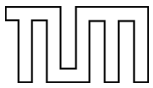

Technische Universität München

Fakultät für Mathematik
Lehrstuhl für Mathematische Statistik

Master's Thesis

Modelling and Estimation of Extremes in Space and Time

Sven Buhl

Themenstellerin:	Prof. Dr. Claudia Klüppelberg
Betreuerin:	Dipl.-Math. Christina Steinkohl
Abgabetermin:	12. April 2013

Hiermit erkläre ich, dass ich die Masterarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

Garching, den 12. April 2013

Zusammenfassung

In dieser Arbeit werden auf dem stetigen Produktraum $\mathbb{R}^d \times [0, \infty)$ definierte *max-stabile Prozesse* verwendet, um extreme Ereignisse in Ort und Zeit zu modellieren und statistisch zu analysieren.

Max-stabile Prozesse können als unendlichdimensionale Verallgemeinerung multivariater Extremwertverteilungen betrachtet werden. Diese Arbeit behandelt hauptsächlich den max-stabilen Brown-Resnick Prozess, der als Grenzwert geeignet reskalierter punktweiser Maxima unabhängiger Gauss-Prozesse gewonnen werden kann, deren Korrelationsfunktion eine zentrale Rolle einnimmt. Es werden keine räumlich isotropen Korrelationsfunktionen vorausgesetzt, welche zu isotropen max-stabilen Prozessen führen. In dieser Arbeit ist es dem Prozess und damit den lokalen Extremen allgemeiner erlaubt, gewisse Richtungen vorzuziehen, was in realistischen Situationen vernünftig erscheint.

Im weiteren Verlauf wird eine an die anisotropen Voraussetzungen angepasste paarweise Likelihood Methode verwendet, um die Parameter eines Brown-Resnick Prozesses zu schätzen. Es wird die Konsistenz und asymptotische Normalität der resultierenden Schätzer unter gewissen Regularitätsbedingungen bewiesen. Mittels einer Simulationsstudie auf einem regulären Gitter wird demonstriert, wie sich die paarweisen Likelihood Schätzer verbessern, wenn die Seitenlänge des Gitters und die Anzahl der Zeitpunkte größer werden.

Die Arbeit endet mit einer Anwendung der paarweisen Likelihood Methode auf Radar-Regendaten aus Florida und zeigt, dass Anisotropie der Regenwetterfronten statistisch signifikant ist.

Die Arbeit erweitert die Theorie von Davis, Klüppelberg und Steinkohl [2012a,b] und Steinkohl [2012].

Abstract

In this thesis we use *max-stable processes* defined on the continuous domain $\mathbb{R}^d \times [0, \infty)$ to model and statistically analyse extreme events in space and time.

Max-stable processes can be viewed as the infinite-dimensional generalization of multivariate extreme value distributions. The thesis mainly considers the max-stable Brown-Resnick process, which can be obtained as a limit of appropriately rescaled pointwise maxima of independent Gaussian processes, whose correlation function plays a central role. We do not assume spatially isotropic correlation functions leading to isotropic max-stable processes. Instead, we more generally allow the process and therefore the spatial extremes to have directional preferences, which seems reasonable in realistic situations.

We use a pairwise likelihood estimation procedure adapted to the anisotropic setting in order to estimate the parameters of a Brown-Resnick process and prove consistency and asymptotic normality of the resulting estimates under some regularity conditions.

We demonstrate by a simulation study on a regular grid, how the pairwise likelihood estimates improve, when the size of the grid and the number of time points increase.

We finish the thesis by applying the pairwise likelihood method to radar rainfall measurements taken in Florida and show that the extreme rainfall weather fronts are statistically significantly anisotropic.

This thesis extends the theory of Davis, Klüppelberg and Steinkohl [2012a,b] and Steinkohl [2012].

Acknowledgments

First of all, I would like to thank Prof. Dr. Claudia Klüppelberg for offering me the possibility to write my Master's Thesis at the Chair of Mathematical Statistics of the Munich University of Technology.

Furthermore, my special thanks goes to Claudia Klüppelberg and Christina Steinkohl who both supported me whenever needed. Although short in time they answered all my questions elaborately and in detail.

Moreover, I would like to thank Chin Man Mok and the Southwest Florida Water Management District (SWFWMD) for providing radar measurement data of rainfall in Florida. These data made it possible for me to apply my theory in a real world example.

Last but not least, I am grateful to all my friends and my family who often had to bear my mood changes strongly depending on the progress of this thesis.

Contents

Chapter 1

Introduction

1.1 Max-stable Processes to model Extremes in Space and Time

In recent years, several damaging catastrophes have caught our attention. The pictures of Japan (2011) after the sea earthquake causing a devastating tsunami which in turn brought about a core meltdown in the nuclear power station of Fukushima are still in our minds. Furthermore, we remember the residents of New Orleans (2005) who had to be evacuated and brought to the local football arena after hurricane Katrina had caused disastrous damages.

Clearly, mathematicians and statisticians feel motivated to try and find ways to model rarely occurring extreme events in order to be better prepared in the future. The theory of univariate and multivariate extreme value distributions (limit distributions of appropriately centred and scaled maxima of, for example, a sequence of independent and identically distributed (i.i.d.) random variables) has been well developed during the past century. A central problem nowadays, however, is to model extremal behaviour and accounting for dependence structures in both space and time. A natural generalization of uni- and multivariate distributions are *max-stable processes*. Max-stable processes can be a useful help when the aim is to model extremes on the continuous domain $\mathbb{R}^d \times [0, \infty)$ for some $d \in \mathbb{N}$ (the spatial dimension).

As in Davis, Klüppelberg and Steinkohl [2012a,b] and Steinkohl [2012], our focus lies on the so-called *Brown-Resnick* process, which can be obtained as the limit of rescaled pointwise maxima of independent Gaussian space-time processes whose correlation function satisfies certain regularity conditions. However, we skip the assumption that the correlation function is spatially isotropic as this suggests that the dependence between two locations decreases equally, no matter if the second location is in the north of the first one or any other cardinal direction. In a natural context, it might often seem more reasonable to assume anisotropy; for instance, in many regions of the world, wind and therefore clouds and the spread of rainfall have a preferred direction. This impression is also confirmed by Figure 1.1, which shows the daily maxima of radar rainfall measurements in inches in some region in Florida (cf. Figure 6.1). The rainfall does not move "in circles", but rather elliptically. A question that arises in the context of this data example is: "can

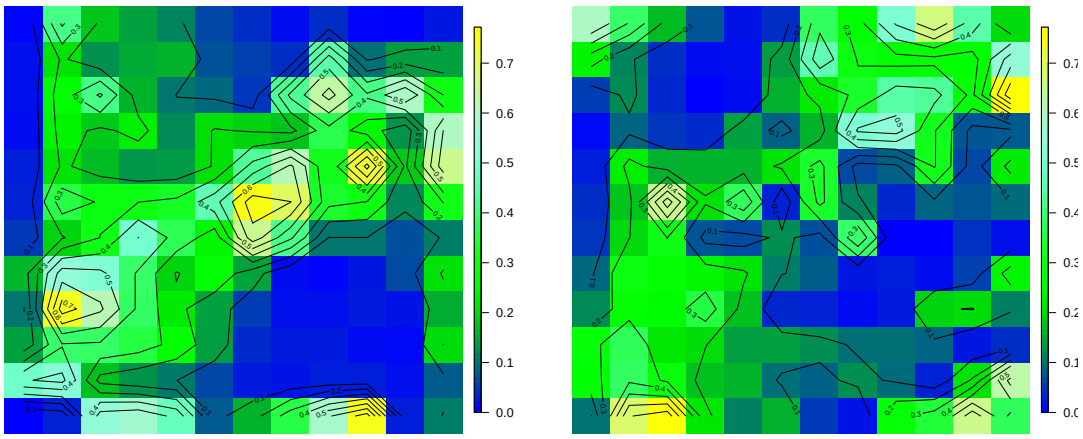

Figure 1.1: Daily maxima of rainfall measurements in Florida, September 6 and 7, 1999.

we even reject that extremal rainfall in the observation area behaves isotropically?"

1.2 Outline of the Thesis

The thesis is organized as follows:

In Chapter 2 we summarize basic theoretical statements and statistical tools of univariate and multivariate extreme value theory, so readers do not need specific knowledge and experience in that area. We introduce the *generalized extreme value distribution*, the *domain of attraction*, the *extreme value index* and present methods to estimate it. Further, we consider representations of multivariate extreme value distributions and ways to model and quantify the degree of extremal marginal dependence. Important in this context is the *Pickands dependence function*.

Chapter 3 introduces max-stable processes and presents examples, such as *Smith's storm profile model* and the construction of the Brown-Resnick process, which the rest of this thesis deals with. Its connection to Gaussian space-time processes is pointed out.

In the first part of Chapter 4 we discuss different models of the correlation function of the Gaussian space-time processes underlying the construction of the max-stable Brown-Resnick process. We particularly focus on anisotropic models and specify necessary assumptions and requirements. The second part of Chapter 4 considers methods of estimating spatial and temporal parameters of the Brown-Resnick process which are originally based on the correlation function of the Gaussian processes. The observation area of the data is assumed to be a regular grid. We adapt the procedure of *pairwise likelihood estimation*, introduced for isotropic models in Davis et al. [2012b], to the anisotropic setting. Furthermore we show that the properties of strong consistency and asymptotic normality of the pairwise likelihood estimates for an increasing number of grid and time points still hold.

On the basis of a simulation study, we show in Chapter 5 that the pairwise likelihood estimates improve when augmenting the number of time points and the side length of the grid the processes are simulated on. We use line plots, normal qq-plots and boxplots to confirm the theory of Chapter 4.

The last chapter, Chapter 6, treats real data consisting of radar rain measurements in Florida from 1999 to 2004. The same data set was already examined by Steinkohl [2012]. Applying our adapted pairwise likelihood method to daily maxima and hourly measurements on a chosen region of 120km×120km, we are interested in the question if the results change considerably compared to to the isotropic estimation. Based on *subsampling* methods, we construct asymptotic confidence intervals for the parameters and can therefore create a hypothesis test stating that isotropy in the extremal rainfall behaviour can be rejected. On top of that, we present and therefore "update" *conditional probability fields* and *return level plots* of certain reference locations chosen by Steinkohl [2012]. The plots try to answer questions like: "If the amount of rainfall in Tampa exceeds some level x, what is the probability that the amount of rainfall in Brandon exceeds the level y?"

1.3 Outlook and further Extensions

As stated in Section 1.2, we focus on anisotropic models for the correlation function of the Gaussian space-time processes underlying the construction of the max-stable Brown-Resnick process, that is, we allow directional preferences of the dependence structure. In this thesis, we assume and model anisotropy independently for each spatial dimension. However, a possible extension might be to permit interaction between the different directions. In Section 4.3, we get back to that issue and explain it in more detail.

Chapter 2

Introduction to Extreme Value Theory

This chapter provides a short introduction into both univariate and multivariate extreme value theory and, therefore, forms the mathematical basis of max-stable processes, which are dealt with in Chapter 3. The theoretical statements in this chapter are mainly based on Embrechts, Klüppelberg and Mikosch [1997] in the univariate case and Beirlant, Goegebeur, Segers and Teugels [2004], Sections 8 and 9, in the multivariate part.

2.1 Univariate Extreme Value Theory

2.1.1 Basic Setting

Let $X_1, \ldots, X_n$ be an i.i.d. sample with common distribution function F. The goal of extreme value theory is to find the limiting distribution of the sample maximum M_n, that is, we want to find real normalizing and centering sequences $(a_n)_{n\in\mathbb{N}}$, $a_n > 0$ for all n, and $(b_n)_{n\in\mathbb{N}}$ such that

$$a_n^{-1}(M_n - b_n) \xrightarrow{\mathcal{D}} Y, \qquad n \to \infty, \tag{2.1}$$

where $Y \sim G$ for a non-degenerate distribution function G. By Definition A.1 of convergence in distribution, one can rewrite (2.1) as

$$\mathbb{P}(a_n^{-1}(M_n - b_n) \leq x) \to G(x), \qquad n \to \infty, \tag{2.2}$$

or as

$$F^n(a_n x + b_n) \to G(x), \qquad n \to \infty, \tag{2.3}$$

for all points $x \in \text{supp}(F) := \{y \in \mathbb{R} : 0 < F(y) < 1\}$, using that $\mathbb{P}(M_n \leq x) = P(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x) = F^n(x)$.

Definition 2.1 (Univariate Extreme Value Distribution)**.** *If (2.1), (2.2) or (2.3) are satisfied we say that G is a* univariate extreme value distribution *and F is in its* maximum domain of attraction *and write $F \in \text{MDA}(G)$.*

The two fundamental questions are given as follows: What are possible limit distributions G in (2.2) and what conditions have to be imposed on F such that (2.2) holds?

2.1.2 Important Results and Theorems

Very important in the context of extreme value theory are *max-stable* distributions.

Definition 2.2 (Max-stable distributions, cf. Embrechts et al. [1997], Definition 3.2.1)**.** *For arbitrary $n \geq 2$ let X, $X_1, \ldots, X_n$ be i.i.d. with common distribution function G. Then G is called* max-stable *if there exist sequences (a_n) and $(b_n) \in \mathbb{R}$ with $a_n > 0$ for all n such that*

$$a_n^{-1}(M_n - b_n) \stackrel{\mathcal{D}}{=} X. \tag{2.4}$$

A fundamental result is that the set of extreme value distributions and max-stable distributions coincide. The Fisher-Tippett theorem (see for instance Embrechts et al. [1997], Theorem 3.2.3), which goes back to Fisher and Tippet [1928], states the possible types of an extreme value distribution function G.

Theorem 2.3 (Fisher-Tippett)**.** *An extreme value distribution G belongs, up to scaling and centering constants, to one of the following three families of distribution functions:*

(I) Fréchet distribution*:*

$$\Phi_\alpha(x) = \begin{cases} 0, & \text{if } x \leq 0, \\ \exp\left\{-x^{-\alpha}\right\}, & \text{if } x > 0 \end{cases}, \ \alpha > 0. \tag{2.5}$$

(II) Weibull distribution*:*

$$\Psi_\alpha(x) = \begin{cases} \exp\left\{-(-x)^{\alpha}\right\}, & \text{if } x \leq 0, \\ 1, & \text{if } x > 0 \end{cases}, \ \alpha > 0. \tag{2.6}$$

(III) Gumbel distribution*:*

$$\Lambda(x) = \exp\{-\mathrm{e}^{-x}\}, \qquad x \in \mathbb{R}. \tag{2.7}$$

For a univariate extreme value distribution G, there exists a finite-dimensional parametrization, the *Jenkinson-von Mises representation* (von Mises [1936] and Jenkinson [1955]), which covers all possible types listed in the Fisher-Tippett theorem above.

$$G(x) = G_\gamma(x) = \begin{cases} \exp\left\{-(1+\gamma x)^{-\frac{1}{\gamma}}\right\}, & \text{if } \gamma \neq 0, \\ \exp\left\{-\mathrm{e}^{-x}\right\}, & \text{if } \gamma = 0, \end{cases} \tag{2.8}$$

provided that $1 + \gamma x > 0$. Depending on γ, the support of G_γ is given by

$$\operatorname{supp}(G_\gamma) = \begin{cases} (-\gamma^{-1}, \infty), & \text{if } \gamma > 0, \\ (-\infty, -\gamma^{-1}), & \text{if } \gamma < 0, \\ \mathbb{R}, & \text{if } \gamma = 0. \end{cases} \tag{2.9}$$

The parameter γ is called the *extreme value index.* The case $\gamma = 0$ leads to the Gumbel family, if $\gamma > 0$ we speak of the Fréchet family with index $\alpha := \gamma^{-1}$ and the case $\gamma < 0$ results in the Weibull family with index $\alpha := (-\gamma)^{-1}$. The Jenkinson-von Mises representation is basically the answer to the first fundamental question about possible classes of an extreme value distribution. The second one, the domain-of-attraction problem, is solved by the *Pickands-Balkema-de Haan Theorem* (Pickands [1975] and Balkema and de Haan [1974]; see also Theorem 3.4.5 in Embrechts et al. [1997]).

Theorem 2.4 (Pickands-Balkema-de Haan)**.** *A distribution function F is in the maximum domain of attraction of an extreme value distribution function G_γ for $\gamma \in \mathbb{R}$ if and only if there exists a positive measurable function a such that for $1 + \gamma x > 0$,*

$$\lim_{u \to x_\star} \frac{1 - F(u + xa(u))}{1 - F(u)} = \begin{cases} (1+\gamma x)^{-\frac{1}{\gamma}}, & \text{if } \gamma \neq 0 \\ \mathrm{e}^{-x}, & \text{if } \gamma = 0, \end{cases} = 1 - \mathrm{GPD}_\gamma(x), \tag{2.10}$$

where $x_\star := \sup\{y \in \mathbb{R} : F(y) < 1\}$ is the right-end point of the distribution F and GPD_γ *is the* generalized Pareto distribution *with index $\gamma \in \mathbb{R}$.*

If $F \in \mathrm{MDA}(G_\gamma)$ then the normalizing and centering sequences (a_n) and (b_n) in (2.1) are chosen in dependence from γ. We have:

(I) If $\gamma > 0$ (Fréchet family), then $a_n = F^{\leftarrow}(1 - \frac{1}{n})$ and $b_n = 0$ for all $n \in \mathbb{N}$. The sequence (a_n) can be replaced with any sequence $(\tilde{a}_n)$ such that $1 - F(\tilde{a}_n) \sim \frac{1}{n}$ as $n \to \infty$.

(II) If $\gamma < 0$ (Weibull family), then $a_n = x_\star - F^{\leftarrow}(1 - \frac{1}{n})$ and $b_n = x_\star$ for all $n \in \mathbb{N}$. The sequence (a_n) can be replaced with any sequence $(\tilde{a}_n)$ such that $1 - F(x_\star - \tilde{a}_n) \sim \frac{1}{n}$ as $n \to \infty$.

(III) If $\gamma = 0$ (Gumbel family), then $a_n = a(b_n)$ and $b_n = F^{\leftarrow}(1 - \frac{1}{n})$ for all $n \in \mathbb{N}$. Here a is the function in (2.10).

For the respective proofs, we refer to Embrechts et al. [1997], Section 3.3.

2.1.3 Statistics of Univariate Extremes

In the following we present statistical methods in univariate extreme value theory. The first part of this subsection deals with estimators of the extreme value index γ of the extreme value distribution G_γ. The second part presents tail estimators of a distribution function F that is assumed to be in the maximum domain of attraction of G_γ.

Estimators of the extreme value index

Method of block maxima. Consider a sample of identically distributed (but not necessarily independent) random variables $X_1, \ldots, X_n$ for $n \in \mathbb{N}$. The method of block maxima

suggests to divide the set into k different groups of size m and take the maximum Y_i over each group $i = 1, \dots, k$, that is,

$$Y_i = \bigvee_{r=(i-1)m+1}^{im} X_r, \qquad i = 1, \dots, k, \tag{2.11}$$

where $m \in \mathbb{N}$ is chosen such that $km = n$ holds. For instance, if $m = 365$ and X_r is the amount of rainfall on day r for $r \in \{1, \dots, n\}$ at a certain fixed location then Y_i, $i \in \{1, \dots, k\}$, represents the maximum amount of rainfall in year i. The sample $Y_1, \dots, Y_k$ is assumed to be independent and identically distributed with an extreme value distribution function $G_{\gamma,\mu,\sigma}$ for $\mu, \gamma \in \mathbb{R}$ and $\sigma > 0$, given by (cf. (2.8))

$$G_{\gamma,\mu,\sigma}(x) = \exp\left\{-(1 + \gamma\frac{x-\mu}{\sigma})^{-\frac{1}{\gamma}}\right\}, \qquad x \in \mathbb{R}, \tag{2.12}$$

where $1+\gamma\frac{x-\mu}{\sigma} > 0$. The parameters μ and σ are scaling parameters. We set $\boldsymbol{\theta} := (\gamma, \mu, \sigma)$. Now assume to have observed a set of realizations $\{y_1, \dots, y_k\}$ of $Y_1, \dots, Y_k$ and let $g_{\boldsymbol{\theta}}$ denote the density of $G_{\boldsymbol{\theta}}$. Then, the likelihood L is equal to

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{k} g_{\boldsymbol{\theta}}(y_i)\mathbf{1}_{\{1+\gamma\frac{y_i-\mu}{\sigma}>0\}},$$

and the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n$ is obtained by maximizing L with respect to $\boldsymbol{\theta} \in \Theta$ for a suitable parameter space Θ. As the support of $L(\boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$ itself, the estimation is non-regular. However, Smith [1985] shows that $\hat{\boldsymbol{\theta}}_n$ is consistent and asymptotically normal if $\gamma > -\frac{1}{2}$.

Pickands' estimator. For $n \in \mathbb{N}$, consider a sample of i.i.d. random variables $X_1, \dots, X_n$ with common distribution function F. The non-parametric *Pickands' estimator* (Pickands [1975]) relies on the fact that if F is in the maximum domain of attraction of an extreme value distribution G_γ, then we have

$$\lim_{t\to\infty} \frac{F^{\leftarrow}(1-\frac{1}{2t}) - F^{\leftarrow}(1-\frac{1}{t})}{F^{\leftarrow}(1-\frac{1}{t}) - F^{\leftarrow}(1-\frac{2}{t})} = 2^{\gamma}, \tag{2.13}$$

see for instance Embrechts et al. [1997], Theorem 3.4.5. (c). Under the assumption that F is continuous and replacing F by the empirical distribution function $\hat{F}_n$, given by

$$\begin{aligned}\hat{F}_n(t) &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \le t\}}, \qquad t \in \mathbb{R} \\ &= X_{(k)}, \qquad t \in (\frac{k-1}{n}, \frac{k}{n}],\end{aligned}$$

where $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ denotes the ordered sample, equation (2.13) yields

$$\gamma \approx \frac{1}{\log 2}\log\frac{\hat{F}_n^{\leftarrow}(1-\frac{1}{2}\frac{2k}{n}) - \hat{F}_n^{\leftarrow}(1-\frac{2k}{n})}{\hat{F}_n^{\leftarrow}(1-\frac{2k}{n}) - \hat{F}_n^{\leftarrow}(1-2\frac{2k}{n})} = \frac{1}{\log 2}\log\frac{X_{(n-k)} - X_{(n-2k)}}{X_{(n-2k)} - X_{n-4k)}},$$

where t is replaced by $\frac{n}{2k}$ for $k = k_n < n$ such that $\frac{k}{n} \to 0$ as $n \to \infty$. For stability reasons Pickands' estimator is chosen as

$$\hat{\gamma}_k^{(P)} := \frac{1}{\log 2} \log \frac{X_{(n-k+1)} - X_{(n-2k+1)}}{X_{(n-2k+1)} - X_{n-4k+1)}}. \tag{2.14}$$

For debates on how to choose k, see for example Embrechts et al. [1997], Section 6.4. If $k = k_n \to \infty$ and $\frac{k}{n} \to 0$ as $n \to \infty$, Pickands' estimator $\hat{\gamma}_k^{(P)}$ fulfills the properties of consistency and asymptotic normality, see for instance Embrechts et al. [1997], Theorem 6.4.1.

Tail Estimators

In the following, we suppose again that the distribution function F is in the maximum domain of attraction of an extreme value distribution G_γ. The purpose of estimating the tail $\bar{F} := 1 - F$ of F is that it is accompanied by a p-quantile estimator x_p, $p \in (0, 1)$, through the equation $1 - p = \bar{F}(x_p)$. We present a very popular method, the *Peaks Over Threshold* (POT) method.

Peaks Over Threshold. The starting point of the POT method is equation (2.10) of the Pickands-Balkema-de Haan Theorem: For a random variable $X \sim F$ and a positive measurable function a we have

$$\lim_{u \to x_\star} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \lim_{u \to x_\star} \mathbb{P}\left(\frac{X - u}{a(u)} > x | X > u \right) = \lim_{u \to x_\star} \bar{F}_u(a(u)x) \overset{(2.10)}{=} \overline{\mathrm{GPD}}_\gamma(x), \tag{2.15}$$

where $F_u(y) := \mathbb{P}(X \leq u + y | X > u)$ for $y > 0$ is the *excess distribution function.* Notice that $\bar{F}_u(y) = \mathbb{P}(X > u + y | X > u) = \frac{\mathbb{P}(X > u+y)}{\mathbb{P}(X>u)} = \frac{\bar{F}(u+y)}{\bar{F}(u)}$, and therefore

$$\bar{F}(u + y) = \bar{F}(u)\bar{F}_u(y). \tag{2.16}$$

Now the POT procedure is as follows:

(i) Choose a rather high threshold u in such a way that $\bar{F}(u)$ can still be estimated by the empirical tail

$$\hat{\bar{F}}_n(u) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i > u\}} =: \frac{N_u}{n}. \tag{2.17}$$

(ii) Use the *GPD approximation* based on (2.15):

$$\begin{aligned} 0 &= \lim_{u \to x_\star} \sup_{0 < xa(u) < x_\star - u} \left| \bar{F}_u(xa(u)) - \overline{\mathrm{GPD}}_\gamma(x) \right| \\ &\overset{y := xa(u)}{=} \lim_{u \to x_\star} \sup_{y < x_\star - u} \left| \bar{F}_u(y) - \overline{\mathrm{GPD}}_\gamma\left(\frac{y}{a(u)} \right) \right| \end{aligned} \tag{2.18}$$

As u and therefore $a(u)$ are fixed, set $\beta := a(u)$ and interpret β as a scaling parameter. This yields

$$\overline{\mathrm{GPD}}_\gamma\left(\frac{y}{a(u)}\right) = \begin{cases} \left(1+\gamma\frac{y}{a(u)}\right)^{-\frac{1}{\gamma}} = \left(1+\gamma\frac{y}{\beta}\right)^{-\frac{1}{\gamma}}, & \text{if } \gamma \neq 0, \\ \exp\left\{-\frac{y}{a(u)}\right\} = \exp\left\{-\frac{y}{\beta}\right\}, & \text{if } \gamma = 0. \end{cases} = \overline{\mathrm{GPD}}_{\gamma,\beta}(y). \tag{2.19}$$

(iii) For $i \in \{1,\dots,n\}$ such that $X_i > u$ (i.e. the ith observation is an exceedance), define $Y_i := X_i - u$ as the excess. As N_u denotes the number of exceedances, the excesses are given by $Y_1,\dots,Y_{N_u}$. Based on the GPD approximation, one now assumes they are exactly GPD distributed and estimates the parameters of $\overline{\mathrm{GPD}}_{\gamma,\beta}$ by maximum likelihood estimation. We denote the corresponding estimators by $\hat{\gamma}$ and $\hat{\beta}$.

(iv) Using the estimators obtained in (iii), it follows from (2.16) and (2.17) that an estimator of the tail $\bar{F}$ is given by

$$\hat{\bar{F}}(u+y) = \frac{N_u}{n}\overline{\mathrm{GPD}}_{\hat{\gamma},\hat{\beta}}(y) \text{ for } y > 0. \tag{2.20}$$

(v) A remaining problem is the choice of u. One possibility is to consider the *mean excess function* of $\overline{\mathrm{GPD}}_{\gamma,\beta}$.

Definition 2.5 (Mean Excess Function, cf. Embrechts et al. [1997], Def. 6.2.3)**.** *Let H be a distribution function. Then its* mean excess function e_H *is given by*

$$\mathrm{e}_H(u) := \mathbb{E}_H\left[X-u | X > u\right], \qquad 0 \leq u < \sup\{y \in \mathbb{R} : 0 < H(y) < 1\}, \tag{2.21}$$

where X is a random variable with distribution function H.

The mean excess function of $\overline{\mathrm{GPD}}_{\gamma,\beta}$ turns out to be linear in u. As for large u it holds that $\bar{F}_u(y) \approx \overline{\mathrm{GPD}}_{\gamma,\beta}(y)$, one can choose u so high that the empirical mean excess function e_n, given by

$$\mathrm{e}_n(u) = \frac{1}{N_u}\sum_{i=1}^{n}(X_i-u)\mathbf{1}_{\{X_i>u\}},$$

is approximately linear for values larger than u.

The following section deals with the generalization of extreme value theory of one-dimensional to multivariate samples. The main change is due to the structure of the dependence between the respective margins of the multivariate distribution under consideration.

2.2 Multivariate Extreme Value Theory

For the definitions of vector operations, such as the componentwise maximum of two or more vectors, we refer to the list of notations and conventions that can be found in Section A.1.

2.2.1 Basic Setting

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be an i.i.d. d-variate sample with common distribution function F. As in the univariate case, the aim of multivariate extreme value theory is to find the limiting distribution of the pointwise sample maximum $\boldsymbol{M}_n$, that is, we want to find normalizing and centering sequences $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$, $\boldsymbol{a}_n > \boldsymbol{0}$ for all n, and $(\boldsymbol{b}_n)_{n\in\mathbb{N}}$ such that

$$\boldsymbol{a}_n^{-1}(\boldsymbol{M}_n - \boldsymbol{b}_n) \xrightarrow{\mathcal{D}} \boldsymbol{Y}, \qquad n \to \infty, \tag{2.22}$$

where $\boldsymbol{Y} \sim G$ for a non-degenerate d-variate distribution function G. By the definition of convergence in distribution and exploiting that G is continuous because its margins are continuous, one can rewrite (2.22) as

$$\mathbb{P}(\boldsymbol{a}_n^{-1}(\boldsymbol{M}_n - \boldsymbol{b}_n) \leq \boldsymbol{x}) \to G(\boldsymbol{x}), \qquad n \to \infty, \tag{2.23}$$

or as

$$F^n(\boldsymbol{a}_n\boldsymbol{x} + \boldsymbol{b}_n) \to G(\boldsymbol{x}), \qquad n \to \infty, \tag{2.24}$$

for all points $\boldsymbol{x} \in \text{supp}(F) = \{\boldsymbol{y} \in \mathbb{R}^d : 0 < F(\boldsymbol{y}) < 1\}$, using that $\mathbb{P}(\boldsymbol{M}_n \leq \boldsymbol{x}) = F^n(\boldsymbol{x})$.

Definition 2.6 (Multivariate Extreme Value Distribution)**.** *If (2.22), (2.23) or (2.24) are satisfied we say that G is a* multivariate extreme value distribution *and F is in its* maximum domain of attraction *and write as $F \in \text{MDA}(G)$.*

As the margins of a converging multivariate distribution function necessarily converge to the respective margins of the limit function, one should note that if $F \in \text{MDA}(G)$ then it follows that for all $j = 1 \ldots d$ we have $F_j \in \text{MDA}(G_j)$ and G_j is a univariate extreme value distribution. Here F_j and G_j denote the marginal distribution functions of F and G, respectively. Recall that one can therefore parametrize G_j for all $j = 1 \ldots d$ using the usual univariate parametrization (see for example Embrechts et al. [1997], page 294):

$$G_j(x_j) = G_j^{\gamma_j}(x_j) = \begin{cases} \exp\left(-(1+\gamma_j\frac{x_j-\mu_j}{\sigma_j})^{-\frac{1}{\gamma_j}}\right), & \text{if } \gamma_j \neq 0 \\ \exp\left(-\exp(-\frac{x_j-\mu_j}{\sigma_j})\right), & \text{if } \gamma_j = 0, \end{cases} \qquad j = 1, \ldots, d, \tag{2.25}$$

provided that $1 + \gamma_j x > 0$. Here $\mu_j \in \mathbb{R}$ and $\sigma_j > 0$ denote the centering and normalizing parameters, respectively. According to the respective distributions, the case $\gamma_j = 0$ leads to the Gumbel family, if $\gamma_j > 0$ we speak of the Fréchet family and the case $\gamma_j < 0$ results in the Weibull family. As in univariate extreme value theory, an important concept in the multivariate case is the concept of *max-stability*, see for instance Beirlant et al. [2004], Section 8.2.1.

Definition 2.7 (Max-Stability)**.** *A $d-$variate distribution function G is called* max-stable *if the following condition is satisfied: For all $k \in \mathbb{N}_+$ there exist vectors $\boldsymbol{\alpha}_k > \boldsymbol{0}$ and $\boldsymbol{\beta}_k$ such that*

$$G^k(\boldsymbol{\alpha}_k\boldsymbol{x} + \boldsymbol{\beta}_k) = G(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^d. \tag{2.26}$$

The interpretation is that for an i.i.d. sample $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_k$ with common max-stable distribution function G there are centering and normalizing vectors $\boldsymbol{\beta}_k$ and $\boldsymbol{\alpha}_k$ such that applying them to $\boldsymbol{M}_k(\boldsymbol{Y}) := \bigvee_{i=1}^{k} \boldsymbol{Y}_i$ we obtain a random vector whose distribution is again G, i.e.

$$\boldsymbol{\alpha}_k^{-1} \left(\boldsymbol{M}_k(\boldsymbol{Y}) - \boldsymbol{\beta}_k \right) \stackrel{\mathcal{D}}{=} \boldsymbol{Y}_1, \qquad \text{for all } k \in \mathbb{N}_+.$$

Recall that in the univariate case the classes of max-stable distributions and of non-degenerate extreme value distributions coincide, see for example Embrechts et al. [1997], Theorem 3.2.2. This also holds for the multivariate setting, for a proof see Beirlant et al. [2004], Section 8.2.2.
The difference between univariate and multivariate extreme value statistics mainly lies in the fact that in addition to modelling the distribution of the one-dimensional margins (for which the univariate techniques can be used), one has to model the class of limiting dependence structures between them, which is not possible with the help of a finite-dimensional parametric family like the one in (2.25). The following subsection presents representations for multivariate extreme value distributions, using models that try to capture the extremal dependence structure.

2.2.2 Representations of Multivariate Extreme Value Distributions

For convenience let G be a d-variate extreme value distribution with standard Fréchet margins, i.e. for all $j = 1 \ldots d$ we have $\mathbb{P}(G_j \leq z_j) = \mathrm{e}^{-\frac{1}{z_j}}$ for $0 < z_j < \infty$. If a d-variate extreme value distribution $\tilde{G}$ does not have standard Fréchet margins, one can standardize those by setting

$$G(\boldsymbol{z}) = \tilde{G}\left(\tilde{G_1}^{\leftarrow}\left(\mathrm{e}^{-\frac{1}{z_1}} \right), \ldots, \tilde{G_d}^{\leftarrow}\left(\mathrm{e}^{-\frac{1}{z_d}} \right) \right), \qquad \boldsymbol{z} = (z_1, \ldots, z_d) \in (\boldsymbol{0}, \boldsymbol{\infty}). \tag{2.27}$$

This standardization relies on the so-called *probability integral transform* method, according to which for a random variable Y with continuous distribution function F and a random variable U that is uniformly distributed on the interval $(0, 1)$ we have:

(i) $F(Y) \stackrel{\mathcal{D}}{=} U$ and

(ii) $F^{\leftarrow}(U) \stackrel{\mathcal{D}}{=} Y$.

To obtain (2.27) one has to apply the method as follows:
Let $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ be a random vector with $Y_j \sim \tilde{G}_j$ and set $U_j := \tilde{G}_j(Y_j) \stackrel{\text{(i)}}{\sim}$ Unif(0,1) for all j. The inverse of the standard Fréchet distribution function is given by $t \mapsto -\log(t)^{-1}$ for $t \in (0, 1)$. So by (ii) we have that the distribution of the random vector $(-\log(U_1)^{-1}, \ldots, -\log(U_d)^{-1})$ has standard Fréchet margins. Furthermore observe

that for $\boldsymbol{z} = (z_1, \ldots, z_d) \in (\mathbf{0}, \boldsymbol{\infty})$ we get

$$\begin{aligned}
&\mathbb{P}(-\log(U_1)^{-1} \leq z_1, \ldots, -\log(U_d)^{-1} \leq z_d) \\
&= \mathbb{P}\left[-\log\left\{\tilde{G}_1(Y_1)\right\}^{-1} \leq z_1, \ldots, -\log\left\{\tilde{G}_d(Y_d)\right\}^{-1} \leq z_d\right] \\
&= \mathbb{P}\left[Y_1 \leq \exp\left(-\frac{1}{z_1}\right), \ldots, Y_d \leq \exp\left(-\frac{1}{z_d}\right)\right] \\
&= G(\boldsymbol{z})
\end{aligned}$$

with $G(\boldsymbol{z})$ in (2.27).
A max-stable distribution function (i.e. an extreme value distribution) can be represented via an *exponent measure*. A proof for the two-dimensional case can be found in Balkema and Resnick [1977] (Theorem 3).

Theorem 2.8. *A d-variate distribution function G with standard Fréchet margins is max-stable if and only if there exists a σ-finite measure μ on $[\mathbf{0}, \boldsymbol{\infty})\backslash\{\mathbf{0}\}$ such that*

$$G(\boldsymbol{z}) = \exp\left\{-\mu([\mathbf{0}, \boldsymbol{z}]^c)\right\}, \qquad \boldsymbol{z} \in [\mathbf{0}, \boldsymbol{\infty}]. \tag{2.28}$$

where for a set $E \subset \mathbb{R}^d$ we denote by E^c its complement in $\mathbb{R}^d$.

Definition 2.9 (Exponent Measure)**.** *The measure μ in (2.28) is called* exponent measure.

Remark 2.10 (Homogeneity relation of the exponent measure)**.** *From the max-stability of G one can conclude the following homogeneity relation of μ:*

$$\mu(sB) = s^{-1}\mu(B), \qquad 0 < s < \infty,\ B \in \mathcal{B}([\mathbf{0}, \boldsymbol{\infty})\backslash\{\mathbf{0}\}), \tag{2.29}$$

where $\mathcal{B}([\mathbf{0}, \boldsymbol{\infty})\backslash\{\mathbf{0}\})$ denotes the set of all Borel subsets of $[\mathbf{0}, \boldsymbol{\infty})\backslash\{\mathbf{0}\}$.

By means of the exponent measure we can define the so-called *stable tail dependence function* (see for example Beirlant et al. [2004], page 257).

Definition 2.11 (Stable Tail Dependence Function)**.** *The* stable tail dependence function *of a multivariate extreme value distribution G with exponent measure μ is defined by*

$$\begin{aligned}
l(\boldsymbol{v}) =& \mu([\mathbf{0}, \boldsymbol{v}^{-1}]^c) && (2.30)\\
=& -\log G(\boldsymbol{v}^{-1}), \qquad \boldsymbol{v} = (v_1, \ldots, v_d) \in [\mathbf{0}, \boldsymbol{\infty}]. && (2.31)
\end{aligned}$$

Notice that it follows from (2.27) that in terms of the unstandardized distribution function $\tilde{G}$ the stable tail dependence function l can be expressed by

$$l(\boldsymbol{v}) = -\log \tilde{G}\left(\tilde{G}_1^{\leftarrow}(\exp(-v_1)), \ldots, \tilde{G}_d^{\leftarrow}(\exp(-v_d))\right), \qquad \boldsymbol{v} = (v_1, \ldots, v_d) \in [\mathbf{0}, \boldsymbol{\infty}]. \tag{2.32}$$

Proposition 2.12 (Properties of the stable tail dependence function)**.** *The stable tail dependence function l as defined in (2.30) has the following properties (cf. Beirlant et al. [2004]):*

(i) $l(s\mathbf{v}) = sl(\boldsymbol{v})$ *for* $0 < s < \infty$, $\boldsymbol{v} \in [\mathbf{0}, \boldsymbol{\infty}]$.

(ii) $l(\boldsymbol{e}_j) = 1$ *for* $j = 1 \dots d$ *where* $\boldsymbol{e}_j$ *is the jth unit vector in* $\mathbb{R}^d$.

(iii) $\bigvee_{j=1}^{d} v_j \leq l(\boldsymbol{v}) \leq \sum_{j=1}^{d} v_j$ *for* $\boldsymbol{v} \in [\mathbf{0}, \boldsymbol{\infty})$.

(iv) l *is convex.*

If l has the form of the lower bound in property (iii) then the margins of G are completely dependent and if l has the form of the upper bound then they are independent. In the following we will use pseudo-polar coordinates to introduce the *spectral measure* of the exponent measure μ of the multivariate extreme value distribution G. We stick close to Beirlant et al. [2004], Section 8.2.3. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two arbitrary norms on $\mathbb{R}^d$ and let $\mathcal{S}_2 := \{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{u}\|_2 = 1\}$ denote the unit sphere with respect to $\|\cdot\|_2$. Furthermore we denote by $r := \|\boldsymbol{z}\|_1$ and $\boldsymbol{u} := \frac{\boldsymbol{z}}{\|\boldsymbol{z}\|_2}$ the radial and the angular part of the vector $\boldsymbol{z} \in \mathbb{R}^d \backslash \{\mathbf{0}\}$, respectively, and we define T to be the function that maps $\boldsymbol{z}$ onto r and $\boldsymbol{u}$, i.e.

$$T : \mathbb{R}^d \backslash \{\mathbf{0}\} \to (0, \infty) \times \mathcal{S}_2, \quad \boldsymbol{z} \mapsto (r, \boldsymbol{u}). \tag{2.33}$$

The mapping T is invertible and the inverse is given by $T^{-1}(r, \boldsymbol{u}) = \frac{r\mathbf{u}}{\|\boldsymbol{u}\|_1}$. In the context of this setting we can now define the spectral measure S on the set $\mathscr{U} := \mathcal{S}_2 \cap [\mathbf{0}, \boldsymbol{\infty})$.

Definition 2.13 (Spectral Measure)**.** *The spectral measure* S *on* $\mathscr{U}$ *is defined by*

$$S(B) := \mu\left(\left\{\boldsymbol{z} \in [\mathbf{0}, \boldsymbol{\infty}) : \|\boldsymbol{z}\|_1 \geq 1, \frac{\boldsymbol{z}}{\|\boldsymbol{z}\|_2} \in B\right\}\right), \qquad B \in \mathcal{B}(\mathscr{U}). \tag{2.34}$$

Example 2.14. *Let* $d = 2$ *and* $\|\cdot\|_1 = \|\cdot\|_2 = \|\cdot\|_\sum$ *where* $\|\cdot\|_\sum$ *is the sum-norm, that is* $\|\cdot\|_\sum(\boldsymbol{x}) = \sum_{j=1}^{d} x_j$ *for* $\boldsymbol{x} \in \mathbb{R}^d$. *Then the set* $\mathscr{U}$ *is given by*

$$\mathscr{U} = \{\boldsymbol{u} \in [\mathbf{0}, \boldsymbol{\infty}) : u_1 + u_2 = 1\} = S_2,$$

where S_2 *is called the unit simplex.*

An important representation of the multivariate extreme value distribution G and its unstandardized analogue $\tilde{G}$ is based on the spectral measure S.

Theorem 2.15. *A multivariate distribution function* G *with standard Fréchet margins is a multivariate extreme value distribution function if anf only if* G *possesses the representation*

$$G(\boldsymbol{z}) = \exp\left\{-\int_{\mathscr{U}} \bigvee_{j=1}^{d} \left(\frac{u_j}{\|\boldsymbol{u}\|_1} \frac{1}{z_j}\right) S(\mathrm{d}\mathbf{u})\right\}, \qquad \boldsymbol{z} \in [\mathbf{0}, \boldsymbol{\infty}], \tag{2.35}$$

with $\mathscr{U} = \mathcal{S}_2 \cap [\mathbf{0}, \boldsymbol{\infty}]$ *and* S *defined in (2.34) such that*

$$\int_{\mathscr{U}} \frac{u_j}{\|\boldsymbol{u}\|_1} S(\mathrm{d}\mathbf{u}) = 1, \qquad j = 1 \dots d. \tag{2.36}$$

This representation was proved by de Haan and Resnick [1977].

Remark 2.16. *In terms of the unstandardized distribution $\tilde{G}$, (2.35) reads as*

$$\tilde{G}(\boldsymbol{x}) = \exp\left\{ \int_{\mathscr{U}} \bigvee_{j=1}^{d} \left(\frac{u_j}{\|\boldsymbol{u}\|_1} \log \tilde{G}_j(x_j) \right) S(\mathrm{d}\mathbf{u}) \right\}, \qquad \boldsymbol{x} \in \mathbb{R}^d. \tag{2.37}$$

*Proof of the necessary part (**Sketch**, cf. Beirlant et al. [2004], pages 258/259).* First we derive the *spectral decomposition* (2.38) (de Haan and Resnick [1977]) of the exponent measure μ. From the homogeneity relation (2.29) we can conclude that

$$\mu\left(\left\{ \boldsymbol{z} \in [\mathbf{0}, \boldsymbol{\infty}) : \|\boldsymbol{z}\|_1 \geq r, \frac{\boldsymbol{z}}{\|\boldsymbol{z}\|_2} \in B \right\} \right) = r^{-1} S(B), \qquad 0 < r < \infty,\ B \in \mathcal{B}(\mathscr{U}).$$

This implies that μ can be written in terms of polar coordinates as a product of measures:

$$\mu \circ T^{-1}(\mathrm{d}r, \mathrm{d}\boldsymbol{u}) = r^{-2} \mathrm{d}r S(\mathrm{d}\boldsymbol{u}). \tag{2.38}$$

It follows that for any measurable function $g : [\mathbf{0}, \boldsymbol{\infty}) \backslash \{\mathbf{0}\} \to \mathbb{R}$ we have:

$$\begin{aligned} \int_{[\mathbf{0},\boldsymbol{\infty})\backslash\{\mathbf{0}\}} g(\boldsymbol{z}) \mu(\mathrm{d}\boldsymbol{z}) &= \int_{\mathscr{U}} \int_{0}^{\infty} g\left(\frac{r\mathbf{u}}{\|\boldsymbol{u}\|_1} \right) r^{-2} \mathrm{d}r S(\mathrm{d}\boldsymbol{u}) \\ &= \int_{\mathscr{U}} \int_{0}^{\infty} g(r\mathbf{u}) r^{-2} \mathrm{d}r \|\boldsymbol{u}\|_1^{-1} S(\mathrm{d}\boldsymbol{u}). \end{aligned} \tag{2.39}$$

A consequence is that

$$\begin{aligned} -\log G(\boldsymbol{z}) &= \int_{[\mathbf{0},\boldsymbol{\infty})\backslash\{\mathbf{0}\}} \mathbf{1}\left\{ \bigvee_{j=1}^{d} \frac{y_j}{z_j} > 1 \right\} \mu(\mathrm{d}\boldsymbol{y}) \\ &\overset{(2.39)}{=} \int_{\mathscr{U}} \bigvee_{j=1}^{d} \left(\frac{u_j}{\|\boldsymbol{u}\|_1} \frac{1}{z_j} \right) S(d\mathbf{u}), \qquad \boldsymbol{z} \in [\mathbf{0}, \boldsymbol{\infty}], \end{aligned}$$

where

$$\int_{\mathscr{U}} \frac{u_j}{\|\boldsymbol{u}\|_1} S(\mathrm{d}\boldsymbol{u}) = 1, \qquad \text{for all } j = 1, \ldots, d.$$

□

Right at the beginning of this section we could have chosen other margins for the multivariate extreme value distribution G, for instance standard extremal Weibull margins,

that is $G_j(z_j) = \mathrm{e}^{z_j}$ for $z_j < 0$, $j = 1, \ldots, d$. That choice leads to the following *Pickands* representation of G (see Pickands [1981]):

$$G(\boldsymbol{z}) = \exp\left\{ \int_{\mathscr{U}} \bigvee_{j=1}^{d} u_j z_j S(\mathrm{d}\boldsymbol{u}) \right\}, \qquad \boldsymbol{z} < \boldsymbol{0}, \tag{2.40}$$

such that

$$\int_{\mathscr{U}} w_j S(\mathrm{d}\boldsymbol{u}) = 1, \qquad \text{for all } j = 1, \ldots, d.$$

2.2.3 Independence and Complete Dependence

As mentioned at the beginning of this chapter, an important topic in multivariate extreme value theory is the degree of dependence between the margins of a multivariate extreme value distribution G.

Recall that G is independent if it factorizes into its margins G_j, $j = 1, \ldots, d$, i.e.

$$G(\boldsymbol{x}) = \prod_{j=1}^{d} G_j(x_j), \qquad \boldsymbol{x} \in \mathbb{R}^d.$$

In terms of the spectral measure S of G (cf. (2.37)) this means that for any S-integrable function $f : \mathscr{U} \to \mathbb{R}$ we have

$$\int_{\mathscr{U}} f(\boldsymbol{u}) S(\mathrm{d}\boldsymbol{u}) = \sum_{j=1}^{d} \|\boldsymbol{e}_j\|_1 f\left(\frac{\boldsymbol{e}_j}{\|\boldsymbol{e}_j\|_2} \right), \tag{2.41}$$

that is, S is made up of point masses of size $\|\boldsymbol{e}_j\|_1$ at points $\frac{\boldsymbol{e}_j}{\|\boldsymbol{e}_j\|_2}$ (see Beirlant et al. [2004], Section 8.2.3).
Complete dependence of G occurs if

$$G(\boldsymbol{x}) = \bigvee_{j=1}^{d} G_j(x_j), \qquad \boldsymbol{x} \in \mathbb{R}^d.$$

Let $\boldsymbol{u}_0 = (u_0, \ldots, u_0) \in \mathscr{U} \cap \{\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d : x_j = x_k \text{ for all } j, k = 1, \ldots, d\}$. For example, taking $d = 2$ and $\|\cdot\|_1 = \|\cdot\|_2 = \|\cdot\|_\sum$ (see Example 1 above) yields $\boldsymbol{u}_0 = (\frac{1}{2}, \frac{1}{2})$. Then, independence in terms of S expresses as

$$\int_{\mathscr{U}} f(\boldsymbol{u}) S(\mathrm{d}\boldsymbol{u}) = \frac{\|\boldsymbol{u}_0\|_1}{\boldsymbol{u}_0} f(\boldsymbol{u}_0) \tag{2.42}$$

for a function f as in (2.41). In that way S results in one single point mass of size $\frac{\|\boldsymbol{u}_0\|_1}{\boldsymbol{u}_0}$ at the point $\boldsymbol{u}_0$ (see again Beirlant et al. [2004], Section 8.2.3).

2.2.4 The bivariate Case

In the following we consider the special case $d = 2$, that is, the bivariate case. Furthermore we choose the norms $\|\cdot\|_1 = \|\cdot\|_2 = \|\cdot\|_\sum$ as in Example 1. Recall that under these conditions the set $\mathscr{U}$ is equal to the unit simplex $S_2 = \{(t_1, t_2) \in [0,1]^2 : t_1 + t_2 = 1\}$. A useful tool to measure the dependence between the margins of a bivariate extreme value distribution is its *Pickands dependence function*. It can be regarded as the restriction of the stable tail dependence function, defined in Definition 2.11, to S_2.

Definition 2.17 (Pickands Dependence Function (Pickands [1981]))**.** *Let G be a bivariate extreme value distribution function with standard Fréchet margins and l its stable tail dependence function given by*

$$l(v_1, v_2) = -\log G\left(\frac{1}{v_1}, \frac{1}{v_2}\right), \qquad 0 \le v_1, v_2 \le \infty.$$

The Pickands dependence function A of G is defined as

$$A(t) = l(1-t, t), \qquad t \in [0,1]. \tag{2.43}$$

Observe that not only is A determined by l, but the opposite holds as well: Let $0 \le v_1, v_2 \le \infty$ such that $v_1 + v_2 > 0$. Then

$$\begin{aligned} l(v_1, v_2) &\overset{\text{Prop. 2.12 } (i)}{=} (v_1 + v_2) l\left(\frac{v_1}{v_1+v_2}, \frac{v_2}{v_1+v_2}\right) \\ &= (v_1 + v_2) l\left(1 - \frac{v_2}{v_1+v_2}, \frac{v_2}{v_1+v_2}\right) \\ &= (v_1 + v_2) A\left(\frac{v_2}{v_1+v_2}\right). \end{aligned} \tag{2.44}$$

Remark 2.18. *The Pickands dependence function can alternatively be defined as*

$$A(t) = l(t, 1-t), \qquad t \in [0,1]. \tag{2.45}$$

From the properties of the stable tail dependence function (see Proposition 2.12) we can conclude the following properties of a Pickands dependence function.

Proposition 2.19 (Properties of a Pickands Dependence Function, cf. Beirlant et al. [2004], page 268)**.** *A Pickands dependence function A satisfies the following two properties:*

(i) $(1-t) \vee t \le 1$ *for* $t \in [0,1]$.

(ii) A is convex.

Again, the upper bound and the lower bound in (i) are related to independence and complete dependence of the margins of the underlying extreme value distribution, respectively. The representation of G in terms of the stable tail dependence function l can be

used to derive a representation in terms of the Pickands dependence function A. We have that for $0 \leq z_1, z_2 \leq \infty$

$$\begin{aligned} G(z_1, z_2) &\overset{(2.31)}{=} \exp\left\{-l\left(\frac{1}{z_1}, \frac{1}{z_2}\right)\right\} \\ &= \exp\left\{-\frac{z_1 + z_2}{z_1 z_2} A\left(\frac{z_1}{z_1 + z_2}\right)\right\}. \end{aligned} \tag{2.46}$$

The relation (2.27) can be used to express (2.46) in terms of the unstandardized distribution function $\tilde{G}$:

$$\begin{aligned} \tilde{G}(x_1, x_2) &\overset{(2.27)}{=} G\left(-\frac{1}{\log \tilde{G}_1(x_1)}, -\frac{1}{\log \tilde{G}_2(x_2)}\right) \\ &\overset{(2.46)}{=} \exp\left\{\left[\log \tilde{G}_1(x_1) + \log \tilde{G}_2(x_2)\right] A\left(\frac{\log \tilde{G}_2(x_2)}{\log \tilde{G}_1(x_2) + \log \tilde{G}_2(x_2)}\right)\right\} \\ &= \exp\left\{\log\left[\tilde{G}_1(x_1)\tilde{G}_2(x_2)\right] A\left(\frac{\log \tilde{G}_2(x_2)}{\log\left[\tilde{G}_1(x_2)\tilde{G}_2(x_2)\right]}\right)\right\}. \end{aligned} \tag{2.47}$$

Theorem 2.20 (Connection between the Pickands dependence function and the spectral measure)**.** *In the current context, let G be a multivariate extreme value distribution function with standard Fréchet margins, S the corresponding spectral measure and l and A its stable tail dependence function and Pickands dependence function, respectively. Then A possesses the following representation:*

$$A(t) = 1 - t + \int_{(0,t]} S([0, u])\mathrm{d}u, \qquad t \in [0, 1]. \tag{2.48}$$

Proof. Recall that l is given by $l(v_1, v_2) = -\log G(v_1^{-1}, v_2^{-1})$ for $0 \leq v_1, v_2 \leq \infty$. So by the de Haan-Resnick representation (2.35) we conclude

$$\begin{aligned} l(v_1, v_2) &= \int_{S_2} \frac{u_1}{u_1 + u_2} v_1 \vee \frac{u_2}{u_1 + u_2} v_2 \; S(\mathrm{d}(u_1, u_2)) \\ &= \int_{S_2} \frac{1}{u_1 + u_2} \left(u_1 v_1 \vee u_2 v_2\right) \; S(\mathrm{d}(u_1, u_2)) \end{aligned} \tag{2.49}$$

where, according to (2.36), the spectral measure S fulfills

$$\int_{S_2} \frac{u_j}{u_1 + u_2} S(\mathrm{d}(u_1, u_2)) = \int_{S_2} u_j S(\mathrm{d}(u_1, u_2)) = 1 \text{ for } j \in \{1, 2\}.$$

In the bivariate setting we can identify $(u_1, u_2) \in S_2 = \{(u_1, u_2) \in [0, 1]^2 : u_1 + u_2 = 1\}$ via $(u, 1 - u)$ with $u = u_1$ (or $u = u_2$). So (2.49) turns into

$$l(v_1, v_2) = \int_{[0,1]} \left(u v_1 \vee (1 - u) v_2\right) S(\mathrm{d}u).$$

where S satisfies

$$\int_{[0,1]} uS(\mathrm{d}u) = \int_{[0,1]} (1-u)S(\mathrm{d}u) = 1. \tag{2.50}$$

Now the definition of the Pickands dependence function A yields

$$A(t) = l(1-t,t) = \int_{[0,1]} (u(1-t) \vee (1-u)t)\, S(\mathrm{d}u) \tag{2.51}$$

$$= t \int_{[0,1]} (1-u)S(\mathrm{d}u) + (1-t) \int_{[0,1]} uS(\mathrm{d}u), \qquad t \in [0,1], \tag{2.52}$$

and by (2.50) we have

$$\begin{aligned}
\int_{(t,1]} uS(\mathrm{d}u) &= \int_{(t,1]} S(\mathrm{d}u) - \int_{(t,1]} S(\mathrm{d}u) + \int_{(t,1]} uS(\mathrm{d}u) \\
&= \int_{[0,1]} u + (1-u)S(\mathrm{d}u) - \int_{[0,t]} S(\mathrm{d}u) - \int_{(t,1]} (1-u)S(\mathrm{d}u) \\
&= \int_{[0,1]} uS(\mathrm{d}u) + \int_{[0,1]} (1-u)S(\mathrm{d}u) - S([0,t]) - \int_{(t,1]} (1-u)S(\mathrm{d}u) \\
&\overset{(2.50)}{=} (2 - S([0,t])) - \left\{ 1 - \int_{[0,t]} (1-u)S(\mathrm{d}u) \right\} \\
&= 1 - S([0,t]) + \int_{[0,t]} (1-u)S(\mathrm{d}u), \qquad t \in [0,1].
\end{aligned}$$

Combining this with (2.52) results in

$$\begin{aligned}
A(t) &= t \int_{[0,t]} (1-u)S(\mathrm{d}u) + (1-t) \int_{(t,1]} uH(\mathrm{d}u) \\
&= t \int_{[0,t]} (1-u)S(\mathrm{d}u) + (1-t) \left\{ 1 - S([0,t]) + \int_{[0,t]} (1-u)S(\mathrm{d}u) \right\} \\
&= \int_{[0,t]} (1-u)S(\mathrm{d}u) + (1-t)\{1 - S([0,t])\}.
\end{aligned}$$

Finally, using that

$$\begin{aligned}\int_{[0,t]} (1-u)S(\mathrm{d}u) &= \int_{[0,t]} \int_{(u,1]} \mathrm{d}y S(\mathrm{d}u) \\ &= \int_{(0,1]} \int_{[0,y\wedge t]} S(\mathrm{d}u)\mathrm{d}y \\ &= \int_{(0,t]} S([0,u])\mathrm{d}u + (1-t)S([0,t]),\end{aligned}$$

we end up with (2.48). □

Remark 2.21. *In a similar way one obtains for a general choice of norms $\|\cdot\|_1$ and $\|\cdot\|_2$ the representation*

$$A(t) = \int_{\mathscr{U}} \left\{(1-t)\frac{u_1}{\|(u_1,u_2\|_1}\right\} \vee \left\{t\frac{u_2}{\|(u_1,u_2)\|_2}\right\} S(\mathrm{d}(u_1,u_2)). \tag{2.53}$$

2.2.5 Summary Measures for Extremal Dependence

In this chapter we give a brief overview over summary measures of extremal dependence between the margins G_j, $j = 1,\dots,d$, of a multivariate extreme value distribution G. The objective is to summarize the central characteristics of extremal dependence in a rather small number of parameters that might be more convenient to handle than the infinite-dimensional measures described in the sections before. Here are some examples described for instance in Beirlant et al. [2004], Section 8.2.7:

(i) *Extremal coefficients.* Let S be the spectral measure of G with respect to $\|\cdot\|_1$, $\|\cdot\|_2$ and $\mathscr{U}$ as introduced in (2.34) and let l be its stable tail dependence function. For a non-empty set $M \subset \{1,\dots,d\}$ define $\boldsymbol{e}_M := \left(\mathbf{1}_{\{1\in M\}}, \mathbf{1}_{\{2\in M\}}, \dots, \mathbf{1}_{\{d\in M\}}\right)$. The extremal coefficient of G with respect to M is defined as

$$\theta_M := l(\boldsymbol{e}_M) = -\log G(\boldsymbol{e}_M^{-1}) \overset{(2.35)}{=} \int_{\mathscr{U}} \bigvee_{j\in M} \left(\frac{u_j}{\|\boldsymbol{u}\|_1}\right) S(\mathrm{d}\boldsymbol{u}). \tag{2.54}$$

We have

$$\mathbb{P}\left(Y_j \le G_j^{\leftarrow}(p) \text{ for all } j \in M\right) = p^{\theta_M} \text{ for } p \in (0,1),$$

where $\boldsymbol{Y} = (Y_1,\dots,Y_d)$ has distribution function G, that is, the smaller the extremal coefficients, the stronger the extremal dependence, and vice versa.

(ii) *Kendall's tau and Spearman's rho for $d = 2$.* Let (X_1,Y_1) and (X_2,Y_2) be two independent replications of a random vector (X,Y) with distribution function F. Then Kendall's tau (Kendall [1938]) is given by

$$\tau := \mathbb{P}[(X_1-X_2)(Y_1-Y_2) > 0] - \mathbb{P}[(X_1-X_2)(Y_1-Y_2) < 0], \tag{2.55}$$

that is, if X and Y are comonotone (i.e. they tend to increase and decrease together) then τ is positive and otherwise negative.
Next, we denote by F_X and F_Y the distribution functions of X and Y, respectively. Recall that according to the probability integral transform method (see Section 2.1.2), $U := F_X(X)$ and $V := F_Y(Y)$ are uniformly distributed on the interval (0,1). Spearman's rho (Spearman [1904]) is defined as

$$\rho_S := \text{Corr}[U, V] = 12\mathbb{E}[UV] - 3. \tag{2.56}$$

(iii) *Tail dependence coefficient for d=2.* The following definition is rooted in Geffroy [1958], Geffroy [1959] and Sibuya [1960].

Definition 2.22 (Tail Dependence Coefficient)**.** *Let (X, Y) be a random vector with distribution function F. Assume that F_X and F_Y, the respective marginal distribution functions of X and Y, are continuous. Then the tail dependence coefficient of X and Y is given by*

$$\mathcal{X} := \lim_{u \to 1} \mathbb{P}\left(F_X(X) > u | F_Y(Y) > u\right). \tag{2.57}$$

As a short remark, recall that the random variables $F_X(X)$ and $F_Y(Y)$ are uniformly distributed on the interval $(0, 1)$. We have $0 \leq \mathcal{X} \leq 1$. If $\mathcal{X} = 0$ then we say that X and Y are *asymptotically* independent and if $\mathcal{X} = 1$ we say that they are *asymptotically* completely dependent. For a more detailed definition of those two boundary cases we refer to Section 2.2.6, Paragraph (vii).

Remark 2.23 (Tail dependence coefficient of an extreme value distribution)**.** *The result of a short computation using the representation (2.47) of an arbitrary bivariate extreme value distribution function $\tilde{G}$ in terms of its Pickands dependence function A shows that its tail dependence coefficient is given by*

$$\mathcal{X} = 2\left[1 - A\left(\frac{1}{2}\right)\right]. \tag{2.58}$$

If one has an estimate of A (see for instance Section 2.2.7 for some examples), this nice and closed form can be used to obtain an estimate of $\mathcal{X}$.

2.2.6 The Domain-of-Attraction Problem

Recall the domain-of-attraction equation from the beginning of this chapter: We say that a d-variate distribution function F is in the maximum domain of attraction (MDA) of a d-variate extreme value distribution G, if for all $\boldsymbol{x} \in \text{supp}(F)$

$$F^n(\boldsymbol{a}_n \boldsymbol{x} + \boldsymbol{b}_n) \to G(\boldsymbol{x}), \qquad n \to \infty, \tag{2.59}$$

where $\boldsymbol{a}_n \in \mathbb{R}^d_+$ is a positive normalizing sequence and $\boldsymbol{b}_n \in \mathbb{R}^d$ is a centering sequence. Having found various representations for the extreme value distribution G in the previous subsections, the objective is now to find conditions on F such that (2.59) holds. We list some of them in the following, where we assume that the margins F_j of F are continuous.

(i) ***Tail analysis.*** We start with an *analysis of the tail* $\bar{F}$ of F, which is defined as $\bar{F} = 1 - F$. Using the max-stability of G it can be shown (see for instance Beirlant et al. [2004], Section 8.3.1) that there exist vectors $\boldsymbol{\alpha}_n > \mathbf{0}$ and $\boldsymbol{\beta}_n$ such that (2.59) is equivalent to

$$\bar{F}(\boldsymbol{a}_n \boldsymbol{x} + \boldsymbol{b}_n) \sim -\log G(\boldsymbol{\alpha}_n \boldsymbol{x} + \boldsymbol{\beta}_n) \sim \bar{G}(\boldsymbol{\alpha}_n \boldsymbol{x} + \boldsymbol{\beta}_n), \qquad n \to \infty, \tag{2.60}$$

for $\boldsymbol{x} \in \operatorname{supp}(G)$. These equations can be used as a basis for statistical modelling and inference on $F(\boldsymbol{x})$ for $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$ such that its margins $F_j(x_j)$ are close to 1 for all $j = 1, \ldots, d$: From (2.60) one concludes

$$F(\boldsymbol{x}) \approx G(\boldsymbol{\alpha}_n \boldsymbol{a}_n^{-1} \boldsymbol{x} - \boldsymbol{\alpha}_n \boldsymbol{a}_n^{-1} \boldsymbol{b}_n + \boldsymbol{\beta}_n) =: G^\star(\boldsymbol{x}),$$

where $G^\star$ is another extreme value distribution of the same type as G and therefore possesses the same stable tail dependence function l, which in turn yields

$$F(\boldsymbol{x}) \approx G^\star(\boldsymbol{x}) = \exp\left\{-l(\boldsymbol{v})\right\}, \tag{2.61}$$

where $\boldsymbol{v} = (v_1, \ldots, v_d) = (-\log G_1^\star(x_1), \ldots, -\log G_d^\star(x_d))$ (cf. (2.32)). There are plenty of parametric models for the stable tail dependence function l, see Section 2.2.7. If one uses those in connection with the general univariate parametrization (2.25) for the margins $G_j^\star$, one obtains a fully parametric model for $F(\boldsymbol{x}) \in [\boldsymbol{u}, \boldsymbol{\infty}]$ where $\boldsymbol{u} \in \mathbb{R}^d$ is chosen such that $F_j(x_j)$ is sufficiently close to 1 for all $\boldsymbol{x} = (x_1, \ldots, x_d) \in [\boldsymbol{u}, \boldsymbol{\infty}]$ and all $j = 1, \ldots, d$.

(ii) ***Exponent measure.*** It is also possible to relate condition (2.59) to the *exponent measure* μ of the multivariate extreme value distribution G. Again we assume that G has standard Fréchet margins so that $\mu([\mathbf{0}, \boldsymbol{x}]^c) = -\log G(\boldsymbol{x})$ for $\boldsymbol{x} \geq \mathbf{0}$. For $\boldsymbol{X} \sim F$ and $n \in \mathbb{N}$ define the measures

$$\mu_n(\cdot) := n\mathbb{P}(\boldsymbol{a}_n^{-1}(\boldsymbol{X} - \boldsymbol{b}_n) \vee \mathbf{0} \in \cdot). \tag{2.62}$$

Then we have that (2.59) holds if and only if μ_n *converges vaguely* (see for instance Resnick [1987]) to μ as $n \to \infty$ on $[\mathbf{0}, \boldsymbol{\infty}] \backslash \{\mathbf{0}\}$, that is,

$$\int\limits_{[\mathbf{0}, \boldsymbol{\infty}] \backslash \{\mathbf{0}\}} f(\boldsymbol{x}) \mu_n(\mathrm{d}\boldsymbol{x}) \to \int\limits_{[\mathbf{0}, \boldsymbol{\infty}] \backslash \{\mathbf{0}\}} f(\boldsymbol{x}) \mu(\mathrm{d}\boldsymbol{x}), \qquad n \to \infty \tag{2.63}$$

for all continuous non-negative functions $f : [\mathbf{0}, \boldsymbol{\infty}] \backslash \{\mathbf{0}\} \to \mathbb{R}_+$ for which there is a compact set $K \subset [\mathbf{0}, \boldsymbol{\infty}] \backslash \{\mathbf{0}\}$ such that $f(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in ([\mathbf{0}, \boldsymbol{\infty}] \backslash \{\mathbf{0}\}) \backslash K$. We use the notation $\mu_n(\cdot) \xrightarrow{v} \mu(\cdot)$, $n \to \infty$.

(iii) ***Point Processes***. A profound introduction to point processes and Poisson random measures can be found in Resnick [1986], Resnick [1987] and Embrechts et al. [1997], Section 5. To get an intuitive overview over point processes, consider a sequence of random vectors ("points") $(\boldsymbol{X}_n)_{n \in \mathbb{N}}$ with state space $E \subset \mathbb{R}^d$ which is equipped with the $\sigma-$algebra $\mathcal{E} := \mathcal{B}(E)$. A point process N counts the number of points in

subsets of E, i.e. $N(A) = \sum_i \mathbf{1}_{\{\boldsymbol{X}_i \in A\}}$ for $A \subset E$. Formally, a point process on E is a measurable map $N : (\Omega, \mathcal{F}, \mathbb{P}) \to (M_p(E), \mathcal{M}_p(E))$, where $M_p(E)$ is the set of all point measures on E and $\mathcal{M}_p(E)$ is an appropriate $\sigma-$algebra. A special kind of point process is the *Poisson process* or the *Poisson random measure*.

Definition 2.24 (Poisson Random Measure, cf. Embrechts et al. [1997], Definiton 5.1.9)**.** *Let ν be a Radon measure on $\mathcal{E}$, that is $\nu(K) < \infty$ for compact sets $K \subset E$. A point process N on E is called* Poisson process *or* Poisson random measure *with* mean measure ν *(i.e. $\mathbb{E}[N(A)] = \nu(A)$ for all $A \subset E$) if the following holds:*

- *For $A \in \mathcal{E}$, $k \in \mathbb{N}$,*

$$\mathbb{P}(N(A) = k) = \begin{cases} \exp\{-\nu(A)\}\frac{(\nu(A))^k}{k!}, & \text{if } \nu(A) < \infty \\ 0, & \text{if } \nu(A) = \infty. \end{cases} \tag{2.64}$$

- *For $m \in \mathbb{N}_+$ and $A_1, \ldots, A_m \in \mathcal{E}$ with $A_i \cap A_\ell = \emptyset$ for all $i, \ell \in \{1, \ldots, m\}$, the random variables $N(A_1), \ldots, N(A_m)$ are independent.*

In the current context of the domain-of-attraction problem (2.59), the theory of point processes can be used as follows: For $i \in \mathbb{N}_+$ set $\boldsymbol{X}_{i,n} := \boldsymbol{a}_n^{-1}(\boldsymbol{X}_i - \boldsymbol{b}_n) \vee \mathbf{0}$, where $\boldsymbol{X}_i$ has distribution function F. Define by

$$N_n(\cdot) := \sum_{i=1}^{\infty} \mathbf{1}_{\{(\frac{i}{n}, \boldsymbol{X}_{i,n}) \in \cdot\}}$$

a point process on $[0, \infty) \times [\mathbf{0}, \boldsymbol{\infty}]$. Then (2.59) is equivalent to

$$N_n \xrightarrow{\mathcal{D}} \text{ Poisson random measure } N \text{ with mean measure } \nu \ , \quad n \to \infty, \tag{2.65}$$

where the measure ν is defined in terms of the exponent measure μ of the extreme value distribution G with standard Fréchet margins as

$$\begin{aligned} \mathbb{E}\left[N\left((s_1, s_2] \times [\boldsymbol{v}_1, \boldsymbol{v}_2]\right)\right] = \nu\left((s_1, s_2] \times [\boldsymbol{v}_1, \boldsymbol{v}_2]\right) &:= \int_{[\boldsymbol{v}_1, \boldsymbol{v}_2]} \int_{s_1}^{s_2} \mathrm{d}t \ \mu(\mathrm{d}\boldsymbol{x}) \\ &= (s_2 - s_1) \cdot \mu([\boldsymbol{v}_1, \boldsymbol{v}_2]). \end{aligned}$$

for a subset $(s_1, s_2] \times [\boldsymbol{v}_1, \boldsymbol{v}_2]$ of $[0, \infty) \times [\mathbf{0}, \boldsymbol{\infty}]$. For a proof we refer to Resnick [1987], Proposition 3.21. The weak convergence in (2.65) is interpreted as follows: For arbitrary $m \in \mathbb{N}_+$ let $A_1, \ldots, A_m$ be subsets of $[0, \infty) \times [\mathbf{0}, \boldsymbol{\infty}]$ such that $\mathbb{P}(N(\partial A_i) = 0) = 1$ for all i, where ∂A_i is the boundary of A_i. Then we have that

$$\mathbb{P}\left\{N_n(A_1) = k_1, \ldots, N_n(A_m) = k_m\right\} \to \mathbb{P}\left\{N(A_1) = k_1, \ldots, N(A_m) = k_m\right\}, \ n \to \infty,$$

for $k_i \in \mathbb{N}$, $i = 1, \ldots, m$.

(iv) ***Continuous Index.*** For $t \in \mathbb{R}$, let $\lfloor t \rfloor := \max\{n \in \mathbb{N} : n \leq t\}$ denote its integer part. As $\frac{t}{\lfloor t \rfloor} \to 1$ as $t \to \infty$ we can generalize equations (2.59), (2.60) and the vague convergence of the exponent measure (cf. (2.62) and (2.63)) as follows:
We have that F is in the maximum domain of attraction of G if the following conditions hold as t tends to infinity:

$$F^t\left(\boldsymbol{a}_{\lfloor t \rfloor}\boldsymbol{x} + \boldsymbol{b}_{\lfloor t \rfloor}\right) \to G(\boldsymbol{x}), \tag{2.66}$$

$$\bar{F}\left(\boldsymbol{a}_{\lfloor t \rfloor}\boldsymbol{x} + \boldsymbol{b}_{\lfloor t \rfloor}\right) \sim -\log G\left(\boldsymbol{\alpha}_{\lfloor t \rfloor}\boldsymbol{x} + \boldsymbol{\beta}_{\lfloor t \rfloor}\right) \sim \bar{G}\left(\boldsymbol{\alpha}_{\lfloor t \rfloor}\boldsymbol{x} + \boldsymbol{\beta}_{\lfloor t \rfloor}\right), \tag{2.67}$$

$$\mu_t(\cdot) = t\mathbb{P}\left(\boldsymbol{X}_{1,\lfloor t \rfloor} \in \cdot\right) \xrightarrow{v} \mu(\cdot), \tag{2.68}$$

where the normalizing and centering sequences are chosen as described in the respective paragraphs.

(v) ***Multivariate regular variation.***

Definition 2.25 (Multivariate regular variation, cf. Beirlant et al. [2004], Section 8.4)**.** *Let H be a d-variate distribution function with* $\operatorname{supp}(H)=[\mathbf{0}, \boldsymbol{\infty})$. *Set* $\boldsymbol{e} := (1, \ldots, 1) \in \mathbb{R}^d$. *Then H is* multivariate regularly varying *on* $(\mathbf{0}, \boldsymbol{\infty})$ *if there exists a function* $h : (\mathbf{0}, \boldsymbol{\infty}) \to (0, \infty)$ *satisfying*

$$\frac{\bar{H}(t\boldsymbol{x})}{\bar{H}(t\boldsymbol{e})} \to h(\boldsymbol{x}), \qquad t \to \infty, \tag{2.69}$$

for $\boldsymbol{x} \in (\mathbf{0}, \boldsymbol{\infty})$.

Now assume one considers an arbitrary distribution function F that is in the maximum domain of attraction of an arbitrary mulitvariate extreme value distribution $\tilde{G}$. For convenience one can transform $\tilde{G}$ into a distribution function G with standard Fréchet margins as done in (2.27). Basically the same approach can be used for F: One obtains a distribution function $F_\star$ with standard Fréchet margins by setting

$$F_\star(\boldsymbol{z}) := F\left(F_1^{\leftarrow}\left(\exp\left[-\frac{1}{z_1}\right]\right), \ldots, F_d^{\leftarrow}\left(\exp\left[-\frac{1}{z_d}\right]\right)\right), \tag{2.70}$$

where $\boldsymbol{z} = (z_1, \ldots, z_d) \in (\mathbf{0}, \boldsymbol{\infty})$. Clearly, the standard Fréchet distribution is in its own domain of attraction, i.e. $F_\star \in \mathrm{MDA}(G)$. Choosing the centering sequences $\boldsymbol{a}_{\lfloor t \rfloor} = (t, \ldots, t)$ and $\boldsymbol{b}_{\lfloor t \rfloor} = \mathbf{0}$ it holds that (see (2.67))

$$\bar{F}_\star(t\boldsymbol{z}) \sim -\log G(\boldsymbol{z}), \qquad \boldsymbol{z} \in (\mathbf{0}, \boldsymbol{\infty}),$$

as t $\to \infty$, which in turn implies (see for instance Beirlant et al. [2004], Section 8.3.2)

$$\frac{\bar{F}_\star(t\boldsymbol{z})}{\bar{F}_\star(t\boldsymbol{e})} \to \frac{-\log G(\boldsymbol{z})}{-\log G(\boldsymbol{e})}, \qquad t \to \infty, \tag{2.71}$$

for $\boldsymbol{e} = (1, \ldots, 1) \in \mathbb{R}^d$ and $\boldsymbol{z} \in [\mathbf{0}, \boldsymbol{\infty}]$. Equation (2.71) means that $\bar{F}_\star$ is multivariate regularly varying on $(\mathbf{0}, \boldsymbol{\infty})$, which is an interesting analogue to the univariate case. One can even show that there is in fact equivalence between (2.71) and $F_\star$ being in the maximum domain of attraction of G.

(vi) ***Tail dependence function.*** The following definition of the *tail dependence function* of the distribution function F can be found for example in Beirlant et al. [2004], Section 8.3.2.

Definition 2.26 (Tail Dependence Function)**.** *For $\boldsymbol{u} = (u_1, \ldots, u_d) \in [\mathbf{0}, \mathbf{1}]$ the* tail dependence function *of $F = (F_1, \ldots, F_d)$ is given by*

$$D_F(\boldsymbol{u}) = 1 - F\left(F_1^{\leftarrow}(1-u_1), \ldots, F_d^{\leftarrow}(1-u_d)\right). \tag{2.72}$$

The tail dependence function can be linked to the domain-of-attraction condition (2.59) in the following way:

Proposition 2.27. *The d-variate distribution function $F = (F_1, \ldots, F_d)$ is in the maximum domain of attraction of the d-variate distribution function $G = (G_1, \ldots, G_d)$ if and only if the following two conditions hold:*

(a) For every $j = 1, \ldots, d$ there are real sequences $(a_{n,j})_n$ and $(b_{n,j})_n$ with $a_{n,j} > 0$ such that for $\boldsymbol{x} = (x_1, \ldots, x_d) \in \operatorname{supp}(F)$ we have

$$F_j^n(a_{n,j}x_j + b_{n,j}) \to G_j(x_j), \qquad n \to \infty. \tag{2.73}$$

(b) For $\boldsymbol{v} \geq \mathbf{0}$ we have

$$\lim_{s \downarrow 0} s^{-1} D_F(s\boldsymbol{v}) = l(\boldsymbol{v}),$$

where l is the stable tail dependence function of G.

(vii) ***Asymptotic independence and complete dependence.*** This paragraph examines the two extreme properties *asymptotic independence* and *asymptotic complete dependence* of F from the point of view of its tail dependence function D_F defined in (2.72).

Definition 2.28 (Asymptotic Independence and Complete Dependence, cf. Beirlant et al. [2004], Section 8.3.2)**.** *A d-variate distribution function F with tail dependence function D_F is said to be* asymptotically independent *if*

$$\lim_{s \downarrow 0} s^{-1} D_F(s\boldsymbol{v}) = \sum_{j=1}^{d} v_j, \qquad \boldsymbol{v} = (v_1, \ldots, v_d) \geq \mathbf{0}. \tag{2.74}$$

On the other hand, it is called asymptotically completely dependent *if*

$$\lim_{s \downarrow 0} s^{-1} D_F(s\boldsymbol{v}) = \bigvee_{j=1}^{d} v_j, \qquad \boldsymbol{v} = (v_1, \ldots, v_d) \geq \mathbf{0}. \tag{2.75}$$

If F is asymptotically independent and its margins fulfill (2.73) then F is in the maximum domain of attraction of the extreme value distribution G given by $G(\boldsymbol{x}) = \prod_{j=1}^{d} G_j(x_j)$ for $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. On the other hand, if F is asymptotically completely dependent it is in the maximum domain of attraction of G where $G(\boldsymbol{x}) = \bigwedge_{j=1}^{d} G_j(x_j)$

2.2.7 Statistics of Multivariate Extremes

This subsection, which sticks to Beirlant et al. [2004], Section 9, summarizes some basic tools of statistical modelling of multivariate extremes. Both non-parametric and parametric approaches are presented. The first part of the subsection deals with the method of *block maxima*, the respectively largest ones of several groups of observations. The second part follows with an approach based on *threshold exceedances* where not only the largest observation is taken into account, but all observations where at least one component exceeds the respective threshold component.

Method of Block Maxima

Consider a sample of d-variate observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ for $n \in \mathbb{N}_+$. The method of block maxima suggests to divide the set of observations into k different groups and take the maximum $\boldsymbol{Y}_i$ over each group $i = 1, \ldots, k$, that is,

$$\boldsymbol{Y}_i = \bigvee_{r=(i-1)m+1}^{im} \boldsymbol{X}_r, \qquad i = 1, \ldots, k, \tag{2.76}$$

where $m \in \mathbb{N}_+$ is chosen such that $km = n$ holds. For instance, if $m = 365$ and $X_{r,j}$ is the amount of rainfall on day r for $r \in \{1, \ldots, n\}$ at spatial location $j \in \{1, \ldots, d\}$ then the component $Y_{i,j}$ of the block maxima $\boldsymbol{Y}_i$, $i \in \{1, \ldots, k\}$ represents the maximum amount of rainfall in year i at location j. The sample $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_k$ is assumed to be i.i.d. with a d-variate extreme value distribution function $\tilde{G}$ which is allowed to have arbitrary margins $\tilde{G}_1, \ldots \tilde{G}_d$. Recall from (2.32) that $\tilde{G}$ can be expressed in terms of its stable tail dependence function l as

$$\tilde{G}(\boldsymbol{y}) = \exp\left\{-l[-\log \tilde{G}_1(y_1), \ldots, -\log \tilde{G}_d(y_d)]\right\}, \qquad \boldsymbol{y} = (y_1, \ldots, y_d) \in [-\boldsymbol{\infty}, \boldsymbol{\infty}]. \tag{2.77}$$

The task is to model and estimate the stable tail dependence function l which leads to an estimate of $\tilde{G}$. We discuss both non-paramteric and parametric approaches to achieve this.

Non-parametric methods. For simplicity we assume $d = 2$, that is, the bivariate case. The sample of block maxima $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_k\}$ consists of random vectors $\boldsymbol{Y}_i = (Y_{i,1}, Y_{i,2})$ which are independent replicates of a random variable $\boldsymbol{Y} = (Y_1, Y_2)$ with distribution function $\tilde{G}$. From (2.44) and (2.47) we conclude that instead of modelling the stable tail dependence function we can model the Pickands dependence function of $\tilde{G}$, based on the representation

$$\tilde{G}(y_1, y_2) = \exp\left\{\log\left[\tilde{G}_1(y_1)\tilde{G}_2(y_2)\right] A\left(\frac{\log \tilde{G}_2(y_2)}{\log\left[\tilde{G}_1(y_2)\tilde{G}_2(y_2)\right]}\right)\right\}.$$

Setting $\xi := -\log \tilde{G}_1(Y_1)$ and $\eta := -\log \tilde{G}_2(Y_2)$ we have that for $t \in [0, 1]$ the random variable $\min\left\{\frac{\xi}{1-t}, \frac{\eta}{t}\right\}$ is exponentially distributed with mean $\frac{1}{A(t)}$. So it is quite convenient to estimate $A(t)$ via the sample mean. The resulting estimator is called *Pickands'*

estimator (Pickands [1981] and Pickands [1989]):

$$\frac{1}{\hat{A}_k^{\mathrm{P}}(t)} := \frac{1}{k}\sum_{i=1}^{k} \min\left\{\frac{\hat{\xi}_i}{1-t}, \frac{\hat{\eta}_i}{t}\right\}, \tag{2.78}$$

with $\hat{\xi}_i = -\log \hat{\tilde{G}}_1(Y_{i,1})$ and $\hat{\eta}_i = -\log \hat{\tilde{G}}_2(Y_{i,2})$ for $i = 1, \ldots, k$ where $\hat{\tilde{G}}_1$ and $\hat{\tilde{G}}_2$ are estimators of $\tilde{G}_1$ and $\tilde{G}_2$, for example the modified empirical distribution functions, given by

$$\hat{\tilde{G}}_j(y) = \frac{1}{k+1}\sum_{i=1}^{k} \mathbf{1}_{\{Y_{i,j}\leq y\}}, \qquad y \in [-\infty, \infty],\ j \in \{1,2\}. \tag{2.79}$$

However, this version of the Pickands estimator does not fulfill the properties of a Pickands dependence function listed in Proposition 2.19, which is a serious drawback. For this reason reason there exist a couple of extensions of (2.78), for example the *Deheuvels estimator* $\hat{A}_k^{\mathrm{D}}$ (Deheuvels [1991]) and the *Hall and Tajvidi estimator* $\hat{A}_k^{\mathrm{HT}}$ (Hall and Tajvidi [2000]):

$$\frac{1}{\hat{A}_k^{\mathrm{D}}(t)} := \frac{1}{k}\sum_{i=1}^{k} \min\left\{\frac{\hat{\xi}_i}{1-t}, \frac{\hat{\eta}_i}{t}\right\} - (1-t)\bar{\xi}_k - t\bar{\eta}_k + 1 \tag{2.80}$$

$$\frac{1}{\hat{A}_k^{\mathrm{HT}}(t)} := \frac{1}{k}\sum_{i=1}^{k} \min\left\{\frac{\hat{\xi}_i/\bar{\xi}_k}{1-t}, \frac{\hat{\eta}_i/\bar{\eta}_k}{t}\right\} \tag{2.81}$$

for $t \in [0,1]$. Here $\bar{\xi}_k = \frac{1}{k}\sum_{i=1}^{k}\hat{\xi}_i$ and $\bar{\eta}_k = \frac{1}{k}\sum_{i=1}^{k}\hat{\eta}_i$ denote the sample means. Still, those approaches lead to estimators that do not have the property of complexity (cf. Proposition 2.19) and therefore one final extension can be to replace them by their convex minorants. Another estimator of the Pickands dependence function A is based on a different approach by Capéràa, Fougères and Genest [1997] . We have that the expectation of the random variable $\log\max\{t\xi, (1-t)\eta\}$ is given by

$$\mathbb{E}[\log\max\{t\xi, (1-t)\eta\}] = \log A(t) + \int_0^\infty \log(x)\mathrm{e}^{-x}\mathrm{d}x$$

for $t \in [0,1]$, so a quite natural procedure is to take the empirical expectation, resulting in

$$\log \hat{A}_k^{\mathrm{CFG}}(t) = \frac{1}{k}\sum_{i=1}^{k} \log\max\left\{t\hat{\xi}_i, (1-t)\hat{\eta}_i\right\} - \int_0^\infty \log(x)\mathrm{e}^{-x}\mathrm{d}x. \tag{2.82}$$

Parametric methods. We can equally model the stable tail dependence function l or the Pickands dependence function A (for $d = 2$) in a parametric way. There are plenty of different models in the literature. We list three of them in Table 2.1. The arguments $\boldsymbol{v}$ and t are elements of the intervals $[\mathbf{0}, \boldsymbol{\infty}]$ and $[0,1]$, respectively. We denote the parameter vector that l depends on by $\boldsymbol{\theta}$ and write $l(\boldsymbol{v}) = l(\boldsymbol{v}; \boldsymbol{\theta})$ in a general way. Then a fully parametric model is obtained by connecting the representation of $\tilde{G}$ in terms of l (2.77) and the general univariate parametrization of its margins $\tilde{G}_j$, $j = 1, \ldots, d$:

Model name	**Dependence function**	**Reference**
Symmetric logistic model	$l(\boldsymbol{v}) = \left(\sum_{j=1}^{d} v_j^{1/\alpha} \right)^{\alpha}$, $\alpha \in (0,1]$	Gumbel [1960a] and Gumbel [1960b]
Asymmetric logistic model	$l(\boldsymbol{v}) = \sum_{c \in C_d} \left\{ \sum_{j \in c} (\psi_{c,j} v_j)^{1/\alpha_c} \right\}^{\alpha_c}$, $C_d = \{c \neq \emptyset : c \subset \{1,\dots,d\}\}$, $\alpha_c \in (0,1]$, $\psi_{c,j} \geq 0$, $\sum_{j \in c} \psi_{c,j} = 1$	Tawn [1990]
Polynomial model	$A(t) = \sum_{u=0}^{m} \psi_u t^u$, $m \in \mathbb{N}_+$, certain restriction on the ψ_u to ensure the properties in Prop. 2.19.	Klüppelberg and May [2001]

Table 2.1: Three parametric models for the stable tail dependence function or Pickands dependence function

(1) $\tilde{G}(\boldsymbol{y}) = \exp\left\{ -l[-\log \tilde{G}_1(y_1), \dots, -\log \tilde{G}_d(y_d)] \right\}$

(2) $\tilde{G}_j(y_j) = \exp\left\{ -\left(1 + \gamma_j \frac{y_j - \mu_j}{\sigma_j}\right)_+^{-\frac{1}{\gamma_j}} \right\}$, $j = 1, \dots, d$,

where $\boldsymbol{y} = (y_1, \dots, y_d) \in [-\boldsymbol{\infty}, \boldsymbol{\infty}]$, γ_j is the marginal extreme value index and $\mu_j \in \mathbb{R}$ and $\sigma_j > 0$ are the marginal centering and normalizing parameters, respectively. Now, based on the observed sample $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_k$ where the $\boldsymbol{Y}_i$ are assumed to i.i.d. d-variate random vectors with common distribution function $\tilde{G} = (\tilde{G}_1, \dots, \tilde{G}_d)$, one can estimate the marginal parameters γ_j, μ_j and σ_j and the dependence parameter vector $\boldsymbol{\theta}$ simultaneously by maximum likelihood. Notice, however, that there are certain conditions ensuring regularity of the estimation, see for example Beirlant et al. [2004], Section 9.3.2. Another approach, which separates the estimation procedures of the marginal and the dependence structure, is based on *extreme value copulas*. For further details, see again Beirlant et al. [2004], Section 9.3.2.

Threshold exceedances

As already mentioned, the method of block maxima above has the disadvantage of only considering the largest values of the blocks. Besides, as the maxima are taken componentwise, they do not necessarily belong to the sample themselves. The method of threshold exceedances tries to cope with this. One does not only take into consideration the largest observed values, but all observations where at least one component exceeds a high chosen threshold. The setting is as follows: Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent d-dimensional random vectors with common distribution function F. We assume that F is in the maximum domain of attraction of a d-variate extreme value distribution function $\tilde{G}$. The aim is to estimate $\bar{F}(\boldsymbol{x})$ for $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ such that $\bar{F}_j(x_j)$ is of order $o(\frac{1}{n})$. We can

approximate F in terms of the stable tail dependence function l of $\tilde{G}$ by

$$F(\boldsymbol{x}) \approx \exp\left\{-l\left(-\log F_1(x_1), \ldots, -\log F_d(x_d)\right)\right\}$$

(see Beirlant et al. [2004], Section 8.3.2.). So by estimating l we get an estimator $\hat{F}$ of F,

$$\hat{F}(\boldsymbol{x}) = \exp\left\{-\hat{l}\left(-\log \hat{F}_1(x_1), \ldots, -\log \hat{F}_d(x_d)\right)\right\}, \tag{2.83}$$

where $\hat{F}_j$ is an estimator of the margin F_j for all $j = 1, \ldots, d$, obtained for instance by using the methods of the univariate case, for example the GPD approach: Let $X_{(1),j} \leq X_{(2),j} \leq X_{(n),j}$ denote the ordered version of the marginal sample $X_{1,j}, \ldots, X_{n,j}$. For a positive integer $k \leq n$ (chosen according to the height of the threshold $\boldsymbol{u} = (u_1, \ldots, u_d)$) set $u_j := X_{(n-k),j}$. Then an estimator for F_j is obtained by fitting a generalized Pareto distribution:

$$\hat{F}_j(x_j) = 1 - \frac{k}{n}\left(1 + \hat{\gamma}_j \frac{x_j - u_j}{\hat{\sigma}_j}\right)^{-\frac{1}{\hat{\gamma}_j}}.$$

For the methods of estimating the GPD parameters γ_j and σ_j see Section 2.1 (POT).

Non-parametric methods. Recall the definition of the tail dependence function D_F of F (2.72) and the limit relation to the stable tail dependence function l of $\tilde{G}$:

$$\lim_{s \downarrow 0} s^{-1} D_F(s\boldsymbol{v}) = l(\boldsymbol{v}).$$

This yields approximately

$$l(\boldsymbol{v}) \approx s^{-1}\mathbb{P}\left[\exists\, j \in \{1, \ldots, d\} : F_j(X_j) > 1 - sv_j\right].$$

Setting $s = \frac{k}{n}$ for $k = k_n \to \infty$ and $\frac{k}{n} \to 0$ as $n \to \infty$, a possible estimator $\tilde{l}$ of l is given by

$$\begin{aligned}\tilde{l}(\boldsymbol{v}) &= \frac{1}{k}\sum_{i=1}^{k} \mathbf{1}_{\left\{\exists\, \in\{1,\ldots,d\}:\hat{F}_j(X_{i,j}) > 1 - \frac{k}{n}v_j\right\}} \\ &= \frac{1}{k}\sum_{i=1}^{k} \mathbf{1}_{\left\{\frac{k}{n}\hat{\boldsymbol{X}}_{\star i} \not\leq \boldsymbol{v}^{-1}\right\}},\end{aligned} \tag{2.84}$$

where the components of $\hat{\boldsymbol{X}}_{\star i}$ are given by

$$\hat{X}_{\star i,j} := \frac{1}{1 - \hat{F}_j(X_{i,j})}, \qquad j = 1, \ldots, d.$$

A problem arises because the marginal tail probabilities $\bar{F}_j(x_j)$ are or order $\mathrm{o}(\frac{1}{n})$: The estimator $\tilde{l}$ in (2.84) involves a region of the sample space with only little data. To cope with this, one can extend $\tilde{l}$ by setting

$$\hat{l}(\boldsymbol{v}) := |\boldsymbol{v}|\tilde{l}\left(\frac{\boldsymbol{v}}{|\boldsymbol{v}|}\right), \qquad \boldsymbol{v} \in [\boldsymbol{0}, \boldsymbol{\infty})\backslash\{\boldsymbol{0}\}, \tag{2.85}$$

using the homogeneity of the stable tail dependence function (cf. Proposition 2.12(i)). The estimator $\hat{l}$ can now be plugged into (2.83).

Parametric methods. In this paragraph we present the *censored likelihood* method (see for example Beirlant et al. [2004], Section 9.4.2). Consider once again the different models $l(\cdot;\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the respective parameter vector, of the stable tail dependence function presented in Table 2.1. Furthermore, let $\boldsymbol{u} = (u_1, \dots, u_d)$ be a threshold such that $F_j(u_j) = \exp(-\lambda_j)$ for a small positive λ_j. From equation (2.61) we know that we can express the d-variate distribution function F approximately in terms of $l(\cdot;\boldsymbol{\theta})$ and a d-variate extreme value distribution $G^\star$ of the same type as $\tilde{G}$ as

$$\begin{aligned} F(\boldsymbol{x}) &\approx \exp\{-l(\boldsymbol{v};\boldsymbol{\theta})\}, \qquad \boldsymbol{x} \geq \boldsymbol{u}, \\ v_j &= -\log G_j^\star(x_j) \\ &\overset{(2.25)}{=} \left(1 + \gamma_j \frac{x_j - \mu_j^\star}{\sigma_j^\star}\right)_+^{-\frac{1}{\gamma_j}} = \lambda_j \left(1 + \gamma_j \frac{x_j - u_j}{\sigma_j}\right)_+^{-\frac{1}{\gamma_j}}, \end{aligned}$$

where $\sigma_j = \sigma_j^\star + \gamma_j(u_j - \mu_j^\star)$. For the margins F_j it follows that

$$F_j(x_j) \approx \exp\left\{-\lambda_j \left(1 + \gamma_j \frac{x_j - u_j}{\sigma_j}\right)_+^{-\frac{1}{\gamma_j}}\right\}, \qquad x_j \geq u_j.$$

When estimating the marginal parameters $(\lambda_j, \gamma_j, \sigma_j)$ and the dependence parameter vector $\boldsymbol{\theta}$ jointly by maximum likelihood, one has to be careful because the model is only defined for observations $\boldsymbol{x} \in [\boldsymbol{u}, \boldsymbol{\infty})$. If a component x_j of an observation $\boldsymbol{x}$ satisfies $x_j < u_j$, it has to be censored from below at u_j. The resulting likelihood contribution $L(\boldsymbol{x})$ is proportional to

$$\begin{aligned} L(\boldsymbol{x}) &\propto \mathbb{P}[X_j \in \mathrm{d}x_j, j \in \{1,\dots,d\} : x_j > u_j; X_j \leq u_j, j \in \{1,\dots,d\} : x_j \leq u_j] \\ &\propto \frac{\partial^m F}{\partial x_{j_1} \cdots \partial x_{j_m}}(\boldsymbol{x} \vee \boldsymbol{u}), \end{aligned}$$

where $\boldsymbol{X} = (X_1, \dots, X_d) \sim F$, $m \in \{1, \dots, d\}$ and $\{j_1, \dots, j_m\}$ is the set of indices where the corresponding components of the observation $\boldsymbol{x}$ exceed the respective threshold components.

Chapter 3

Max-stable Processes

In this chapter, we examine a generalization of finite-dimensional multivariate extreme value distributions to infinite dimensions. The methods explained in Section 2.2 can only be used for statistical analysis of extreme values and interaction of a finite number of margins of a random vector. For example, it is possible to model extremal rainfall at only a finite number of places, i.e. a discrete grid, and a finite number of time points. In order to overcome this disadvantage, one can use *max-stable processes.* In the literature, there exist plenty of different families of max-stable processes. Profound introductions can be found for example in Brown and Resnick [1977], Deheuvels [1983], de Haan [1984], de Haan and Pickands [1986], Kabluchko [2009] or Schlather [2002]. In the following section we present some basic tools and properties of *Gaussian space-time processes* needed in the course of the rest of this thesis. We stick close to Davis, Klüppelberg and Steinkohl [2012a].

3.1 Gaussian Space-Time Processes

Let T be an arbitrary index set, for example $T = [0, \infty)$. Recall that a standard Gaussian process $\boldsymbol{X} = \{X_t : t \in T\}$ is a stochastic process with multivariate normally distributed finite dimensional margins. If one thinks of T as a set of time points, a *space-time process* can be regarded as the extension of a stochastic process indexed by T to one with a second index set $S \subset \mathbb{R}^d$ for $d \geq 1$, which might for instance consist of d-dimensional spatial locations $\boldsymbol{s}$. A *Gaussian* space-time process $\boldsymbol{Z} = \{Z_{\boldsymbol{s},t} : \boldsymbol{s} \in S, t \in T\}$ is a space-time process such that all finite-dimensional margins $(Z_{\boldsymbol{s}_j,t_j})_{j=1,\dots,k}$ for all $k \in \mathbb{N}$, $\boldsymbol{s}_1, \dots, \boldsymbol{s}_k \in S$ and $t_1, \dots, t_k \in T$ are multivariate Gaussian. The correlation function $\mathscr{C}$ of $\boldsymbol{Z}$ plays a central role in the connection between Gaussian space-time processes and max-stable processes.

Definition 3.1 (Correlation Function of a Space-Time Process)**.** *The correlation function of a space-time process* $\boldsymbol{Z} = \{Z_{\boldsymbol{s},t} : \boldsymbol{s} \in S, t \in T\}$ *is defined as*

$$\mathscr{C}(\boldsymbol{s}_1, t_1; \boldsymbol{s}_2, t_2) = \frac{\mathrm{Cov}\left[Z_{\boldsymbol{s}_1,t_1}, Z_{\boldsymbol{s}_2,t_2}\right]}{\sqrt{\mathrm{Var}[Z_{\boldsymbol{s}_1,t_1}]\mathrm{Var}[Z_{\boldsymbol{s}_2,t_2}]}}, \qquad \boldsymbol{s}_1, \boldsymbol{s}_2 \in S,\ t_1, t_2 \in T. \tag{3.1}$$

The correlation function $\mathscr{C}$ is called *separable* if it can be written as the product or

the sum of a spatial correlation function $\mathscr{C}_1$ and a temporal correlation function $\mathscr{C}_2$, i.e.

$$\mathscr{C}(\boldsymbol{s}_1, t_1; \boldsymbol{s}_2, t_2) = \mathscr{C}_1(\boldsymbol{s}_1, \boldsymbol{s}_2)\mathscr{C}_2(t_1, t_2)$$

or

$$\mathscr{C}(\boldsymbol{s}_1, t_1; \boldsymbol{s}_2, t_2) = \mathscr{C}_1(\boldsymbol{s}_1, \boldsymbol{s}_2) + \mathscr{C}_2(t_1, t_2).$$

for $\boldsymbol{s}_1, \boldsymbol{s}_2 \in S$ and $t_1, t_2 \in T$. From now on we mostly assume $\boldsymbol{S} = \mathbb{R}^d$ (the set of spatial locations) for some $d \geq 1$ and $T = [0, \infty)$ (the time). Furthermore, we suppose for convenience that $\mathrm{Var}[Z(\boldsymbol{s}, t)] = 1$ for all $\boldsymbol{s} \in \mathbb{R}^d$ and $t \in [0, \infty)$ and that $\boldsymbol{Z}$ is stationary, that is, its correlation function $\mathscr{C}$ only depends on the spatial lag $\boldsymbol{h} := \boldsymbol{s}_1 - \boldsymbol{s}_2$ and the temporal lag $u := t_1 - t_2$:

$$\mathscr{C}(\boldsymbol{s}_1, t_1; \boldsymbol{s}_2, t_2) = \mathscr{C}(\boldsymbol{s}_1 - \boldsymbol{s}_2, t_1 - t_2; \boldsymbol{0}, 0) =: \gamma(\boldsymbol{h}, u)$$

for all $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{R}^d$ and $t_1, t_2 \in [0, \infty)$. If γ in turn only depends on the *absolute* spatial lag $\|\boldsymbol{h}\|$ and the *absolute* temporal lag $|u|$ in such a way that there exists another correlation function $\tilde{\gamma}$ that satisfies

$$\gamma(\boldsymbol{h}, u) = \tilde{\gamma}(\|\boldsymbol{h}\|, |u|), \qquad \boldsymbol{h} \in \mathbb{R}^d,\ u \in \mathbb{R},$$

then γ is called *isotropic*, otherwise *anisotropic*. Figures 3.1 and 3.2 show an isotropic and an anisotropic simulated Gaussian space-time process, respectively. The simulation was achieved in `R` using the package `RandomFields` by Schlather [2001].

3.2 Definition of Max-Stable Processes

The definition of max-stable processes is reminiscent of that of max-stable distributions (see Definition 2.7).

Definition 3.2 (Max-Stable Processes in Space and Time)**.** *Let $S = \mathbb{R}^d$, $T = [0, \infty)$ for some $d \geq 1$ and $\boldsymbol{X} = \{X_{\boldsymbol{s},t} : \boldsymbol{s} \in S, t \in T\}$ be a stationary stochastic process. Furthermore, let $\boldsymbol{X}^{(i)} = \left\{X^{(i)}_{\boldsymbol{s},t} : \boldsymbol{s} \in S, t \in T\right\}$ for $i = 1 \ldots n$ be independent copies of $\boldsymbol{X}$. Then X is called* max-stable*, if the process $\boldsymbol{M}^{(n)} = \left\{M^{(n)}_{\boldsymbol{s},t} : \boldsymbol{s} \in S, t \in T\right\}$, defined by*

$$M^{(n)}_{\boldsymbol{s},t} := \bigvee_{i=1}^{n} X^{(i)}_{\boldsymbol{s},t}, \qquad \boldsymbol{s} \in S,\ t \in T,$$

satisfies the following condition: There exist sequences $\left(a^{(n)}_{\boldsymbol{s},t}\right)$ and $\left(b^{(n)}_{\boldsymbol{s},t}\right)$ for $n \in \mathbb{N}$, $\boldsymbol{s} \in S$ and $t \in T$ with $a^{(n)}_{\boldsymbol{s},t} > 0$ for all n, $\boldsymbol{s}$ and t such that it holds that for all n,

$$\frac{M^{(n)}_{\boldsymbol{s},t} - b^{(n)}_{\boldsymbol{s},t}}{a^{(n)}_{\boldsymbol{s},t}} \overset{\mathcal{D}}{=} X_{\boldsymbol{s},t}, \qquad \boldsymbol{s} \in S,\ t \in T. \tag{3.2}$$

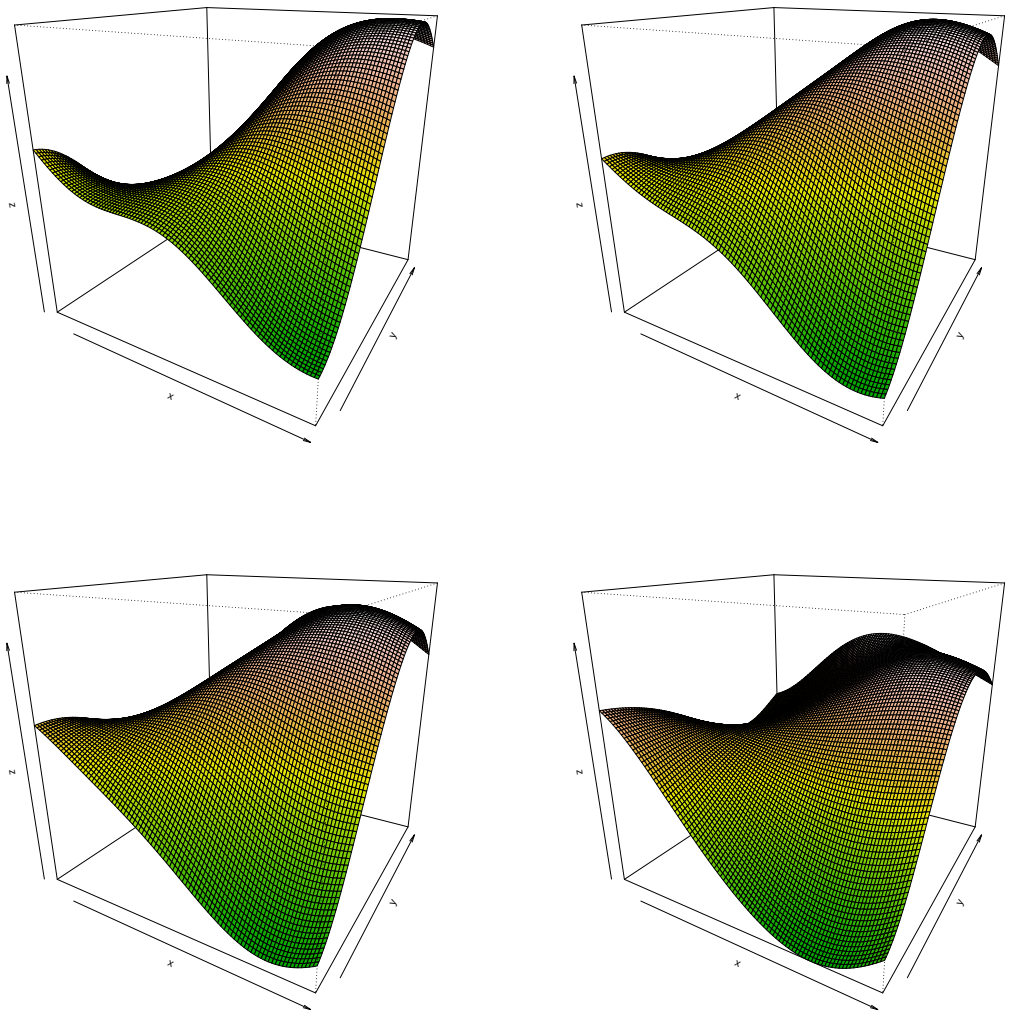

Figure 3.1: Simulated isotropic Gaussian space-time process. The picture shows four consecutive time steps from the top left to the bottom right.

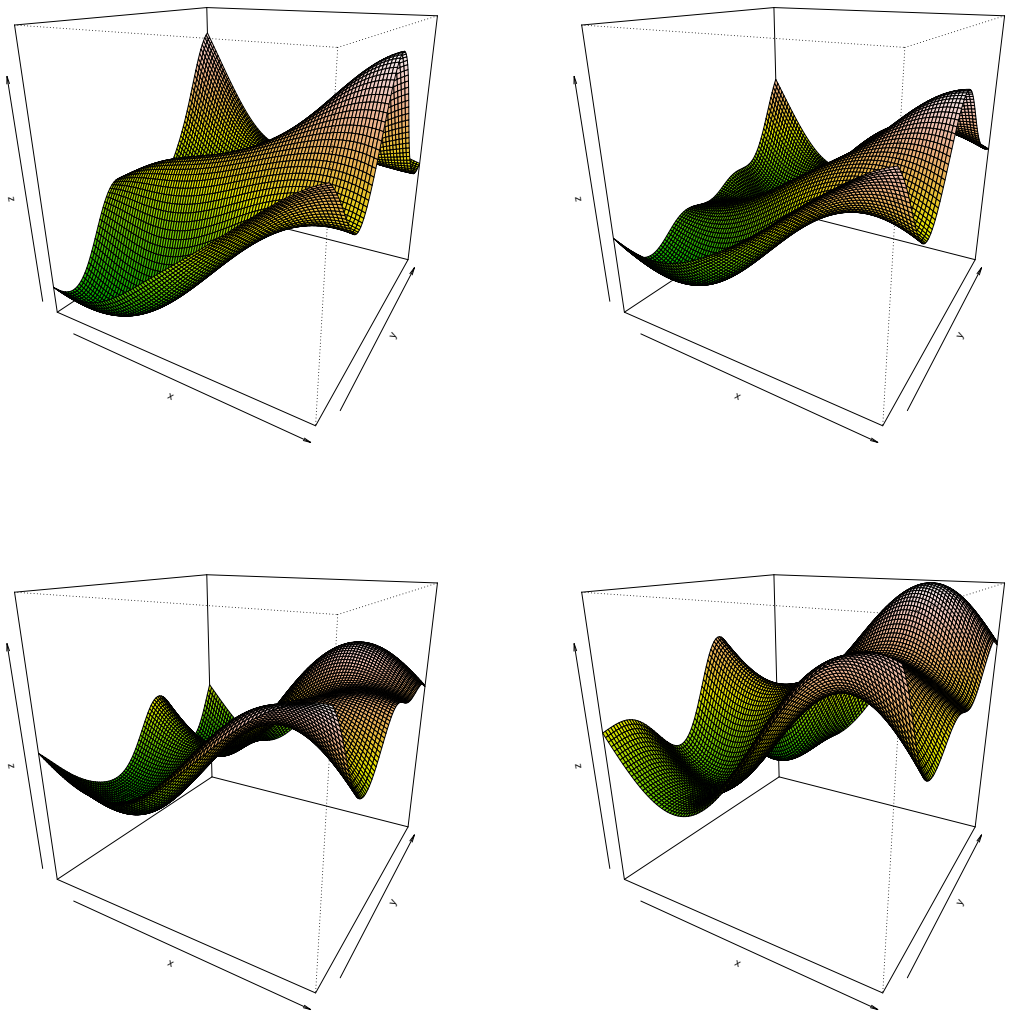

Figure 3.2: Simulated anisotropic Gaussian space-time process. The picture shows four consecutive time steps from the top left to the bottom right.

The equality in distribution (3.2) has to be understood in the following way: For arbitrary $k_1, k_2 \in \mathbb{N}$ let $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{k_1} \in S$ and $t_1, \ldots, t_{k_2} \in T$ be arbitrarily chosen index points. Then the finite-dimensional version of (3.2) holds:

$$\left(\frac{M^{(n)}_{\boldsymbol{s}_i,t_j} - b^{(n)}_{\boldsymbol{s}_i,t_j}}{a^{(n)}_{\boldsymbol{s}_i,t_j}} \right)_{\substack{i=1,\ldots,k_1,\\ j=1,\ldots,k_2}} \stackrel{\mathcal{D}}{=} (X_{\boldsymbol{s}_i,t_j})_{\substack{i=1,\ldots,k_1,\\ j=1,\ldots,k_2}}.$$

In a nutshell, the finite-dimensional margins of max-stable processes are max-stable and therefore follow extreme value distributions. As in Section 2.2, we assume in the following that the marginal distributions are standard Fréchet, that is, for all $\boldsymbol{s} \in \mathbb{R}^d$ and $t \in [0, \infty)$ we have

$$F(x) := \mathbb{P}(X_{\boldsymbol{s},t} \leq x) = \mathrm{e}^{-\frac{1}{x}}, \qquad x > 0.$$

In this case, in the definition of max-stable processes (3.2), the sequences $\left(a^{(n)}_{\boldsymbol{s},t}\right)$ and $\left(b^{(n)}_{\boldsymbol{s},t}\right)$ are chosen as $a^{(n)}_{\boldsymbol{s},t} = n$ and $b^{(n)}_{\boldsymbol{s},t} = 0$ for all $n \in \mathbb{N}$, $\boldsymbol{s} \in \mathbb{R}^d$ and $t \in [0, \infty)$. This is in accordance with the theory of univariate extremes (cf. the remarks after Theorem 2.4) because this choice leads to $\bar{F}\left(a^{(n)}_{\boldsymbol{s},t}\right) = 1 - \mathrm{e}^{-\frac{1}{n}} \sim \frac{1}{n}$ as $n \to \infty$. To prove that a stochastic process $\boldsymbol{X}$ as above with standard Fréchet margins is max-stable, one needs to show that (cf. de Haan [1984], equation (1))

$$\begin{aligned}
&\mathbb{P}\left[X_{\boldsymbol{s}_1,t_1} \leq x_1, \ldots, X_{\boldsymbol{s}_{k_1},t_{k_2}} \leq x_{k_1k_2}\right] \\
&= \mathbb{P}\left[\frac{M^{(n)}_{\boldsymbol{s}_1,t_1}}{n} \leq x_1, \ldots, \frac{M^{(n)}_{\boldsymbol{s}_{k_1},t_{k_2}}}{n} \leq x_{k_1k_2}\right] \\
&= \mathbb{P}\left[\frac{1}{n}X^{(1)}_{\boldsymbol{s}_1,t_1} \leq x_1, \ldots, \frac{1}{n}X^{(n)}_{\boldsymbol{s}_1,t_1} \leq x_1, \ldots, \frac{1}{n}X^{(1)}_{\boldsymbol{s}_{k_1},t_{k_2}} \leq x_{k_1k_2}, \ldots, \frac{1}{n}X^{(n)}_{\boldsymbol{s}_{k_1},t_{k_2}} \leq x_{k_1k_2}\right] \\
&= \mathbb{P}\left[\frac{1}{n}X^{(1)}_{\boldsymbol{s}_1,t_1} \leq x_1, \ldots, \frac{1}{n}X^{(1)}_{\boldsymbol{s}_{k_1},t_{k_2}} \leq x_{k_1k_2}, \ldots, \frac{1}{n}X^{(n)}_{\boldsymbol{s}_1,t_1} \leq x_1, \ldots, \frac{1}{n}X^{(n)}_{\boldsymbol{s}_{k_1},t_{k_2}} \leq x_{k_1k_2}\right] \\
&= \prod_{i=1}^{n} \mathbb{P}\left[X^{(i)}_{\boldsymbol{s}_1,t_1} \leq nx_1, \ldots, X^{(i)}_{\boldsymbol{s}_{k_1},t_{k_2}} \leq nx_{k_1k_2}\right] \\
&= \mathbb{P}^n\left[X_{\boldsymbol{s}_1,t_1} \leq nx_1, \ldots, X_{\boldsymbol{s}_{k_1},t_{k_2}} \leq nx_{k_1k_2}\right] \qquad (3.3)
\end{aligned}$$

for all n, k_1, $k_2 \in \mathbb{N}$ and all $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{k_1} \in \mathbb{R}^d$, $t_1, \ldots, t_{k_2} \in [0, \infty)$. In the second last and the last step we use that the $\boldsymbol{X}^{(i)}$ are independent copies of $\boldsymbol{X}$.

Example 3.3 (Smith's storm profile model)**.** *An example of a process that satisfies (3.3) is constructed as follows: Let $\{Y_k, T_k,\ k \in \mathbb{N}\}$ be the points of a Poisson random measure on $\mathbb{R}_+ \times [0,1]$ with intensity $(\mathrm{d}y/y^2)\mathrm{d}t$. Let T be some index set and $\{f_t(\cdot),\ t \in T\}$ be a set of nonnegative, measurable functions on $[0,1]$ such that*

$$\int_0^1 f_t(y)\mathrm{d}y < \infty, \qquad t \in T.$$

Then the process $\boldsymbol{X} = (X_t)_{t\in T}$ *with*

$$X_t := \sup_{k\geq 1} f_t(T_k)Y_k, \qquad t \in T, \tag{3.4}$$

is max-stable. To prove that (3.3) holds, one uses the properties of Poisson point processes (see Section 2.2.6 (iii)). The process in (3.4) was introduced by de Haan [1984] and later extended by Smith [1990].

In the following, our main focus lies on the max-stable *Brown-Resnick process*, introduced in Brown and Resnick [1977] and extended in Kabluchko, Schlather and de Haan [2009].

Definition 3.4 (Brown-Resnick Process)**.** *Let* $\{\xi_i,\ i \geq 1\}$ *be the points of a Poisson random measure on* $[0,\infty)$ *with intensity* $\xi^{-2}\mathrm{d}\xi$*. Let* $\boldsymbol{W} = \{W_t : t \in [0,\infty)\}$ *be a Gaussian process with stationary increments. Furthermore, let* $\boldsymbol{Y}^{(i)} = \left\{Y^{(i)}_{s,t} : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\right\}$*,* $i \in \mathbb{N}$*, be independent copies of a space-time process* $\boldsymbol{Y} = \{Y_{s,t} : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\}$ *satisfying*

$$\mathbb{E}[Y_{\boldsymbol{s},t}] < \infty \text{ and } Y_{\boldsymbol{s},t} \geq 0 \text{ almost surely for all } \boldsymbol{s} \in \mathbb{R}^d \text{ and } t \in [0,\infty).$$

Then the Brown-Resnick process *is defined as*

$$\eta_{\boldsymbol{s},t} := \bigvee_{i=1}^{\infty} \xi_i Y^{(i)}_{\boldsymbol{s},t}, \qquad \boldsymbol{s} \in \mathbb{R}^d,\ t \in [0,\infty). \tag{3.5}$$

The process $\boldsymbol{\eta} := \{\eta_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\}$ *is max-stable and has Fréchet margins (see Davis et al. [2012a]).*

The next section presents a detailed instruction of how to construct the Brown-Resnick process and how it can be used as a model for extremes observed in space and time.

3.3 Max-Stable Processes as Models for Extremes observed in Space and Time

In the first part of this section we show how to construct the Brown-Resnick process (3.5) as the limit of a sequence of point-wise maxima of independent Gaussian space-time processes, following Davis et al. [2012a].

3.3.1 Construction of the Brown-Resnick Process

Let $\boldsymbol{Z} = \{Z_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\}$ be a stationary Gaussian space-time process with mean 0, variance 1 and correlation function $\gamma = \gamma(\boldsymbol{h}, u)$, $\boldsymbol{h} \in \mathbb{R}^d$, $u \in \mathbb{R}$, as introduced in Section 3.1. A central assumption on γ is the following smoothness condition:

Assumption 3.1 (Smoothness Condition for γ, cf. Davis et al. [2012a], Ass. 2.1)**.** *There exist nonnegative sequences* $(s_{n,m})_{n\in\mathbb{N}}$, $m = 1,\dots,d$, *and* $(t_n)_{n\in\mathbb{N}}$ *that satisfy* $s_{n,m} \to 0$ *and* $t_n \to 0$ *as* $n \to \infty$, *and a nonnegative function* δ *such that*

$$(\log n)\,(1-\gamma(s_{n,1}h_1, s_{n,2}h_2, \dots, s_{n,d}h_d, t_n u)) \to \delta(h_1, h_2, \dots, h_d, u) > 0, \qquad n \to \infty, \tag{3.6}$$

where $\boldsymbol{h} = (h_1,\dots,h_d) \in \mathbb{R}^d\backslash\{\mathbf{0}\}$, $u \in \mathbb{R}$.

For the sake of simpler notation, set

$$S_n := \begin{pmatrix} s_{n,1} & 0 & 0 & \cdots & 0 \\ 0 & s_{n,2} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & 0 \\ 0 & 0 & \cdots & s_{n,d-1} & 0 \\ 0 & 0 & \cdots & 0 & s_{n,d} \end{pmatrix} \in \mathbb{R}^{d\times d}.$$

The following theorem (cf. Theorem 2.2 in Davis et al. [2012a]) states a possible way of constructing a Brown-Resnick process (see Definition 3.4), exploiting Assumption 3.1. It is originally based on Theorem 1 in Hüsler and Reiss [1989] and Theorem 17 in Kabluchko et al. [2009].

Theorem 3.5 (Construction of a Brown-Resnick Process)**.**
Let $\boldsymbol{Z}^{(i)} = \left\{Z^{(i)}_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\right\}$, $i \in \mathbb{N}$, *be independent copies of the Gaussian space-time process* $\boldsymbol{Z}$. *Suppose that its correlation function* γ *satisfies Assumption 3.1 with a nonnegative limit function* δ. *Furthermore, assume the existence of a metric* D *on* $\mathbb{R}^d \times [0,\infty)$ *such that the following condition holds:*

$$\delta(\boldsymbol{s}_1 - \boldsymbol{s}_2, t_1 - t_2) \le D((\boldsymbol{s}_1, t_1), (\boldsymbol{s}_2, t_2))^2 \text{ for all } \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{R}^d,\ t_1, t_2 \in [0,\infty). \tag{3.7}$$

Let Φ *be the standard normal distribution function. Setting*

$$\eta^{(n)}_{\boldsymbol{s},t} := \frac{1}{n}\bigvee_{i=1}^{n} -\frac{1}{\log\left(\Phi(Z^{(i)}_{S_n\boldsymbol{s},t_n t})\right)}, \qquad \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty),$$

we obtain

$$\eta^{(n)}_{\boldsymbol{s},t} \xrightarrow{\mathcal{L}} \eta_{\boldsymbol{s},t} := \bigvee_{i=1}^{\infty} \xi_i Y^{(i)}_{\boldsymbol{s},t}, \qquad n \to \infty, \tag{3.8}$$

where $\{\xi_i : i \in \mathbb{N}\}$ *is the set of points of a Poisson random measure as in Definition 3.4 and the* $\boldsymbol{Y}^{(i)} = \left\{Y^{(i)}_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\right\}$ *are independent copies of the process* $\boldsymbol{Y} = \{Y_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\}$ *with* $Y_{\boldsymbol{s},t} := \exp\{W_{\boldsymbol{s},t} - \delta(\boldsymbol{s},t)\}$. *The stochastic process* $\boldsymbol{W} = \{W_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\}$ *has mean 0, variance 1 and covariance function*

$$\mathrm{Cov}[W_{\boldsymbol{s}_1,t_1}, W_{\boldsymbol{s}_2,t_2}] = \delta(\boldsymbol{s}_1,t_1) + \delta(\boldsymbol{s}_2,t_2) - \delta(\boldsymbol{s}_1-\boldsymbol{s}_2, t_1-t_2).$$

The symbol "$\xrightarrow{\mathcal{L}}$*" in (3.8) stands for weak convergence on* $\mathcal{C}(\mathbb{R}^d \times [0,\infty))$.

Remark 3.6 (Bivariate distribution functions)**.** *For* $(\boldsymbol{s}_1, t_1)$ *and* $(\boldsymbol{s}_2, t_2) \in \mathbb{R}^d \times [0, \infty)$*, the bivariate distribution functions of the Brown-Resnick process* $\boldsymbol{\eta} = \{\eta_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{R}^d, t \in [0, \infty)\}$ *constructed in (3.8) are given by*

$$\begin{aligned} G(y_1, y_2) &:= \mathbb{P}(\eta_{\boldsymbol{s}_1,t_1} \leq y_1, \eta_{\boldsymbol{s}_2,t_2} \leq y_2) \\ &= \exp\left\{ -\frac{1}{y_1}\Phi\left[\frac{\log\left(\frac{y_2}{y_1}\right)}{2\sqrt{\delta(\boldsymbol{h},u)}} + \sqrt{\delta(\boldsymbol{h},u)} \right] - \frac{1}{y_2}\Phi\left[\frac{\log\left(\frac{y_1}{y_2}\right)}{2\sqrt{\delta(\boldsymbol{h},u)}} + \sqrt{\delta(\boldsymbol{h},u)} \right] \right\}, \end{aligned} \tag{3.9}$$

where $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$ *and* $u = t_1 - t_2$*.*

The proofs of Theorem 3.5 and Remark 3.6 can be found in Kabluchko et al. [2009] and Davis et al. [2012a]. We achieved a simulation of both isotropic and anisotropic max-stable random fields based on Theorem 3.5 in `R`, again using the package `RandomFields`. Figures 3.3 to 3.6 visualize the results. In the following subsections we apply concepts and methods of multivariate extreme value theory (see Section 2.2) to the finite-dimensional margins of the Brown-Resnick process constructed in Theorem 3.5.

3.3.2 Stable tail and Pickands Dependence Functions

Recall that the finite-dimensional margins of a max-stable process $\boldsymbol{\eta} = \{\eta_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{R}^d, t \in [0, \infty)\}$ follow extreme value distributions. So for $k \in \mathbb{N}$, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k \in \mathbb{R}^d$, $t_1, \ldots, t_k \in [0, \infty)$ and $y_1, \ldots, y_k \in [0, \infty]$ it holds that

$$\mathbb{P}\left(\eta_{\boldsymbol{s}_1,t_1} \leq y_1, \ldots, \eta_{\boldsymbol{s}_2,t_2} \leq y_k\right) =: G(y_1, \ldots, y_k) \overset{(2.31)}{=} \exp\left\{ -l\left(\frac{1}{y_1}, \ldots, \frac{1}{y_k}\right) \right\},$$

where l is the stable tail dependence function of G defined in (2.31). In particular, this applies for the Brown-Resnick space-time process $\boldsymbol{\eta}$ constructed in (3.8) and the two-dimensional ($k = 2$) random vectors $(\eta_{\boldsymbol{s}_1,t_1}, \eta_{\boldsymbol{s}_2,t_2})$:

$$\mathbb{P}\left(\eta_{\boldsymbol{s}_1,t_1} \leq y_1, \eta_{\boldsymbol{s}_2,t_2} \leq y_2\right) = G(y_1, y_2) = \exp\left\{ -l\left(\frac{1}{y_1}, \frac{1}{y_2}\right) \right\}, \tag{3.10}$$

where in this case, using the bivariate distribution functions of $\boldsymbol{\eta}$ given in (3.9), l can be expressed as

$$\begin{aligned} l\left(\frac{1}{y_1}, \frac{1}{y_2}\right) &\overset{(2.31)}{=} -\log G(y_1, y_2) \\ &\overset{(3.9)}{=} \frac{1}{y_1}\Phi\left[\frac{\log\left(\frac{y_2}{y_1}\right)}{2\sqrt{\delta(\boldsymbol{h},u)}} + \sqrt{\delta(\boldsymbol{h},u)} \right] + \frac{1}{y_2}\Phi\left[\frac{\log\left(\frac{y_1}{y_2}\right)}{2\sqrt{\delta(\boldsymbol{h},u)}} + \sqrt{\delta(\boldsymbol{h},u)} \right], \end{aligned} \tag{3.11}$$

where $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$ and $u = |t_1 - t_2|$. Obviously l depends on $\delta(\boldsymbol{h}, u)$, so we write $l(y_1, y_2) = l(y_1, y_2; \delta(\boldsymbol{h}, u))$. Next recall from Section 2.2.4 the role of the Pickands dependence function A of G in the bivariate case and its tight connection to the stable tail

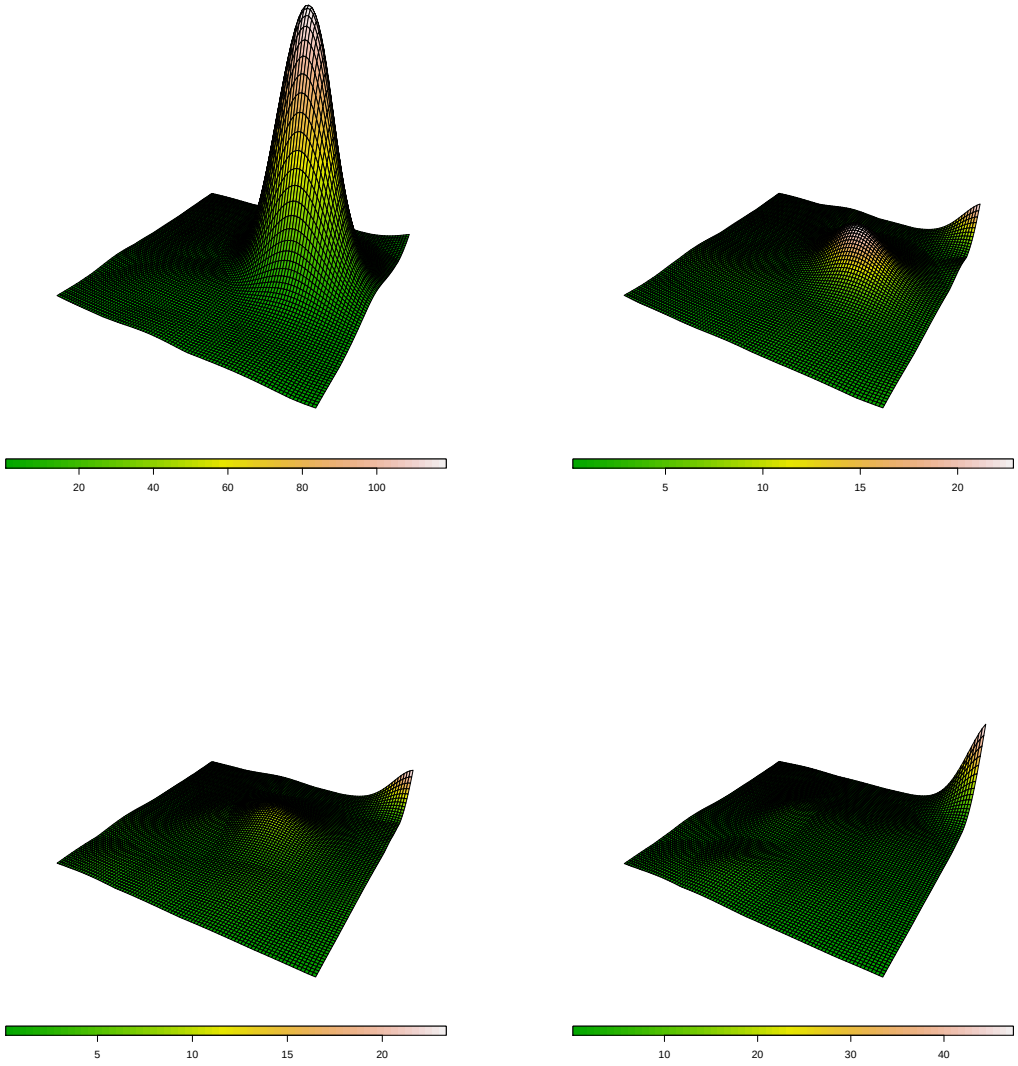

Figure 3.3: Perspective plot of a simulated isotropic max-stable random field. The picture shows four consecutive time steps from the top left to the bottom right.

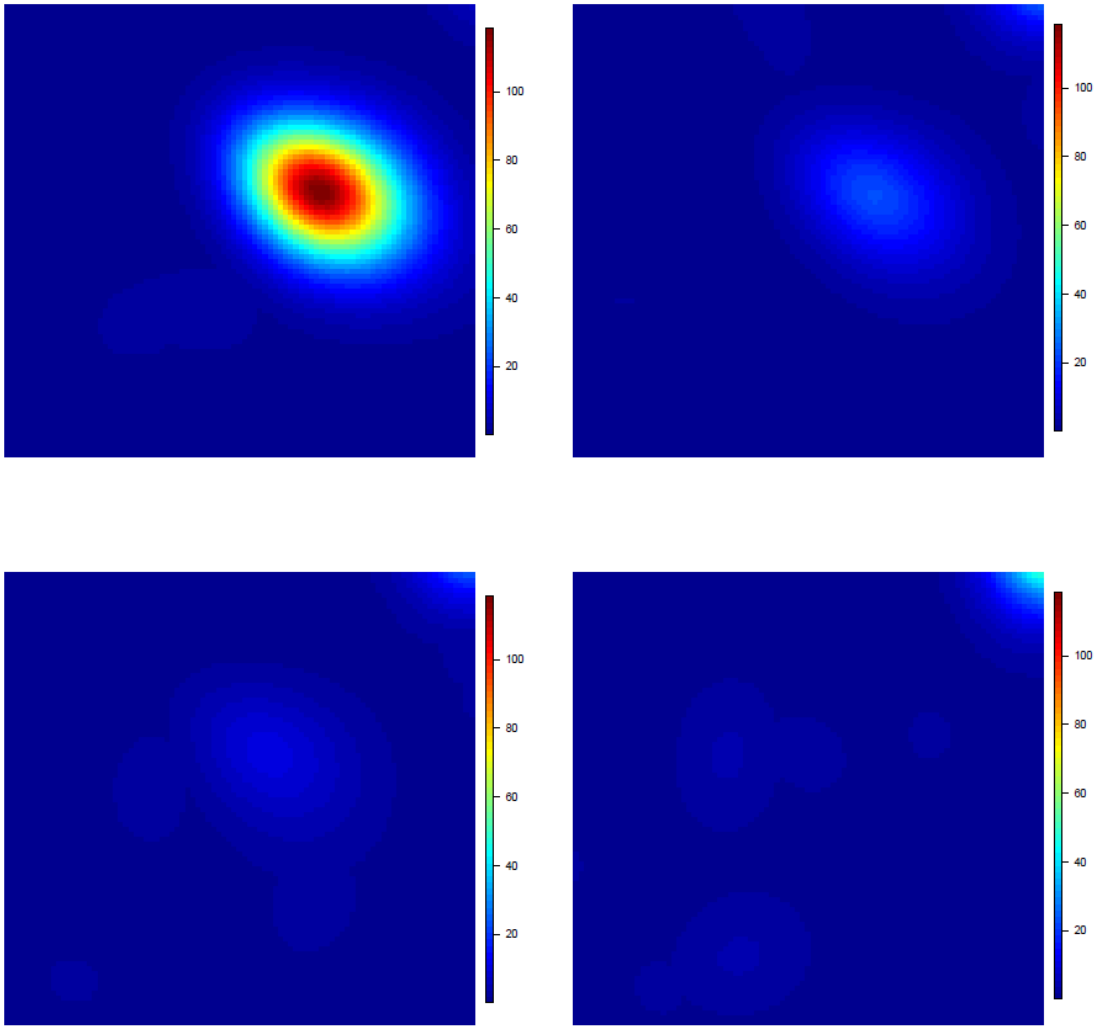

Figure 3.4: Image plot of the simulated isotropic max-stable random field (cf. Figure 3.3). The picture shows four consecutive time steps from the top left to the bottom right.

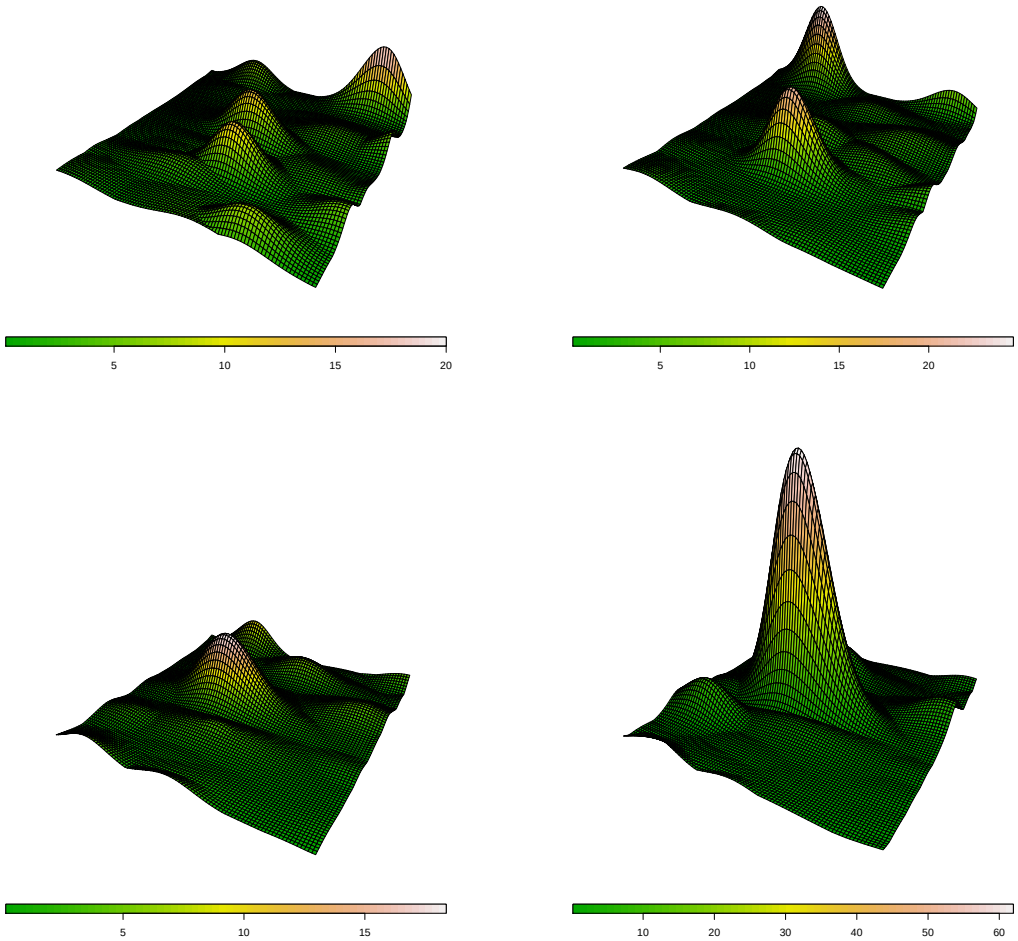

Figure 3.5: Perspective plot of a simulated anisotropic max-stable random field. The picture shows four consecutive time steps from the top left to the bottom right.

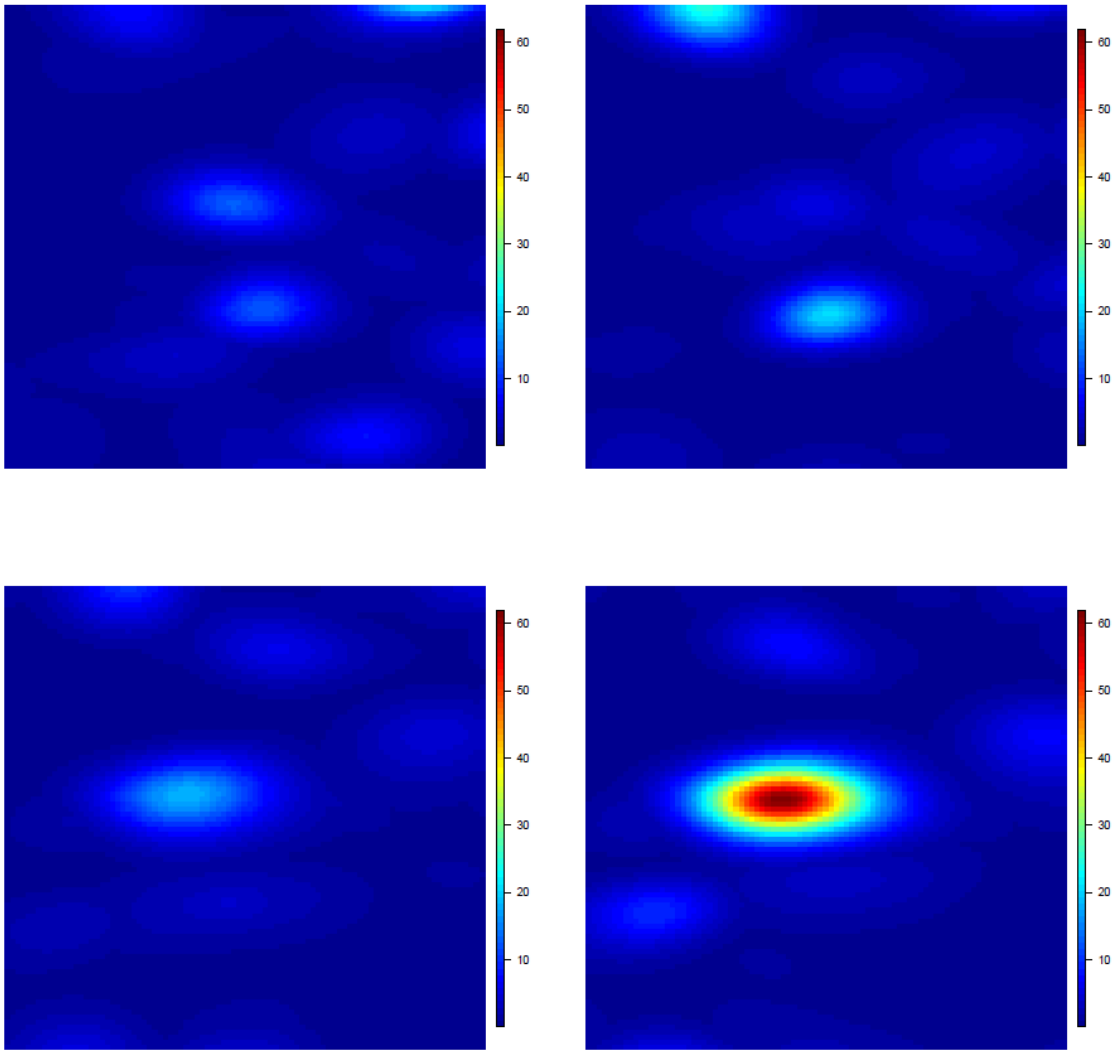

Figure 3.6: Image plot of the simulated anisotropic max-stable random field (cf. Figure 3.5). The picture shows four consecutive time steps from the top left to the bottom right.

dependence function l. Using equation (2.46) we find

$$G(y_1, y_2) = \exp\left\{-l\left(\frac{1}{y_1}, \frac{1}{y_2}; \delta(\boldsymbol{h}, u)\right)\right\} \overset{(2.46)}{=} \exp\left\{-\frac{y_1 + y_2}{y_1 y_2} A\left(\frac{y_1}{y_1 + y_2}; \delta(\boldsymbol{h}, u)\right)\right\} \tag{3.12}$$

Using the original definition of the Pickands dependence function (cf. (2.45)) we can write

$$\begin{aligned}
&A\left(\frac{y_1}{y_1 + y_2}; \delta(\boldsymbol{h}, u)\right) \\
&\overset{(2.45)}{=} l\left(\frac{y_1}{y_1 + y_2}, 1 - \frac{y_1}{y_1 + y_2}\right) \\
&= l\left(\frac{y_1}{y_1 + y_2}, \frac{y_2}{y_1 + y_2}\right) \\
&= l\left(\frac{1}{\frac{y_1}{y_1+y_2}}, \frac{1}{\frac{y_2}{y_1+y_2}}\right) \\
&\overset{(3.11)}{=} \frac{y_1 + y_2}{y_1}\Phi\left[\frac{\log\left(\frac{y_1}{y_2}\right)}{2\sqrt{\delta(\boldsymbol{h}, u)}} + \sqrt{\delta(\boldsymbol{h}, u)}\right] + \frac{y_1 + y_2}{y_2}\Phi\left[\frac{\log\left(\frac{y_2}{y_1}\right)}{2\sqrt{\delta(\boldsymbol{h}, u)}} + \sqrt{\delta(\boldsymbol{h}, u)}\right].
\end{aligned} \tag{3.13}$$

3.3.3 Tail Dependence Coefficient

An important summary measure for extremal dependence introduced in Chapter 2 is the tail dependence coefficient, defined in (2.57). In the current context, for the 2-dimensional random vectors $(\eta_{\boldsymbol{s}_1,t_1}, \eta_{\boldsymbol{s}_2,t_2})$, it is given by

$$\mathcal{X}(\boldsymbol{h}, u) = \lim_{x \to 1} \mathbb{P}\left[\eta_{\boldsymbol{s}_1,t_1} > G^{\leftarrow}_{\eta_{\boldsymbol{s}_1,t_1}}(x) | \eta_{\boldsymbol{s}_2,t_2} > G^{\leftarrow}_{\eta_{\boldsymbol{s}_2,t_2}}(x)\right], \quad (\boldsymbol{s}_1, t_1),\ (\boldsymbol{s}_2, t_2) \in \mathbb{R}^d \times [0, \infty), \tag{3.14}$$

where $G^{\leftarrow}_{\eta_{\boldsymbol{s}_j,t_j}}$ for $j \in \{1, 2\}$ is the generalized inverse of the distribution function of $\eta_{\boldsymbol{s}_j,t_j}$, $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$ and $u = t_1 - t_2$. If $\mathcal{X}(\boldsymbol{h}, u) = 0$ then $\eta_{\boldsymbol{s}_1,t_1}$ and $\eta_{\boldsymbol{s}_2,t_2}$ are asymptotically independent, if $\mathcal{X}(\boldsymbol{h}, u) = 1$, they are asymptotically completely dependent. For the Brown-Resnick process in (3.8) it turns out that

$$\mathcal{X}(\boldsymbol{h}, u) = 2(1 - \Phi(\sqrt{\delta(\boldsymbol{h}, u)})), \tag{3.15}$$

so if $\delta(\boldsymbol{h}, u) \to \infty$ when $\boldsymbol{h}$ (the spatial lag) and u (the temporal lag) change, the components tend to be asymptotically independent, and if $\delta(\boldsymbol{h}, u) \to 0$, they tend to be asympotically completely dependent.

Chapter 4

Anisotropic Models for the Correlation Function γ of the underlying Gaussian Space-Time Process and Estimation Techniques

This chapter is divided into three sections: Section 4.1 examines in more detail the correlation function γ of the underlying Gaussian space-time process $\boldsymbol{Z}$ that plays a central role in the construction of the Brown-Resnick process $\boldsymbol{\eta}$ in Theorem 3.5. It is based on Davis et al. [2012a], Section 4, where plenty of possible correlation functions are presented. However, their focus lies on spatial isotropic correlation functions and they only present a way of how to adapt those if the anisotropic context is wished. In this thesis we do not restrict to spatial isotropy but always consider the general case. Section 4.2 deals with a pairwise likelihood method to estimate the parameters of the correlation model that underlies the max-stable process: We adapt the pairwise likelihood method introduced in Davis, Klüppelberg and Steinkohl [2012b] to the anisotropic setting and show that the theorems of consistency and asymptotic normality still hold.

4.1 Anisotropic Models

We start with a condition ensuring that the smoothness Assumption 3.1 holds, which is needed for the construction of the Brown-Resnick process in Theorem 3.5.

Assumption 4.1 (cf. Assumption 4.1 in Davis et al. [2012a])**.** *Let $\boldsymbol{h} = (h_1, \dots, h_d) \in \mathbb{R}^d$ and $u \in \mathbb{R}$. The correlation function γ of $\boldsymbol{Z}$ possesses the following expansion around $(\mathbf{0}, 0)$:*

$$\gamma(\boldsymbol{h}, u) = 1 - \sum_{m=1}^{d} C_m |h_m|^{\alpha_m} - C_{d+1}|u|^{\alpha_{d+1}} + \mathcal{O}(\sum_{m=1}^{d} |h_m|^{\alpha_m} + |u|^{\alpha_{d+1}}),$$

where $0 < \alpha_m \leq 2$ and $C_m > 0$ for all $m = 1, \dots, d+1$. The constants C_m are independent of $\boldsymbol{h}$ and u.

The parameters α_m in Assumption 4.1 are directly related to the smoothness of the sample paths of the process $\boldsymbol{Z}$, see Davis et al. [2012a], Section 4. For instance in the case $d = 2$, if $\alpha_1 = \alpha_2 = \alpha_3 = 2$, then the process is mean-square differentiable, meaning that all second-order partial derivatives of $\gamma(h_1, h_2, u)$ exist in $(0, 0, 0)$. On top of that, Assumption 4.1 ensures almost sure continuity. Another advantage of the proposed expansion is the fact that the function δ, which arised in Assumption 3.1, turns out to be expressable in closed form. Under Assumption 4.1 we have for the nonnegative sequences $s_{n,m} \to 0$ and $t_n \to 0$:

$$\begin{aligned}
&(\log n)(1 - \gamma(s_{n,1}h_1, \dots, s_{n,d}h_d, t_n u)) \\
&= (\log n) \left[\sum_{m=1}^{d} C_m (s_{n,m}|h_m|)^{\alpha_m} + C_{d+1}(t_n|u|)^{\alpha_{d+1}} + \mathcal{O}\left(\sum_{m=1}^{d} (s_{n,m}|h_m|)^{\alpha_m} + (t_n|u|)^{\alpha_{d+1}} \right) \right] \\
&= (\log n) \left[\sum_{m=1}^{d} C_m s_{n,m}^{\alpha_m} |h_m|^{\alpha_m} + C_{d+1} t_n^{\alpha_{d+1}} |u|^{\alpha_{d+1}} + \mathcal{O}\left(\sum_{m=1}^{d} s_{n,m}^{\alpha_m} |h_m|^{\alpha_m} + t_n^{\alpha_{d+1}} |u|^{\alpha_{d+1}} \right) \right].
\end{aligned}$$

Solving for $s_{n,m}$ and t_n in the equations $s_{n,m}^{\alpha_m} = (\log n)^{-1}$ and $t_n^{\alpha_{d+1}} = (\log n)^{-1}$, respectively, we obtain $s_{n,m} = (\log n)^{-\frac{1}{\alpha_m}}$, $m = 1, \dots, d$, $t_n = (\log n)^{-\frac{1}{\alpha_{d+1}}}$ and thus,

$$(\log n)(1 - \gamma(s_{n,1}h_1, \dots, s_{n,d}h_d, t_n u)) \to \sum_{m=1}^{d} C_m |h_m|^{\alpha_m} + C_{d+1}|u|^{\alpha_{d+1}} =: \delta(h_1, \dots, h_d, u), \tag{4.1}$$

as $n \to \infty$. The tightness condition (3.7) is satisfied by choosing the metric D as

$$\begin{aligned}
&D((s_1^{(1)}, \dots, s_1^{(d)}, t_1), (s_2^{(1)}, \dots, s_2^{(d)}, t_2)) \\
&= \max\left\{ |s_1^{(1)} - s_2^{(1)}|^{\frac{\alpha_1}{2}}, |s_1^{(2)} - s_2^{(2)}|^{\frac{\alpha_2}{2}}, \dots, |s_1^{(d)} - s_2^{(d)}|^{\frac{\alpha_d}{2}}, |t_1 - t_2|^{\frac{\alpha_2}{2}} \right\},
\end{aligned} \tag{4.2}$$

where $\boldsymbol{s}_1 = (s_1^{(1)}, \dots, s_1^{(d)})$ and $\boldsymbol{s}_2 = (s_2^{(1)}, \dots, s_2^{(d)}) \in \mathbb{R}^d$, t_1 and $t_2 \in [0, \infty)$.

Examples

In the following, some examples of anisotropic correlation functions satisfying Assumption 4.1 are presented.

(1) *The "stable" class.* For $\boldsymbol{h} = (h_1, \dots, h_d)$, consider the correlation function

$$\begin{aligned}
\gamma(\boldsymbol{h}, u) &= \exp\left\{ - \left\| \left(C_1^{\frac{1}{2}} h_1, C_2^{\frac{1}{2}} h_2, \dots, C_d^{\frac{1}{2}} h_d, C_{d+1}^{\frac{1}{2}} u \right) \right\|^2 \right\} \\
&= \exp\left\{ - \sum_{m=1}^{d} C_m h_m^2 - C_{d+1} u^2 \right\},
\end{aligned}$$

where $C_1, \dots, C_{m+1}$ are some positive constants. If one insists on $C_1 = C_2 = \cdots = C_d$ the result is $\gamma(\boldsymbol{h}, u) = \exp\{-C_1 \|\boldsymbol{h}\|^2 - C_{d+1} u^2\}$ and γ is spatially isotropic. Notice

that the choice intuitively makes sense because if the norm of the vector $(\boldsymbol{h}, u)$ tends to 0, i.e. the distance between the points $(\boldsymbol{s}_1, t_1)$ and $(\boldsymbol{s}_2, t_2)$ is small, then the correlation tends to 1 at a certain rate. On the other hand, if the norm gets larger, the correlation tends to 0. Cleary, the correlation function γ allows for the following Taylor expansion:

$$\gamma(\boldsymbol{h}, u) = 1 - \sum_{m=1}^{d} C_m h_m^2 - C_{d+1} h_{d+1}^2 + \mathcal{O}(\sum_{m=1}^{d} h_m^2 + u^2).$$

So Assumption 4.1 is satisfied with $\alpha_m = 2$, $m = 1, \ldots, (d+1)$, i.e. this is a very smooth example of a correlation function. As in (4.1) we choose $s_{n,m} = t_n = (\log n)^{-\frac{1}{2}}$, obtaining

$$\delta(\boldsymbol{h}, u) = \sum_{m=1}^{d} C_m h_m^2 + C_{d+1} u^2$$

(2) *Iaco-Cesare.* The basic version of the Iaco-Cesare correlation function is as follows:

$$\gamma(\boldsymbol{h}, u) = \left(1 + \theta_1 \|\boldsymbol{h}\|^{\nu} + \theta_2 |u|^{\lambda}\right)^{-C},$$

where $\nu, \lambda \in [1, 2]$, $C \geq \frac{d+1}{2}$ and $\theta_1, \theta_2 \geq 0$ (cf. for example the documentation of the `R`-package `RandomFields` by Schlather [2001]). However, the correlation function is still isotropic. So our intention is to generalize it in the following: Consider a correlation function γ of the form

$$\gamma(\boldsymbol{h}, u) = \left(1 + \sum_{m=1}^{d} \frac{C_m}{C} |h_m|^{\nu_m} + \frac{C_{d+1}}{C} |u|^{\lambda}\right)^{-C}, \boldsymbol{h} = (h_1, \ldots, h_d) \in \mathbb{R}^d,\ u \in [0, \infty), \tag{4.3}$$

where ν_m, $m = 1, \ldots, d \in [1, 2]$, $\lambda \in [1, 2]$ and $C \geq \frac{d+1}{2}$. Taylor expansion around 0 of the function $f : [0, \infty) \to (0, 1]$, $x \mapsto \frac{1}{(1+x)^C}$ yields

$$f(x) = 1 - Cx + \mathcal{O}(x),$$

leading to

$$\begin{aligned}
\gamma(\boldsymbol{h}, u) &= \left(1 + \sum_{m=1}^{d} \frac{C_m}{C} |h_m|^{\nu_m} + \frac{C_{d+1}}{C} |u|^{\lambda}\right)^{-C} \\
&= f\left(\sum_{m=1}^{d} \frac{C_m}{C} |h_m|^{\nu_m} + \frac{C_{d+1}}{C} |u|^{\lambda}\right) \\
&= 1 - C\left(\sum_{m=1}^{d} \frac{C_m}{C} |h_m|^{\nu_m} + \frac{C_{d+1}}{C} |u|^{\lambda}\right) + \mathcal{O}\left(\sum_{m=1}^{d} \frac{C_m}{C} |h_m|^{\nu_m} + \frac{C_{d+1}}{C} |u|^{\lambda}\right) \\
&= 1 - \left(\sum_{m=1}^{d} C_m |h_m|^{\nu_m} + C_{d+1} |u|^{\lambda}\right) + \mathcal{O}\left(\sum_{m=1}^{d} |h_m|^{\nu_m} + |u|^{\lambda}\right),
\end{aligned}$$

so Assumption 4.1 is satisfied with $\alpha_m = \nu_m$, $m = 1, \dots, d$, $\alpha_{d+1} = \lambda$ and as in (4.1) we have, as $n \to \infty$,

$$(\log n)(1 - \gamma(s_{n,1}h_1, \dots, s_{n,d}h_d, t_n u)) \to \sum_{m=1}^{d} C_m |h_m|^{\alpha_m} + C_{d+1}|u|^{\alpha_{d+1}} = \delta(\boldsymbol{h}, u)$$

for $s_{n,m} = (\log n)^{-\frac{1}{\alpha_m}}$, $m = 1, \dots, d$ and $t_n = (\log n)^{-\frac{1}{\alpha_{d+1}}}$. The function γ as defined in (4.3) is an example of a non-separable anisotropic correlation function.

Example 4.1 (Iaco-Cesare for $d = 2$, $\nu_1 = 2$, $\nu_2 = 2$, $\lambda = 1$, $C = 2$, $C_1 = 0.03$, $C_2 = 0.06$, $C_3 = 0.05$)**.** *Within the given context we have:*

- $\gamma(h_1, h_2, u) = \left(1 + \frac{0.03}{2}|h_1|^2 + \frac{0.06}{2}|h_2|^2 + \frac{0.05}{2}|u|\right)^{-2}$
- $\delta(h_1, h_2, u) = 0.03|h_1|^2 + 0.06|h_2|^2 + 0.05|u|$.

The left columns of Figures 4.1 and 4.2 show different contour plots of correlation functions of the stable and the Iaco-Cesare class, respectively. The plots were constructed using the `R`-functions `image.plot` and `contour` of the package `fields`. For the sake of simplicity and better comparability we choose $d = 2$. The parameters of the isotropic stable class correlation function take the values $\alpha_1 = \alpha_2 = \alpha_3 = 2$, $C_1 = C_2 = 0.03$ and $C_3 = 0.05$. For the anisotropic case we set $C_2 = 0.04$. The values of the Iaco-Cesare parameters are chosen as $\nu_1 = \nu_2 = 2$, $\lambda = 1$, $C_1 = C_2 = 0.03$, $C_3 = 0.05$ and $C = 2$ in the isotropic case and in the anisotropic case we slightly change C_2 to 0.06. Where h_1 is plotted against h_2, the temporal lag u is set equal to 0. What should particular be noticed is the difference between the isotropic and the anisotropic plots: Whereas in the isotropic case, the plots look like a quarter of a circle, the anisotropic plots rather yield ellipses. The former circles are stretched into one direction, which is also remarked in Davis et al. [2012a], Section 4.2. This also makes sense in a natural context because for instance, wind turns out to follow a certain direction and not to spread equally into all directions. The right columns of both figures show the corresponding tail dependence coefficients $\mathcal{X}(\boldsymbol{h}, u) = 2(1 - \Phi(\sqrt{\delta(\boldsymbol{h}, u)}))$ (cf. (3.15)) for the respective classes of correlation functions. One can see that they have basically the same shape as the corresponding correlation functions.

4.2 Pairwise Likelihood Methods

This section is based upon Davis et al. [2012b], Sections 3-5 and Steinkohl [2012], Chapter 4.

4.2.1 Introduction to Pairwise Likelihood Estimation

Recall that we assume the limit function δ, which arised in (3.1) and is directly connected to the underlying correlation function γ of the Gaussian space-time process, to be of the following shape:

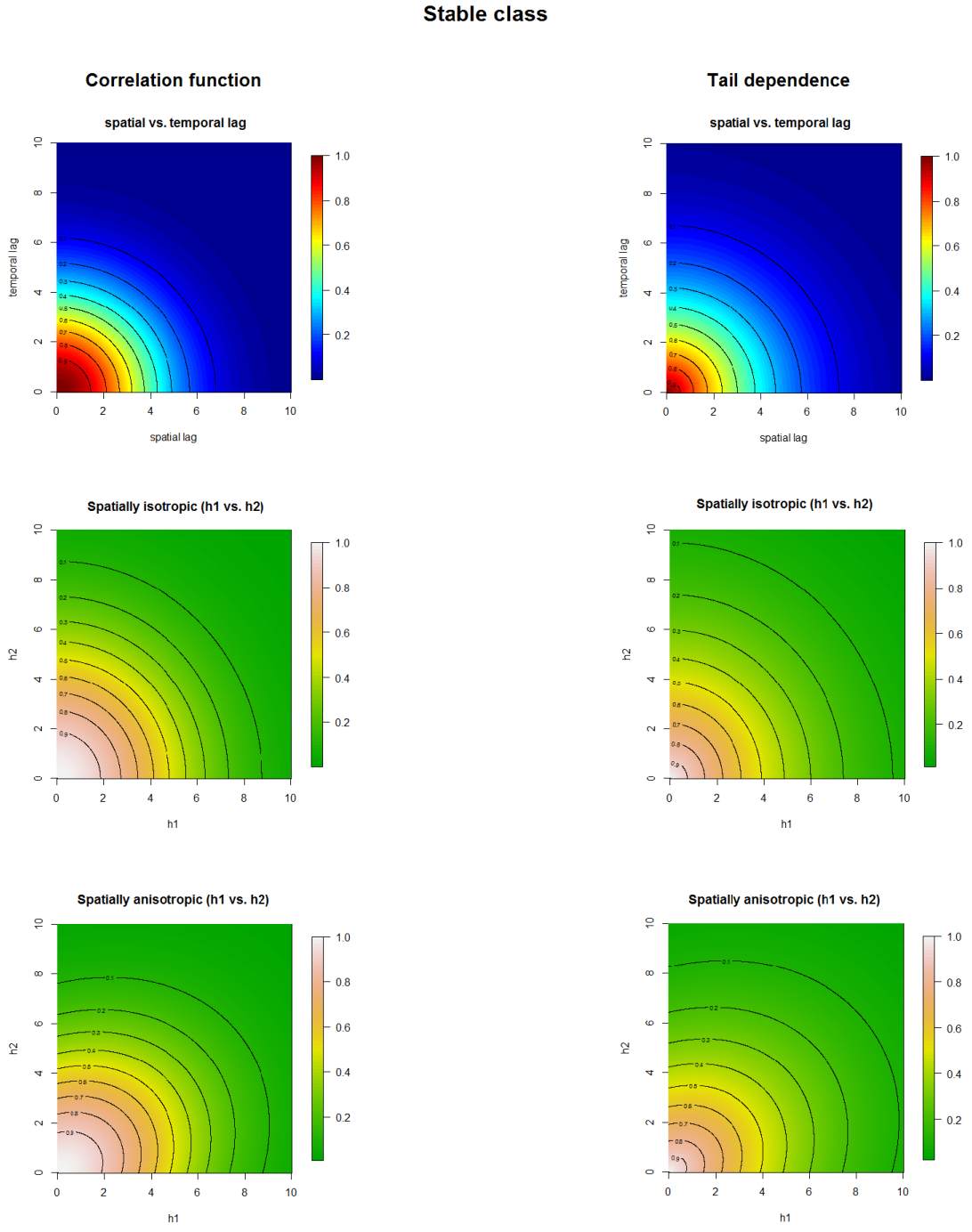

Figure 4.1: Stable class correlation function and tail dependence coefficient for $d = 2$, isotropic and anisotropic case.

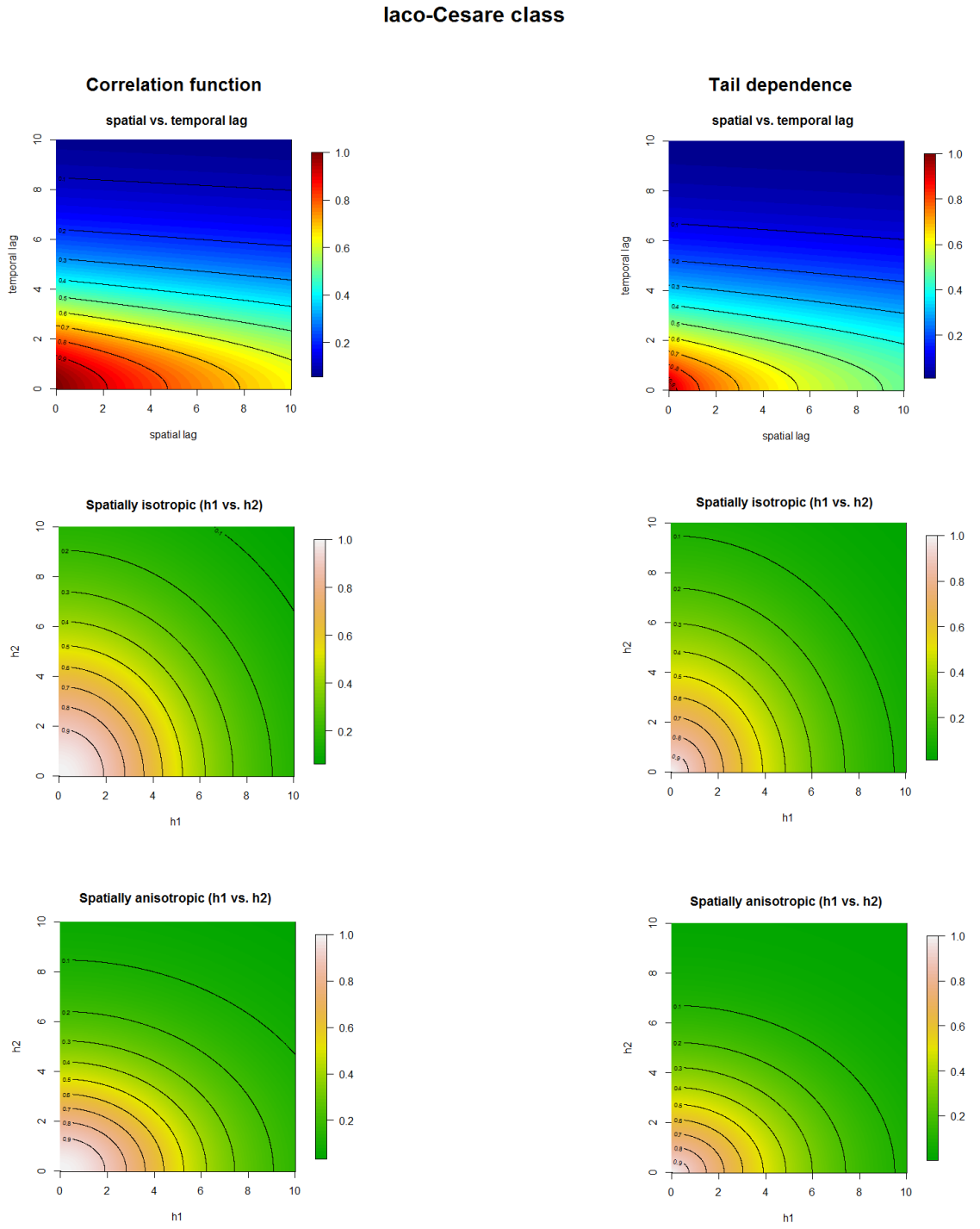

Figure 4.2: Iaco-Cesare class correlation function and tail dependence coefficient for $d = 2$, isotropic and anisotropic case.

$$\delta(\boldsymbol{h},u)=\delta(h_1,\dots,h_d,u)=C_1|h_1|^{\alpha_1}+\cdots+C_d|h_d|^{\alpha_d}+C_{d+1}|u|^{\alpha_{d+1}},\qquad \boldsymbol{h}\in\mathbb{R}^d,u\in\mathbb{R},\tag{4.4}$$

where $C_m>0$ and $0<\alpha_m\leq 2$ for all $m=1,\dots,d+1$. As stated in Remark 3.6, the bivariate distribution functions of the max-stable Brown-Resnick process $\boldsymbol{\eta}=\left\{\eta_{\boldsymbol{s},t}:\boldsymbol{s}\in\mathbb{R}^d,t\in[0,\infty)\right\}$ constructed in Theorem 3.5 depend on δ and are given by

$$\begin{aligned}G(y_1,y_2)&=\mathbb{P}(\eta_{\boldsymbol{s}_1,t_1}\leq y_1,\eta_{\boldsymbol{s}_2,t_2}\leq y_2)\\&=\exp\left\{-\frac{1}{y_1}\Phi\left[\frac{\log\left(\frac{y_2}{y_1}\right)}{2\sqrt{\delta(\boldsymbol{h},u)}}+\sqrt{\delta(\boldsymbol{h},u)}\right]-\frac{1}{y_2}\Phi\left[\frac{\log\left(\frac{y_1}{y_2}\right)}{2\sqrt{\delta(\boldsymbol{h},u)}}+\sqrt{\delta(\boldsymbol{h},u)}\right]\right\},\end{aligned}$$

for $\boldsymbol{h}=\boldsymbol{s}_1-\boldsymbol{s}_2$, $u=t_1-t_2$, $\boldsymbol{s}_1$, $\boldsymbol{s}_2\in\mathbb{R}^d$ and t_1, $t_2\in[0,\infty)$. Obviously, for fixed $(\boldsymbol{s}_1,t_1)$ and $(\boldsymbol{s}_2,t_2)$, the distribution function of $(\eta_{\boldsymbol{s}_1,t_1},\eta_{\boldsymbol{s}_2,t_2})$ is independent of the function δ, so we can adopt the results of Davis et al. [2012b], Section 3.2, who computed the bivariate log-density g, because the difference to the setting here only consists of the different δ-functions. The log-density g is given by:

$$\log g(y_1,y_2)=-V(y_1,y_2)+\log\left(V_1(y_1,y_2)V_2(y_1,y_2)-V_{12}(y_1,y_2)\right),\tag{4.5}$$

where

$$V(y_1,y_2)=\frac{1}{y_1}\Phi\left(\frac{\log\frac{y_2}{y_1}}{2\sqrt{\delta(\boldsymbol{h},u)}}+\sqrt{\delta(\boldsymbol{h},u)}\right)+\frac{1}{y_2}\Phi\left(\frac{\log\frac{y_1}{y_2}}{2\sqrt{\delta(\boldsymbol{h},u)}}+\sqrt{\delta(\boldsymbol{h},u)}\right),\tag{4.6}$$

$$\begin{aligned}V_1(y_1,y_2)&=\frac{\partial V(y_1,y_2)}{\partial y_1}\\&=-\frac{1}{y_1^2}\Phi\left(\frac{\log\frac{y_2}{y_1}}{2\sqrt{\delta(\boldsymbol{h},u)}}+\sqrt{\delta(\boldsymbol{h},u)}\right)-\frac{1}{2\sqrt{\delta(\boldsymbol{h},u)}y_1^2}\varphi\left(\frac{\log\frac{y_2}{y_1}}{2\sqrt{\delta(\boldsymbol{h},u)}}+\sqrt{\delta(\boldsymbol{h},u)}\right)\\&\quad+\frac{1}{2\sqrt{\delta(\boldsymbol{h},u)}y_1y_2}\varphi\left(\frac{\log\frac{y_1}{y_2}}{2\sqrt{\delta(\boldsymbol{h},u)}}+\sqrt{\delta(\boldsymbol{h},u)}\right),\end{aligned}\tag{4.7}$$

$$\begin{aligned}V_2(y_1,y_2)&=\frac{\partial V(y_1,y_2)}{\partial y_2}\\&=-\frac{1}{y_2^2}\Phi\left(\frac{\log\frac{y_1}{y_2}}{2\sqrt{\delta(\boldsymbol{h},u)}}+\sqrt{\delta(\boldsymbol{h},u)}\right)-\frac{1}{2\sqrt{\delta(\boldsymbol{h},u)}y_2^2}\varphi\left(\frac{\log\frac{y_1}{y_2}}{2\sqrt{\delta(\boldsymbol{h},u)}}+\sqrt{\delta(\boldsymbol{h},u)}\right)\\&\quad+\frac{1}{2\sqrt{\delta(\boldsymbol{h},u)}y_2y_1}\varphi\left(\frac{\log\frac{y_2}{y_1}}{2\sqrt{\delta(\boldsymbol{h},u)}}+\sqrt{\delta(\boldsymbol{h},u)}\right)\end{aligned}\tag{4.8}$$

and

$$\begin{aligned} V_{12}(y_1,y_2) &= \frac{\partial^2 V(y_1,y_2)}{\partial y_1 \partial y_2} \\ &= -\frac{2\sqrt{\delta(\boldsymbol{h},u)} - \frac{\log\frac{y_2}{y_1}}{2\sqrt{\delta(\boldsymbol{h},u)}} + \sqrt{\delta(\boldsymbol{h},u)}}{4\delta(\boldsymbol{h},u)y_1^2 y_2}\varphi\left(\frac{\log\frac{y_2}{y_1}}{2\sqrt{\delta(\boldsymbol{h},u)}} + \sqrt{\delta(\boldsymbol{h},u)}\right) \\ &\quad - \frac{2\sqrt{\delta(\boldsymbol{h},u)} - \frac{\log\frac{y_1}{y_2}}{2\sqrt{\delta(\boldsymbol{h},u)}} + \sqrt{\delta(\boldsymbol{h},u)}}{4\delta(\boldsymbol{h},u)y_2^2 y_1}\varphi\left(\frac{\log\frac{y_1}{y_2}}{2\sqrt{\delta(\boldsymbol{h},u)}} + \sqrt{\delta(\boldsymbol{h},u)}\right). \end{aligned} \tag{4.9}$$

The greek letter φ denotes the density of the univariate standard normal distribution. Given data that are assumed to be generated by a max-stable process η as above, the aim is to estimate the parameters C_m and α_m, $m = 1,\dots,d+1$ of the function δ. As standard maximum likelihood estimation is known to be computationally intractable for max-stable processes, one may use *composite likelihood methods* instead. Composite likelihood estimation has its origin in Besag [1974] and Lindsay [1988]. The pairwise likelihood estimation that is presented in the following is a special kind of the composite likelihood method. For the sake of convenience we first present the derivation of the pairwise likelihood function for the case of $d = 2$ dimensions, where the function δ has the anisotropic form

$$\delta(h_1,h_2,u) = \delta_{\boldsymbol{\theta}}(h_1,h_2,u) = C_1|h_1|^{\alpha_1} + C_2|h_2|^{\alpha_2} + C_3|u|^{\alpha_3}.$$

The parameter vector $\boldsymbol{\theta} = (C_1,C_2,C_3,\alpha_1,\alpha_2,\alpha_3)$ is part of the parameter space

$$\Theta := \{\boldsymbol{\theta} = (C_1,C_2,C_3,\alpha_1,\alpha_2,\alpha_3) : C_m \in (0,\infty),\ \alpha_m \in (0,2],\ m = 1,2,3\}.$$

The space-time setting is modelled as follows: We assume to consider M^2 spatial locations at T different time points, where M and T are positive integers. The spatial locations lie on a grid and can be summarized in the set $\{\boldsymbol{s} = (z_{i_1}, z_{i_2}) : i_1, i_2 \in \{1,\dots,M\}\}$. We use the notation (z_{i_1}, z_{i_2}) to determine the nature of the grid. For instance, if $(z_{i_1}, z_{i_2}) = (i_1, i_2)$, we assume the grid to be a regular squarish one. The time points are enumerated in the ordered way, that is, $0 \le t_1 < t_2 < \cdots < t_T$. In the context of this setting, the pairwise log-likelihood function can be written in its general form as a function of $\boldsymbol{\theta} \in \Theta$ as

$$\begin{aligned} &PL^{(M,T)}(\boldsymbol{\theta}) = \\ &\sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{j_1=i_1}^{M}\sum_{j_2=\mathbf{1}_{\{j_1=i_1\}}i_2+1}^{M}\sum_{k=1}^{T-1}\sum_{l=k+1}^{T} w_{i_1,j_1}^{(M)} w_{i_2,j_2}^{(M)} w_{k,l}^{(T)} \log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(z_{i_1},z_{i_2}),t_k}, \eta_{(z_{j_1},z_{j_2}),t_l}\right)\right\}, \end{aligned} \tag{4.10}$$

where $\log g_{\boldsymbol{\theta}}$ is the log-density computed in (4.5), which depends on the function $\delta = \delta_{\boldsymbol{\theta}}$. The quantities $w_{i_1,j_1}^{(M)} \ge 0$ and $w_{i_2,j_2}^{(M)} \ge 0$ denote spatial weights that depend on the first and the second components of the spatial pairs $((z_{i_1},z_{i_2}),(z_{j_1},z_{j_2}))$, respectively, whereas the temporal weights $w_{k,l}^{(T)} \ge 0$ depend on the pairs of time points (t_k,t_l). It seems natural to only give weight to space-time pairs whose space-time lag $(|z_{i_1} - z_{j_1}|, |z_{i_2} -$

$z_{j_2}|, |t_k - t_l|)$ is componentwise smaller than a certain chosen maximal space-time lag $(r_1, r_2, p) \in \mathbb{N}^3$, because observations that lie sufficiently far away in space and time do not have a noteworthy influence on each other. So the weight of observations where at least one of the lags exceeds its corresponding maximal lag is set equal to 0, leading to the following choice of weights:

$$w_{i_1,j_1}^{(M)} = \mathbf{1}_{\{|z_{i_1} - z_{j_1}| \leq r_1\}}, \; w_{i_2,j_2}^{(M)} = \mathbf{1}_{\{|z_{i_2} - z_{j_2}| \leq r_2\}}$$

and

$$w_{k,l}^{(T)} = \mathbf{1}_{\{|t_k - t_l| \leq p\}}.$$

Plugging those weights into (4.10) yields

$$PL^{(M,T)}(\boldsymbol{\theta}) = \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{\substack{j1=i1 \\ |z_{i_1} - z_{j_1}| \leq r_1}}^{M} \sum_{\substack{j_2=\mathbf{1}_{\{j_1=i_1\}} i_2+1 \\ |z_{i_2} - z_{j_2}| \leq r_2}}^{M} \sum_{k=1}^{T-p} \sum_{l=k+1}^{\min\{k+p,T\}} \log \left\{ g_{\boldsymbol{\theta}} \left(\eta_{(z_{i_1}, z_{i_2}),t_k}, \eta_{(z_{j_1}, z_{j_2}),t_l} \right) \right\}. \tag{4.11}$$

for $\boldsymbol{\theta} \in \Theta$. The pairwise likelihood estimate is then given by

$$\hat{\boldsymbol{\theta}} = (\hat{C}_1, \hat{\alpha}_1, \hat{C}_2, \hat{\alpha}_2, \hat{C}_3, \hat{\alpha}_3) = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} \, PL^{(M,T)}(\boldsymbol{\theta}). \tag{4.12}$$

In the following subsections, we show consistency and asymptotic normality of the pairwise likelihood estimates. However, we assume a regular grid, that is $(z_{i_1}, z_{i_2}) = (i_1, i_2)$ for all pairs (i_1, i_2) in $\{1, \ldots, M\} \times \{1, \ldots, M\}$. On top of that, we assume equidistant time points with distance 1; without loss of generality we set $t_k = k$ for $k = 0, \ldots, T$. In this case one can rewrite the pairwise likelihood function in a similar way as done in Davis et al. [2012b], Section 3.2: In their isotropic setting, they use the design mask (Nott and Rydén [1999])

$$\mathcal{H}_r = (\mathbb{N}^2 \cap B((0,0), r) \backslash \{(0,0)\},$$

where r is a positive integer and

$$B((0,0), r) = \left\{ (h_1, h_2) \in \mathbb{R}^2 : \|(h_1, h_2)\| \leq r \right\}$$

is the positive part of the ball in $\mathbb{R}^2$ with radius r around the origin. However, in the anisotropic setting we are dealing with here we have to focus on the absolute values of the single components of the spatial lags rather than on the norm of the whole vector. We adapt the design mask and come up with a design mask given by

$$\mathcal{H}_{r_1,r_2} = \left\{ (h_1, h_2) \in \mathbb{N}^2 \backslash \{(0,0)\} : h_1 \leq r_1, \; h_2 \leq r_2 \right\} \tag{4.13}$$

For example, for $r_1 = 3$ and $r_2 = 2$ we have

$$\mathcal{H}_{r_1,r_2} = \{(0,1), (0,2), (1,0), (1,1), (1,2), (2,0), (2,1), (2,2), (3,0), (3,1), (3,2)\}.$$

Using this design mask and the assumption of a regular grid and equidistant time points, one can rewrite the pairwise log-likelihood function in (4.11) in turn as

$$\begin{aligned} PL^{(M,T)}(\boldsymbol{\theta}) &= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{t=1}^{T} \sum_{\substack{(h_1,h_2)\in\mathcal{H}_{r_1,r_2}\\(i_1+h_1,i_2+h_2)\le(M,M)}} \sum_{\substack{u=1\\t+u\le T}}^{p} \log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(i_1,i_2),t},\eta_{(i_1+h_1,i_2+h_2),t+u}\right)\right\} \\ &= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{t=1}^{T} q_{\boldsymbol{\theta}}(i_1,i_2,t;r_1,r_2,p) - \mathcal{R}^{(M,T)}(\boldsymbol{\theta}), \qquad \boldsymbol{\theta}\in\Theta, \end{aligned} \tag{4.14}$$

where

$$q_{\boldsymbol{\theta}}(i_1,i_2,t;r_1,r_2,p) = \sum_{(h_1,h_2)\in\mathcal{H}_{r_1,r_2}} \sum_{u=1}^{p} \log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(i_1,i_2),t},\eta_{(i_1+h_1,i_2+h_2),t+u}\right)\right\}$$

and

$$\mathcal{R}^{(M,T)}(\boldsymbol{\theta}) = \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{t=1}^{T} \sum_{\substack{(h_1,h_2)\in\mathcal{H}_{r_1,r_2}\\(i_1,i_2)+(h_1,h_2)>(M,M)}} \sum_{\substack{u=1\\t+u>T}}^{p} \log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(i_1,i_2),t},\eta_{(i_1+h_1,i_2+h_2),t+u}\right)\right\}.$$

In order to formulate the pairwise log-likelihood for a general number of dimensions d, we slightly need to extend the design mask again. For a vector $\boldsymbol{r}=(r_1,\dots,r_d)\in\mathbb{N}^d$ set

$$\mathcal{H}_{\boldsymbol{r}} = \left\{(h_1,\dots,h_d)\in\mathbb{N}^d\backslash\{\mathbf{0}\} : h_1\le r_1,\dots,h_d\le r_d\right\}. \tag{4.15}$$

The parameter vector $\boldsymbol{\theta}=(C_1,\dots,C_d,C_{d+1},\alpha_1,\dots,\alpha_d,\alpha_{d+1})$ consists of the parameters of the function

$$\delta(h_1,\dots,h_d,u) = C_1|h_1|^{\alpha_1}+\cdots C_d|h_d|^{\alpha_d}+C_{d+1}|u|^{\alpha_{d+1}}.$$

The parameter space is

$$\Theta = \{\boldsymbol{\theta}=(C_m,\alpha_m,\ m=1,\dots,d+1) : C_m\in(0,\infty),\ \alpha_m\in(0,2],\ m=1,\dots,d+1\}.$$

The pairwise log-likelihood function turns out to be

$$\begin{aligned} &PL^{(M,T)}(\boldsymbol{\theta}) \\ &= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\cdots\sum_{i_d=1}^{M}\sum_{t=1}^{T} \sum_{\substack{(h_1,\dots,h_d)\in\mathcal{H}_{\boldsymbol{r}}\\(i_1+h_1,\dots,i_d+h_d)\le(M,\dots,M)}} \sum_{\substack{u=1\\t+u\le T}}^{p} \log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(i_1,\dots,i_d),t},\eta_{(i_1+h_1,\dots,i_d+h_d),t+u}\right)\right\} \\ &= \sum_{i_1=1}^{M}\cdots\sum_{i_d=1}^{M}\sum_{t=1}^{T} q_{\boldsymbol{\theta}}(i_1,\dots,i_d,t;\boldsymbol{r},p) - \mathcal{R}^{(M,T)}(\boldsymbol{\theta}), \qquad \boldsymbol{\theta}\in\Theta, \end{aligned} \tag{4.16}$$

where, similarly as in the case $d=2$, $q_{\boldsymbol{\theta}}(\cdot)$ and $\mathcal{R}^{(M,T)}$ are given by

$$q_{\boldsymbol{\theta}}(i_1,\dots,i_d,t;\boldsymbol{r},p) = \sum_{(h_1,\dots,h_d)\in\mathcal{H}_{\boldsymbol{r}}} \sum_{u=1}^{p} \log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(i_1,\dots,i_d),t},\eta_{(i_1+h_1,\dots,i_d+h_d),t+u}\right)\right\} \tag{4.17}$$

and

$$\mathcal{R}^{(M,T)}(\boldsymbol{\theta}) = \sum_{i_1=1}^{M} \cdots \sum_{i_d=1}^{M} \sum_{t=1}^{T} \sum_{\substack{(h_1,\ldots,h_d)\in\mathcal{H}_{\boldsymbol{r}} \\ (i_1+h_1,\ldots,i_d+h_d)>(M,\ldots,M)}} \sum_{\substack{u=1 \\ t+u>T}}^{p} \log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(i_1,\ldots,i_d),t}, \eta_{(i_1+h_1,\ldots,i_d+h_d),t+u}\right)\right\}. \tag{4.18}$$

4.2.2 Strong Consistency of the Pairwise Likelihood Estimates

There has been done some recent work in the field of composite likelihood methods and especially pairwise likelihood estimation. For instance, Davis and Yau [2011] show strong consistency of composite likelihood estimates for univariate time series. Padoan, Ribatet and Sisson [2010] show consistency and asymptotic normality of pairwise likelihood estimates for max-stable random fields with independent replications in time. Davis et al. [2012b] get rid of the assumption of independence in time and prove strong consistency of pairwise likelihood estimates of the Brown-Resnick process if the space-time domain increases jointly. However, they consider isotropic correlation functions of the underlying Gaussian process. In this subsection, we closely follow their proofs and show that the properties of strong consistency and asymptotic normality still hold if anisotropic correlation functions are admitted. A central assumption is that we suppose that the Brown-Resnick process is *mixing* in time and space.

Definition 4.2 (Mixing processes, cf. Def. 4.1 in Davis et al. [2012b])**.** *A strictly stationary space-time process* $\boldsymbol{\eta} := \left\{\eta_{\boldsymbol{s},t}, \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\right\}$ *is called* mixing *if for all sets* A *and* $B \in \sigma(\boldsymbol{\eta})$ *the following is satisfied:*

$$\lim_{n\to\infty} \mathbb{P}\left(A \cap \tau_{s_{n,1},\ldots,s_{n,d},t_n}(B)\right) = \mathbb{P}(A)\mathbb{P}(B) \tag{4.19}$$

for all sequences $\{(s_{n,1},\ldots,s_{n,d},t_n),\ n\in\mathbb{N}\}$ *with* $\max\{|s_{n,1}|,\ldots,|s_{n,d}|,|t_n|\} \to \infty$ *as* $n\to\infty$*. Here,* $\sigma(\boldsymbol{\eta})$ *is the* σ*-algebra generated by the process* $\boldsymbol{\eta}$ *and* $\tau_{\boldsymbol{h},u}(\cdot)$*,* $\boldsymbol{h} \in \mathbb{R}^d$*,* $u \in \mathbb{R}$*, denotes the multiparameter shift operator.*

In Section 4.1, Davis et al. [2012b] summarize and extend theoretical statements of Wang, Roy and Stoev [2009] , Krengel [1985], Stoev and Taqqu [2005] and Wang and Stoev [2010], which are based on extremal integral representations and ergodic properties of max-stable processes and the Brown-Resnick process in particular, and end up with the following sufficient condition for a Brown-Resnick processes to be mixing in space and time:

Proposition 4.3 (cf. Prop. 4.3 in Davis et al. [2012b])**.** *If the correlation function* γ *of the underlying Gaussian space-time process in the construction of the Brown-Resnick process* $\boldsymbol{\eta} = \left\{\eta_{\boldsymbol{s},t}, \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\right\}$ *(Theorem 3.5) satisfies Assumption 3.1, then* $\boldsymbol{\eta}$ *is mixing in space and time.*

Furthermore, the strong law of large numbers holds: For every measurable function $f : \mathbb{R} \to \mathbb{R}$ *satisfying* $\mathbb{E}\left[\left|f\left(\eta_{(1,\dots,1),1}\right)\right|\right] < \infty$ *we have*

$$\frac{1}{M^d T}\sum_{i_1=1}^{M}\cdots\sum_{i_d=1}^{M}\sum_{t=1}^{T} f\left(\eta_{(i_1,\dots,i_d),t}\right) \to \mathbb{E}\left[f\left(\eta_{(1,\dots,1),1}\right)\right] \text{ almost surely as } M,T\to\infty. \tag{4.20}$$

We can now formulate and prove the analogue of Theorem 4.4 in Davis et al. [2012b]:

Theorem 4.4 (Strong Consistency for large M and T)**.** *Let*

$$\boldsymbol{\eta} = \left\{\eta_{\boldsymbol{s},t} : \boldsymbol{s} \in \{1,\dots,M\}^d, t \in \{1,\dots,T\}\right\}$$

be a Brown-Resnick process constructed as in Theorem 3.5 with realizations on a regular d-dimensional grid. Let the correlation function γ *of the underlying Gaussian space-time process satisfy Assumption 4.1 with parameter vector*

$$\boldsymbol{\theta} = (C_1,\dots,C_{d+1},\alpha_1,\dots,\alpha_{d+1}).$$

Furthermore, assume that the true parameter vector

$$\boldsymbol{\theta}^\star = (C_1^\star,\dots,C_{d+1}^\star,\alpha_1^\star,\dots,\alpha_{d+1}^\star)$$

is part of a compact set

$$\Theta^\star \subset \{(C_1,\dots,C_{d+1},\alpha_1,\dots,\alpha_{d+1}) : C_m \in (c,\infty), \alpha_m \in (0,2],\ m=1,\dots,d+1\}\setminus\{\mathbf{0}\} \tag{4.21}$$

for some $c > 0$*. On top of that, suppose that the following identifiability condition holds:*

$$\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}} \text{ if and only if } g_{\boldsymbol{\theta}}\left(\eta_{\boldsymbol{s}_1,t_1},\eta_{\boldsymbol{s}_2,t_2}\right) = g_{\tilde{\boldsymbol{\theta}}}\left(\eta_{\boldsymbol{s}_1,t_1},\eta_{\boldsymbol{s}_2,t_2}\right) \tag{4.22}$$

for all $(\boldsymbol{s}_1,t_1)$ *and* $(\boldsymbol{s}_2,t_2)$*. Then, the pairwise likelihood estimate*

$$\hat{\boldsymbol{\theta}}^{(M,T)} = \underset{\boldsymbol{\theta}\in\Theta^\star}{\operatorname{argmax}}\, PL^{(M,T)}(\boldsymbol{\theta})$$

for realizations of $\boldsymbol{\eta}$ *is strongly consistent, that is,*

$$\hat{\boldsymbol{\theta}}^{(M,T)} \to \boldsymbol{\theta}^\star \text{ almost surely as } M,T\to\infty.$$

Proof. We follow the steps of Davis et al. [2012b]. The proof uses the method of Wald [1946]. The aim is to show that for some chosen space-time lag $(\boldsymbol{r},p)$ with $\boldsymbol{r} = (r_1,\dots,r_d) \in \mathbb{N}^d$ and $p \in \mathbb{N}$,

$$\begin{aligned}&\frac{1}{M^d T} PL^{(M,T)}(\boldsymbol{\theta})\\ &= \frac{1}{M^d T}\left(\sum_{i_1=1}^{M}\cdots\sum_{i_d=1}^{M}\sum_{t=1}^{T} q_{\boldsymbol{\theta}}(i_1,\dots,i_d,t;\boldsymbol{r},p) - \mathcal{R}^{(M,T)}(\boldsymbol{\theta})\right) \to \mathbb{E}[(q_{\boldsymbol{\theta}}(1,\dots,1,1;\boldsymbol{r},p)]\end{aligned}$$

almost surely as $M,T\to\infty$. This is done by verifying the following three statements:

(A)

$$\frac{1}{M^d T}\sum_{i_1=1}^{M}\cdots\sum_{i_d=1}^{M}\sum_{t=1}^{T} q_{\boldsymbol{\theta}}(i_1,\ldots,i_d,t;\boldsymbol{r},p) \to \mathbb{E}[(q_{\boldsymbol{\theta}}(1,\ldots,1,1;\boldsymbol{r},p)]$$

almost surely as $M, T \to \infty$, uniformly on the compact parameter space $\Theta^\star$.

(B)

$$\frac{1}{M^d T}\mathcal{R}^{(M,T)}(\boldsymbol{\theta}) \to 0$$

almost surely as $M, T \to \infty$.

(C) The limit function $\mathbb{E}(q_{\boldsymbol{\theta}}(1,\ldots,1,1;\boldsymbol{r},p)]$ is uniquely maximized at the true parameter vector $\boldsymbol{\theta}^\star \in \Theta^\star$.

Proof of (A):
The convergence holds because $q_{\boldsymbol{\theta}}(\cdot)$ is a measurable function of lagged versions of $\eta_{\boldsymbol{s},t}$ for $\boldsymbol{s} \in \{1,\ldots,M\}^d$, $t \in \{1,\ldots,T\}$. One can therefore directly apply Proposition 4.3. What remains to be shown is that the convergence is uniform on the compact parameter space $\Theta^\star$. To this end, for $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \{1,\ldots,M\}^d$ and $t_1, t_2 \in \{1,\ldots,T\}$, recall the definition of the log-density $\log g_{\boldsymbol{\theta}}$ (4.5) of the joint distribution of $\eta_{\boldsymbol{s}_1,t_1}$ and $\eta_{\boldsymbol{s}_2,t_2}$. For $y_1, y_2 > 0$ we have

$$\log g_{\boldsymbol{\theta}}(y_1,y_2) = -V(y_1,y_2) + \log\left\{V_1(y_1,y_2)V_2(y_1,y_2) - V_{12}(y_1,y_2)\right\}$$

with V, V_1, V_2, V_{12} defined in (4.6)-(4.9). It follows (with $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$ and $u = t_1 - t_2$)

$$\begin{aligned}
|\log g_{\boldsymbol{\theta}}(y_1,y_2)| &\le |-V(y_1,y_2)| + |\log\left\{V_1(y_1,y_2)V_2(y_1,y_2) - V_{12}(y_1,y_2)\right\}| \\
&\le \frac{1}{y_1}\left|\Phi\left(\frac{\log\frac{y_2}{y_1}}{2\sqrt{\delta(\boldsymbol{h},u)}}\right) + \sqrt{\delta(\boldsymbol{h},u)}\right| + \frac{1}{y_2}\left|\Phi\left(\frac{\log\frac{y_1}{y_2}}{2\sqrt{\delta(\boldsymbol{h},u)}}\right) + \sqrt{\delta(\boldsymbol{h},u)}\right| \\
&\quad + |V_1(y_1,y_2)V_2(y_1,y_2) - V_{12}(y_1,y_2)| \qquad (4.23)
\end{aligned}$$

By standard multiplication and using that $\Phi(x) \le 1$ and $\varphi(x) < 1$ for all $x \in \mathbb{R}$, we

continue (4.23) and find

$$
\begin{aligned}
&|\log g_{\boldsymbol{\theta}}(y_1, y_2)| \\
&\leq \frac{1}{y_1} + \frac{1}{y_2} + \frac{1}{y_1^2 y_2^2} + \frac{1}{2\sqrt{\delta(\boldsymbol{h}, u)}} \left(\frac{2}{y_1^2 y_2^2} + \frac{1}{y_1^3 y_2} + \frac{1}{y_1 y_2^3} + \frac{1}{y_1^2 y_2} + \frac{1}{y_1 y_2^2} \right) \\
&+ \frac{1}{4\delta(\boldsymbol{h}, u)} \left(\frac{2}{y_1^2 y_2^2} + \frac{1}{y_1^3 y_2} + \frac{1}{y_1 y_2^3} \right) \\
&+ \left| \left(\frac{\log \frac{y_2}{y_1}}{8(\delta(\boldsymbol{h}, u))^{\frac{3}{2}}} + \frac{1}{4\sqrt{\delta(\boldsymbol{h}, u)}} \right) \frac{1}{y_1^2 y_2} + \left(\frac{\log \frac{y_1}{y_2}}{8(\delta(\boldsymbol{h}, u))^{\frac{3}{2}}} + \frac{1}{4\sqrt{\delta(\boldsymbol{h}, u)}} \right) \frac{1}{y_1 y_2^2} \right| \\
&\leq \frac{1}{y_1} + \frac{1}{y_2} + \frac{1}{y_1^2 y_2^2} + \frac{1}{2\sqrt{\delta(\boldsymbol{h}, u)}} \left(\frac{2}{y_1^2 y_2^2} + \frac{1}{y_1^3 y_2} + \frac{1}{y_1 y_2^3} + \frac{1}{y_1^2 y_2} + \frac{1}{y_1 y_2^2} \right) \\
&+ \frac{1}{4\delta(\boldsymbol{h}, u)} \left(\frac{2}{y_1^2 y_2^2} + \frac{1}{y_1^3 y_2} + \frac{1}{y_1 y_2^3} \right) \\
&+ \frac{1}{8(\delta(\boldsymbol{h}, u))^{\frac{3}{2}}} \left(\frac{1}{y_1^3 + \frac{1}{y_2^3}} \right) + \frac{1}{4\sqrt{\delta(\boldsymbol{h}, u)}} \left(\frac{1}{y_1^2 y_2} + \frac{1}{y_1 y_2^2} \right). \qquad (4.24)
\end{aligned}
$$

The fact that for fixed $\boldsymbol{s} \in \{1, \ldots, M\}^d$ and $t \in \{1, \ldots, T\}$, the random variable $\eta_{\boldsymbol{s},t}$ is standard Fréchet ensures that $\frac{1}{\eta_{\boldsymbol{s},t}}$ follows an exponential(1) distribution which has finite moments. We can therefore apply Hölder's inequality and obtain

$$
\mathbb{E}\left[\left|\log g_{\boldsymbol{\theta}}\left(\eta_{\boldsymbol{s}_1,t_1}, \eta_{\boldsymbol{s}_2,t_2}\right)\right|\right] \leq K_1 + \frac{K_2}{2\sqrt{\delta(\boldsymbol{h}, u)}} + \frac{K_3}{4\delta(\boldsymbol{h}, u)} + \frac{K_4}{8(\delta(\boldsymbol{h}, u))^{\frac{3}{2}}}, \qquad (4.25)
$$

where K_1, K_2, K_3 and K_4 are positive finite constants. Using that $\Theta^\star$ is assumed to be compact, we can bound δ away from 0:

$$
\begin{aligned}
\delta(\boldsymbol{h}, u) &\geq \min\{C_1, \ldots, C_{d+1}\} \left(|h_1|^{\alpha_1} + \cdots + |h_d|^{\alpha_d} + |u|^{\alpha_{d+1}}\right) \\
&> c \left(|h_1|^{\alpha_1} + \cdots + |h_d|^{\alpha_d} + |u|^{\alpha_{d+1}}\right) \\
&> \tilde{c}, \qquad (4.26)
\end{aligned}
$$

where $\tilde{c} > 0$ is independent of the parameters. From (4.25) we therefore conclude

$$
\mathbb{E}\left[\left|\log g_{\boldsymbol{\theta}}\left(\eta_{\boldsymbol{s}_1,t_1}, \eta_{\boldsymbol{s}_2,t_2}\right)\right|\right] < K_1 + \frac{K_2}{2\tilde{c}} + \frac{K_3}{4\tilde{c}} + \frac{K_4}{8\tilde{c}^{\frac{3}{2}}} =: K_5 < \infty. \qquad (4.27)
$$

As K_5 is independent of the parameter vector $\boldsymbol{\theta}$ and $(\boldsymbol{s}_1, t_1)$ and $(\boldsymbol{s}_2, t_2)$ were chosen arbitrarily, (4.26) and (4.27) yield

$$
\mathbb{E}\left[\sup_{\boldsymbol{\theta} \in \Theta^\star} \left|\log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(1,\ldots,1),1}, \eta_{(1,\ldots,1)+(h_1,\ldots,h_d),1+u}\right)\right\}\right|\right] < \infty \text{ for all } (h_1, \ldots, h_d) \in \mathbb{R}^d
$$

and therefore, by the definition of $q_{\boldsymbol{\theta}}$,

$$
\mathbb{E}\left[\sup_{\boldsymbol{\theta} \in \Theta^\star} |q_{\boldsymbol{\theta}}(1, \ldots, 1, 1; \boldsymbol{r}, p)|\right] < \infty.
$$

By Theorem 2.7 in Straumann [2004] we conclude that the convergence in (A) is uniform, so (A) is proved.

Proof of (B):
Using the definition 4.18 of the boundary term $\mathcal{R}^{(M,T)}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta^\star$, we find

$$
\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{M^d T}\mathcal{R}^{(M,T)}(\boldsymbol{\theta})\right|\right] \\
&= \\
&\frac{1}{M^d T}\mathbb{E}\left[\left|\sum_{i_1=1}^{M}\cdots\sum_{i_d=1}^{M}\sum_{t=1}^{T}\sum_{\substack{(h_1,\ldots,h_d)\in\mathcal{H}_{\boldsymbol{r}}\\(i_1+h_1,\ldots,i_d+h_d)>(M,\ldots,M)}}\sum_{\substack{u=1\\t+u>T}}^{p}\log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(i_1,\ldots,i_d),t},\eta_{(i_1+h_1,\ldots,i_d+h_d),t+u}\right)\right\}\right|\right] \\
&\leq \\
&\frac{1}{M^d T}\mathbb{E}\left[\sum_{i_1=1}^{M}\cdots\sum_{i_d=1}^{M}\sum_{t=1}^{T}\sum_{\substack{(h_1,\ldots,h_d)\in\mathcal{H}_{\boldsymbol{r}}\\(i_1+h_1,\ldots,i_d+h_d)>(M,\ldots,M)}}\sum_{\substack{u=1\\t+u>T}}^{p}\left|\log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(i_1,\ldots,i_d),t},\eta_{(i_1+h_1,\ldots,i_d+h_d),t+u}\right)\right\}\right|\right] \\
&= \\
&\frac{1}{M^d T}\sum_{i_1=1}^{M}\cdots\sum_{i_d=1}^{M}\sum_{t=1}^{T}\sum_{\substack{(h_1,\ldots,h_d)\in\mathcal{H}_{\boldsymbol{r}}\\(i_1+h_1,\ldots,i_d+h_d)>(M,\ldots,M)}}\sum_{\substack{u=1\\t+u>T}}^{p}\mathbb{E}\left[\left|\log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(i_1,\ldots,i_d),t},\eta_{(i_1+h_1,\ldots,i_d+h_d),t+u}\right)\right\}\right|\right] \\
&\leq \\
&\frac{1}{M^d T}\sum_{i_1=1}^{M}\cdots\sum_{i_d=1}^{M}\sum_{t=1}^{T}\sum_{\substack{(h_1,\ldots,h_d)\in\mathcal{H}_{\boldsymbol{r}}\\(i_1+h_1,\ldots,i_d+h_d)>(M,\ldots,M)}}\sum_{\substack{u=1\\t+u>T}}^{p}K_5, \qquad (4.28)
\end{aligned}
$$

where in the last step we used relation (4.27). The number of space-time points used in the boundary term $\mathcal{R}$ is of order M^{d-1}, independent of T. Therefore, there exists a positive constant K_6, independent of M and T, such that

$$
\sum_{i_1=1}^{M}\cdots\sum_{i_d=1}^{M}\sum_{t=1}^{T}\sum_{\substack{(h_1,\ldots,h_d)\in\mathcal{H}_{\boldsymbol{r}}\\(i_1+h_1,\ldots,i_d+h_d)>(M,\ldots,M)}}\sum_{\substack{u=1\\t+u>T}}^{p}K_5 \leq K_5 K_6 M^{d-1}.
$$

All in all, continuing (4.28),

$$
\mathbb{E}\left[\left|\frac{1}{M^d T}\mathcal{R}^{(M,T)}(\boldsymbol{\theta})\right|\right] \leq \frac{K_5 K_6}{MT} \to 0,
$$

as $M, T \to \infty$.
As a next step and for convenience, we rewrite the boundary term $\mathcal{R}^{(M,T)}(\boldsymbol{\theta})$. For $\boldsymbol{h} =$

$(h_1, \ldots, h_d) \in \mathbb{R}^d$ and $u \in \mathbb{R}$ define the set of boundary points $\mathcal{G}_{M,T}(\boldsymbol{h}, u)$ as

$$\begin{aligned}&\mathcal{G}_{M,T}(\boldsymbol{h}, u) := \\ &\left\{(i_1, \ldots, i_d) \in \{1, \ldots, M\}^d : (i_1, \ldots, i_d) + (h_1, \ldots, h_d) > (M, \ldots, M)\right\} \times \\ &\left\{t \in \{1, \ldots, T\} : t + u > T\right\}. \end{aligned} \tag{4.29}$$

Then $\mathcal{R}^{(M,T)}(\boldsymbol{\theta})$ is given by

$$\mathcal{R}^{(M,T)}(\boldsymbol{\theta}) = \sum_{\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}}} \sum_{u=1}^{p} \sum_{(i_1, \ldots, i_d, t) \in \mathcal{G}_{M,T}(\boldsymbol{h}, u)} \log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(i_1, \ldots, i_d), t}, \eta_{(i_1+h_1, \ldots, i_d+h_d), t+u}\right)\right\}. \tag{4.30}$$

Using Proposition 4.3 and (4.27) we have that, uniformly on $\Theta^\star$,

$$\begin{aligned}&\sum_{\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}}} \sum_{u=1}^{p} \frac{1}{|\mathcal{G}_{M,T}(\boldsymbol{h}, u)|} \sum_{(i_1, \ldots, i_d, t) \in \mathcal{G}_{M,T}(\boldsymbol{h}, u)} \log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(i_1, \ldots, i_d), t}, \eta_{(i_1+h_1, \ldots, i_d+h_d), t+u}\right)\right\} \\ &\to \mathbb{E}\left[\sum_{\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}}} \sum_{u=1}^{p} \sum_{(i_1, \ldots, i_d, t) \in \mathcal{G}_{M,T}(\boldsymbol{h}, u)} \log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(1, \ldots, 1), 1}, \eta_{(1+h_1, \ldots, 1+h_d), 1+u}\right)\right\}\right]\end{aligned}$$

almost surely as $M, T \to \infty$. So, using that $|\mathcal{G}_{M,T}(\boldsymbol{h}, u)| \leq K_6 M^{d-1}$ for all $\boldsymbol{h} \in \mathbb{R}^d$ and $u \in \mathbb{R}$, it follows that

$$\begin{aligned}&\frac{1}{M^d T} \mathcal{R}^{(M,T)}(\boldsymbol{\theta}) \\ &\leq \frac{K_6}{MT} \sum_{\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}}} \sum_{u=1}^{p} \frac{1}{|\mathcal{G}_{M,T}(\boldsymbol{h}, u)|} \sum_{(i_1, \ldots, i_d, t) \in \mathcal{G}_{M,T}(\boldsymbol{h}, u)} \log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(i_1, \ldots, i_d), t}, \eta_{(i_1+h_1, \ldots, i_d+h_d), t+u}\right)\right\} \\ &\to 0\end{aligned}$$

almost surely as $M, T \to \infty$, since

$$\mathbb{E}\left[\left|\log\left\{g_{\boldsymbol{\theta}}\left(\eta_{(i_1, \ldots, i_d), t}, \eta_{(i_1+h_1, \ldots, i_d+h_d), t+u}\right)\right\}\right|\right] < \infty.$$

This is what we want to show.

Proof of (C):
Let $\boldsymbol{\theta} \neq \boldsymbol{\theta}^\star$. For $i_1, \ldots, i_d \in \{1, \ldots, M\}$ and $t \in \{1, \ldots, T\}$, Jensen's inequality and the

identifiability condition (4.22) lead to

$$\begin{aligned}
&\mathbb{E}\left[\log\left\{\frac{g_{\boldsymbol{\theta}}\left(\eta_{(i_1,\dots,i_d),t},\eta_{(i_1+h_1,\dots,i_d+h_d),t+u}\right)}{g_{\boldsymbol{\theta}^\star}\left(\eta_{(i_1,\dots,i_d),t},\eta_{(i_1+h_1,\dots,i_d+h_d),t+u}\right)}\right\}\right]\\
&< \log\left\{\mathbb{E}\left[\frac{g_{\boldsymbol{\theta}}\left(\eta_{(i_1,\dots,i_d),t},\eta_{(i_1+h_1,\dots,i_d+h_d),t+u}\right)}{g_{\boldsymbol{\theta}^\star}\left(\eta_{(i_1,\dots,i_d),t},\eta_{(i_1+h_1,\dots,i_d+h_d),t+u}\right)}\right]\right\}\\
&= \log\left\{\int\limits_{(0,\infty)^2}\frac{g_{\boldsymbol{\theta}}(y_1,y_2)}{g_{\boldsymbol{\theta}^\star}(y_1,y_2)}g_{\boldsymbol{\theta}^\star}(y_1,y_2)\mathrm{d}(y_1,y_2)\right\}\\
&= \log\left\{\int\limits_{(0,\infty)^2} g_{\boldsymbol{\theta}}(y_1,y_2)\mathrm{d}(y_1,y_2)\right\}\\
&= 0,
\end{aligned}$$

and it directly follows from the definition (4.17) of $q_{\boldsymbol{\theta}}$ that

$$\mathbb{E}[q_{\boldsymbol{\theta}}(1,\dots,1,1;\boldsymbol{r},p)] < \mathbb{E}[q_{\boldsymbol{\theta}^\star}(1,\dots,1,1;\boldsymbol{r},p)].$$

□

As stated by Davis et al. [2012b], there are combinations of maximum space-time lags that lead to non-identifiable parameters, but the theorem can still be applied to the other parameters. Exemplified for the case $d = 2$, Table 4.1 shows which parameters are identifiable for different maximum space-time lags (r_1, r_2, p). Note that the identifiability condition (4.22) depends on the identifiability of the function $\delta(h_1, h_2, u) = C_1|h_1|^{\alpha_1} + C_2|h_2|^{\alpha_2} + C_3|h_3|^{\alpha_3}$.

4.2.3 Asymptotic Normality of the Pairwise Likelihood Estimates

In this section we prove asymptotic normality of the pairwise likelihood estimates defined in (4.12). As in the proof of Theorem 4.4, we follow the lines of Davis et al. [2012b], Section 5, adapting their work to our setting whenever needed. We start with some basic results casually needed throughout the rest of the section.

Lemma 4.5 (cf. Lemma 5.1 in Davis et al. [2012b]). *Let $\boldsymbol{\eta} := \{\eta_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\}$ be the Brown-Resnick process constructed in Theorem 3.5 where the underlying correlation function γ satisfies Assumption 3.1. Assume that all conditions of Theorem 4.4 are satisfied. Then for $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{R}^d$ and $t_1, t_2 \in [0,\infty)$, the following two results hold:*

(i) The gradient of the bivariate log-density satisfies

$$\mathbb{E}\left[\left|\nabla_{\boldsymbol{\theta}} \log g_\theta(\eta_{\boldsymbol{s}_1,t_1},\eta_{\boldsymbol{s}_2,t_2})\right|^3\right] < \infty, \quad \boldsymbol{\theta} \in \Theta^\star.$$

r_1	r_2	p	identifiable parameters
1	0	0	C_1
0	1	0	C_2
0	0	1	C_3
1	1	0	C_1, C_2
1	0	1	C_1, C_3
0	1	1	C_2, C_3
1	1	1	C_1, C_2, C_3
>1	0	0	C_1, α_1
0	>1	0	C_2, α_2
0	0	>1	C_3, α_3
>1	1	0	C_1, α_1, C_2
1	>1	0	C_1, C_2, α_2
>1	0	1	C_1, α_1, C_3
1	0	>1	C_1, C_3, α_3
0	>1	1	C_2, α_2, C_3
0	1	>1	C_2, C_3, α_3
>1	>1	0	C_1, α_1, C_2, α_2
>1	0	>1	C_1, α_1, C_3, α_3
0	>1	>1	C_2, α_2, C_3, α_3
>1	1	1	C_1, α_1, C_2, C_3
1	>1	1	C_1, C_2, α_2, C_3
1	1	>1	C_1, C_2, C_3, α_3
>1	>1	1	C_1, α_1, C_2, α_2, C_3
>1	1	>1	C_1, α_1, C_2, C_3, α_3
1	>1	>1	C_1, C_2, α_2, C_3, α_3
>1	>1	>1	C_1, α_1, C_2, α_2, C_3, α_3

Table 4.1: Identifiable parameters for different maximum space-time lags.

(ii) The Hessian matrix of the bivariate log-density satisfies

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\Theta^\star}\left|\nabla^2_{\boldsymbol{\theta}}\log g_\theta(\eta_{\boldsymbol{s}_1,t_1},\eta_{\boldsymbol{s}_2,t_2})\right|\right]<\infty.$$

The absolute values are taken componentwise.

Proof. Assume identifiability of all parameters C_m, α_m, $m=1,\dots,d+1$. For $y_1,y_2\in(0,\infty)$ and for $\boldsymbol{h}=(h_1,\dots,h_d)\in\mathbb{R}^d$, $u\in\mathbb{R}$ such that $0<\min\{|h_1|,\dots,|h_d|,|u|\}$, $\max\{|h_1|,\dots,|h_d|,|u|\}<\infty$ we have (cf. (4.5)-(4.9)),

$$\nabla_{\boldsymbol{\theta}}\log g_{\boldsymbol{\theta}}(y_1,y_2)=\frac{\partial\log g_{\boldsymbol{\theta}}(y_1,y_2)}{\partial\delta(\boldsymbol{h},u)}\nabla_{\boldsymbol{\theta}}\delta(\boldsymbol{h},u)$$

Furthermore,

$$\frac{\partial\delta(\boldsymbol{h},u)}{\partial C_m} = |h_m|^{\alpha_m},\ \frac{\partial\delta(\boldsymbol{h},u)}{\partial\alpha_m} = C_m|h_m|^{\alpha_m}\log|h_m|,\ m=1,\dots d,$$

and

$$\frac{\partial\delta(\boldsymbol{h},u)}{\partial C_{d+1}} = |u|^{\alpha_{d+1}},\ \frac{\partial\delta(\boldsymbol{h},u)}{\partial\alpha_{d+1}} = C_{d+1}|u|^{\alpha_{d+1}}\log|u|.$$

Using assumption (4.21) we can bound those first partial derivatives as well as the second order partial derivatives from above and below. So it remains to be shown that for $\boldsymbol{s}_1,\boldsymbol{s}_2 \in \mathbb{R}^d$ and $t_1,t_2\in[0,\infty)$,

$$\mathbb{E}_{\boldsymbol{\theta}^\star}\left[\left|\frac{\partial\log\{g_{\boldsymbol{\theta}}(\eta_{\boldsymbol{s}_1,t_1},\eta_{\boldsymbol{s}_2,t_2})\}}{\partial\delta(\boldsymbol{h},u)}\right|^3\right] < \infty$$

and

$$\mathbb{E}_{\boldsymbol{\theta}^\star}\left[\sup_{\boldsymbol{\theta}\in\Theta^\star}\left|\frac{\partial^2\log\{g_{\boldsymbol{\theta}}(\eta_{\boldsymbol{s}_1,t_1},\eta_{\boldsymbol{s}_2,t_2})\}}{\partial\delta(\boldsymbol{h},u)}\right|\right] < \infty,$$

where the function $\delta(\boldsymbol{h},u)=\delta$ can be treated as a constant since one can bound it away from 0 by (4.21). So, from here on, we refer to the proof of Steinkohl [2012], Lemma 4.6, since at this place, the only difference consists of the differences between the respective functions $\delta(\boldsymbol{h},u)$. □

As a next step, we need the definition of the α-mixing property. We follow Davis et al. [2012b], Section 5.2, who use and adapt results of Huser and Davison [2012] and Bolthausen [1982] to their space-time setting.

Definition 4.6 (Mixing Coefficients and α-mixing)**.** *Let $\{\eta_{\boldsymbol{s},t} : \boldsymbol{s}\in\mathbb{Z}^d, t\in\mathbb{N}\}$ be a space-time process. For $\boldsymbol{s}_1=(i_1^{(1)},\dots,i_d^{(1)})$ and $\boldsymbol{s}_2=(i_1^{(2)},\dots,i_d^{(2)})$ set*

$$d((\boldsymbol{s}_1,t_1),(\boldsymbol{s}_2,t_2)) := \max\{\max_{1\le m\le d}|i_m^{(1)}-i_m^{(2)}|,|t_1-t_2|\}.$$

Further, for Λ_1 and $\Lambda_2\subset\mathbb{Z}^d\times\mathbb{N}$ let

$$d(\Lambda_1,\Lambda_2) := \inf\{d((\boldsymbol{s}_1,t_1),(\boldsymbol{s}_2,t_2)),(\boldsymbol{s}_1,t_1)\in\Lambda_1,(\boldsymbol{s}_2,t_2)\in\Lambda_2\}.$$

(i) For $k,\ell,n\ge 0$ the mixing coefficients *are defined as*

$$\alpha_{k,\ell}(n) := \sup\{|\mathbb{P}(A_1\cap A_2)-\mathbb{P}(A_1)\mathbb{P}(A_2)| : A_1\in\mathcal{F}_{\Lambda_1}, A_2\in\mathcal{F}_{\Lambda_2}, |\Lambda_1|\le k, |\Lambda_2|\le\ell, d(\Lambda_1,\Lambda_2)\ge n\}, \tag{4.31}$$

where $\mathcal{F}_{\Lambda_i}=\sigma(\eta_{\boldsymbol{s},t}:(\boldsymbol{s},t)\in\Lambda_i)$ for $i=1,2$.

(ii) The process $\{\eta_{\boldsymbol{s},t}:\boldsymbol{s}\in\mathbb{Z}^d,t\in\mathbb{N}\}$ is called α-mixing *if for all $k,\ell>0$,*

$$\alpha_{k,\ell}(n)\to 0,\quad n\to\infty.$$

In the following, we show and use that $\{\eta_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{Z}^d, t \in \mathbb{N}\}$ is α-mixing. Then by definition (4.17) of $q_{\boldsymbol{\theta}}$, we have that the space-time process $\{\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}}(i_1, \ldots, i_d, t; \boldsymbol{r}, p) : i_m \in \mathbb{Z}, m = 1, \ldots, d, t \in \mathbb{N}\}$, where $\boldsymbol{r}$ and p are the chosen space and time lags, respectively, is α-mixing for all $\boldsymbol{\theta} \in \Theta^\star$. Based on the central limit theorem of Bolthausen [1982], we can formulate the following proposition.

Proposition 4.7. *Let $\boldsymbol{\eta} = \{\eta_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{R}^d, t \in [0, \infty)\}$ be the Brown-Resnick process constructed in Theorem 3.5. Assume the following three conditions to hold:*

(1) The process $\{\eta_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{Z}^d, t \in \mathbb{N}\}$ is α-mixing.

(2) The α-mixing coefficients in (4.31) satisfy:

- $\sum_{n=1}^{\infty} n^d \alpha_{k,\ell}(n) < \infty$ *for* $k + l \leq 4(|\mathcal{H}_{\boldsymbol{r}}| + 1)(p + 1)$
- $\alpha_{(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1),\infty}(n) = o(n^{-(d+1)})$.

(3) There exists a positive β such that

- $\mathbb{E}\left[|\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(i_1, \ldots, i_d, t; \boldsymbol{r}, p)|^{2+\beta}\right] < \infty$ *for* $i_1, \ldots, i_d \in \mathbb{Z}$, $t \in \mathbb{N}$, *and*
- $\sum_{n=1}^{\infty} n^d \alpha_{(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1),(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1)}(n)^{\frac{\beta}{2+\beta}} < \infty$.

Then,

$$\frac{1}{M^{\frac{d}{2}}\sqrt{T}} \sum_{i_1=1}^{M} \cdots \sum_{i_d=1}^{M} \nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(i_1, \ldots, i_d, t; \boldsymbol{r}, p) \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{0}, \Sigma), \quad M, T \to \infty, \tag{4.32}$$

where

$$\Sigma = \sum_{i_1=-\infty}^{\infty} \cdots \sum_{i_d=-\infty}^{\infty} \sum_{t=1}^{\infty} \mathrm{Cov}\left[\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(1, \ldots, 1, 1; \boldsymbol{r}, p), \nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(i_1, \ldots, i_d, t; \boldsymbol{r}, p)\right]. \tag{4.33}$$

In what follows, we show that the conditions of Proposition 4.7 hold. To this end we need the next Lemma, which is based on Corollary 2.2 of Dombry and Eyi-Minko [2012].

Lemma 4.8 (cf. Davis et al. [2012b], Lemma 5.4)**.** *The α-mixing coefficients in (4.31) of a stationary max-stable space-time process $\{\eta_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{Z}^d, t \in \mathbb{N}\}$ with tail dependence coefficient $\mathcal{X}(\boldsymbol{h}, u)$ (see (3.14)) satisfy*

- $\alpha_{k,\ell}(n) \leq k\ell \sup_{\max\{\sqrt{d}|h_1|, \ldots, \sqrt{d}|h_d|, |u|\} \geq n} \mathcal{X}(\boldsymbol{h}, u)$ *and*
- $\alpha_{k,\infty}(n) \leq k \sum_{\max\{\sqrt{d}|h_1|, \ldots, \sqrt{d}|h_d|, |u|\} \geq n} \mathcal{X}(\boldsymbol{h}, u)$

for positive k, ℓ and n.

Proof. The statements are based on Dombry and Eyi-Minko [2012], Corollary 2.2 and Davis et al. [2012b], Lemma 5.4, ensuring that, under the same conditions,

- $\alpha_{k,\ell}(n) \leq k\ell \sup_{\max\{\|\boldsymbol{h}\|,|u|\}\geq n} \mathcal{X}(\boldsymbol{h},u)$ and
- $\alpha_{k,\infty}(n) \leq k \sum_{\max\{\|\boldsymbol{h}\|,|u|\}\geq n} \mathcal{X}(\boldsymbol{h},u)$.

Note that for $\boldsymbol{h} = (h_1,\ldots,h_d) \in \mathbb{R}^d$ we have

$$\|\boldsymbol{h}\| = \sqrt{|h_1|^2+\cdots+|h_d|^2} \leq \sqrt{d\max\{|h_1|^2,\ldots,|h_d|^2\}} = \sqrt{d}\max\{|h_1|,\ldots,|h_d|\}$$

Therefore, for $n \in \mathbb{N}$, presuming $\max\{\|\boldsymbol{h}\|,|u|\} \geq n$ results in $\max\{\sqrt{d}|h_1|,\ldots,\sqrt{d}|h_d|,|u|\} \geq n$ and finally we have

- $\alpha_{k,\ell}(n) \leq k\ell \sup_{\max\{\|\boldsymbol{h}\|,|u|\}\geq n} \mathcal{X}(\boldsymbol{h},u) \leq k\ell \sup_{\max\{\sqrt{d}|h_1|,\ldots,\sqrt{d}|h_d|,|u|\}\geq n} \mathcal{X}(\boldsymbol{h},u)$ and
- $\alpha_{k,\infty}(n) \leq k \sum_{\max\{\|\boldsymbol{h}\|,|u|\}\geq n} \mathcal{X}(\boldsymbol{h},u) \leq k \sum_{\max\{\sqrt{d}|h_1|,\ldots,\sqrt{d}|h_d|,|u|\}\geq n} \mathcal{X}(\boldsymbol{h},u)$.

□

The task is now to show that the conditions of Proposition 4.7 hold for the Brown-Resnick process $\boldsymbol{\eta}$ constructed in Theorem 3.5. Recall from (3.15) that, for $\boldsymbol{h} = (h_1,\ldots,h_d) \in \mathbb{R}^d$ and $u \in \mathbb{R}$, its tail dependence coefficient is given by

$$\mathcal{X}(\boldsymbol{h},u) = 2\left(1-\Phi\left(\sqrt{C_1|h_1|^{\alpha_1}+\cdots+C_d|h_d|^{\alpha_d}+C_{d+1}|u|^{\alpha_{d+1}}}\right)\right).$$

Proposition 4.9. *For the Brown-Resnick process $\boldsymbol{\eta} = \{\eta_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\}$, conditions (1)-(3) of Proposition 4.7 hold.*

Proof. (1) Using $1-\Phi(x) \leq \exp(-\frac{1}{2}x^2)$ for $x > 0$ we find, applying Lemma 4.8,

$$\begin{aligned}
\alpha_{k,\ell}(n) &\leq 2k\ell \sup_{\max\{\sqrt{d}|h_1|,\ldots,\sqrt{d}|h_d|,|u|\}\geq n} \left(1-\Phi(\sqrt{\delta(h_1,\ldots,h_d,u)}\right) \\
&\leq 2k\ell \sup_{\max\{\sqrt{d}|h_1|,\ldots,\sqrt{d}|h_d|,|u|\}\geq n} \exp\left\{-\frac{\delta(h_1,\ldots,h_d,u)}{2}\right\} \\
&= 2k\ell \sup_{\max\{\sqrt{d}|h_1|,\ldots,\sqrt{d}|h_d|,|u|\}\geq n} \\
&\quad \exp\left\{-\frac{1}{2}\left[C_1|h_1|^{\alpha_1}+\ldots+C_d|h_d|^{\alpha_d}+C_{d+1}|u|^{\alpha_{d+1}}\right]\right\} \\
&\leq 2k\ell \sup_{\max\{\sqrt{d}|h_1|,\ldots,\sqrt{d}|h_d|,|u|\}\geq n} \\
&\quad \exp\left\{-\frac{1}{2}\min\{|C_1|,\ldots,|C_{d+1}|\}\max\{|h_1|,\ldots,|h_d|,|u|\}^{\min\{\alpha_1,\ldots,\alpha_{d+1}\}}\right\} \\
&\to 0 \text{ as } n\to\infty \text{ for all } k,\ell \geq 0.
\end{aligned}$$

By Definition 4.6, $\boldsymbol{\eta}$ is α-mixing.

(2) • For $k+\ell \le 4(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1)$ we have, again applying Lemma 4.8,

$$\begin{aligned}
&\sum_{n=1}^{\infty} n^d \alpha_{k,\ell}(n) \\
&\le 2k\ell \sum_{n=1}^{\infty} n^d \sup_{\max\{\sqrt{d}|h_1|,\ldots,\sqrt{d}|h_d|,|u|\}\ge n} \\
&\quad \exp\left\{-\frac{1}{2}\left(C_1|h_1|^{\alpha_1}+\ldots+C_d|h_d|^{\alpha_d}+C_{d+1}|u|^{\alpha_{d+1}}\right)\right\} \\
&\le 2k\ell \sum_{n=1}^{\infty} n^d \exp\left\{-\frac{1}{2}\min\{C_1,\ldots,C_{d+1}\}\left(\frac{n}{\sqrt{d}}\right)^{\min\{\alpha_1,\ldots,\alpha_{d+1}\}}\right\} \\
&< \infty
\end{aligned}$$

• Use Lemma 4.8 to obtain

$$\begin{aligned}
&n^{d+1}\alpha_{(|\mathcal{H}_{bsyr}|+1)(p+1),\infty}(n) \\
&\le n^{d+1}(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1)\sum_{x\ge\frac{n}{\sqrt{d}}} \exp\left\{-\frac{1}{2}\min\{C_1,\ldots,C_{d+1}\}x^{\min\{\alpha_1,\ldots,\alpha_{d+1}\}}\right\} \\
&\to 0 \text{ as } n\to\infty.
\end{aligned}$$

(3) With $\beta = 1$, we obtain

• $\mathbb{E}\left[\left|\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(i_1,\ldots,i_d,t;\boldsymbol{r},p)\right|^{(2+\beta)}\right] < \infty$ because of Lemma 4.5 (i).

• $\sum_{n=1}^{\infty} n^d \alpha_{(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1),(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1)}(n)^{\frac{\beta}{2+\beta}}$
$\le (2[(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1)]^2)^{\frac{1}{3}} \sum_{n=1}^{\infty} n^d \exp\left\{-\frac{1}{6}\min\{C_1,\ldots,C_{d+1}\}\left(\frac{n}{\sqrt{d}}\right)^{\min\{\alpha_1,\ldots,\alpha_{d+1}\}}\right\}$
$< \infty$.

□

Now we formulate the main result in this section.

Theorem 4.10. *Assume the same conditions as in Theorem 4.4. Then*

$$\sqrt{M^d T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, F^{-1}\Sigma(F^{-1})'), \text{ as } M,T\to\infty, \tag{4.34}$$

where

$$F = \mathbb{E}\left[-\nabla^2_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(1,\ldots,1,1;\boldsymbol{r},p)\right]$$

and

$$\Sigma = \sum_{i_1=-\infty}^{\infty}\cdots\sum_{i_d=-\infty}^{\infty}\sum_{t=1}^{\infty} \mathrm{Cov}\left[\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(1,\ldots,1,1;\boldsymbol{r},p), \nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(i_1,\ldots,i_d,t;\boldsymbol{r},p)\right].$$

Proof. By Propositions 4.7 and 4.9 we have that the central limit theorem for $\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}}$ holds in the following sense:

$$\frac{1}{M^{\frac{d}{2}}\sqrt{T}} \sum_{i_1=1}^{M} \cdots \sum_{i_d=1}^{M} \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(i_1, \ldots, i_d, t; \boldsymbol{r}, p) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \Sigma) \text{ as } M, T \to \infty.$$

Taylor expansion around the true parameter vector $\boldsymbol{\theta}^\star$ of the score function $\nabla_{\boldsymbol{\theta}} PL^{(M,T)}(\boldsymbol{\theta})$ yields for some $\tilde{\boldsymbol{\theta}} \in [\boldsymbol{\theta}^\star, \hat{\boldsymbol{\theta}}]$:

$$\mathbf{0} = \nabla_{\boldsymbol{\theta}} PL^{(M,T)}(\hat{\boldsymbol{\theta}}) = \nabla_{\boldsymbol{\theta}} PL^{(M,T)}(\boldsymbol{\theta}^\star) + \nabla^2_{\boldsymbol{\theta}} PL^{(M,T)}(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star).$$

Therefore,

$$\begin{aligned}
M^{\frac{d}{2}}\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) &= -\left(\frac{1}{M^d T} \nabla^2_{\boldsymbol{\theta}} PL^{(M,T)}(\tilde{\boldsymbol{\theta}})\right)^{-1} \left(\frac{1}{M^{\frac{d}{2}} T} \nabla_{\boldsymbol{\theta}} PL^{(M,T)}(\boldsymbol{\theta}^\star)\right) \\
&= -\left(\frac{1}{M^d T} \sum_{i_1=1}^{M} \cdots \sum_{i_d=1}^{M} \sum_{t=1}^{M} \nabla^2_{\boldsymbol{\theta}} q_{\tilde{\boldsymbol{\theta}}}(i_1, \ldots, i_d, t; \boldsymbol{r}, p) - \frac{1}{M^d T} \nabla^2_{\boldsymbol{\theta}} \mathcal{R}^{(M,T)}(\tilde{\boldsymbol{\theta}})\right)^{-1} \\
&\quad \left(\frac{1}{M^{\frac{d}{2}}\sqrt{T}} \sum_{i_1=1}^{M} \cdots \sum_{i_d=1}^{M} \sum_{t=1}^{M} \nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(i_1, \ldots, i_d, t; \boldsymbol{r}, p) - \frac{1}{M^{\frac{d}{2}}\sqrt{T}} \nabla_{\boldsymbol{\theta}} \mathcal{R}^{(M,T)}(\boldsymbol{\theta}^\star)\right) \\
&=: -(I_1 - I_2)^{-1}(J_1 - J_2).
\end{aligned}$$

Note the following:

- By Proposition 4.7 it follows that J_1 converges to $\mathcal{N}(\mathbf{0}, \Sigma)$ in distribution as M, T tend to infinity.
- By assumption and similar arguments as in the proof of strong consistency (Theorem 4.4), the term J_2 converges to $\mathbf{0}$ in probability.
- As $\{\eta_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{Z}^d, t \in \mathbb{N}\}$ is α-mixing, the process

  $$\{\nabla^2_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}}(i_1, \ldots, i_d, t; \boldsymbol{r}, p) : i_m \in \mathbb{Z}, m = 1, \ldots d, t \in \mathbb{N}\}$$

  is α-mixing as a set of measurable functions of mixing lagged processes. Uniform convergence holds because of Lemma 4.5 which states that

  $$\mathbb{E}\left[\sup_{\boldsymbol{\theta} \in \Theta^\star} \left|\nabla^2_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}}(1, \ldots, 1, 1; \boldsymbol{r}, p)\right|\right] < \infty.$$

  Furthermore, as $\tilde{\boldsymbol{\theta}} \in [\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}^\star]$ and $\hat{\boldsymbol{\theta}}$ is strongly consistent, we have that I_1 converges to $-F$ almost surely as M and T converge to infinity.
- Concerning I_2, the law of large numbers applied on $\{\nabla^2_{\boldsymbol{\theta}} \log\{g_{\boldsymbol{\theta}}(\eta_{\boldsymbol{s},t}, \eta_{\boldsymbol{s}+\boldsymbol{h},t+u})\}\}$ results in the fact that, similarly as in the proof of Theorem 4.4, I_2 converges to $\mathbf{0}$ almost surely as M and T raise to infinity.

Finally, summarizing those results, Slutzky's Lemma yields

$$M^{\frac{d}{2}}\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, F^{-1}\Sigma(F^{-1})') \text{ as } M, T \to \infty.$$

□

4.3 Further Extensions to the Correlation Models

Throughout this thesis we mostly assume the limit function δ in Assumption 3.1 to be of the form

$$\delta(h_1, \dots, h_m, u) = C_1|h_1|^{\alpha_1} + \cdots + C_d|h_d|^{\alpha_d} + C_{d+1}|u|^{\alpha_{d+1}}$$

with $C_m > 0$ and $\alpha_m \in (0, 2]$ for $m = 1, \dots, d+1$. One can say that the spatial anisotropy is modelled independently for each dimension. However, one might be motivated to do further research on anisotropic models that allow for interaction between the different directions. One could generalize our model and assume δ to be (here for notational ease for the case $d = 2$)

$$\delta(h_1, h_2, u) = C_1|h_1|^{\alpha_1} + C_2|h_2|^{\alpha_2} + C_3|h_1 h_2|^{\alpha_3} + C_4|u|^{\alpha_4}.$$

One way to achieve this is by introducing geometric anisotropy, as done in Davis et al. [2012a], Section 4.2. Consider the isotropic stable class correlation function

$$\tilde{\gamma}(h_1, h_2, u) = \exp\left\{ - \left\| \left(\theta_1^{\frac{1}{2}} h_1, \theta_1^{\frac{1}{2}} h_2, \theta_2^{\frac{1}{2}} u \right) \right\|^2 \right\}.$$

Set

$$R := \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}, \; T := \begin{pmatrix} \frac{1}{a_{\max}} & 0 \\ 0 & \frac{1}{a_{\min}} \end{pmatrix},$$

where R is a rotation matrix and T a distance matrix in $\mathbb{R}^2$ with parameters $\varphi \in [0, 2\pi)$ and $0 < a_{\min} \leq a_{\max}$, respectively. Then the spatial anisotropy matrix $\tilde{A}$ is given through R and T as

$$\tilde{A} = TR = \begin{pmatrix} \frac{\cos\varphi}{a_{\max}} & \frac{\sin\varphi}{a_{\max}} \\ -\frac{\sin\varphi}{a_{\min}} & \frac{\cos\varphi}{a_{\min}} \end{pmatrix}.$$

Now define

$$A := \begin{pmatrix} TR & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} = \begin{pmatrix} \frac{\cos\varphi}{a_{\max}} & \frac{\sin\varphi}{a_{\max}} & 0 \\ -\frac{\sin\varphi}{a_{\min}} & \frac{\cos\varphi}{a_{\min}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and the anisotropic correlation function γ by

$$\gamma(h_1, h_2, u) := \tilde{\gamma}(A(h_1, h_2, u)') = \exp\left\{ - \left\| A \left(\theta_1^{\frac{1}{2}} h_1, \theta_1^{\frac{1}{2}} h_2, \theta_2^{\frac{1}{2}} u \right) \right\|^2 \right\},$$

where θ_1 and θ_2 are some positive constants. This yields

$$\begin{aligned}
&\gamma(h_1, h_2, u) \\
&= \exp\left\{ -\theta_1 \left[\left(\left(\frac{\cos\varphi}{a_{\max}} \right)^2 + \left(\frac{\sin\varphi}{a_{\min}} \right)^2 \right) h_1^2 + \left(\left(\frac{\sin\varphi}{a_{\max}} \right)^2 + \left(\frac{\cos\varphi}{a_{\min}} \right)^2 \right) h_2^2 \right.\right. \\
&\quad \left.\left. + \left(2\frac{\cos\varphi \sin\varphi}{a_{\max}^2} - 2\frac{\cos\varphi \sin\varphi}{a_{\min}^2} \right) h_1 h_2 \right] - \theta_2 u^2 \right\} \\
&= \exp\left\{ -\theta_1 \|\tilde{A}\boldsymbol{h}\|^2 - \theta_2 u^2 \right\}
\end{aligned}$$

Choosing as before $s_{n,1} = s_{n,2} = t_n = (\log n)^{-\frac{1}{2}}$, we have again

$$\begin{aligned}
&\lim_{n\to\infty} \{(\log n)(1 - \gamma(s_{n,1}h_1, s_{n,2}h_2, t_n u))\} \\
&= \theta_1 \left[\left(\left(\frac{\cos\varphi}{a_{\max}} \right)^2 + \left(\frac{\sin\varphi}{a_{\min}} \right)^2 \right) h_1^2 + \left(\left(\frac{\sin\varphi}{a_{\max}} \right)^2 + \left(\frac{\cos\varphi}{a_{\min}} \right)^2 \right) h_2^2 \right. \\
&\quad \left. + \left(2\frac{\cos\varphi \sin\varphi}{a_{\max}^2} - 2\frac{\cos\varphi \sin\varphi}{a_{\min}^2} \right) h_1 h_2 \right] + \theta_2 u^2 \\
&= \theta_1 \|\tilde{A}\boldsymbol{h}\|^2 + \theta_2 u^2 \\
&=: \delta(h_1, h_2, u).
\end{aligned}$$

The following two chapters deal with applications of the pairwise likelihood estimation procedure theoretically introduced throughout Section 4.2. Chapter 5 contains a simulation study, whereas Chapter 6 treats real data. The special interest lies in how the results differ from those of Davis et al. [2012b] and Steinkohl [2012].

Chapter 5

Simulation Study

In this chapter we test the pairwise likelihood method introduced in Section 4.2 for the Brown-Resnick process constructed in Theorem 3.5. We assume two spatial dimensions ($d = 2$). For the simulation of the Gaussian space-time processes we use again the `R`-package `RandomFields`. However, in the anisotropic context, the package only allows for simulation of covariance functions γ where the parameters α_1 and α_2 are equal to two and therefore lie on the boundary of the parameter space. This is a problem; for example, estimates above 2 would have to be censored, destroying the normal distribution. So in the following we fix α_1, α_2 and α_3 and show how the method works for the remaining parameters C_1, C_2 and C_3.

5.1 Procedure

The following steps are executed 100 times:

(i) We simulate 1000 independent Gaussian space-time processes $Z_j(S_n(s_1, s_2), t_n t)$, $j = 1, \dots, 1000$ with covariance function $\gamma(S_n(h_1, h_2), t_n u)$ where γ belongs to the stable class and is given by

$$\gamma(h_1, h_2, u) = \exp\left\{-\left(C_1^\star h_1^2 + C_2^\star h_2^2 + C_3^\star u^2\right)\right\}.$$

The true parameters are chosen as $C_1^\star = 0.02$, $C_2^\star = 0.04$ and $C_3^\star = 0.05$. The processes are simulated on a regular grid $\{1, \dots, M\} \times \{1, \dots, M\}$ for some $M \in \mathbb{N}$. The time points are equidistant and given by the set $\{1, 2, \dots, T\}$ for some $T \in \mathbb{N}$. As seen in Section 4.1, Assumption 4.1 is satisfied with corresponding limit function

$$\delta(h_1, h_2, u) = C_1^\star h_1^2 + C_2^\star h_2^2 + C_3^\star u^2.$$

(ii) We transform the simulated processes to standard Fréchet margins by setting

$$\tilde{Z}_j(s_1, s_2, t) := -\frac{1}{\log\left(\Phi(Z_j(s_1, s_2, t))\right\}},$$
$$(s_1, s_2) \in \{1, \dots, M\} \times \{1, \dots, M\},\ t \in \{1, \dots, T\}.$$

For justification recall the probability integral transform method.

(iii) To obtain an approximation of the max-stable Brown-Resnick process, take the pointwise maximum over the generations $\tilde{Z}_j(S_n(s_1, s_2, t_n t)$ in space and time and divide by 1000 (the number of simulated Gaussian space-time processes):

$$\eta(s_1, s_2, t) = \frac{1}{1000} \bigvee_{j=1}^{1000} \tilde{Z}_j(S_n(s_1, s_2), t_n t), \;\; (s_1, s_2) \in \{1, \dots, M\}^2, \;\; t \in \{1, \dots, T\}.$$

(iv) Obtain estimates $(\hat{C}_1, \hat{C}_2, \hat{C}_3)$ by using pairwise likelihood estimation, as described in Section 4.2.

5.2 Results

We did the procedure described in Section 5.1 a couple of times, each time slightly changing the edge length M of the grid and the number of time points T. In the following, we illustrate how the results change and improve when M and T increase. We present the estimates obtained in each setting. We write $M \times M \times T$ when simulating the processes on the regular grid with edge length M during T equidistant time points. What is conspicuous, is the fact that in the anisotropic setting and for the pairwise log-likelihood function $PL^{(M,T)}$ defined in (4.14) we basically need the squared number of grid points as Davis et al. [2012b] need in their estimation procedure because for estimating the spatial parameters C_1 and C_2 we only use data that lie on directional lines, whereas they use the data quadratically. In their simulation study (Chapter 6), Davis et al. [2012b] point out that the results improve if, in the pairwise likelihood estimation, one only uses pairs with rather small spatial and temporal lags. On top of that, when estimating one certain parameter (for instance C_1), one obtains good results if the other lags (here: r_2 and p) are zero. The pairwise likelihood function (4.14) then naturally has a slightly simpler form: For example, if the maximum space-time lags are chosen as $(r_1, r_2, p) = (r_1, 0, 0)$ with $r_1 > 0$, then we have

$$PL^{(M,T)}(\boldsymbol{\theta}) = \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{t=1}^{T} \sum_{\substack{h_1=1 \\ i_1+h_1 \leq M}}^{r_1} \log \left\{ g_{\boldsymbol{\theta}}(\eta_{(i_1,i_2),t}, \eta_{(i_1+h_1,i_2,t)}) \right\}$$

and its form is analogous if we choose $(r_1, r_2, p) = (0, r_2, 0)$ with $r_2 > 0$. Similarly, if we set $(r_1, r_2, p) = (0, 0, p)$ with $p > 0$, we obtain

$$PL^{(M,T)}(\boldsymbol{\theta}) = \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{t=1}^{T} \sum_{\substack{u=1 \\ t+u \leq p}}^{p} \log \left\{ g_{\boldsymbol{\theta}}(\eta_{(i_1,i_2),t}, \eta_{(i_1,i_2),t+u}) \right\}.$$

We took this to heart in the estimation procedure.

(I) $10 \times 10 \times 100$.

The following table summarizes means and root mean squared errors (RMSE) for different spatial and temporal lags up to three. The results always refer to the estimates ($\hat{C}_1$, $\hat{C}_2$ or $\hat{C}_3$) whose corresponding lag is not zero. For example, the mean of the estimates obtained with maximum lags (r_1,r_2,p)=(0,2,0) corresponds to the estimate $\hat{C}_2$.

max. lags	(1,0,0)	(2,0,0)	(3,0,0)	(0,1,0)	(0,2,0)	(0,3,0)	(0,0,1)	(0,0,2)	(0,0,3)
mean	0.1343	0.1082	0.0948	0.1072	0.0888	0.0795	0.0519	0.0505	0.0489
RMSE	0.1156	0.0894	0.0760	0.0689	0.0504	0.0411	0.0107	0.0106	0.0106

Table 5.1: Simulation step $10 \times 10 \times 100$. Means and RMSE for the different parameters and different spatial and temporal lags

Figures 5.1 to 5.2 show line plots, box plots and normal qq-plots of the respective estimates. What one clearly sees is that the spatial estimates $\hat{C}_1$ and $\hat{C}_2$ are far away from the true parameters $C_1^\star$ and $C_2^\star$, whereas the temporal estimates $\hat{C}_3$ of the parameter $C_3^\star$ have a small RMSE and practically no bias. The reason is that for estimating the two spatial parameters in the anisotropic way, we have very few data points, as the estimation procedure does not proceed quadratically. As already mentioned, we need the squared number of grid points as Davis et al. [2012b] in their isotropic simulation study, so a 10×10 grid is probably not enough. The estimation works fine for the temporal parameter C_3 as we have a lot of time points for only one parameter. For the temporal parameters the procedure does not change compared to Davis et al. [2012b]. Therefore our results are similar. What is also noticeable, is the normality of the estimates, which can be concluded from the QQ-plots. Hoping to get better results concerning the spatial parameters C_1 and C_2, we go further and extend the grid successively. To cope with that computationally, we simulate only one single time point $t^\star$ and leave out the estimation of the time parameters. As can be seen above, their estimation already works pleasingly. In this case, the pairwise likelihood function (4.14) simplifies further and turns out to be (here for the maximum space-time lags $(r_1, r_2, p) = (r_1, 0, 0)$ with $r_1 > 0$):

$$PL^{(M,T)}(\boldsymbol{\theta}) = \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{\substack{h_1=1 \\ i_1+h_1 \leq M}}^{r_1} \log \left\{ g_{\boldsymbol{\theta}}(\eta_{(i_1,i_2),t^\star}, \eta_{(i_1+h_1),i_2,t^\star}) \right\}.$$

We present the results of the simulation on grid sizes 70×70 to 100×100, enlarging the side length by 10 in each step.

(II) $70 \times 70 \times 1$.

In the case of a grid of size 70×70 the bias and the root mean squared errors are reduced considerably compared to the case of a 10×10 grid. Table 5.2 and Figures 5.4 to 5.6 summarize and visualize the results. Especially, the improvements reflect in the box plots (Figure 5.5), where the green lines are now inside the boxes.

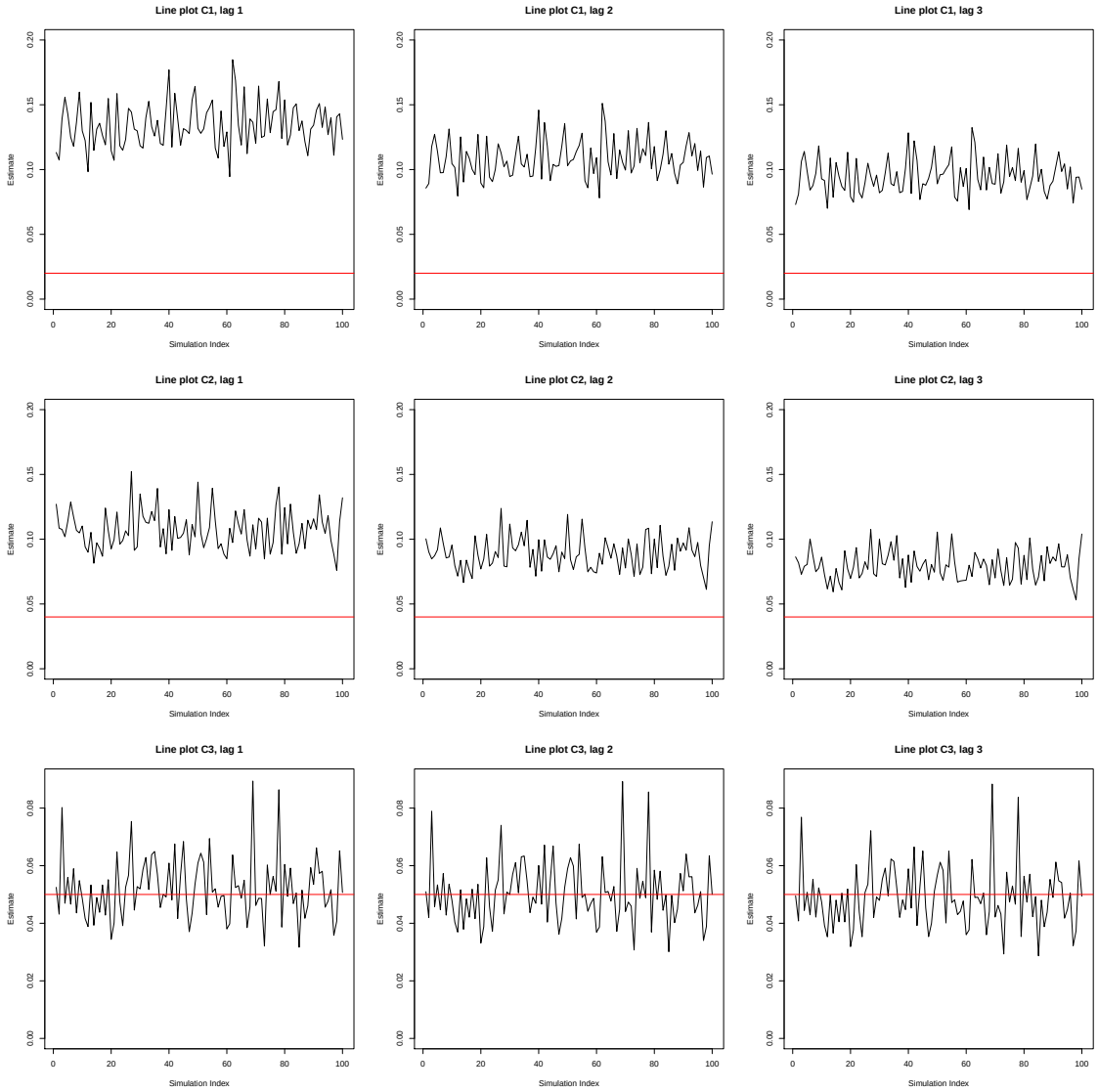

Figure 5.1: Simulation step $10 \times 10 \times 100$. Plots of the estimates $\hat{C}_1$, $\hat{C}_2$ and $\hat{C}_3$ for maximum lags 1 to 3.

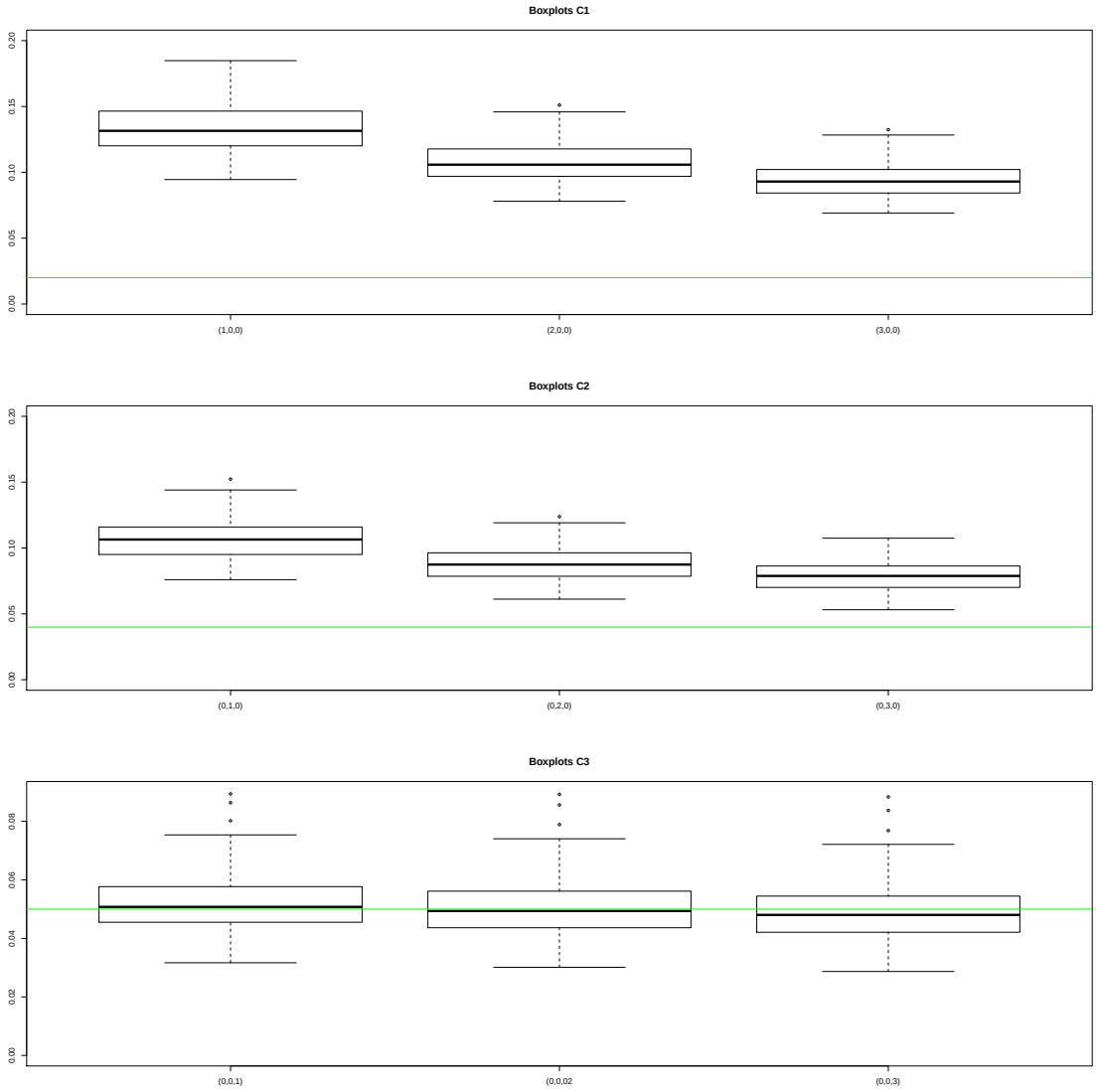

Figure 5.2: Simulation step $10 \times 10 \times 100$. Boxplots of the estimates $\hat{C}_1$, $\hat{C}_2$ and $\hat{C}_3$ for maximum lags 1 to 3.

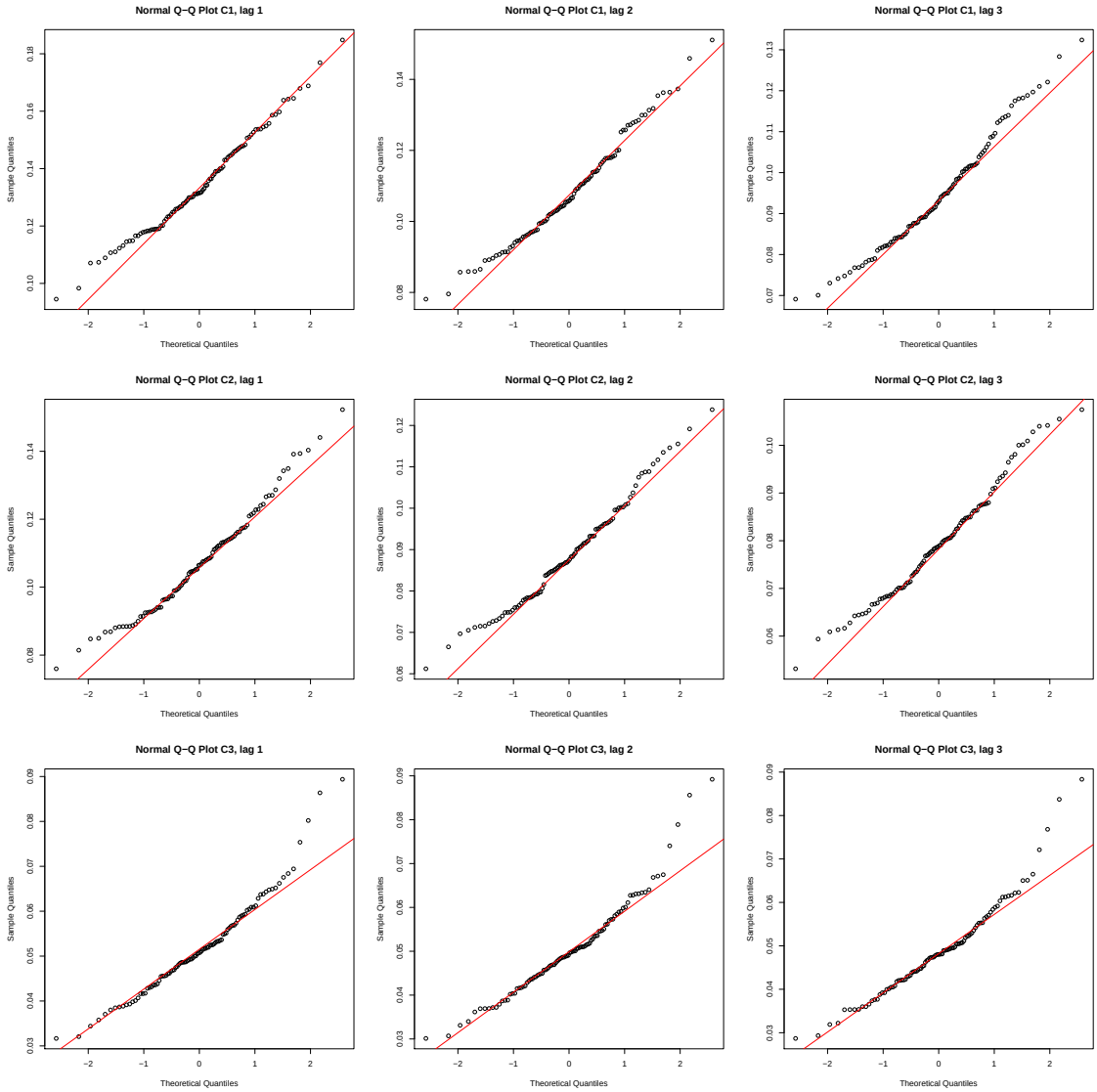

Figure 5.3: Simulation step $10 \times 10 \times 100$. Normal QQ-Plots of the estimates $\hat{C}_1$, $\hat{C}_2$ and $\hat{C}_3$ for maximum lags 1 to 3.

max. lags	(1,0,0)	(2,0,0)	(3,0,0)	(0,1,0)	(0,2,0)	(0,3,0)
mean	0.0189	0.0187	0.0185	0.0375	0.0369	0.0362
RMSE	0.0041	0.0043	0.0044	0.0080	0.0084	0.0088

Table 5.2: Simulation step $70 \times 70 \times 1$. Means and RMSE for the different parameters and different spatial lags

(III) $80 \times 80 \times 1$.

The trend of the pairwise likelihood estimates $\hat{C}_1$ and $\hat{C}_2$ towards their respective true values, which was indicated in the simulation step before, is still recognizable. In particular, the variance is further reduced (cf. Table 5.3). For visualizations see Figures 5.7 to 5.9.

max. lags	(1,0,0)	(2,0,0)	(3,0,0)	(0,1,0)	(0,2,0)	(0,3,0)
mean	0.0190	0.0188	0.0185	0.0373	0.0367	0.0362
RMSE	0.0032	0.0033	0.0034	0.0063	0.0067	0.0072

Table 5.3: Simulation step $80 \times 80 \times 1$. Means and RMSE for the different parameters and different spatial and lags

(IV) $90 \times 90 \times 1$.

Table 5.4 and Figures 5.10 to 5.12 show that, compared to the simulation step before, the results are only slightly improved. The best improvement is reflected in the estimate of the true parameter $C_2^\star$.

max. lags	(1,0,0)	(2,0,0)	(3,0,0)	(0,1,0)	(0,2,0)	(0,3,0)
mean	0.0191	0.0188	0.0186	0.0381	0.0375	0.0369
RMSE	0.0031	0.0032	0.0033	0.0061	0.0064	0.0067

Table 5.4: Simulation step $90 \times 90 \times 1$. Means and RMSE for the different parameters and different spatial lags

(V) $100 \times 100 \times 1$.

In this last simulation step, the results improve considerably again, which is particularly reflected in the root mean squared errors, which decline a lot compared to the steps before.

max. lags	(1,0,0)	(2,0,0)	(3,0,0)	(0,1,0)	(0,2,0)	(0,3,0)
mean	0.0198	0.0196	0.0194	0.0394	0.0388	0.0382
RMSE	0.0021	0.0022	0.0023	0.0036	0.0038	0.0042

Table 5.5: Simulation step $100 \times 100 \times 1$. Means and RMSE for the different parameters and different spatial lags

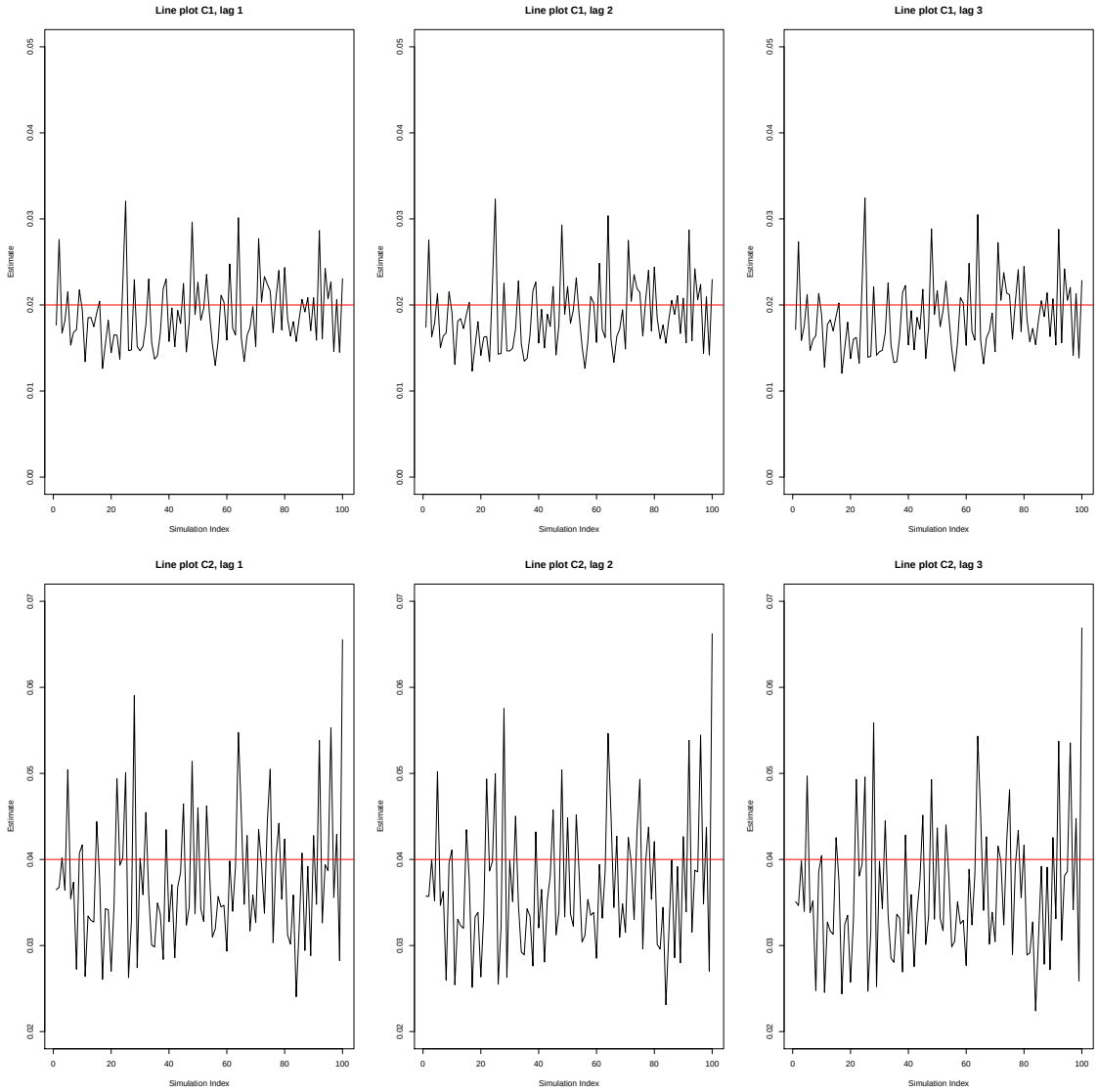

Figure 5.4: Simulation step $70 \times 70 \times 1$. Plots of the estimates $\hat{C}_1$ and $\hat{C}_2$ for maximum lags 1 to 3.

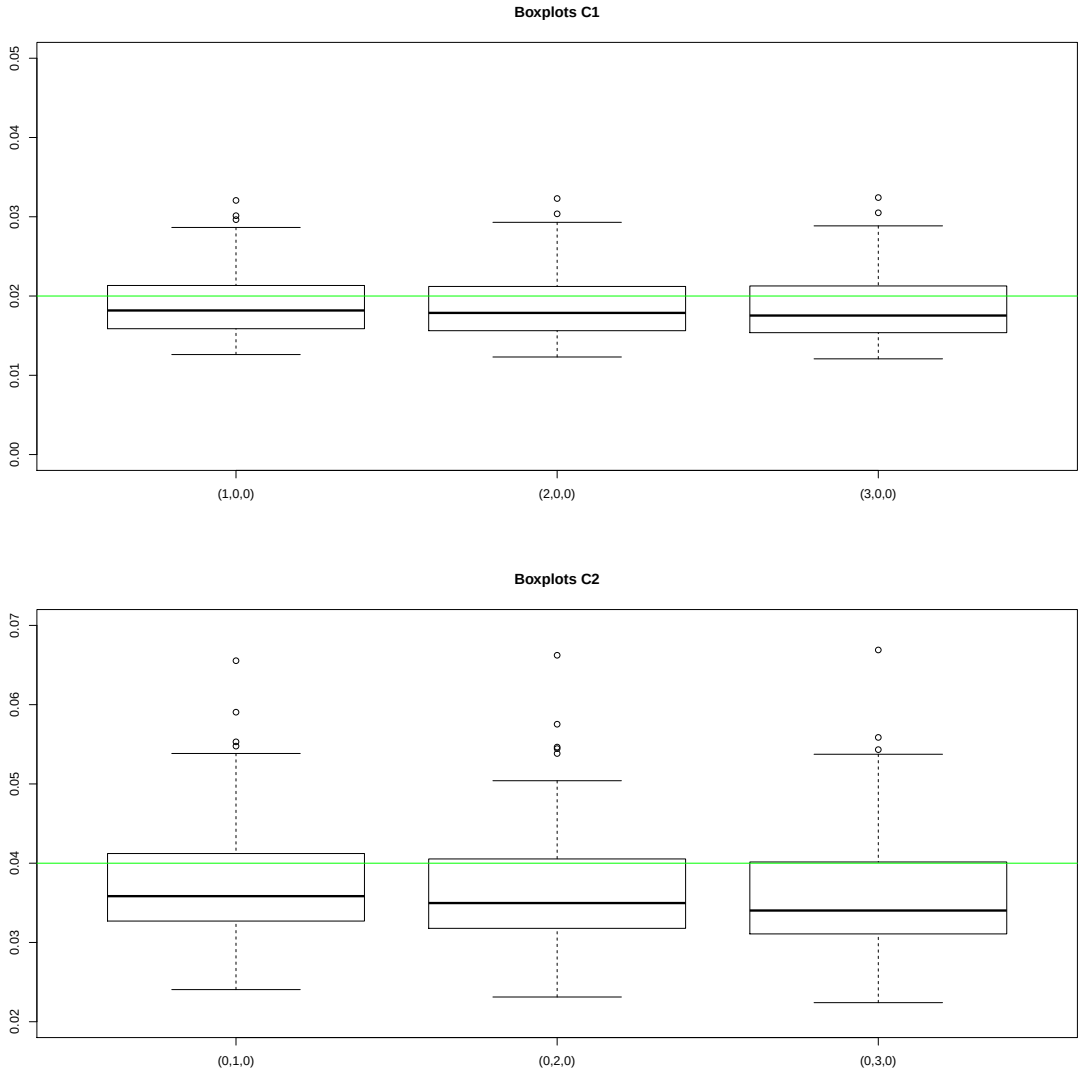

Figure 5.5: Simulation step $70 \times 70 \times 1$. Boxplots of the estimates $\hat{C}_1$ and $\hat{C}_2$ for maximum lags 1 to 3.

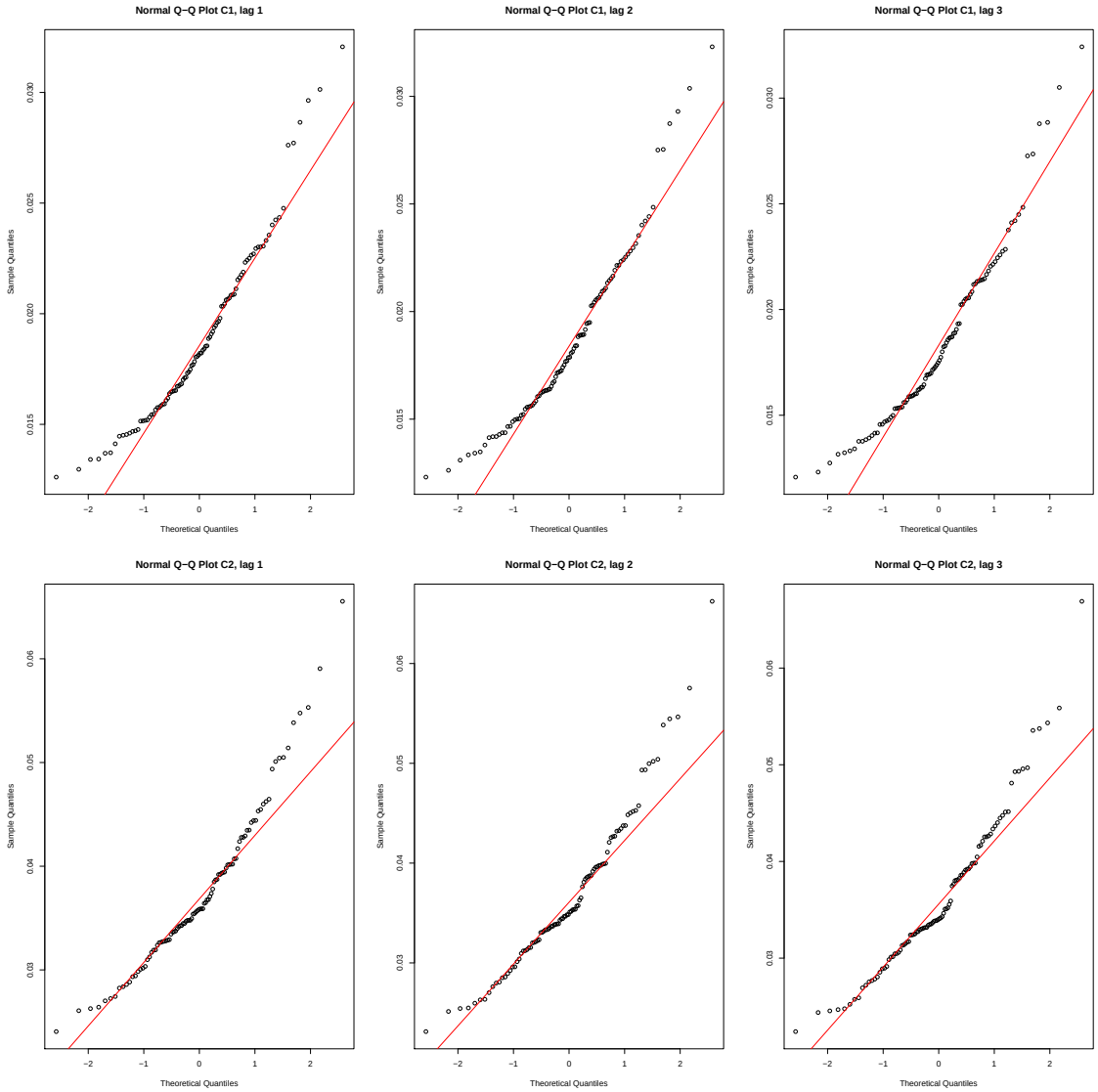

Figure 5.6: Simulation step $70 \times 70 \times 1$. Normal QQ-Plots of the estimates $\hat{C}_1$ and $\hat{C}_2$ for maximum lags 1 to 3.

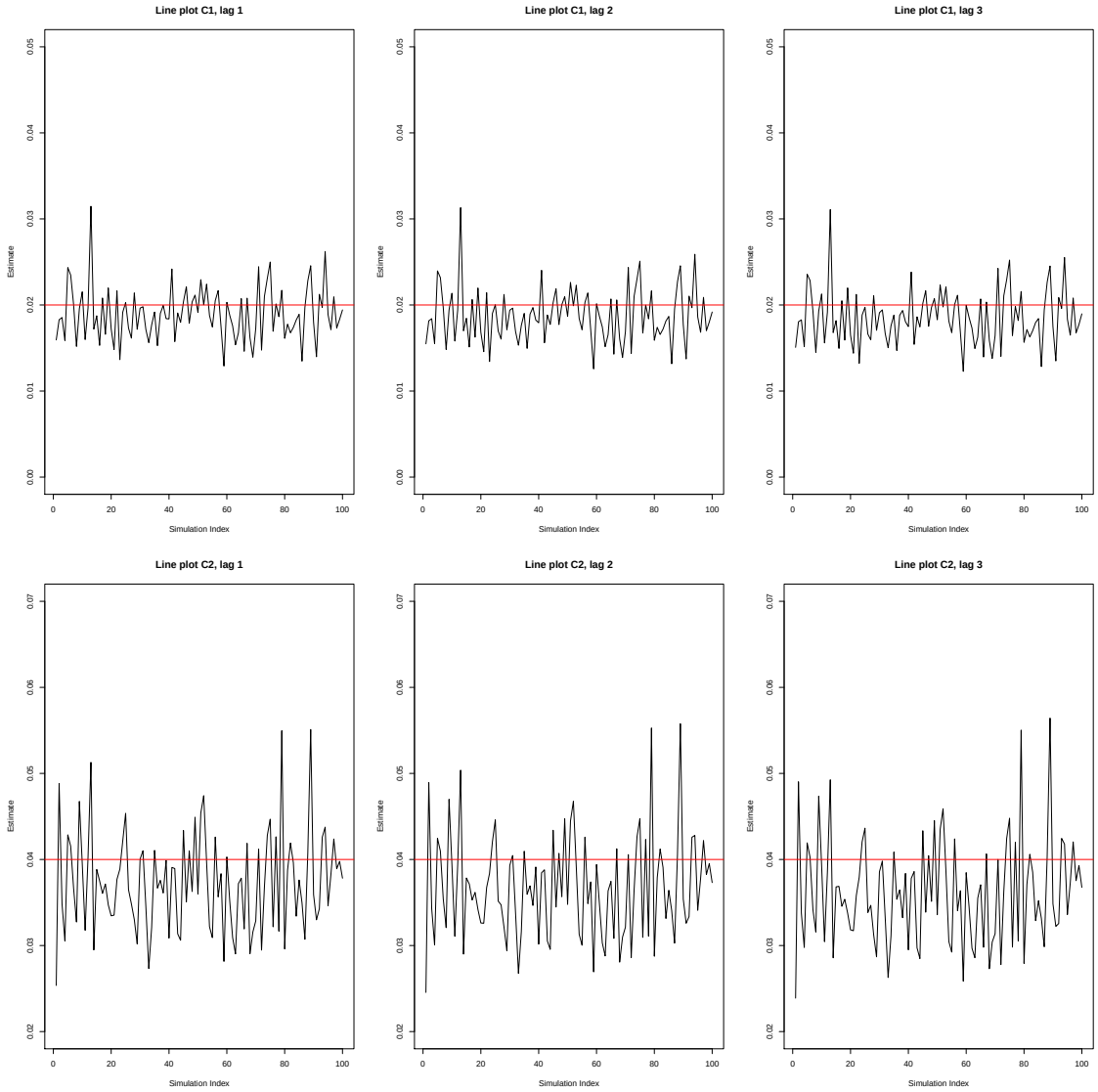

Figure 5.7: Simulation step $80 \times 80 \times 1$. Plots of the estimates $\hat{C}_1$ and $\hat{C}_2$ for maximum lags 1 to 3.

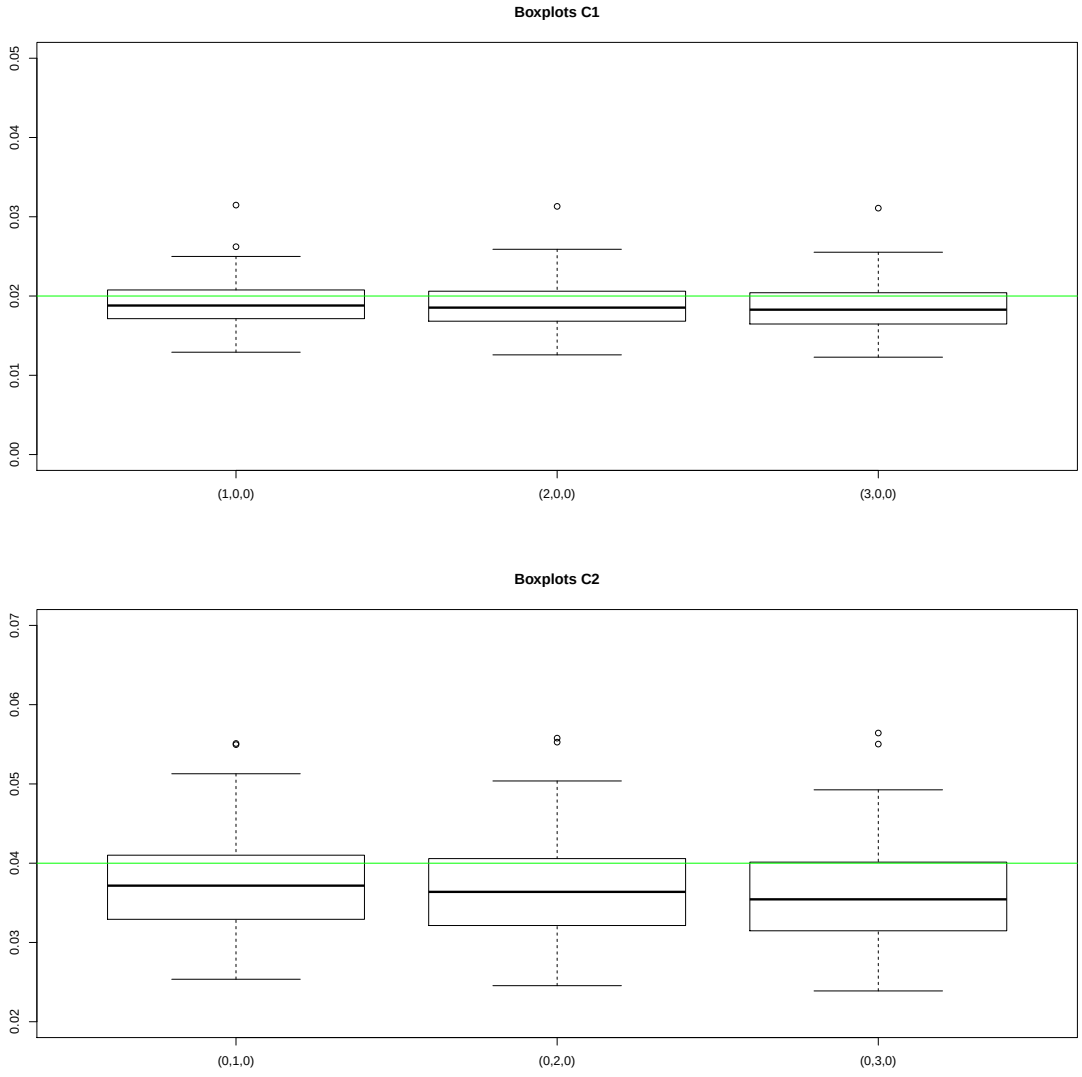

Figure 5.8: Simulation step $80 \times 80 \times 1$. Boxplots of the estimates $\hat{C}_1$ and $\hat{C}_2$ for maximum lags 1 to 3.

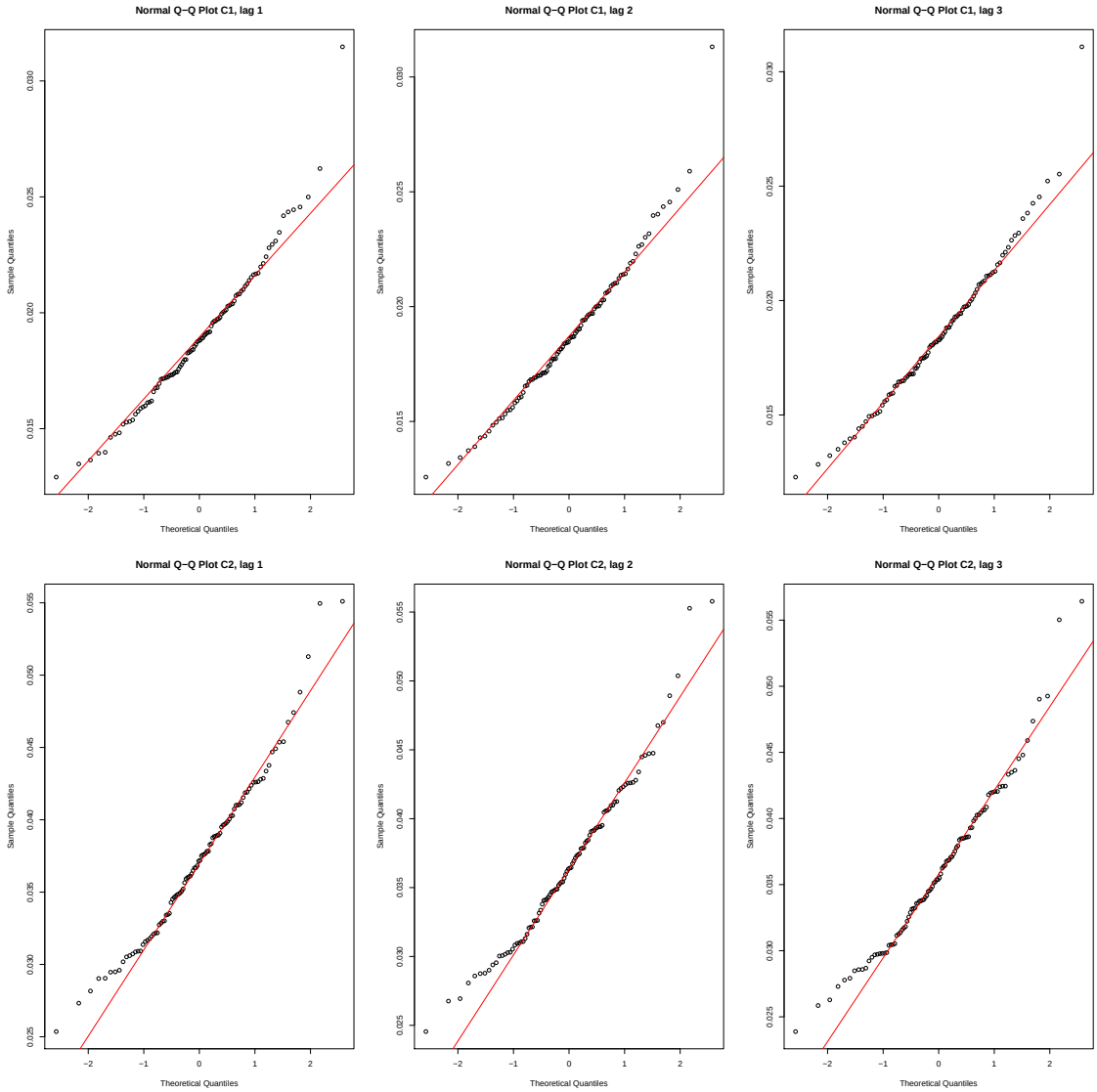

Figure 5.9: Simulation step $80 \times 80 \times 1$. Normal QQ-Plots of the estimates $\hat{C}_1$ and $\hat{C}_2$ for maximum lags 1 to 3.

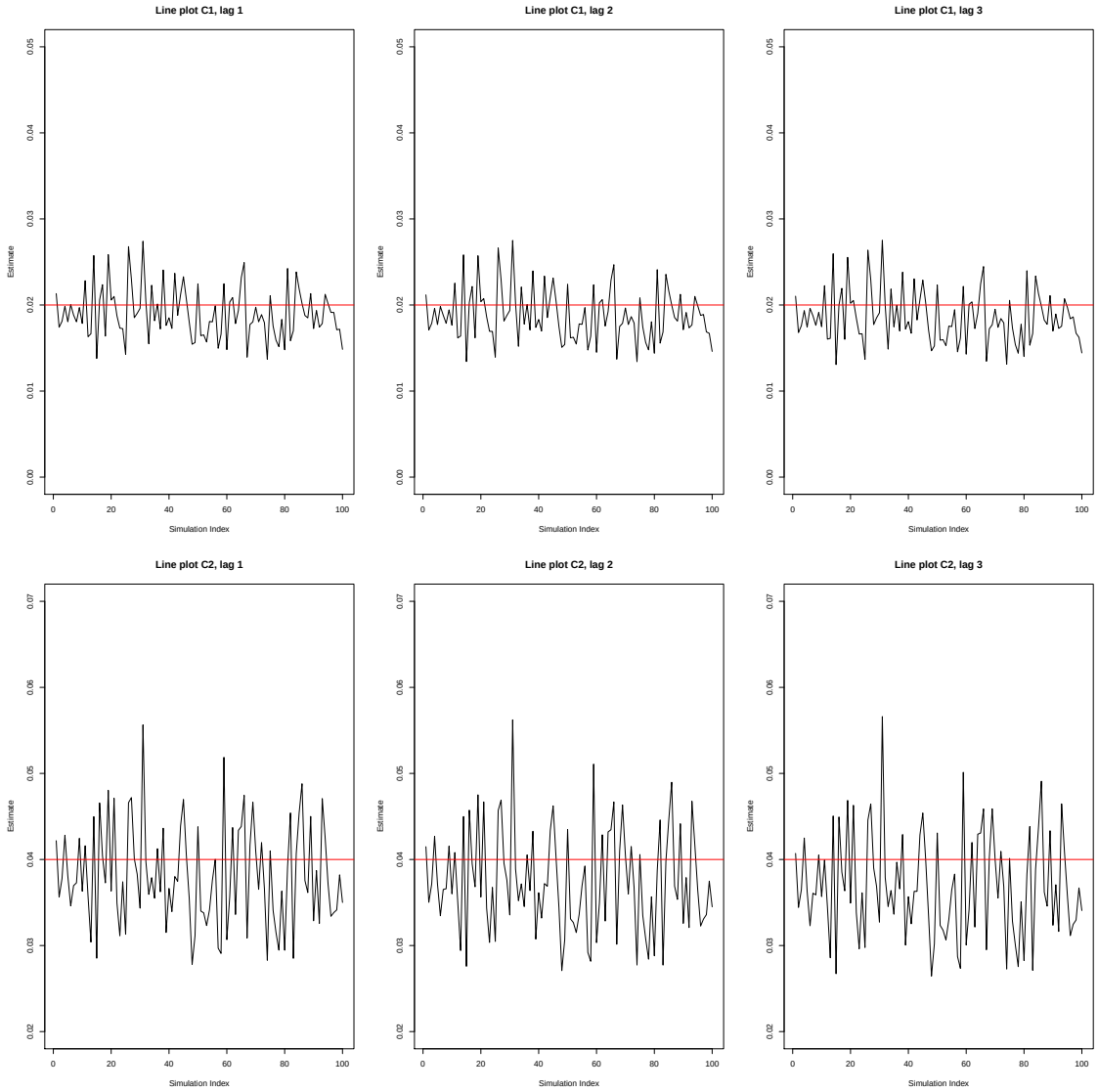

Figure 5.10: Simulation step $90 \times 90 \times 1$. Plots of the estimates $\hat{C}_1$ and $\hat{C}_2$ for maximum lags 1 to 3.

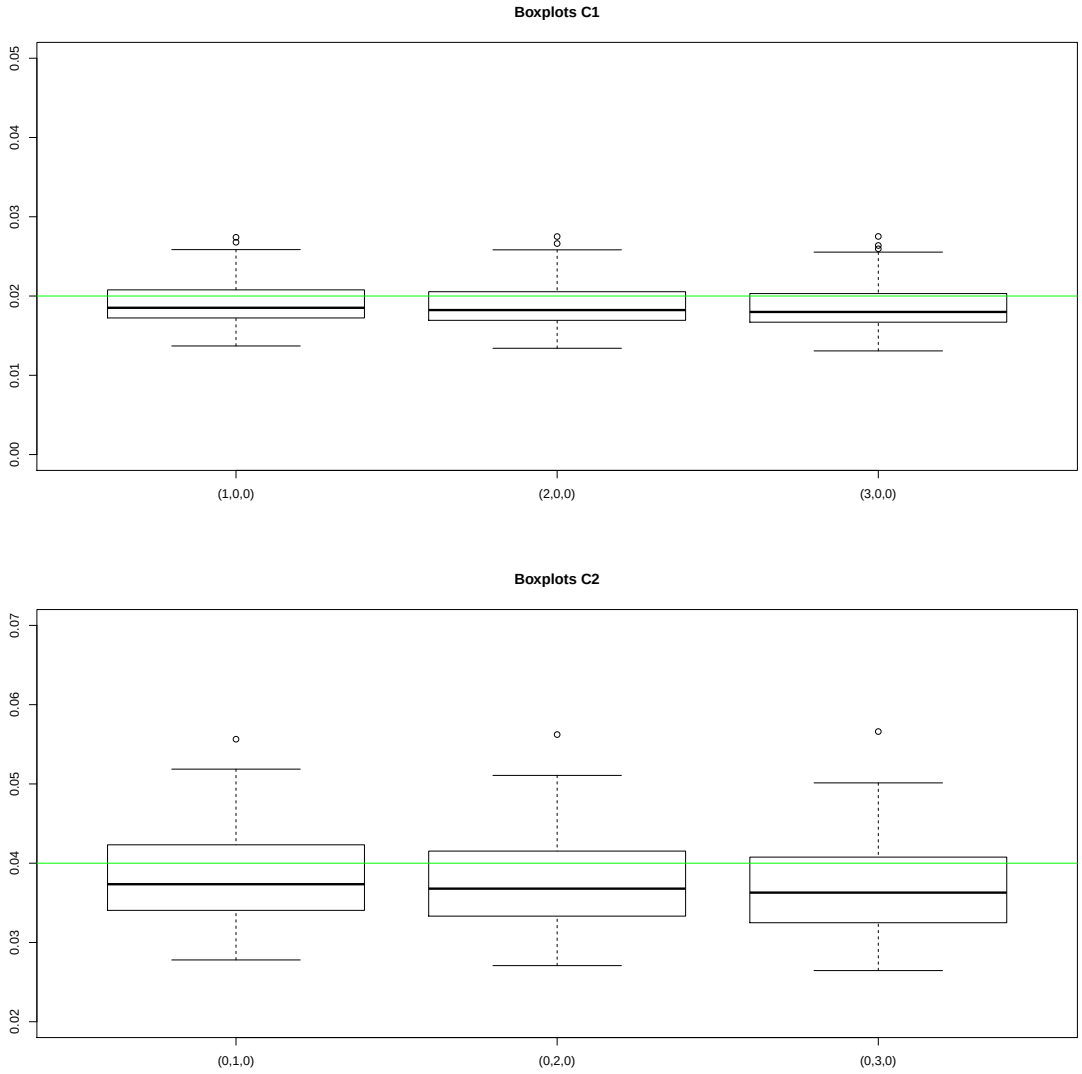

Figure 5.11: Simulation step $90 \times 90 \times 1$. Boxplots of the estimates $\hat{C}_1$ and $\hat{C}_2$ for maximum lags 1 to 3.

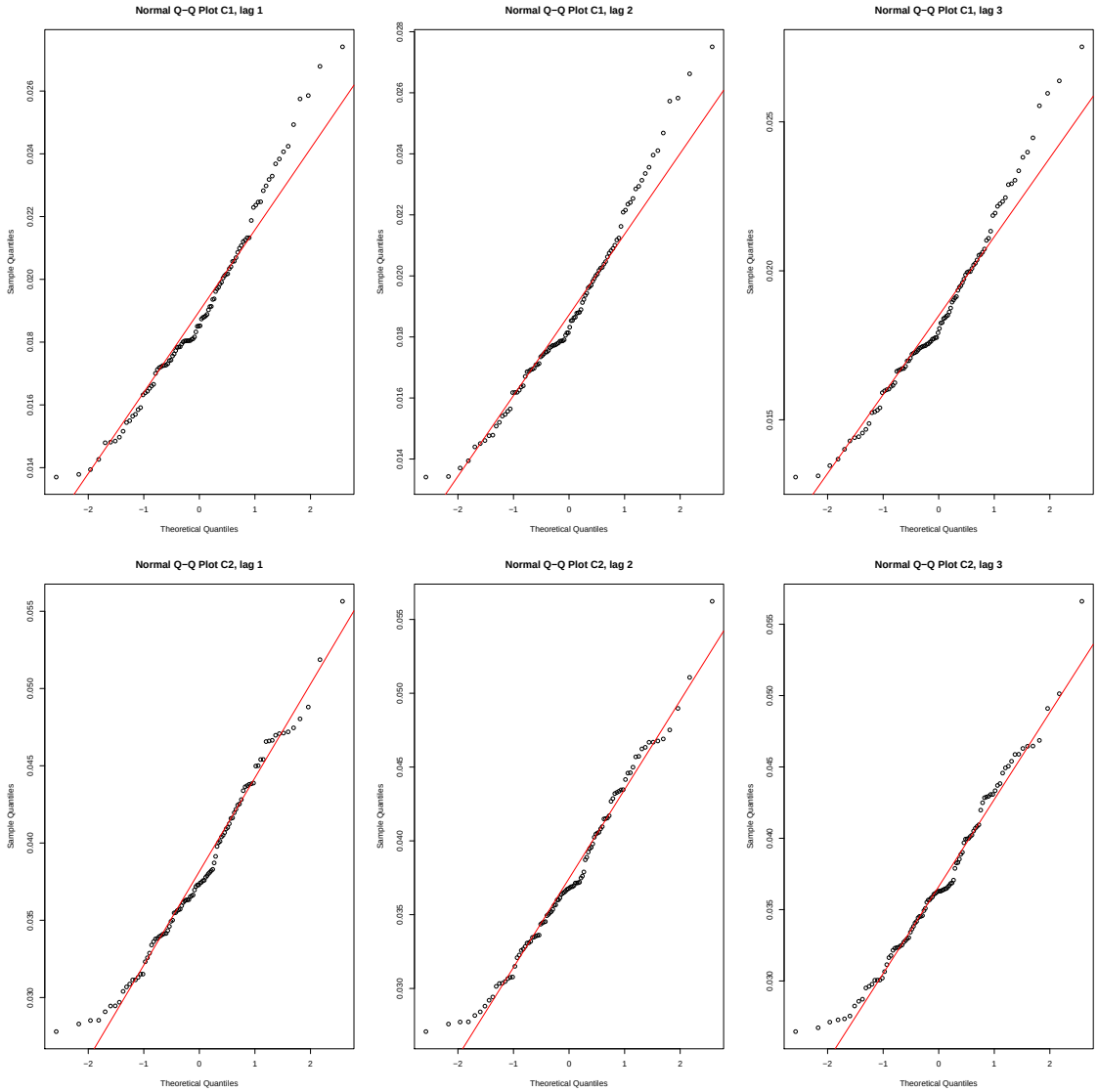

Figure 5.12: Simulation step $90 \times 90 \times 1$. Normal QQ-Plots of the estimates $\hat{C}_1$ and $\hat{C}_2$ for maximum lags 1 to 3.

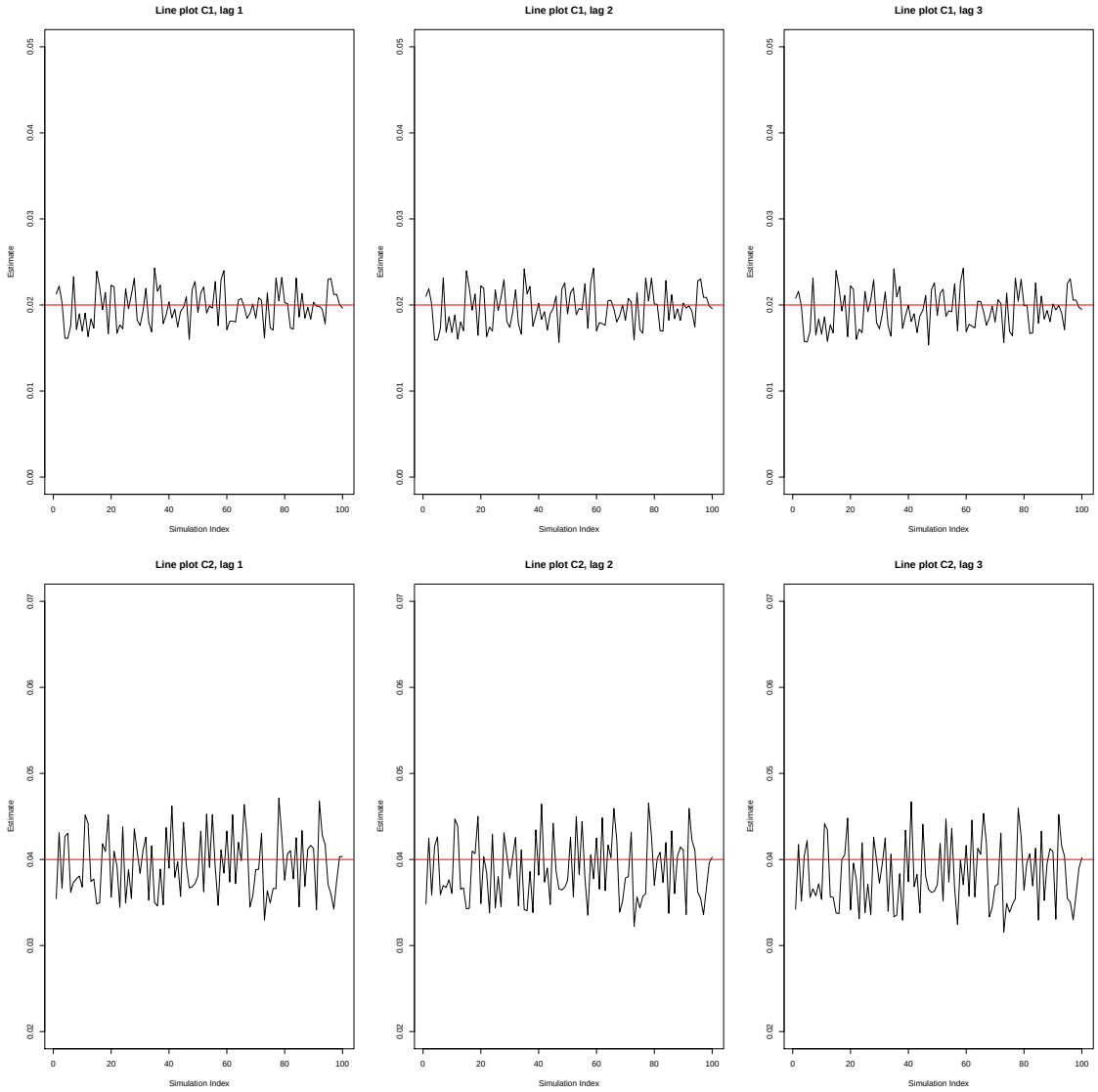

Figure 5.13: Simulation step $100 \times 100 \times 1$. Plots of the estimates $\hat{C}_1$ and $\hat{C}_2$ for maximum lags 1 to 3.

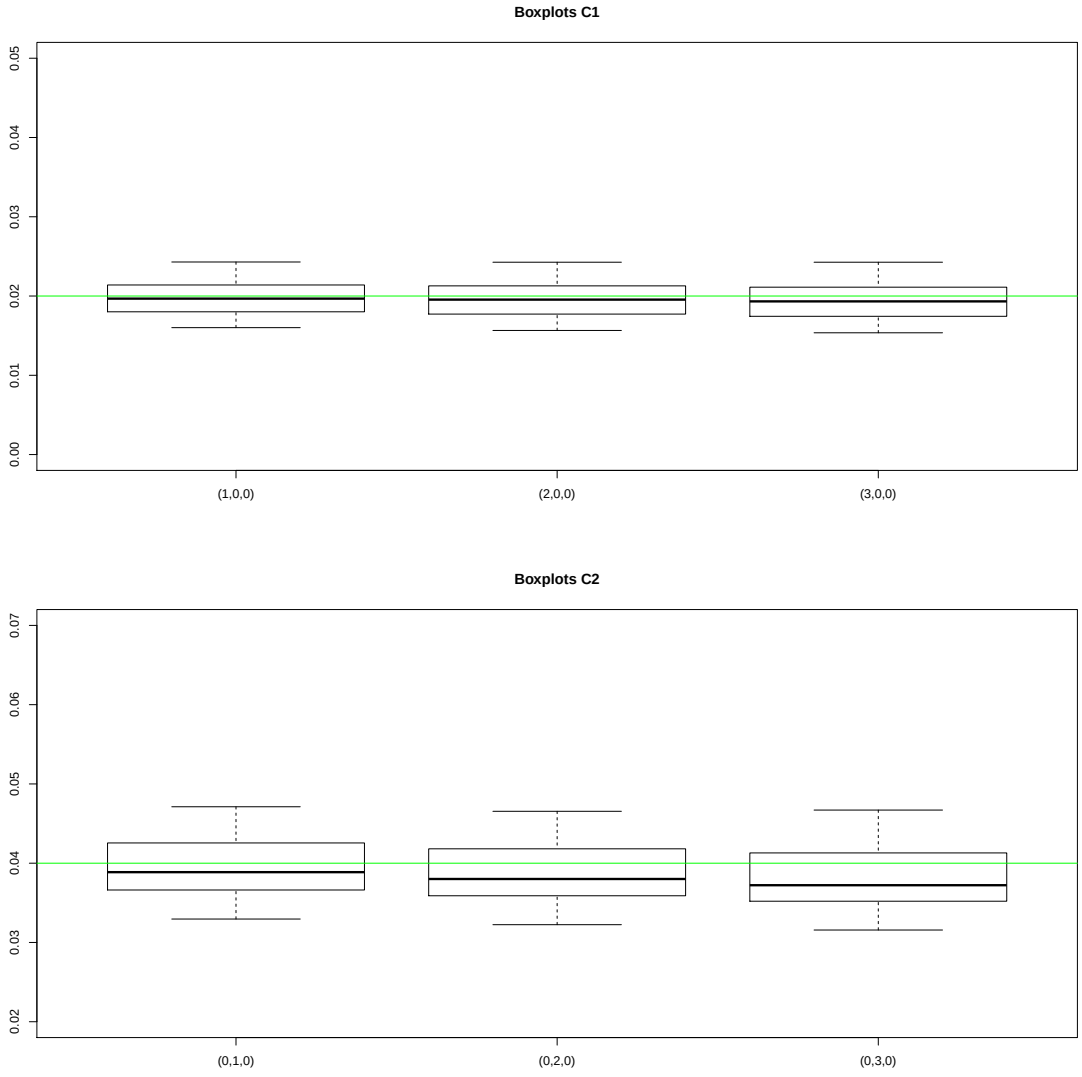

Figure 5.14: Simulation step $100 \times 100 \times 1$. Boxplots of the estimates $\hat{C}_1$ and $\hat{C}_2$ for maximum lags 1 to 3.

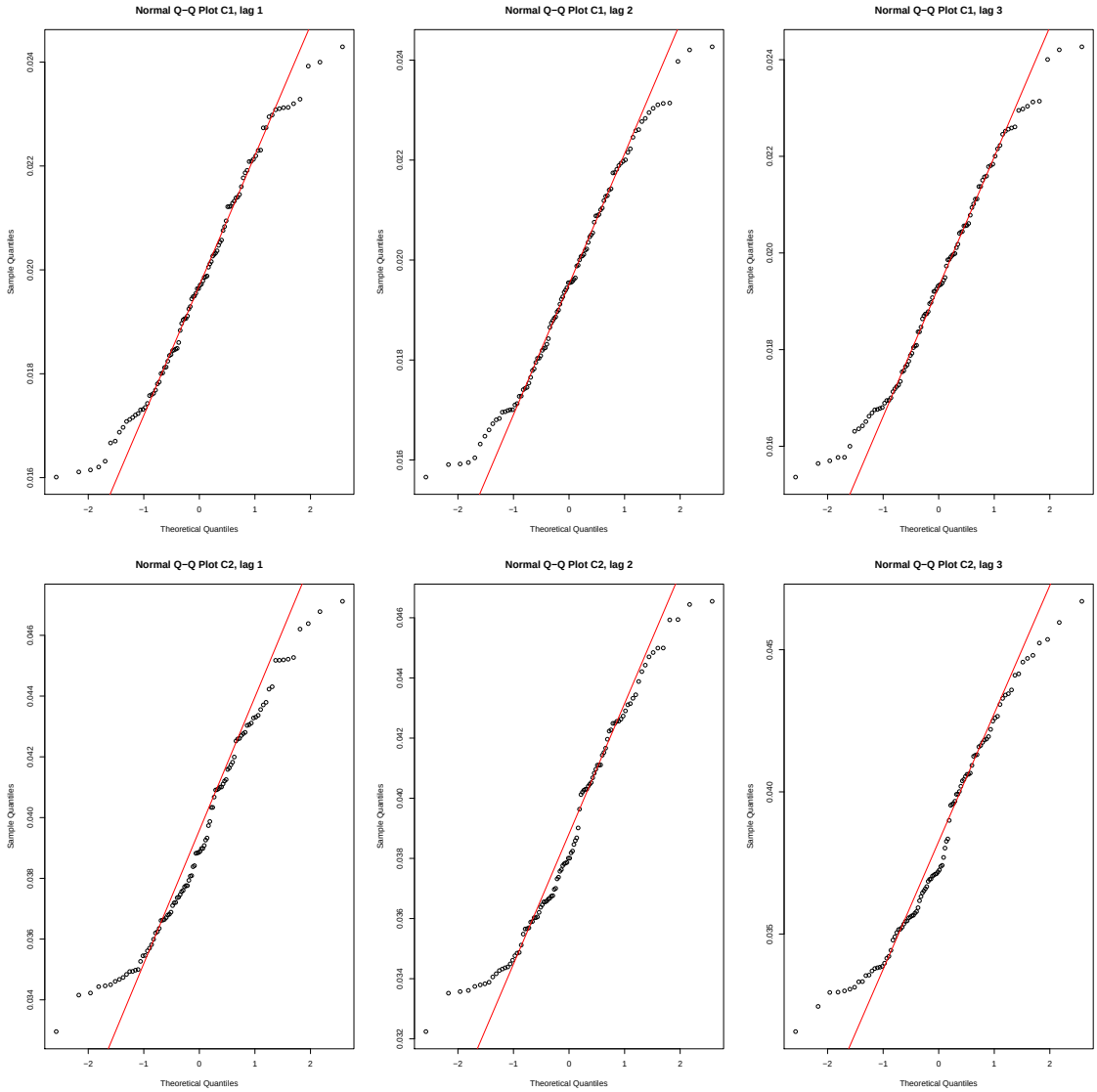

Figure 5.15: Simulation step $100 \times 100 \times 1$. Normal QQ-Plots of the estimates $\hat{C}_1$ and $\hat{C}_2$ for maximum lags 1 to 3.

Summary

We finish this chapter by summarizing the results of the different simulation steps, showing how the spatial parameter estimates $\hat{C}_1$ and $\hat{C}_2$ improve when enlarging the side length of the grid the processes are simulated on. We only compare the results of the estimation procedures based on a maximum spatial lag of value 1, since they seem better than those based on higher lags. Tables 5.6 and 5.7 show means and root mean squared errors of the estimates $\hat{C}_1$ and $\hat{C}_2$, respectively. The trend towards the true parameter values $C_1^\star = 0.02$ and $C_2^\star = 0.04$ is recognizable; in particular, the variance definitely decreases with increasing side length.

side length	10	70	80	90	100
mean	0.1343	0.0189	0.0190	0.0191	0.0198
RMSE	0.1156	0.0041	0.0032	0.0031	0.0021

Table 5.6: Means and RMSE of the parameter estimate $\hat{C}_1$ for maximum lag 1 and different side lengths of the grid

side length	10	70	80	90	100
mean	0.1072	0.0375	0.0373	0.0381	0.0394
RMSE	0.0689	0.0080	0.0063	0.0061	0.0036

Table 5.7: Means and RMSE of the parameter estimate $\hat{C}_2$ for maximum lags 1 and different side lengths of the grid

To confirm those statements, Figure 5.16 presents the different boxplots of the two estimates for side lengths 70-100. One can see that, with increasing side length, the green lines (the true parameter values) move more and more inside the boxes. Moreover, the boxes get smaller, indicating a decreasing variance.

In the next chapter, the pairwise likelihood method is applied to real data, stemming from rain measurements in Florida, USA.

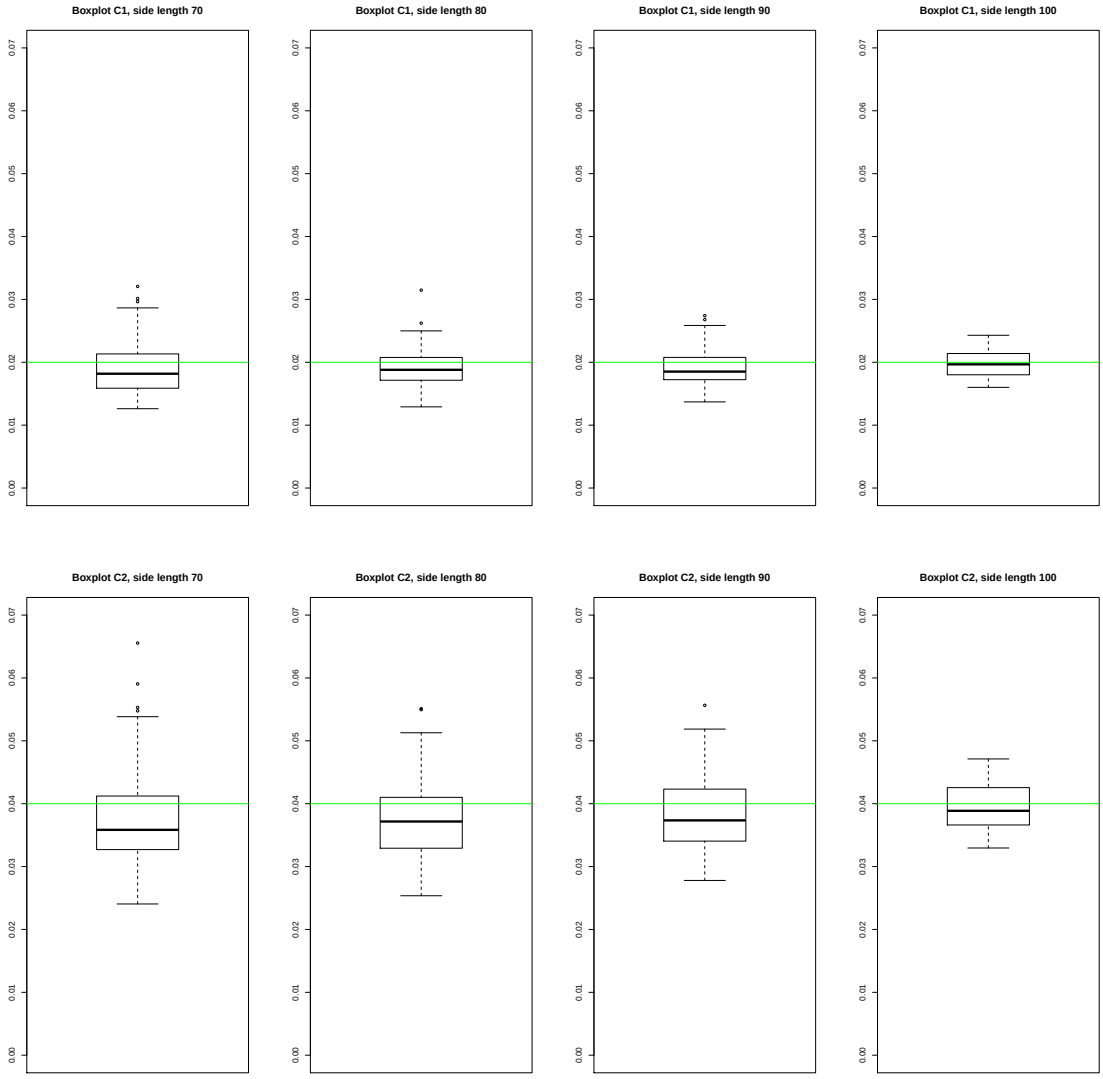

Figure 5.16: Boxplots of the parameter estimates $\hat{C}_1$ and $\hat{C}_2$ for maximum lags 1 and different side lengths of the grid

Chapter 6

Real Data Application

In this chapter we apply the pairwise likelihood estimation method, which was introduced in Chapter 4 and tested within the framework of a simulation study in Chapter 5, to real data. The data set we use in the following is an edited version of a rainfall data base consisting of radar values measured in inches every two kilometres and every 15 minutes in some part of Florida (USA) from 1999-2004. The same data set is used by Steinkohl [2012], Chapter 7, who adapts the set by choosing a square of 120km×120km in the middle of Florida, completely dividing it regularly into smaller squares of size 10km×10km and finally calculating the daily spatial maximum and the cumulative hourly values in every square, ending up with a 12×12 grid for all time points. The time points equal either the 732 days of the six wet seasons June-September 1999-2004 (in case we consider daily maxima) or the 2928 hours of the period June-September 1999 (if we deal with hourly rainfall measurements). Figure 6.1 illustrates the location of the 12×12 grid. In what follows, we analyse the two adapted versions of the data set, which we denote by

$$\left\{\tilde{\tilde{\eta}}_{(i_1,i_2),t} : i_1, i_2 = 1, \ldots, 12, t = 1, \ldots, 732\right\}$$

and

$$\left\{\tilde{\tilde{Z}}_{(i_1,i_2),t} : i_1, i_2 = 1, \ldots, 12, t = 1, \ldots, 2928\right\},$$

respectively. We proceed similarly as Steinkohl [2012], who assumes isotropy in the behaviour of rainfall. The special interest lies in the question if the estimation results change significantly if we allow anisotropy.

6.1 Fitting and Transformation of the Data Set

Daily Maxima

In general, taking maxima in space and time is reasonable since we deal with extreme values. However, before further analysis of the data set described above is possible, we need to achieve some transformations, which can be found in Steinkohl [2012], Section 7.2. For completeness, we describe them here as well.

(i) **Deseasonalizing the data.**

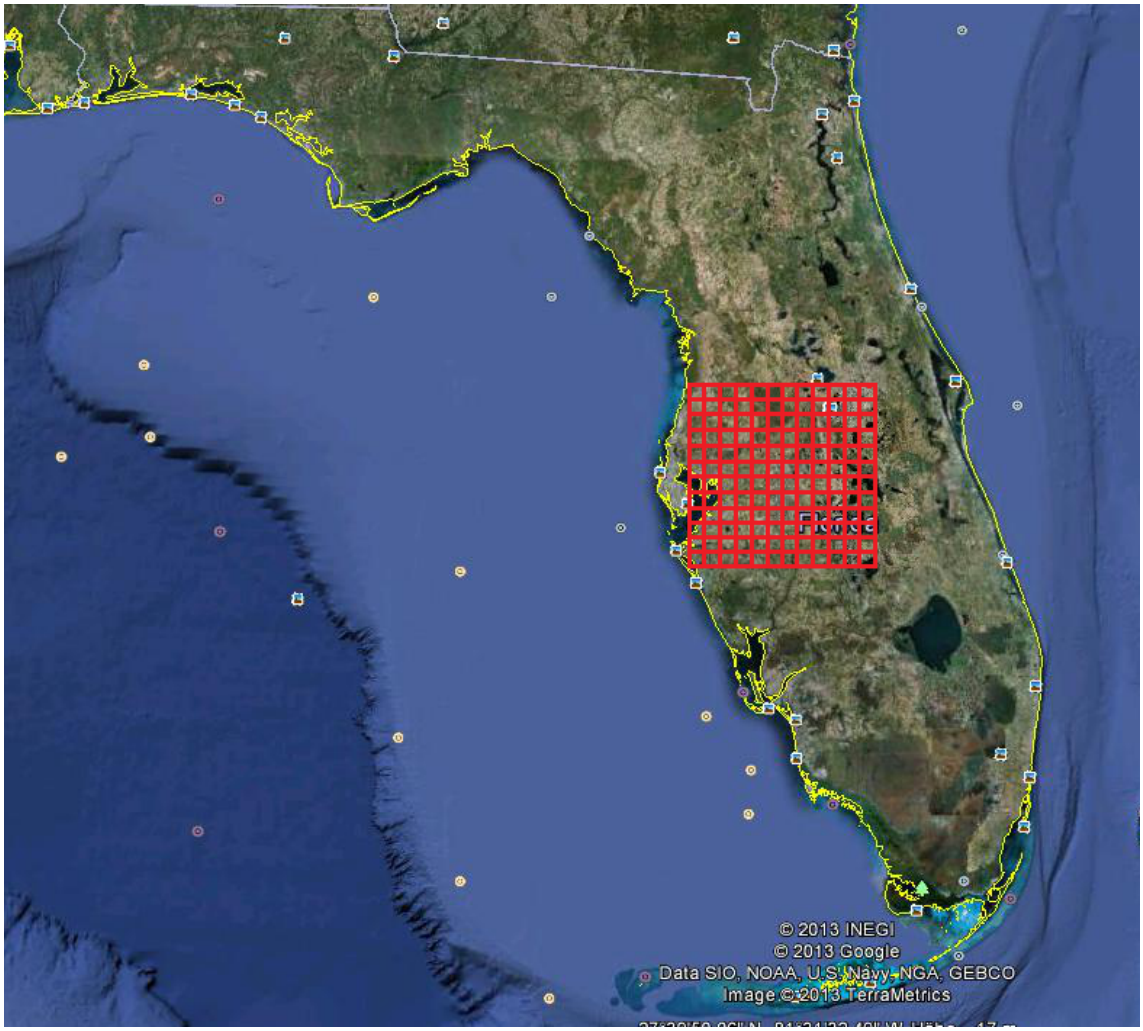

Figure 6.1: Map of Florida. The daily maxima are taken over the little red squares, forming the 12×12 grid our further analysis is dedicated to.

We use a simple moving average to get rid of seasonal effects. One period consists of a total wet season, i.e. of 122 days. For each fixed location $(i_1, i_2) \in \{1, \dots 12\}^2$ set

$$\tilde{\eta}_{(i_1,i_2),t_{k+122(j-1)}} := \tilde{\tilde{\eta}}_{(i_1,i_2),t_{k+122(j-1)}} - \frac{1}{6}\sum_{j=1}^{6} \tilde{\tilde{\eta}}_{(i_1,i_2),t_{k+122(j-1)}}, \quad k = 1, \dots, 122.$$

Examining autocorrelation plots, Steinkohl [2012] found that there is no temporal dependence in the time series $\{\tilde{\eta}_{(i_1,i_2),t} : t = 1, \dots, 732\}$ for a fixed location (i_1, i_2).

(ii) **Fitting a generalized extreme value distribution.**

For each fixed location (i_1, i_2), we fit a generalized extreme value distribution $G_{\gamma,\mu,\sigma}$ (see (2.12)) to the new data $\left\{\tilde{\eta}_{(i_1,i_2),t} : i_1, i_2 = 1, \dots, 12, t = 1, \dots, 732\right\}$, where $\gamma = \gamma(i_1, i_2)$, $\mu = \mu(i_1, i_2) \in \mathbb{R}$ and $\sigma = \sigma(i_1, i_2) > 0$, obtaining estimates $\hat{\gamma} = \hat{\gamma}(i_1, i_2)$, $\hat{\mu} = \hat{\mu}(i_1, i_2)$ and $\hat{\sigma} = \sigma(i_1, i_2)$. Statistical analysis of Steinkohl [2012] suggests to fit a Gumbel distribution, that is, we let $\gamma = 0$. Recall that

$$G_{0,\mu,\sigma}(x) = \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, \quad x \in \mathbb{R}.$$

(iii) **Transformation to Standard Fréchet margins.**

By the probability integral transform method, set for each fixed location (i_1, i_2),

$$\eta_{(i_1,i_2),t} := -\frac{1}{\log\left\{G_{0,\hat{\mu},\hat{\sigma}}(\tilde{\eta}_{(i_1,i_2),t})\right\}}, \quad t = 1, \dots, 732.$$

Hourly Measurements

Similar transformations need to be achieved when dealing with hourly measurements.

(i) **Deseasonalizing the data.**

We use a simple moving average to get rid of seasonal effects. One period consists of one day, i.e. of 24 hours. For each fixed location $(i_1, i_2) \in \{1, \dots 12\}^2$ set

$$\tilde{Z}_{(i_1,i_2),t_{k+24(j-1)}} := \tilde{\tilde{Z}}_{(i_1,i_2),t_{k+24(j-1)}} - \frac{1}{122}\sum_{j=1}^{122} \tilde{\tilde{Z}}_{(i_1,i_2),t_{k+24(j-1)}}, \quad k = 1, \dots, 24.$$

(ii) **Fitting a generalized Pareto distribution.**

Examining autocorrelation plots, Steinkohl [2012] found that, for a fixed location (i_1, i_2), the time series $\{\tilde{\eta}_{(i_1,i_2),t} : t = 1, \dots, 2928\}$ is stationary with a short-rate temporal dependence and can therefore be treated as an IID sequence, since it satisfies additional mixing conditions (see Leadbetter [1974] and Leadbetter [1983]).

For each fixed location (i_1, i_2), we fit a generalized Pareto distribution $\mathrm{GPD}_{\gamma,\beta}$ with $\gamma = \gamma(i_1, i_2) \in \mathbb{R}$ and $\beta = \beta(i_1, i_2) > 0$ using the *Peaks over Threshold* method with a threshold $v = 0.2$ (see Section 2.1.3) to the new data

$$\left\{\tilde{Z}_{(i_1,i_2),t} : i_1, i_2 = 1, \ldots, 12, t = 1, \ldots, 2928\right\},$$

obtaining estimates $\hat{\gamma} = \hat{\gamma}(i_1, i_2)$ and $\hat{\beta} = \hat{\beta}(i_1, i_2)$. Recall that

$$\mathrm{GPD}_{\gamma,\beta}(x) = 1 - \begin{cases} \left(1 + \gamma\frac{x}{\beta}\right)^{-\frac{1}{\gamma}}, & \text{if } \gamma \neq 0, \\ \exp\left\{-\frac{x}{\beta}\right\}, & \text{if } \gamma = 0. \end{cases}$$

(iii) **Transformation to Standard Fréchet margins.**

For each fixed location (i_1, i_2) set

$$Z_{(i_1,i_2),t} := -\frac{1}{\log\left\{\hat{F}_{(i_1,i_2)}(\tilde{Z}_{(i_1,i_2),t})\right\}}, \quad t = 1, \ldots, 2928,$$

where

$$\hat{F}_{(i_1,i_2)}(x) = \begin{cases} 1 - \frac{N_v(i_1,i_2)}{2928}(1 - \mathrm{GPD}_{\hat{\gamma},\hat{\beta}}(x - v)), & \text{if } x > v, \\ \frac{1}{2928}\sum_{t=1}^{2928} \mathbf{1}_{\{\tilde{Z}_{(i_1,i_2),t} \leq x\}}, & \text{if } x \leq v, \end{cases} \tag{6.1}$$

with $N_v(i_1, i_2) = \sum_{t=1}^{2928} \mathbf{1}_{\{\tilde{Z}_{(i_1,i_2),t} > v\}}$.

6.2 Pairwise Likelihood Estimation

Daily Maxima

We want to apply the pairwise likelihood method to the transformed data set

$$\left\{\eta_{(i_1,i_2),t} : i_1, i_2 = 1, \ldots, 12, t = 1, \ldots, 732\right\}.$$

We assume that the observations are realizations from a Brown-Resnick process $\boldsymbol{\eta} = \left\{\eta_{(i_1,i_2),t} : i_1, i_2 \in \mathbb{R}, t \in [0, \infty)\right\}$ like in Theorem 3.5 where the function δ satisfies

$$\delta(h_1, h_2, u) = C_1|h_1|^{\alpha_1} + C_2|h_2|^{\alpha_2} + C_3|u|^{\alpha_3},$$

with $h_1 = i_1^{(1)} - i_1^{(2)}$, $h_2 = i_2^{(1)} - i_2^{(2)}$, $u = t_1 - t_2$, for two spatial locations $(i_1^{(1)}, i_2^{(1)})$ and $(i_1^{(2)}, i_2^{(2)})$ and two time points t_1 and t_2. Our goal is to estimate the parameters $C_1, C_2, C_3 \in (c, \infty)$ for some $c > 0$ and $\alpha_1, \alpha_2, \alpha_3 \in (0, 2]$. We did the estimation procedure for maximum spatial and temporal lags 1 to 3, estimating the respective parameters separately, meaning that if one lag is allowed to take positive values, the others are set

equal to 0. This advancement is justified by Davis et al. [2012a], Chapter 6, who state that the parameters of the underlying space-time correlation function get separated in the extremal setting. That way, if we set one maximum spatial lag or the maximum temporal lag equal to 0, it does not affect the respective other parameters. Tables 6.1 and 6.2 present the results. The notation (a_1, a_2, a_3) for $a_i \in \{0, 1, 2, 3\}, i = 1, \ldots, 3$, relates to estimates of the parameters C_{a_j} or α_{a_j}, where j is the index of the number a_j that is not equal to 0.

max. lags	(1,0,0)	(2,0,0)	(3,0,0)	(0,1,0)	(0,2,0)	(0,3,0)	(0,0,1)	(0,0,2)	(0,0,3)
estimated value	0.3143	0.3143	0.3179	0.3636	0.3636	0.3685	2.4189	2.4189	2.4353

Table 6.1: Estimates of the parameters C_1, C_2 and C_3 for different maximum spatial and temporal lags, respectively.

max. lags	(2,0,0)	(3,0,0)	(0,2,0)	(0,3,0)	(0,0,2)	(0,0,3)
estimated value	0.9437	0.8599	0.9517	0.8521	0.1981	0.1561

Table 6.2: Estimates of the parameters α_1, α_2 and α_3 for different maximum spatial and temporal lags, respectively.

As one would expect, the estimates of the temporal parameters C_3 and α_3 above and the corresponding estimates of Steinkohl [2012] are nearly identical, since in this thesis, we only changed assumptions concerning the spatial parameters C_1, C_2, α_1 and α_2. Their estimates, however, are at least of the same range as the spatial estimates of Steinkohl [2012]. The combination of a rather large parameter estimate $\hat{C}_3$ and a rather small estimate $\hat{\alpha}_3$ implies extremal temporal independence. For a brief explanation based on the so-called *extremal extremogram* (see Davis and Mikosch [2009] and Steinkohl [2012], Definition 5.1), we refer to Steinkohl [2012].

Hourly Measurements

We now apply the pairwise likelihood method to the transformed data set of hourly measurements $\left\{Z_{(i_1,i_2),t} : i_1, i_2 = 1, \ldots, 12, t = 1, \ldots, 2928\right\}$. As for the daily maxima, we assume that the observations are realizations from a Brown-Resnick process and our goal is to estimate the parameters C_1, C_2, C_3 and $\alpha_1, \alpha_2, \alpha_3$ of the function δ. We did the estimation procedure for maximum spatial and temporal lags 1 to 3, again estimating the respective parameters separately. Tables 6.3 and 6.4 present the results. The notation (a_1, a_2, a_3) for $a_i \in \{0, 1, 2, 3\}, i = 1, \ldots, 3$, relates to estimates of the parameters C_{a_j} or α_{a_j}, where j is the index of the number a_j that is not equal to 0.

max. lags	(1,0,0)	(2,0,0)	(3,0,0)	(0,1,0)	(0,2,0)	(0,3,0)	(0,0,1)	(0,0,2)	(0,0,3)
estimated value	0.1538	0.1537	0.1559	0.1501	0.1501	0.1517	0.3532	0.3532	0.3505

Table 6.3: Estimates of the parameters C_1, C_2 and C_3 for different maximum spatial and temporal lags, respectively.

max. lags	(2,0,0)	(3,0,0)	(0,2,0)	(0,3,0)	(0,0,2)	(0,0,3)
estimated value	0.9630	0.8704	0.9021	0.8311	1.0628	1.1252

Table 6.4: Estimates of the parameters α_1, α_2 and α_3 for different maximum spatial and temporal lags, respectively.

The temporal parameter estimates of C_3 and α_3 now clearly indicate dependence, since they reveal the combination of a small C_3 and a large α_3. Normally, scaling of the time domain does not influence the spatial parameters (for an explanation based on the extremal extremogram we refer to Steinkohl [2012], Section 7.3). So the estimated values of C_1 and C_2 based on hourly measurements should more or less accord with those based on daily maxima. A possible explanation for the fact that this is not the case here might be that the hourly measurements have not yet reached the Fréchet limit, i.e. they are maybe not "extreme" enough.

6.3 Conditional Probability Fields and Return Level Maps

Based on the estimation results in Section 6.2, we want to "update" the *conditional probability fields* and *return level plots* of Steinkohl [2012], Section 7.2. We consider again both daily maxima and hourly measurements.

Daily Maxima

For some space-time point $(i_1, i_2, t) \in \{1, \dots, 12\}^2 \times \{1, \dots, 732\}$ and some non-negative rain level z_c we consider probabilities of the form

$$\mathbb{P}\left(\tilde{\eta}_{(i_1,i_2),t} > z_c | \tilde{\eta}_{(i_1^\star,i_2^\star),t^\star} > z_c\right),$$

where $(i_1^\star, i_2^\star, t^\star)$ and $z^\star$ are the reference point and the reference level, respectively. The process $\left\{\tilde{\eta}_{(i_1,i_2),t} : i_1, i_2 = 1, \dots, 12, t = 1, \dots, 732\right\}$ is the deseasonalized process of step (1) above and assumed to be realizations of a Brown-Resnick process with marginal Gumbel

distribution $G_{0,\mu(i_1,i_2),\sigma(i_1,i_2)}$. Setting $p^\star := \mathbb{P}\left(\tilde{\eta}_{(i_1^\star,i_2^\star),t^\star} > z^\star\right)$, we find

$$
\begin{aligned}
&\mathbb{P}\left(\tilde{\eta}_{(i_1,i_2),t} > z_c | \tilde{\eta}_{(i_1^\star,i_2^\star),t^\star} > z^\star\right) \\
&= \frac{1}{p^\star}\mathbb{P}\left(\tilde{\eta}_{(i_1,i_2),t} > z_c, \tilde{\eta}_{(i_1^\star,i_2^\star),t^\star} > z^\star\right) \\
&= \frac{1}{p^\star}\big[1 - \mathbb{P}\left(\tilde{\eta}_{(i_1,i_2),t} \le z_c, \tilde{\eta}_{(i_1^\star,i_2^\star),t^\star} \le z^\star\right) - \mathbb{P}\left(\tilde{\eta}_{(i_1,i_2),t} > z_c, \tilde{\eta}_{(i_1^\star,i_2^\star),t^\star} \le z^\star\right) \\
&\quad - \mathbb{P}\left(\tilde{\eta}_{(i_1,i_2),t} \le z_c, \tilde{\eta}_{(i_1^\star,i_2^\star),t^\star} > z^\star\right)\big] \\
&= \frac{1}{p^\star}\big[1 - \mathbb{P}\left(\tilde{\eta}_{(i_1,i_2),t} \le z_c\right) - \mathbb{P}\left(\tilde{\eta}_{(i_1^\star,i_2^\star),t^\star} \le z^\star\right) + \mathbb{P}\left(\tilde{\eta}_{(i_1,i_2),t} \le z_c, \tilde{\eta}_{(i_1^\star,i_2^\star),t^\star} \le z^\star\right)\big] \\
&= \frac{1}{p^\star}\big[1 - G_{0,\mu(i_1,i_2),\sigma(i_1,i_2)}(z_c) - (1-p^\star) + \mathbb{P}\left(\tilde{\eta}_{(i_1,i_2),t} \le z_c, \tilde{\eta}_{(i_1^\star,i_2^\star),t^\star} \le z^\star\right)\big] \\
&= 1 - \frac{1}{p^\star}G_{0,\mu(i_1,i_2),\sigma(i_1,i_2)}(z_c) \\
&\quad + \frac{1}{p^\star}\mathbb{P}\bigg(- \frac{1}{\log\left\{G_{0,\mu(i_1,i_2),\sigma(i_1,i_2)}(\tilde{\eta}_{(i_1,i_2),t})\right\}} \le -\frac{1}{\log\left\{G_{0,\mu(i_1,i_2),\sigma(i_1,i_2)}(z_c)\right\}}, \\
&\qquad\qquad - \frac{1}{\log\left\{G_{0,\mu(i_1,i_2),\sigma(i_1,i_2)}(\tilde{\eta}_{(i_1,i_2),t})\right\}} \le -\frac{1}{\log\{1-p^\star\}}\bigg) \\
&= 1 - \frac{1}{p^\star}G_{0,\mu(i_1,i_2),\sigma(i_1,i_2)}(z_c) + \frac{1}{p^\star}G\bigg(- \frac{1}{\log\left\{G_{0,\mu(i_1,i_2),\sigma(i_1,i_2)}(z_c)\right\}}, -\frac{1}{\log\{1-p^\star\}}\bigg),
\end{aligned}
$$

where G denotes the bivariate distribution function of the Brown-Resnick process given in (3.9). To estimate the conditional probabilities, we use the estimates $\hat{\mu}(i_1,i_2)$ and $\hat{\sigma}(i_1,i_2)$ of the Gumbel distribution parameters and the parameter estimates $\hat{C}_1$, $\hat{C}_2$, $\hat{C}_3$, $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ for the function G. Figure 6.2 shows four predicted conditional probability fields for the reference points $(1,1,1)$, $(5,6,1)$, $(8,10,1)$ and $(10,7,1)$, i.e. we choose the same as Steinkohl [2012], Section 7.2. Because of temporal independence (see above), we keep the time point fixed. Clearly, there are only little changes, because the spatial parameter estimates only slightly differ and they are more or less of the same range as the isotropic spatial parameter estimate of Steinkohl [2012]. Basically the same statement holds for the return level plots in Figure 6.3 for the same reference locations: For given p_c and $p^\star \in [0,1]$ (in Figure 6.3: $p_c = p^\star = 0.01$), the *conditional return level* z_c with return period $\frac{1}{p_c}$ is obtained by solving for it in the equation

$$\mathbb{P}\left(\tilde{\eta}_{(i_1,i_2),t} > z_c | \tilde{\eta}_{(i_1^\star,i_2^\star),t^\star} > z^\star\right) = p_c,$$

with $z^\star$ such that $\mathbb{P}\left(\tilde{\eta}_{(i_1^\star,i_2^\star),t^\star} > z^\star\right) = p^\star$. To estimate the return levels, we use again the estimated versions of the parameters included in its calculation.

Hourly Measurements

For some space-time point $(i_1,i_2,t) \in \{1,\dots,12\}^2 \times \{1,\dots,2928\}$ and some non-negative

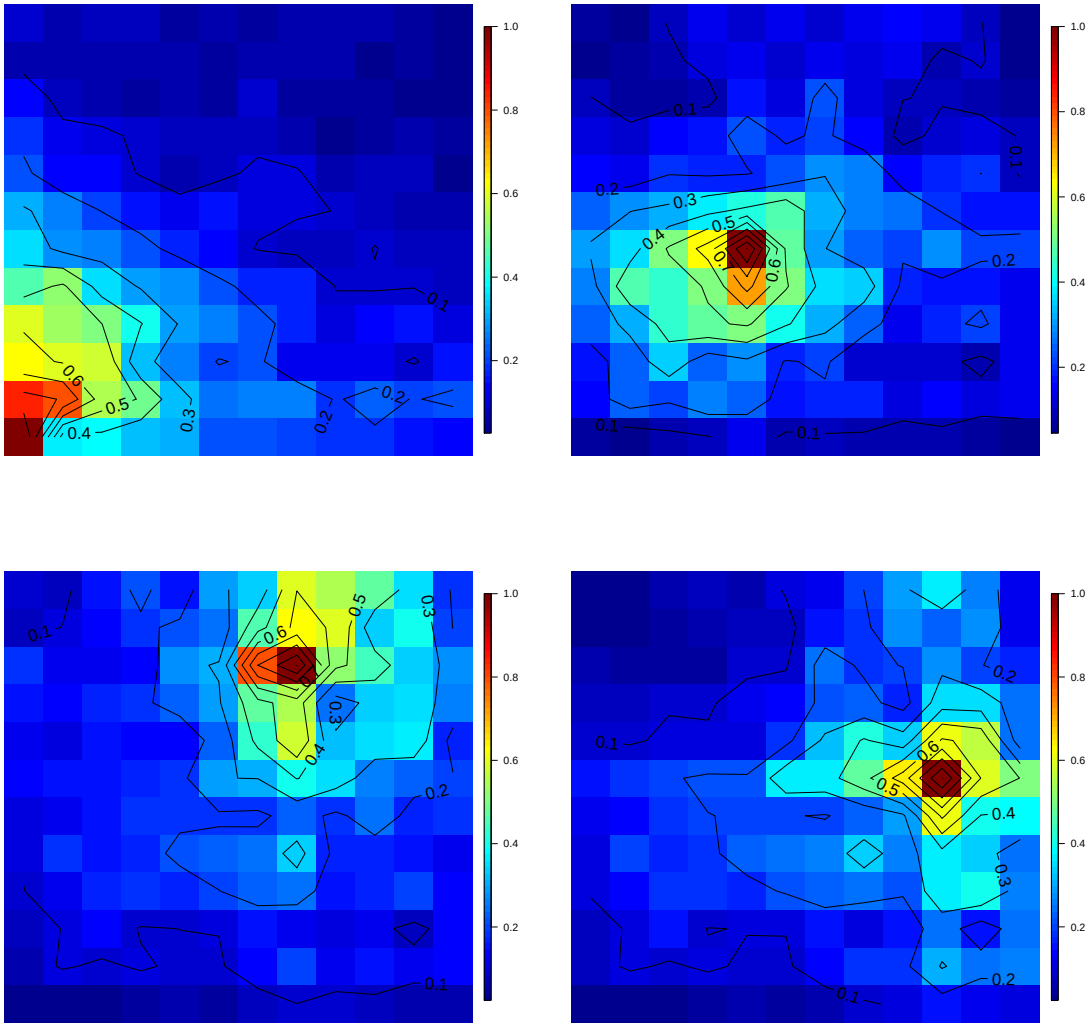

Figure 6.2: Predicted conditional probability fields based on daily maxima for reference points (1,1,1), (5,6,1), (8,10,1) and (10,7,1) (form the top left to the bottom right).

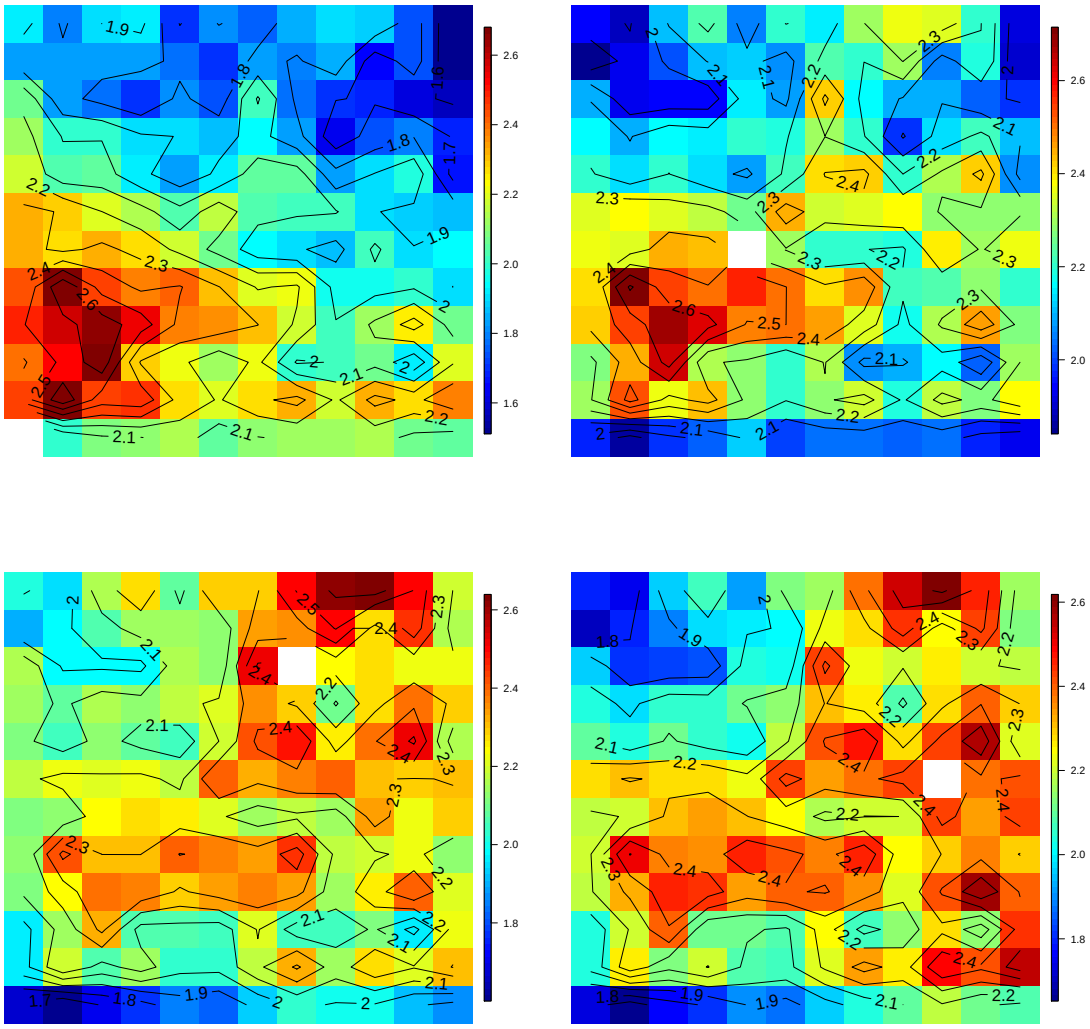

Figure 6.3: Conditional 100-day-return level plots based on daily maxima for reference points (white squares) (1,1,1), (5,6,1), (8,10,1) and (10,7,1) (form the top left to the bottom right).

rain level z_c we consider analogously as in the case of daily maxima conditional probabilities of the form

$$\mathbb{P}\left(\tilde{Z}_{(i_1,i_2),t} > z_c | \tilde{Z}_{(i_1^\star,i_2^\star),t^\star} > z_c\right)$$

and predict them by

$$1 - \frac{1}{p^\star}\hat{F}_{(i_1,i_2)}(z_c) + \frac{1}{p^\star}\hat{G}\bigg(- \frac{1}{\log\left\{\hat{F}_{(i_1,i_2)}(z_c)\right\}}, -\frac{1}{\log\{1-p^\star\}}\bigg),$$

where $(i_1^\star, i_2^\star), t^\star$ and $z^\star$ are the reference point and the reference level, respectively, $p^\star = \mathbb{P}\left(\tilde{Z}_{(i_1^\star,i_2^\star),t^\star} > z^\star\right)$, $\hat{F}_{(i_1,i_2)}(\cdot)$ is given in (6.1) and $\hat{G}(\cdot)$ is the estimate of the bivariate distribution function of the Brown-Resnick process (cf. (3.9)). On top of that, we estimate the conditional return levels z_c obtained by solving for it in the equations

$$\mathbb{P}\left(\tilde{Z}_{(i_1,i_2),t} > z_c | \tilde{Z}_{(i_1^\star,i_2^\star),t^\star} > z_c\right) = p_c$$

for some $p_c \in [0,1]$. Figure 6.4 shows predicted conditional probability fields for the reference locations $(5,6)$ and $(10,7)$. Figure 6.5 shows predicted conditional return level plots for the reference locations $(1,1)$, $(5,6)$ and $(10,7)$. Both plots visualize four different time points in a row. Again we choose the same locations as Steinkohl [2012], Section 7.2.

In the following, we consider the natural question that arose in the previous sections: Can we reject the assumption of isotropy? In other words: Are the parameters C_1 and C_2 or α_1 and α_2 significantly different? The next section deals with that question. However, we only consider the estimation results based on daily maxima because of the remarks at the end of Section 6.2.

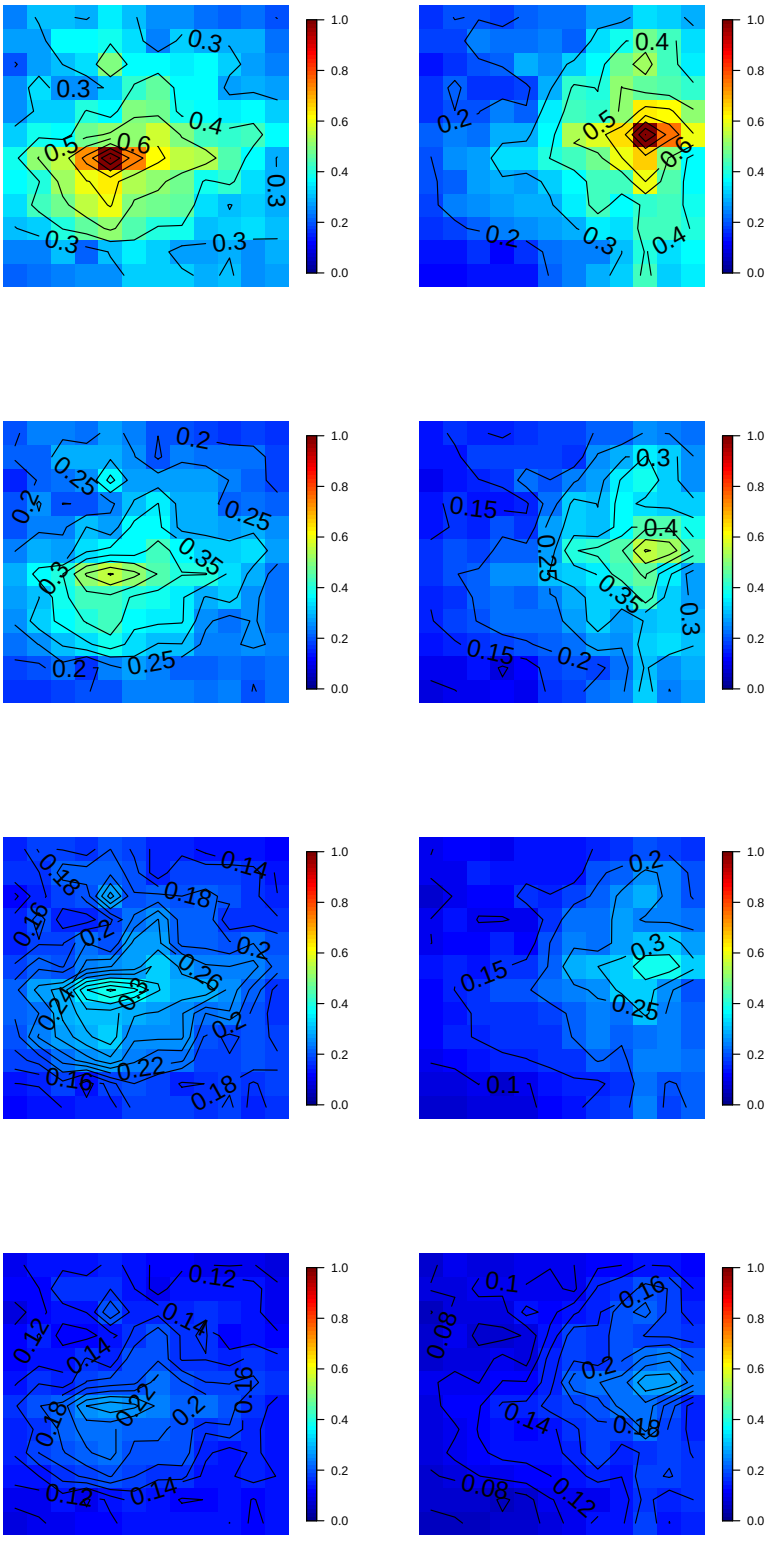

Figure 6.4: Predicted conditional probability fields based on hourly measurement for reference locations (5,6) (left row) and (10,7) (right row) for four consecutive time points (from top to bottom).

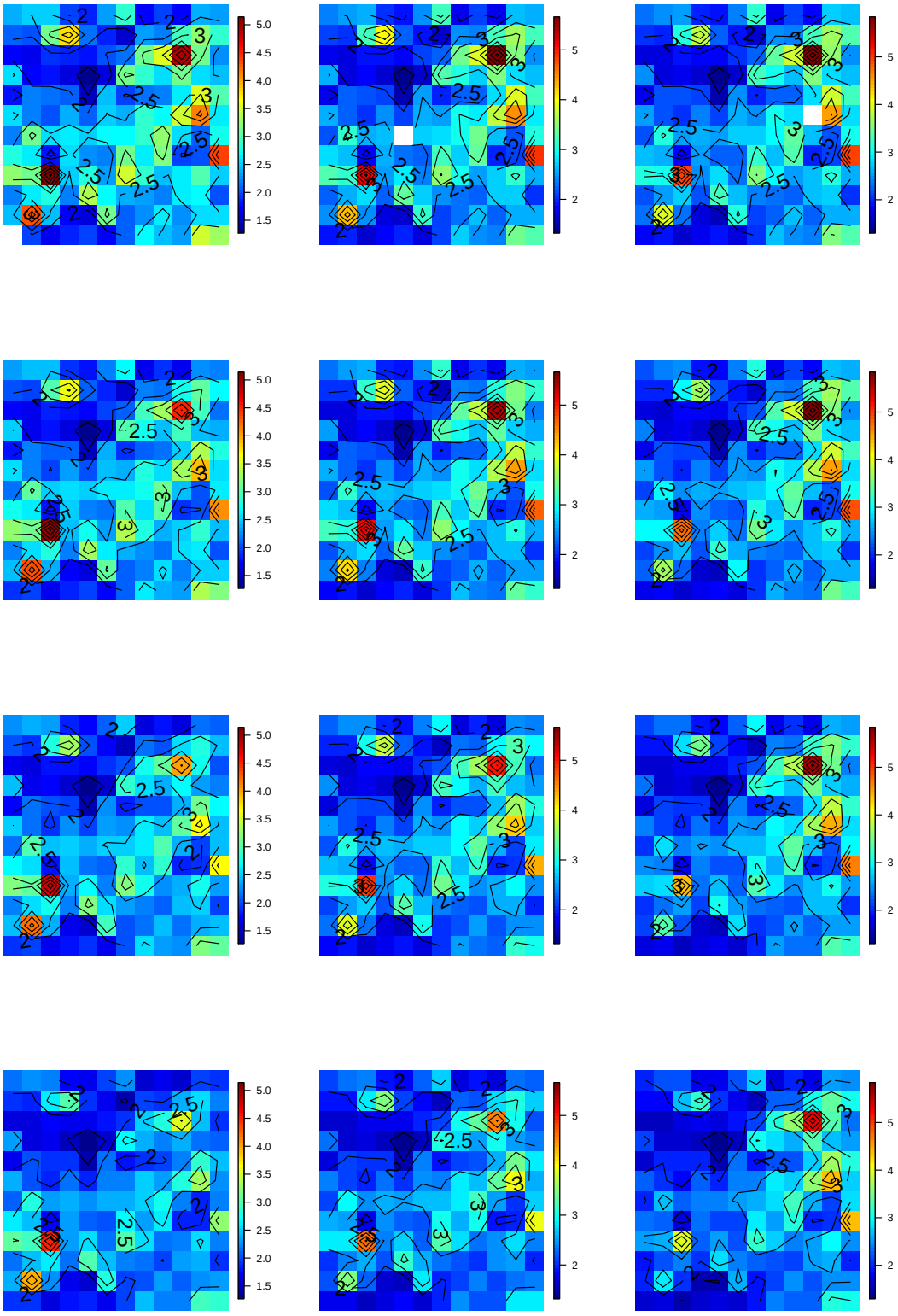

Figure 6.5: Predicted conditional return level plots based on hourly measurement for reference locations (1,1) (left row), (5,6) (middle row) and (10,7) (right row) for four consecutive time points (from top to bottom).

6.4 A Hypothesis Test: Can we reject isotropy of the rainfall weather fronts?

From Chapter 4.2.3 we know that asymptotic normality holds for the pairwise likelihood estimate

$$\hat{\boldsymbol{\theta}} = (\hat{C}_1, \hat{C}_2, \hat{C}_3, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3) = \hat{\boldsymbol{\theta}}^{(M,T)} = (\hat{C}_1^{(M,T)}, \hat{C}_2^{(M,T)}, \hat{C}_3^{(M,T)}, \hat{\alpha}_1^{(M,T)}, \hat{\alpha}_2^{(M,T)}, \hat{\alpha}_3^{(M,T)})$$

in the following sense:

$$M^{\frac{d}{2}}\sqrt{T}\left[\begin{pmatrix}\hat{C}_1 - C_1\\ \hat{C}_2 - C_2\\ \hat{C}_3 - C_3\\ \hat{\alpha}_1 - \alpha_1\\ \hat{\alpha}_2 - \alpha_2\\ \hat{\alpha}_3 - \alpha_3\end{pmatrix}\right] \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \tilde{\Sigma}), \qquad \text{as } M, T \to \infty, \tag{6.2}$$

where $\tilde{\Sigma} \in \mathbb{R}^{6\times 6}$ is the asymptotic covariance matrix. The quantities M and T stand for the side length of the regular grid of d dimensions and the number of time points, respectively.

Proposition 6.1 (Asymptotics)**.** *Set* $\tau_{M,T} := M^{\frac{d}{2}}\sqrt{T}$ *and define* $A_1 := (-1, 1, 0, 0, 0, 0)$ *and* $A_2 := (0, 0, 0, -1, 1, 0)$*. Then:*

$$\tau_{M,T}((\hat{C}_2 - \hat{C}_1) - (C_2 - C_1)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, A_1\tilde{\Sigma}A_1') \tag{6.3}$$

and

$$\tau_{M,T}((\hat{\alpha}_2 - \hat{\alpha}_1) - (\alpha_2 - \alpha_1)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, A_2\tilde{\Sigma}A_2'). \tag{6.4}$$

Proof. We only prove (6.3); (6.4) is proved analogously. Using the *continuous mapping theorem* and (6.2), we obtain

$$\begin{aligned}
&A_1\left\{M^{\frac{d}{2}}\sqrt{T}\left[\begin{pmatrix}\hat{C}_1 - C_1\\ \hat{C}_2 - C_2\\ \hat{C}_3 - C_3\\ \hat{\alpha}_1 - \alpha_1\\ \hat{\alpha}_2 - \alpha_2\\ \hat{\alpha}_3 - \alpha_3\end{pmatrix}\right]\right\}\\
&= M^{\frac{d}{2}}\sqrt{T}(C_1 - \hat{C}_1 - C_2 + \hat{C}_2)\\
&= M^{\frac{d}{2}}\sqrt{T}((\hat{C}_2 - \hat{C}_1) - (C_2 - C_1))\\
&= \tau_{M,T}((\hat{C}_2 - \hat{C}_1) - (C_2 - C_1)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, A_1\tilde{\Sigma}A_1').
\end{aligned}$$

□

Note and recall that $\hat{C}_1$, $\hat{C}_2$, $\hat{\alpha}_1$ and $\hat{\alpha}_2$ all depend on M and T. For convenience, we define $\theta_C := (C_2 - C_1)$, $\hat{\theta}_C^{(M,T)} := (\hat{C}_2 - \hat{C}_1)$, $\theta_\alpha := (\alpha_2 - \alpha_1)$ and $\hat{\theta}_\alpha := (\hat{\alpha}_2 - \hat{\alpha}_1)$. In the following, we test the null hypothesis

$$H_0 : \theta_C = \theta_\alpha = 0$$

versus the alternative hypothesis

$$H_A : \theta_C \neq 0 \text{ or } \theta_\alpha \neq 0.$$

Note that this test scheme is an example of multiple hypothesis testing, as we are about to test two theses simultaneously on the same data set. In order to come up with this, we use the *Bonferroni inequality*.

Proposition 6.2 (Bonferroni Inequality, cf. Lehmann and Romano [2005], Thm. 9.1.1)**.** *For $s \in \mathbb{N}$, let $\{H_k : k = 1, \dots, s\}$ be the set of hypotheses that are about to be tested simultaneously at some overall confidence level $\beta \in (0, 1)$. Further, let $\{H_i : i \in I\}$ for some $I \subset \{1, \dots, s\}$ be the set of the true hypotheses. Suppose that each hypothesis H_k is rejected if $\mathbb{P}(\text{reject } H_k | k \in I) \leq \beta/s$. Then*

$$\mathbb{P}(\text{reject any } H_k \text{ with } k \in I) \leq \beta,$$

i.e., the overall probability of a first type error is limited by β.

Proof. $\mathbb{P}(\text{reject any } H_k \text{ with } k \in I) \leq \sum_{i \in I} \mathbb{P}(\text{reject } H_i) \leq |I| \frac{\beta}{s} \leq \beta.$ □

As $\tilde{\Sigma}$ and therefore the limiting distributions of $\hat{\theta}_C^{(M,T)}$ and $\hat{\theta}_\alpha^{(M,T)}$ are not known in closed form, we need asymptotic confidence intervals to be able to test our two hypotheses. To achieve this, one can use subsampling methods. In the following, we use the method described by Politis, Romano and Wolf [1999], Chapter 5. However, they consider stationary random fields of the form $\left\{X_{\tilde{\boldsymbol{t}}} : \tilde{\boldsymbol{t}} \in \mathbb{G}^{\tilde{d}}\right\}$ where $\mathbb{G} \in \{\mathbb{R}, \mathbb{Z}\}$. To make the theory applicable to our setting and the Brown-Resnick process $\boldsymbol{\eta}$, we choose $\mathbb{G} = \mathbb{Z}$, $\tilde{d} = d + 1$ and write

$$\left\{\eta_{\boldsymbol{s},t} : \boldsymbol{s} \in \mathbb{N}^d, t \in \mathbb{N}\right\} =: \left\{\eta_{\tilde{\boldsymbol{t}}} : \tilde{\boldsymbol{t}} \in \mathbb{N}^{\tilde{d}}\right\},$$

where $\tilde{d} = d + 1$. For $\boldsymbol{m} = (m_1, \dots, m_{\tilde{d}}) \in \mathbb{N}^{\tilde{d}}$ define

$$\begin{aligned} E_{\boldsymbol{m}} :&= \left\{\tilde{\boldsymbol{t}} = (\tilde{t}_1, \dots, \tilde{t}_{\tilde{d}}) \in \mathbb{N}^{\tilde{d}} : \tilde{t}_1 \leq m_1, \dots, \tilde{t}_{\tilde{d}} \leq m_{\tilde{d}}\right\} \\ &= \left\{(\boldsymbol{s}, t) = (s_1, \dots, s_d, t) \in \mathbb{N}^d \times \mathbb{N} : s_1 \leq m_1, \dots, s_d \leq m_d, t \leq m_{d+1}\right\}. \end{aligned}$$

The set of observations is given by $\left\{\eta_{\tilde{\boldsymbol{t}}} : \tilde{\boldsymbol{t}} \in E_{\boldsymbol{n}}\right\}$ with $\boldsymbol{n} = (M, \dots, M, T)$. Given an estimate

$$\hat{\varphi} = \hat{\varphi}_{\boldsymbol{n}} = \hat{\varphi}_{\boldsymbol{n}}\left(\left\{\eta_{\tilde{\boldsymbol{t}}} : \tilde{\boldsymbol{t}} \in E_{\boldsymbol{n}}\right\}\right)$$

of some arbitrary univariate parameter $\varphi \in \mathbb{R}$, the goal of subsampling is to calculate asymptotically correct $(1 - \beta) \cdot 100\%-$confidence intervals for φ for some level β, e.g. $\beta = 0.05$. For an appropriate scaling sequence $(\zeta_{\boldsymbol{n}})$, let $J^{\boldsymbol{n}}$ denote the sampling probability law and $J^{\boldsymbol{n}}(\cdot)$ the corresponding distribution function of $\zeta_{\boldsymbol{n}}(\hat{\varphi}_{\boldsymbol{n}} - \varphi)$. The central and only requirement concerning $J^{\boldsymbol{n}}$ is the following:

Assumption 6.1 (cf. Politis et al. [1999], Assumption 5.3.1)**.** *The sampling distribution $J^{\boldsymbol{n}}$ converges weakly to a limit law J with corresponding distribution function $J(\cdot)$ as $n_j \to \infty$ for $j = 1, \ldots, \tilde{d}$.*

In our case, as $n_j = M$ for $j = 1, \ldots, d$, we can for notational reasons write $\zeta_{M,T}$ instead of $\zeta_{M,\ldots,M,T} = \zeta_{\boldsymbol{n}}$, $\varphi_{M,T}$ instead of $\varphi_{M,\ldots,M,T} = \varphi_{\boldsymbol{n}}$ and $J^{(M,T)}$ instead of $J^{(M,\ldots,M,T)} = J^{\boldsymbol{n}}$. We choose $\varphi = \theta_C$ or $\varphi = \theta_\alpha$, $\hat{\varphi}_{M,T} = \hat{\theta}_C^{(M,T)}$ or $\hat{\varphi}_{M,T} = \hat{\theta}_\alpha^{(M,T)}$ and $\zeta_{M,T} = M^{\frac{d}{2}}\sqrt{T} = \tau_{M,T}$. Then Assumption 6.1 is satisfied by Proposition 6.1. The following steps summarize the subsampling procedure described in Politis et al. [1999]. For notational ease and as the procedure is the same for the parameters θ_C and θ_α, we use the subscript "l" for either C or α.

Subsampling Procedure

(1) For $\boldsymbol{i} = (i_1, \ldots, i_{\tilde{d}})$, $\boldsymbol{b} = (b_1, \ldots, b_{\tilde{d}})$ and $\boldsymbol{e} = (e_1, \ldots, e_{\tilde{d}}) \in \mathbb{N}^{\tilde{d}}$ define the blocks $Y_{\boldsymbol{i}} := \left\{\eta_{\tilde{t}} : \tilde{t} \in E_{\boldsymbol{i},\boldsymbol{b},\boldsymbol{e}}\right\}$, where

$$E_{\boldsymbol{i},\boldsymbol{b},\boldsymbol{e}} = \left\{\boldsymbol{j} = (j_1, \ldots, j_{\tilde{d}}) \in \mathbb{N}^{\tilde{d}} : (i_k - 1)e_k < j_k \leq (i_k - 1)e_k + b_k, k = 1, \ldots, \tilde{d}.\right\}.$$

The vectors $\boldsymbol{b}$ and $\boldsymbol{e}$ determine the side lengths of the blocks and the degree of overlap, respectively. The total number of blocks available is $q = \prod_{j=1}^{\tilde{d}} q_j$, where $q_j = \lfloor \frac{n_j - b_j}{e_j} \rfloor + 1$ for $j = 1, \ldots, \tilde{d}$. Recall from Section 6.1, that within our context of application, we have $d = 2$ (i.e. $\tilde{d} = 3$), $n_1 = n_2 = M = 12$ and $n_3 = T = 732$. Since throughout this thesis, we only deal with regular grids we choose $b_1 = b_2 = 5$. Concerning the number of time points in each subsample, we take $b_3 = 600$. In order to obtain a large number of subsamples, we further choose $e_1 = e_2 = e_3 = 1$, the maximum degree of overlap. The total number of blocks is therefore given by $8 \cdot 8 \cdot 133 = 8512$. Figure 6.6 visualizes this procedure.

(2) Obtain q different estimates $\hat{\theta}_{l,\boldsymbol{b},\boldsymbol{i}}^{(M,T)} = \hat{\theta}_{l,\boldsymbol{b}}^{(M,T)}(Y_{\boldsymbol{i}})$ by estimating the parameter θ_l on the q different block subsamples $Y_{\boldsymbol{i}}$, $i_1 = 1, \ldots, q_1, \ldots, i_{\tilde{d}} = 1, \ldots, q_{\tilde{d}}$.

(3) Define the subsampling approximation to $J^{(M,T)}(x)$ by

$$L_{\boldsymbol{b},l}^{(M,T)}(x) := \frac{1}{q}\sum_{i_1=1}^{q_1} \cdots \sum_{i_{\tilde{d}}=1}^{q_{\tilde{d}}} \mathbf{1}_{\left\{\tau_{\boldsymbol{b}}\left(\hat{\theta}_{l,\boldsymbol{b},\boldsymbol{i}}^{(M,T)} - \hat{\theta}_l^{(M,T)}\right) \leq x\right\}}, \tag{6.5}$$

where $\tau_{\boldsymbol{b}}$ is the analogue of $\tau_{M,T}$ in terms of the smaller subsample. That is, in the special setting described at the end of (1), this yields

$$\tau_{\boldsymbol{b}} = b_1^{\frac{d}{2}}\sqrt{b_{\tilde{d}}} = b_1\sqrt{b_3} = 5\sqrt{600}.$$

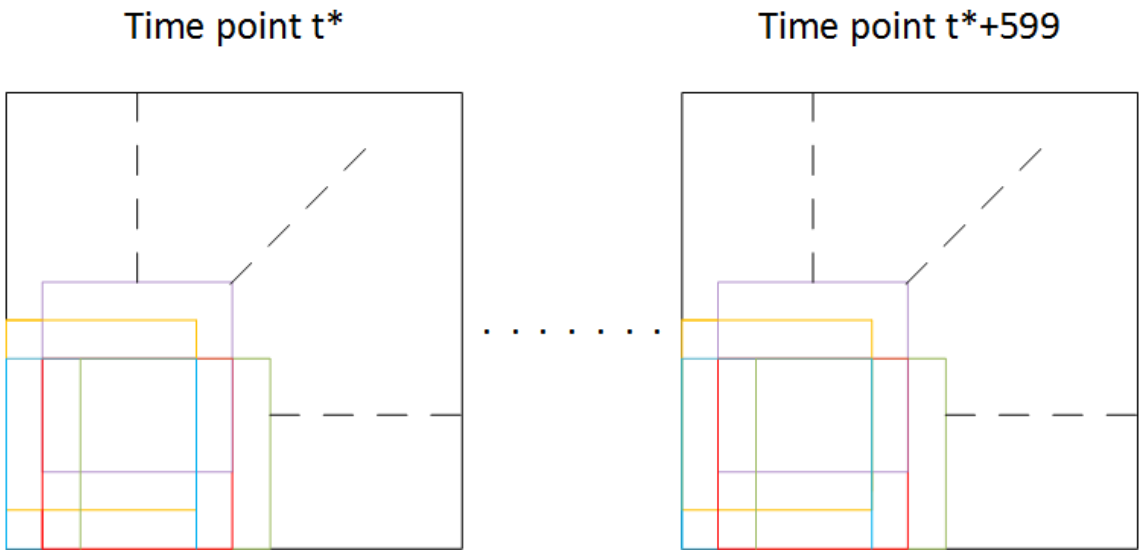

Figure 6.6: The grid of observations and the time line are divided into smaller subsamples. The values of the structure determining vectors are chosen as $\boldsymbol{b} = (5, 5, 600)$ and $\boldsymbol{e} = (1, 1, 1)$. The first picture shows a starting time point $t^\star$ between 1 and 133 and the second picture shows the last time point of the subsample $t^\star + 599$.

(4) For a chosen confidence level β, set

$$c_{\boldsymbol{b},l}^{(M,T)}(1-\beta) := \inf\left\{x : L_{\boldsymbol{b},l}^{(M,T)}(x) \geq 1-\beta\right\}. \tag{6.6}$$

Then the interval

$$\left[\hat{\theta}_l^{(M,T)} - \tau_{M,T}^{-1} c_{\boldsymbol{b},l}^{(M,T)}(1-\beta), \infty\right)$$

turns out to be a one-sided asymptotic $(1-\beta)\cdot 100\%-$confidence interval for the parameter θ_l.

The statement in step (4) is proved by the following theorem.

Theorem 6.3 (One-Sided Asymptotic Confidence Intervals)**.** *Assume that, as M and $T \to \infty$:*

(I) $\frac{\tau_{\boldsymbol{b}}}{\tau_{M,T}} \to 0$,

(II) $b_i \to \infty$, $i = 1, \ldots, \tilde{d}$,

(III) $\frac{b_1}{M} \to 0, \ldots, \frac{b_d}{M} \to 0$,

(IV) $\frac{b_{\tilde{d}}}{T} \to 0$,

(V) $\boldsymbol{e}$ *is constant.*

Then the following statements hold:

(i) $L_{\boldsymbol{b},l}^{(M,T)}(x) \to \Phi_{0,\sigma_l^2}(x)$ *in probability as* M *and* $T \to \infty$ *for all* $x \in \mathbb{R}$*, where* $\Phi_{0,\sigma_l^2}(\cdot)$ *denotes the distribution function of the univariate normal distribution with mean 0 and variance*

$$\sigma_l^2 = \begin{cases} A_1 \tilde{\Sigma} A_1', & \text{if } l = C, \\ A_2 \tilde{\Sigma} A_2', & \text{if } l = \alpha. \end{cases}$$

(ii) $\sup_{x \in \mathbb{R}} \left| L_{\boldsymbol{b},l}^{(M,T)}(x) - J_l^{(M,T)}(x) \right| \to 0$ *in probability as* M *and* $T \to \infty$*, where* $J_l^{(M,T)}(x) = \mathbb{P}\left(\tau_{M,T}(\hat{\theta}_l^{(M,T)} - \theta_l) \le x \right)$.

(iii) For a chosen confidence level β*, it holds that*

$$\mathbb{P}\left(\tau_{M,T}(\hat{\theta}_l^{(M,T)} - \theta_l) \le c_{\boldsymbol{b},l}^{(M,T)}(1-\beta) \right) \to 1 - \beta \text{ as } M \text{ and } T \to \infty. \tag{6.7}$$

It follows that an asymptotic $(1-\beta)\cdot 100\%-$*confidence interval for* θ_l *is given by*

$$\left[\hat{\theta}_l^{(M,T)} - \tau_{M,T}^{-1} c_{\boldsymbol{b},l}^{(M,T)}(1-\beta), \infty \right). \tag{6.8}$$

Proof. We have to verify the conditions of Politis et al. [1999], Theorem 5.3.1 and Corollary 5.3.1. The central assumption, Assumption 6.1 above, is satisfied by Proposition 6.1. The limit distribution is normal and therefore continuous. Assumptions (I)-(V) are also presumed by Politis et al. [1999]. It remains to be shown that

$$\frac{1}{n} \sum_{k=1}^{n^\star} k^{\tilde{d}-1} \hat{\alpha}_b(k) = \frac{1}{n} \sum_{k=1}^{n^\star} k^d \hat{\alpha}_b(k) \to 0 \text{ as } M \text{ and } T \to \infty, \tag{6.9}$$

where $n = \prod_{j=1}^{\tilde{d}} n_j = M^d T$, $n^\star := \max\{M, T\}$, $b := \prod_{j=1}^{\tilde{d}} b_j$ and, for ℓ and $k \in \mathbb{N}$, $\hat{\alpha}_\ell(k)$ is a special kind of mixing coefficient given by

$$\hat{\alpha}_\ell(k) := \sup_{\substack{\Lambda_1 = E\boldsymbol{m}, \Lambda_2 = \Lambda_1 + \boldsymbol{r} \\ \boldsymbol{r} \in \mathbb{Z}^{\tilde{d}}, \boldsymbol{m} \in \mathbb{N}^{\tilde{d}}}} \left\{ |\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| : A_i \in \mathcal{F}(\Lambda_i), i = 1, 2, |A_1| \le \ell, d(\Lambda_1, \Lambda_2) \ge k \right\},$$

where $d(\cdot,\cdot)$ is defined in Definition 4.6. As, for $\ell \in \mathbb{N}$, we obviously have that $\hat{\alpha}_\ell(k) \le \alpha_{\ell,\ell}(k)$ with the latter mixing coefficient defined in Definition 4.6, we can proceed similarly as in the proof of Proposition 4.9, where we use Lemma 4.8, and conclude:

$$\begin{aligned} \sum_{k=1}^{n^\star} k^d \hat{\alpha}_b(k) &\le \sum_{k=1}^{n^\star} k^d \alpha_{b,b}(k) \\ &\le 2b^2 \sum_{k=1}^{n^\star} k^d \sup_{\max\{\sqrt{d}|h_1|,\ldots,\sqrt{d}|h_d|,|u|\}} \exp\left\{ -\frac{1}{2} \left(C_1 |h_1|^{\alpha_1} + \cdots + C_d |h_d|^{\alpha_d} + C_{\tilde{d}} |u|^{\alpha_{\tilde{d}}} \right) \right\} \\ &\le 2b^2 \sum_{k=1}^{n^\star} k^d \exp\left\{ -\frac{1}{2} \min\{C_1, \ldots, C_{\tilde{d}}\} \left(\frac{k}{\sqrt{d}} \right)^{\min\left\{\alpha_1,\ldots,\alpha_{\tilde{d}}\right\}} \right\} \\ &< \infty \text{ for all } n^\star = \max\{M, T\}. \end{aligned}$$

Therefore,

$$\frac{1}{n}\sum_{k=1}^{n^\star} k^d \hat{\alpha}_b(k) = \frac{1}{M^d T}\sum_{k=1}^{\max\{M,T\}} k^d \hat{\alpha}_b(k) \to 0 \text{ as } M \text{ and } T \to \infty.$$

□

Using Theorem 6.3 we can even construct symmetric two-sided asymptotic confidence intervals for the parameters θ_l for $l \in \{C, \alpha\}$.

Corollary 6.4 (Symmetric Two-Sided Asymptotic Confidence Intervals)**.** *Proceed as in steps (1)-(2) of the subsampling procedure above. Further:*

(3') *Define*

$$L^{(M,T)}_{\boldsymbol{b},l,|\cdot|}(x) := \frac{1}{q}\sum_{i_1=1}^{q_1}\cdots\sum_{i_{\tilde{d}}=1}^{q_{\tilde{d}}} \mathbf{1}_{\left\{\tau_{\boldsymbol{b}}\left|\hat{\theta}^{(M,T)}_{l,\boldsymbol{b},\boldsymbol{i}} - \hat{\theta}^{(M,T)}_{l}\right| \le x\right\}}. \tag{6.10}$$

(4') *For a chosen confidence level β, set*

$$c^{(M,T)}_{\boldsymbol{b},l,|\cdot|}(1-\beta) := \inf\left\{x : L^{(M,T)}_{\boldsymbol{b},l,|\cdot|}(x) \ge 1-\beta\right\}. \tag{6.11}$$

Assume all the conditions of Theorem 6.3. Then the following statements hold:

(i) $L^{(M,T)}_{\boldsymbol{b},l,|\cdot|}(x) \to \Phi^{|\cdot|}_{0,\sigma_l^2}(x)$ *in probability as* M *and* $T \to \infty$ *for all* $x \in \mathbb{R}$*, where* $\Phi^{|\cdot|}_{0,\sigma_l^2}(\cdot)$ *denotes the distribution function of* $|Z|$ *where the random variable* Z *follows a univariate normal distribution with mean 0 and variance* σ_l^2*.*

(ii) $\sup_{x\in\mathbb{R}}\left|L^{(M,T)}_{\boldsymbol{b},l,|\cdot|}(x) - J^{(M,T)}_{l,|\cdot|}(x)\right| \to 0$ *in probability as* M *and* $T \to \infty$*, where* $J^{(M,T)}_{l,|\cdot|}(x)(x) = \mathbb{P}\left(\tau_{M,T}|\hat{\theta}^{(M,T)}_l - \theta_l| \le x\right)$*.*

(iii) For a chosen confidence level β, it holds that

$$\mathbb{P}\left(\tau_{M,T}|\hat{\theta}^{(M,T)}_l - \theta_l| \le c^{(M,T)}_{\boldsymbol{b},l,|\cdot|}(1-\beta)\right) \to 1-\beta \text{ as } M \text{ and } T \to \infty. \tag{6.12}$$

It follows that an asymptotic symmetric $(1-\beta)\cdot 100\%-$confidence interval for θ_l is given by

$$\left[\hat{\theta}^{(M,T)}_l - \tau^{-1}_{M,T} c^{(M,T)}_{\boldsymbol{b},l,|\cdot|}(1-\beta), \hat{\theta}^{(M,T)}_l + \tau^{-1}_{M,T} c^{(M,T)}_{\boldsymbol{b},l,|\cdot|}(1-\beta)\right]. \tag{6.13}$$

The corollary directly follows from Theorem 6.3 and the continuous mapping theorem; for further details consult Politis et al. [1999], Corollary 3.2 which treats the analogous subsampling procedure for stationary time series.

Recall the test scheme:

$$H_0 : \theta_C = \theta\alpha = 0,$$
$$H_A : \theta_C \neq 0 \text{ or } \theta_\alpha \neq 0.$$

Under H_0 we have

- $\hat{\theta}_C^{(M,T)} - \theta_C = (\hat{C}_2 - \hat{C}_1) - (C_2 - C_1) = (\hat{C}_2 - \hat{C}_1) = \hat{\theta}_C^{(M,T)}$ and
- $\hat{\theta}_\alpha^{(M,T)} - \theta_\alpha = (\hat{\alpha}_2 - \hat{\alpha}_1) - (\alpha_2 - \alpha_1) = (\hat{\alpha}_2 - \hat{\alpha}_1) = \hat{\theta}_\alpha^{(M,T)}$.

Motivated by (6.12), rejection areas for the test statistics $\tau_{M,T}\hat{\theta}_C^{(M,T)}$ and $\tau_{M,T}\hat{\theta}_\alpha^{(M,T)}$ are given by

$$\text{Rej}_C := (-\infty, -c_{\boldsymbol{b},C,|\cdot|}^{(M,T)}) \cup (c_{\boldsymbol{b},C,|\cdot|}^{(M,T)}, \infty) \text{ and } \text{Rej}_\alpha := (-\infty, -c_{\boldsymbol{b},\alpha,|\cdot|}^{(M,T)}) \cup (c_{\boldsymbol{b},\alpha,|\cdot|}^{(M,T)}, \infty).$$

To perform the test at an overall confidence level $\beta = 5\%$, we use the Bonferroni inequality (Proposition 6.2) and test the two individual hypotheses that $C_1 = C_2$ and $\alpha_1 = \alpha_2$ at confidence levels of $\frac{\beta}{2}$ each. Tables 6.5 and 6.6 present the results of the individual tests at confidence levels $\frac{\beta}{2} = 2.5\%$. We use pairwise likelihood estimates based on maximum lags 2 and 3. The differences $(\hat{C}_2 - \hat{C}_1)$ and $(\hat{\alpha}_2 - \hat{\alpha}_1)$ can be obtained from Tables 6.1 and 6.2.

max. lag	$\tau_{M,T}$	$(\hat{C}_2 - \hat{C}_1)$	$\tau_{M,T}(\hat{C}_2 - \hat{C}_1)$	Rej_C	**asy. 2.5%-CI for** $(C_2 - C_1)$	**Reject** $C_1 = C_2$**?**
2	324.666	0.0492	15.9739	$(-\infty, -14.4016) \cup (14.4016, \infty)$	$[0.0048, 0.0936]$	yes
3	324.666	0.0506	16.4277	$(-\infty, -14.4023) \cup (14.4023, \infty)$	$[0.0064, 0.0948]$	yes

Table 6.5: Test results for parameters C_1 and C_2. All values are rounded.

max. lag	$\tau_{M,T}$	$(\hat{\alpha}_2 - \hat{\alpha}_1)$	$\tau_{M,T}(\hat{\alpha}_2 - \hat{\alpha}_1)$	Rej_α	**asy. 2.5%-CI for** $(\alpha_2 - \alpha_1)$	**Reject** $\alpha_1 = \alpha_2$**?**
2	324.666	0.0080	2.6105	$(-\infty, -25.9455) \cup (25.9455, \infty)$	$[-0.0719, 0.0880]$	no
3	324.666	0.0078	-2.5381	$(-\infty, -25.5573) \cup (25.5573, \infty)$	$[-0.0865, 0.0709]$	no

Table 6.6: Test results for parameters α_1 and α_2. All values are rounded.

Since we can reject the individual hypothesis that $C_1 = C_2$ at a confidence level of 2.5%, we can reject the overall hypothesis H_0 at a confidence level of 5% and conclude that the rain does not behave equally in both directions. So, all in all, we can reject the assumption of isotropy in the extremal rainfall weather fronts.

Appendix

A.1 List of Notations and Conventions

Throughout this thesis we use the following notations.

(i) We denote by $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$ a d-dimensional real-valued vector.

(ii) Unless stated differently, vector operations such as $\boldsymbol{xy}$, $\boldsymbol{x} \leq \boldsymbol{y}$, $\boldsymbol{x}^{-1}$ or $\boldsymbol{x} \vee \boldsymbol{y}$ are defined component-wise. Sometimes we consider vectors $\boldsymbol{z}$ where certain entries z_j for $j \in J \subset \{1, \ldots, d\}$ are equal to 0. In that case we set $z_j^{-1} := \infty$ for $j \in J$.

(iii) If $X_1, , \ldots, , X_d$ are real-valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ then we use the bold notation $\boldsymbol{X} = (X_1, \ldots, X_d)$ for the random vector whose one-dimensional marginals are $X_j, j = 1 \ldots d$.

(iv) Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a sample of d-dimensional observations. Then the sample maximum $\boldsymbol{M}_n = \bigvee_{i=1}^{n} \boldsymbol{X}_i$ is defined in the component-wise way:

$$M_{n,j} := \bigvee_{i=1}^{n} X_{i,j}, j = 1 \ldots d.$$

Note that $\boldsymbol{M}_n$ is not necessarily part of the sample itself.

(v) If F is a distribution function, we use the bar notation $\bar{F} = 1 - F$ for its tail.

(vi) For a non-decreasing function $h : I \subset \mathbb{R} \to \mathbb{R}$ we denote by $h^{\leftarrow}(y) = \inf\{x \in I : h(x) \geq y\}$ for $y \in \mathbb{R}$.

(vii) Let $a, b : \mathbb{R} \to \mathbb{R}$ be two real-valued functions. We write $a(t) \sim b(t)$ as $t \to z$ (mostly $z \in \{0, \infty\}$) if and only if $\lim_{t \to z} \frac{a(t)}{b(t)} = 1$.

(viii) For a subset A of $\mathbb{R}^d$ for some $d \geq 1$ we denote by $\mathcal{C}(A)$ the set of continuous real-valued functions on A. Convergence on $\mathcal{C}(A)$ is defined as uniform convergence on compact subsets K of A.

(ix) We also need the concept of *convergence in distribution* or *weak convergence*.

Definition A.1 (Weak Convergence, cf. Durrett [2010], Section 3.2)**.** *Let* $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ *be a sequence of random vectors with distribution functions* F_n*,* $n \in \mathbb{N}$*, respectively. The sequence is said to converge in distribution or converge weakly to a limit* $\boldsymbol{X}$ *with distribution function* F *if the following holds:*

$$F_n(\boldsymbol{x}) \to F(\boldsymbol{x}), \qquad n \to \infty, \tag{A.1}$$

for all continuity points $\boldsymbol{x}$ *of* F*. We write* $\boldsymbol{X}_n \xrightarrow{\mathcal{D}} \boldsymbol{X}$ *as* $n \to \infty$*.*

Bibliography

August A. Balkema and Laurens de Haan. Residual life time at great age. *Annals of Probability*, 2:792–804, 1974.

August A. Balkema and Sidney Resnick. Max-infinite divisibility. *Journal of Applied Probability*, 14(2):309–319, 1977.

Jan Beirlant, Yuri Goegebeur, Johan Segers, and Jozef Teugels. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, Ltd, 2004.

Julian Ernst Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B*, 34(2):192–236, 1974.

Erwin Bolthausen. On the central limit theorem for stationary mixing random fields. *The Annals of Probability*, 10(4):1047–1050, 1982.

B.M. Brown and Sidney Resnick. Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14(4):732–739, 1977.

P. Capéràa, Anne-Laure Fougères, and Christian Genest. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84:567–577, 1997.

Richard Davis and Thomas Mikosch. The extremogram: A correlogram for extreme events. *Bernoulli*, 15(4):977–1009, 2009.

Richard Davis and Chun Yip Yau. Comments on pairwise likelihood in time series models. *Statistica Sinica*, 21(1):255–277, 2011.

Richard Davis, Claudia Klüppelberg, and Christina Steinkohl. Max-stable processes for extremes of processes observed in space and time. *Preprint, arXiv:1107.4464v1 (stat.ME)*, 2012a.

Richard Davis, Claudia Klüppelberg, and Christina Steinkohl. Statistical inference for max-stable processes in space and time. *Journal of the Royal Statistical Society B, to appear*, 2012b.

Laurens de Haan. A spectral representation for max-stable processes. *Annals of Probability*, 12:1194–1204, 1984.

Laurens de Haan and James Pickands. Stationary min-stable stochastic processes. *Probability Theory and Related Fields*, 72(4):477–492, 1986.

Laurens de Haan and Sidney Resnick. Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40(4):317–337, 1977.

Paul Deheuvels. Point processes and multivariate extreme values. *Journal of Multivariate Analysis*, 13(2):257–272, 1983.

Paul Deheuvels. On the limiting behavior of the pickands estimator for bivariate extreme value distributions. *Statistics & Probability Letters*, 12:429–439, 1991.

Clément Dombry and Frédéric Eyi-Minko. Strong mixing properties of max-infinitely divisible random fields. *Stochastic Processes and their Applications*, 122(11):3790–3811, 2012.

Rick Durrett. *Probability: Theory and Examples, 4th edition.* Cambridge University Press, 2010.

Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling Extremal Events for Insurance and Finance.* Springer Verlag, 1997.

R.A. Fisher and L.H.C. Tippet. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.*, 24:180–190, 1928.

Jean Geffroy. Contributions à la théorie des valeurs extrêmes. *Publications de l'Institute de Statistique de l'Université de Paris*, 7:36–123, 1958.

Jean Geffroy. Contributions à la théorie des valeurs extrêmes. *Publications de l'Institute de Statistique de l'Université de Paris*, 8:3–52, 1959.

Emil Julius Gumbel. Distributions des valeurs extrêmes en plusieurs dimensions. *Publications de l'Institute de Statistique de l'Université de Paris*, 9:171–173, 1960a.

Emil Julius Gumbel. Bivariate exponential distributions. *Journal of the American Statistical Association*, 55:698–707, 1960b.

Peter Hall and Nader Tajvidi. Distribution and dependence-function estimation for bivariate extreme value distributions. *Bernoulli*, 6:835–844, 2000.

Raphaël Huser and Anthony C. Davison. Space-time modelling for extremes. *Preprint*, 2012. arXiv:1201.3245v1 [stat.ME].

Jürg Hüsler and Rolf-Dieter Reiss. Maxima of normal random vectors: between independence and complete dependence. *Statistics and Probability Letters*, 7(4):283–286, 1989.

A.F. Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quart. J. Roy. Meteo. Soc.*, 81:158–171, 1955.

Zakhar Kabluchko. Extremes of space-time gaussian processes. *Stochastic Processes and their Applications*, 119:3962–3980, 2009.

Zakhar Kabluchko, Martin Schlather, and Laurens de Haan. Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 37(5):2042–2065, 2009.

M.G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

Claudia Klüppelberg and A. May. Parametric families for bivariate extreme value distributions: polynomial distributions. *Extreme Value Distributions (eds: S. Kotz and S. Nadarajah)*. Imperial College Press., page 118 ff., 2001.

Ulrich Krengel. *Ergodic Theorems*. de Gruyter, Berlin, 1985.

M. Ross Leadbetter. On extreme values in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 28:289–303, 1974.

M. Ross Leadbetter. Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 65:291–306, 1983.

Erich L. Lehmann and Joseph P. Romano. *Testing Statistical Hypotheses, 3rd edition*. Springer Science+Business Media, LLC, 2005.

Bruce G. Lindsay. Composite likelihood methods. *Contemporary Mathematics*, 80:221–239, 1988.

David John Nott and Tobias Rydén. Pairwise likelihood methods for inference in image models. *Biometrika*, 86(3):661–676, 1999.

Simone A. Padoan, Mathieu Ribatet, and Scott A. Sisson. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489): 263–277, 2010.

James Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3:119–131, 1975.

James Pickands. Multivariate extreme value distributions. *Proceedings of the 43rd Session ISI*, pages 859–878, 1981.

James Pickands. Multivariate negative exponential and extreme value distributions. *Lecture Notes in Statistics (51); Extreme Value Theory: Proceedings, Oberwolfach 1987 (eds. Jürg Hüsler and Rolf-Dieter Reiss)*, pages 262–274, 1989.

Dimitris N. Politis, Joseph P. Romano, and Michael Wolf. *Subsampling*. Springer-Verlag New York, 1999.

Sidney Resnick. Point processes, regular variation and weak convergence. *Advances in Applied Probability*, 18(1):66–138, 1986.

Sidney Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer, 1987.

Martin Schlather. Simulation of stationary and isotropic random fields. *R-News*, 1(2): 18–20, 2001.

Martin Schlather. Models for stationary max-stable random fields. *Extremes*, 5(1):33–44, 2002.

Masaaki Sibuya. Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11:195–210, 1960.

Richard L. Smith. Maximum likelihood estimation in a class of non-regular cases. *Biometrika*, 72:67–90, 1985.

Richard L. Smith. Max-stable processes and spatial extremes. Unpublished manuscript, University of North California, 1990.

Charles Spearman. The proof and and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101, 1904.

Christina Steinkohl. *Statistical Modelling of Extremes in Space and Time using Max-Stable Processes*. Dissertation, 2012.

Stilian A. Stoev and Murad S. Taqqu. Extremal stochastic integrals: a parallel between max-stable processes and α-stable processes. *Extremes*, 8:237–266, 2005.

Daniel Straumann. *Estimation in Conditionally Heteroscedastic Time Series Models*. Lecture Notes in Statistics, Springer, Berlin, 2004.

Jonathan Tawn. Modelling multivariate extreme value distributions. *Biometrika*, 77: 245–253, 1990.

R. von Mises. La distribution de la plus grande de n valeurs. Reprinted in *Selected Papers Volume II, American Mathematical Society,* Providence, R.I. (1954), pages 271–294, 1936.

Abraham Wald. Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, 20(4):595–601, 1946.

Yizao Wang and Stilian A. Stoev. On the structure and representations of max-stable processes. *Advances in Applied Probability*, 42(3):855–877, 2010.

Yizao Wang, Parthanil Roy, and Stilian A. Stoev. Ergodic properties of sum-and max-stable stationary random fields via null and positive group actions. *Technical Report*, 2009.