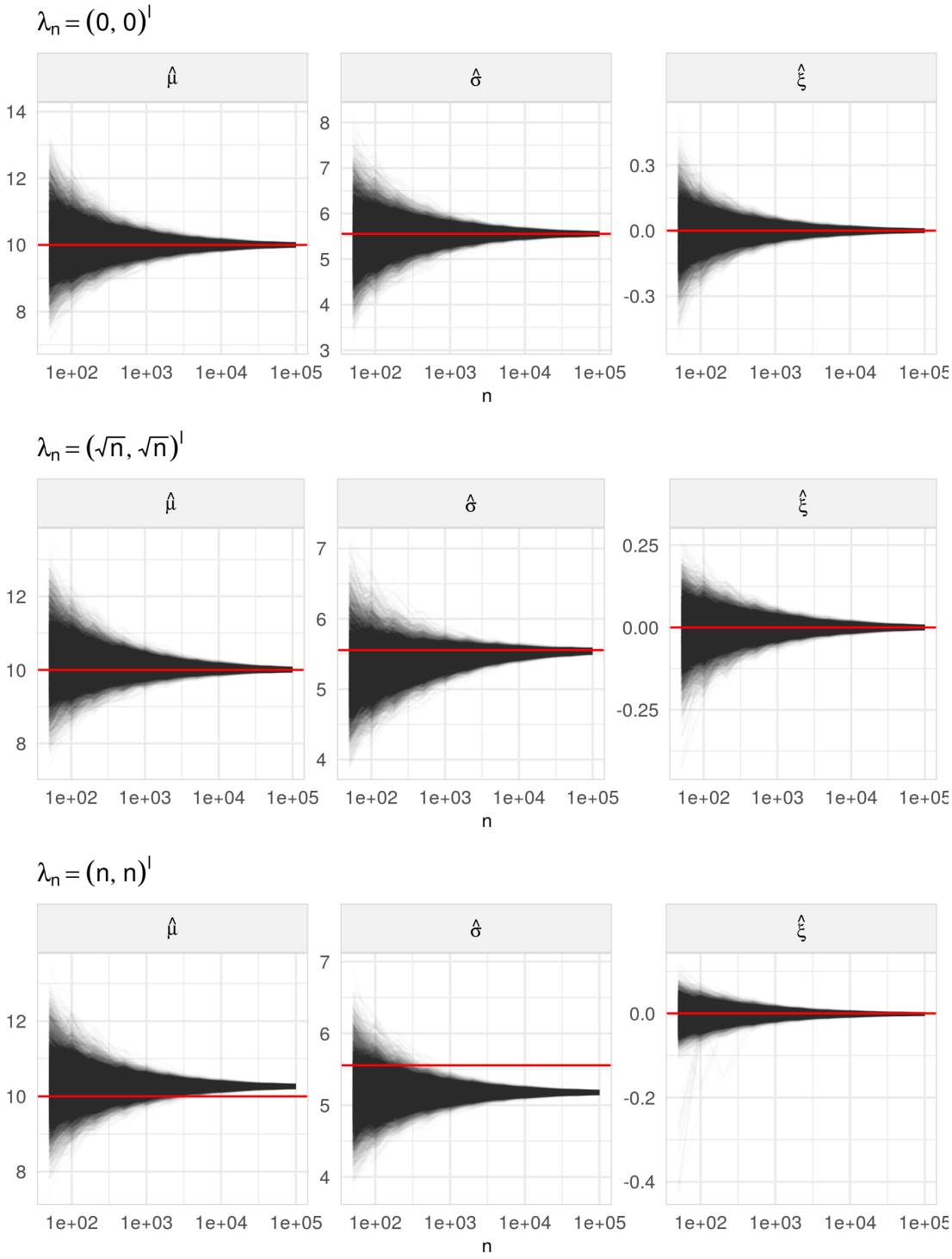

Figure B.2: Behavior of 10 000 estimations $\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}}$ for increasing n and different $\boldsymbol{\lambda}_n$ for $\xi_0 = 0$. Red line marks the true parameter.

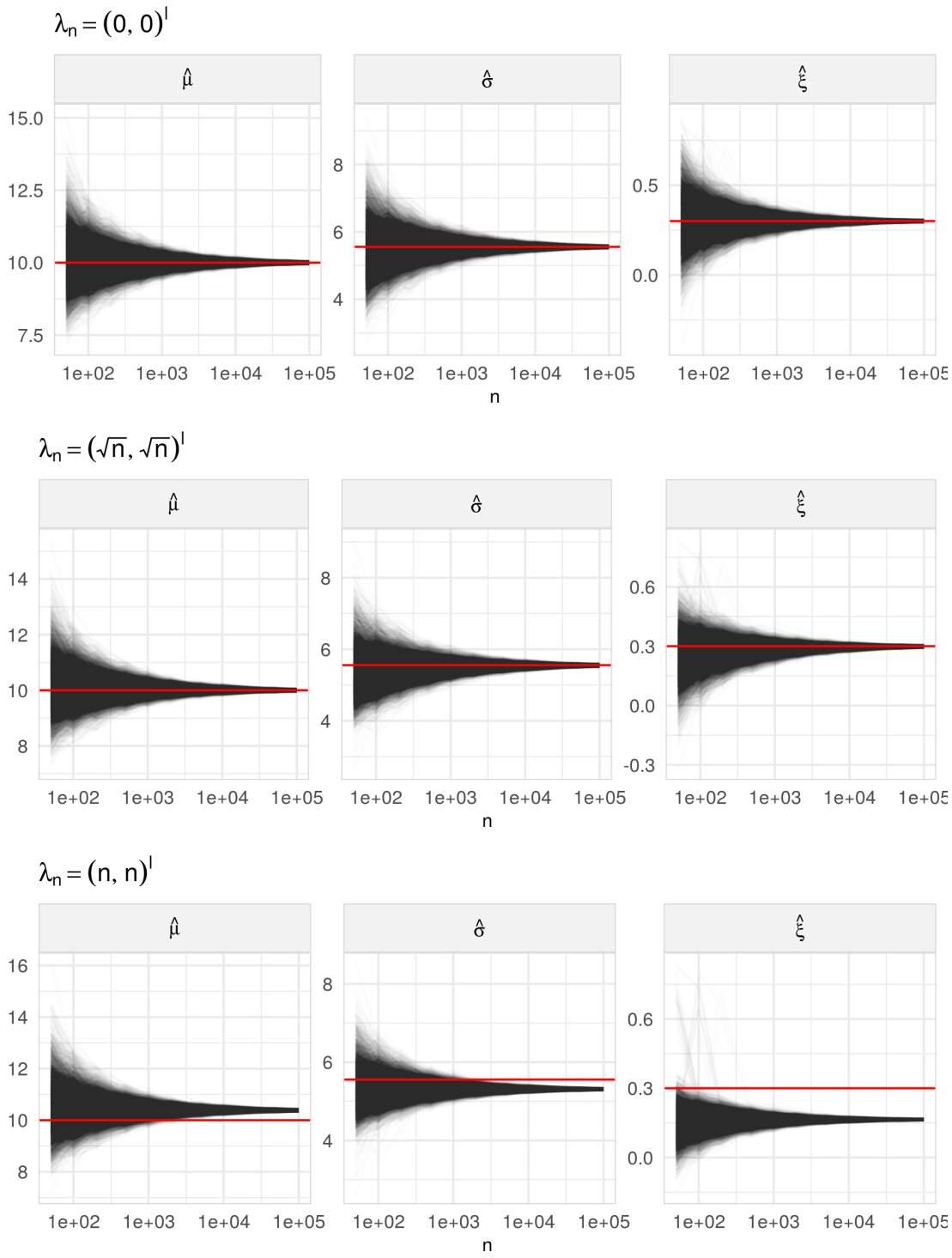

Figure B.3: Behavior of 10 000 estimations $\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}}$ for increasing n and different $\boldsymbol{\lambda}_n$ for $\xi_0 = 0.3$. Red line marks the true parameter.

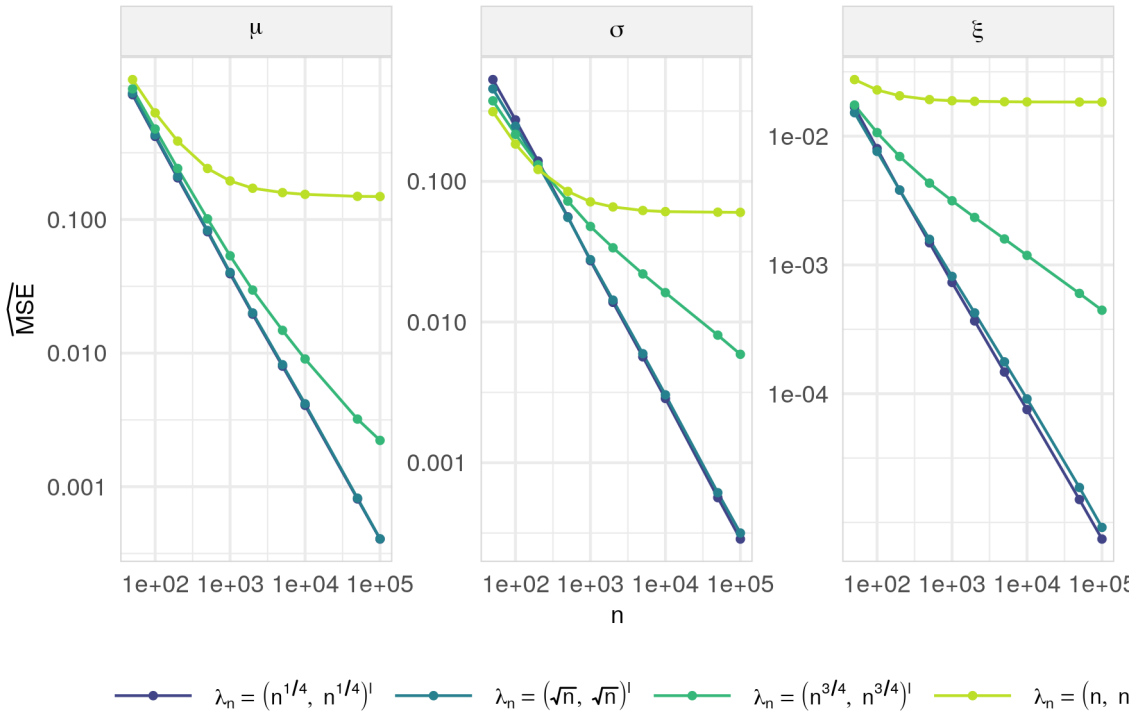

Figure B.4: Empirical MSE of $\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}}$ for increasing n and different $\boldsymbol{\lambda}_n$ for $\xi_0 = 0.3$.

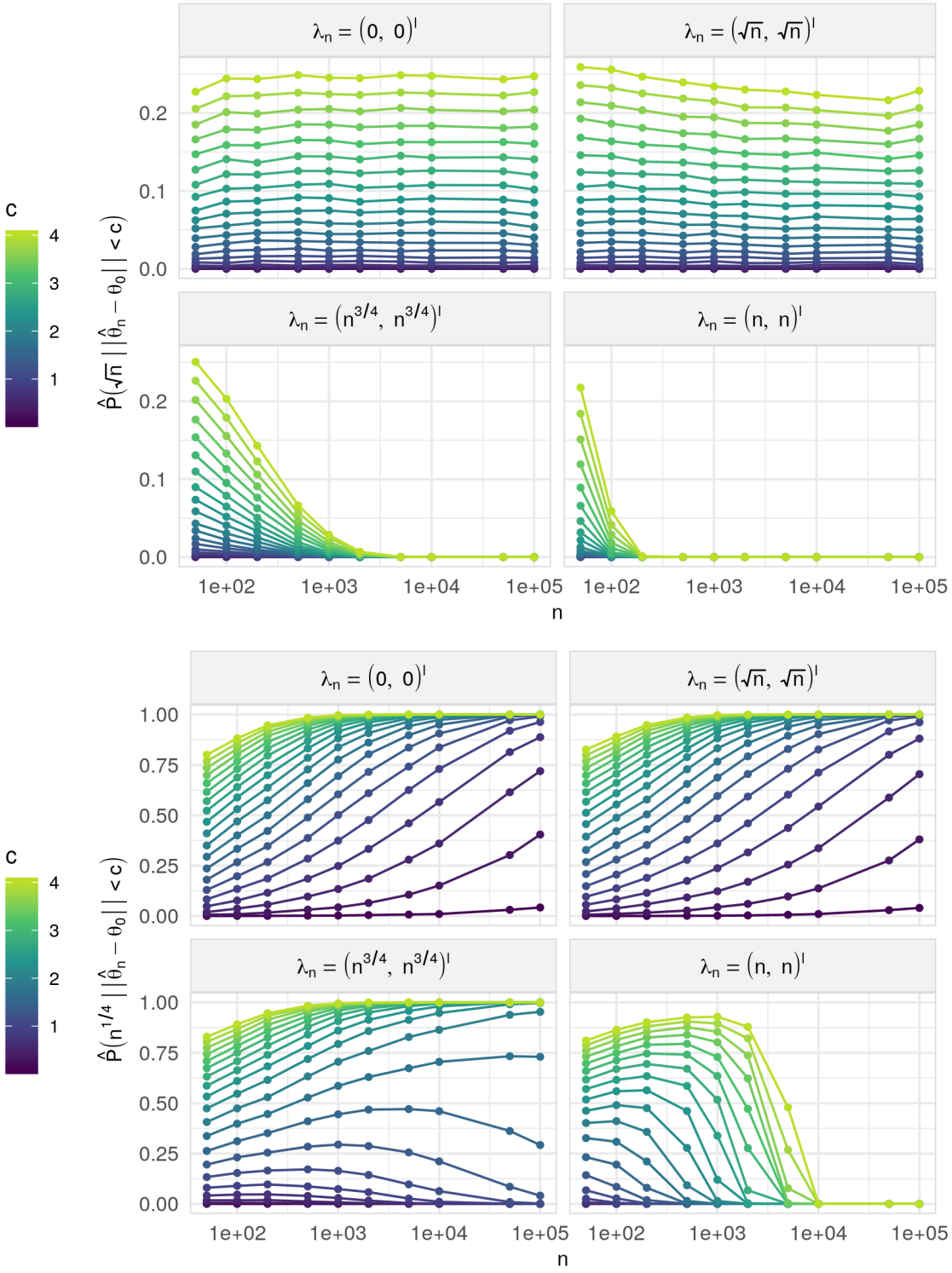

Figure B.5: Non-exceedance probabilities of $\hat{P}(\sqrt{n}||\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}} - \boldsymbol{\vartheta}_0|| < c)$ and $\hat{P}(n^{1/4}||\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}} - \boldsymbol{\vartheta}_0|| < c)$ for different $\boldsymbol{\lambda}_n$ ($\xi_0 = 0.3$).

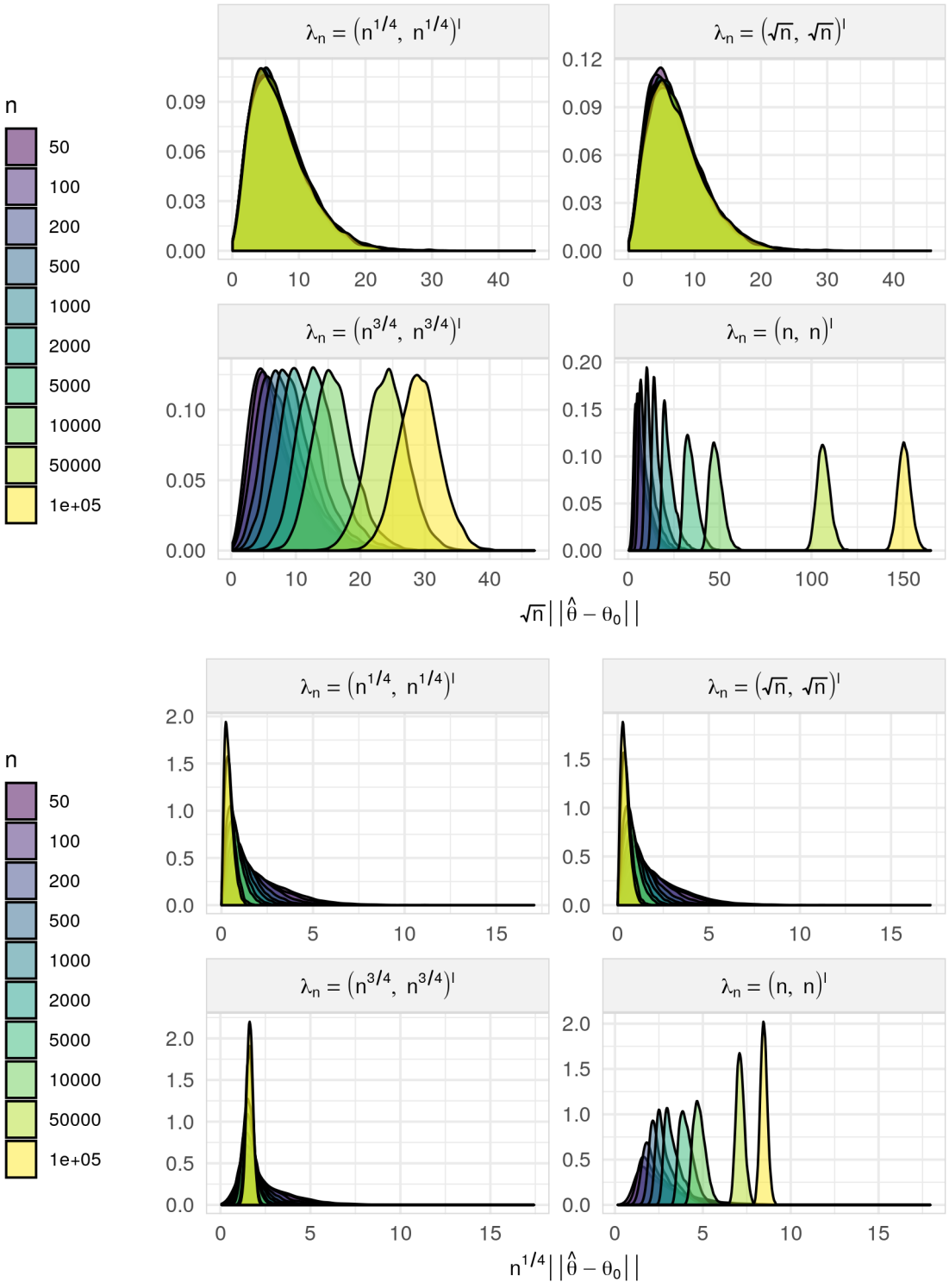

Figure B.6: Density functions of $\sqrt{n}||\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}} - \boldsymbol{\vartheta}_0||$ and $n^{1/4}||\hat{\boldsymbol{\vartheta}}_{n,\boldsymbol{\lambda}} - \boldsymbol{\vartheta}_0||$ for different $\boldsymbol{\lambda}_n$ (true shape $\xi_0 = 0.3$).