Technische Universität München
Zentrum Mathematik
Lehrstuhl für Mathematische Statistik

Statistical Modelling and Estimation of Space-Time Extremes

Sven Buhl

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:		Prof. Dr. Caroline Lasser
Prüfer der Dissertation:	1.	Prof. Dr. Claudia Klüppelberg
	2.	Prof. Dr. Thomas Mikosch (Universität Kopenhagen)
	3.	Prof. Dr. Gernot Müller (Universität Augsburg)

Die Dissertation wurde am 11.04.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 01.06.2017 angenommen.

Zusammenfassung

Diese Arbeit behandelt die statistische Modellierung und Schätzung extremer stochastischer Raum- und Raum-Zeit-Prozesse. Solche Prozesse finden in der Analyse umweltbezogener Daten Anwendung, die den Fokus auf die Bewertung seltener und extremer Ereignisse in Raum und/oder Zeit legen, wie z.B. Stürme, extremer Niederschlag oder Hitzewellen.

Betrachtet werden zwei Klassen in der Extremwerttheorie und -statistik bedeutender stochastischer Prozesse: regulär variierende und deren Teilklasse der langschwänzigen max-stabilen Prozesse, die auf $\mathbb{R}^d$ oder $\mathbb{R}^{d-1} \times [0, \infty)$ definiert sind. Im ersten Fall werden die Prozesse als räumlich und im zweiten als raum-zeitlich aufgefasst. Unter den max-stabilen Prozessen ist das in der Arbeit führende Beispiel der Brown-Resnick-Prozess, der sowohl als räumliches als auch als raum-zeitliches Modell definiert wird. Sowohl regulär variierende als auch max-stabile Prozesse sind durch extreme Abhängigkeitsfunktionen charakterisiert.

Die beiden Hauptziele dieser Arbeit sind die folgenden: einerseits werden neue parametrische Modelle für die extremen Abhängigkeitsfunktionen regulär variierender und max-stabiler Prozesse entwickelt, die für die betrachteten Anwendungen geeignet sind. Andererseits werden neue Inferenzmethoden entwickelt, um die Modelle an beobachtete Daten anzupassen und relevante asymptotische Eigenschaften der Schätzer wie Konsistenz und asymptotische Normalität bewiesen. Die Inferenzmethoden umfassen paarweise Likelihood-Schätzung von neuen anisotropen Brown-Resnick-Prozessen in Raum und Zeit sowie sowohl empirische Verfahren als auch semiparametrische kleinste-Quadrate-Methoden, die auf die weitaus größere Klasse regulär variierender Prozesse angewandt werden können. Sie beruhen auf empirischen Schätzern des räumlichen oder raum-zeitlichen Extremogramms, welches der Kovarianzfunktion von Indikatorvariablen von Überschreitungsereignissen im asymptotischen Sinne entspricht. Um die asymptotischen Eigenschaften der Schätzer nachzuweisen, ist es notwendig, schwache Mischungs-Eigenschaften zu formulieren und zu überprüfen.

Alle Inferenzmethoden funktionieren asymptotisch korrekt in sehr flexiblen Beobachtungsschemata. Das bedeutet beispielsweise im Raum-Zeit-Kontext, dass die Anzahl der räumlichen Beobachtungen im Vergleich zur Anzahl der zeitlichen Beobachtungen sehr klein sein kann, was oft in der Analyse realer Daten anzutreffen ist. Die Resultate dieser Arbeit decken aber auch allgemeinere Schemata ab.

Alle Methoden werden sowohl in Simulationsstudien als auch in einer Analyse extremen Regenfalls in Florida, USA, verwendet und ein sehr gutes Verhalten im Falle endlicher Stichproben nachgewiesen. Eine Reihe von Methoden zur Bewertung der Qualität des Modellfits und zur Modellselektion werden vorgestellt und angewandt. Insbesondere kann die Hypothese verworfen werden, dass die Messungen des Regenfalls in Florida einem räumlich isotropen Brown-Resnick Prozess folgen.

Abstract

This thesis deals with the statistical modelling and estimation of extremal spatial and space-time stochastic processes. Such processes find applications in many areas of environmental analysis with focus on the assessment of rare and extreme events in space and/or time such as wind storms, extremal rainfall or heatwaves.

We consider two classes of stochastic processes that are prominent in extreme value theory and extreme value statistics: regularly varying processes and their subclass of heavy tailed max-stable processes that live on $\mathbb{R}^d$ or $\mathbb{R}^{d-1} \times [0, \infty)$. In the former case, we interpret the processes as spatial and in the latter as space-time processes. Among the max-stable processes, our leading example is the Brown-Resnick process, which we define both as a spatial and as a space-time model. Both regularly varying and max-stable processes are characterised by extremal dependence functions.

The two central goals of this thesis are as follows: firstly, we develop new parametric models for the extremal dependence functions of regularly varying and max-stable processes that are suitable for the considered applications. Secondly, we introduce new inference methods to fit the new models to data and prove relevant asymptotic properties such as consistency and asymptotic normality of the estimates. The inference methods comprise pairwise likelihood estimation of new spatially anisotropic Brown-Resnick space-time models as well as both empirical approaches and semiparametric least squares methods which are applicable to the much larger class of regularly varying processes. They rely on empirical estimates of the spatial or space-time extremogram, which is the covariance function of indicator functions of exceedance events in an asymptotic sense. In order to prove asymptotic properties of the estimates we need to state and verify weak mixing conditions for the processes.

All inference methods work asymptotically correct in very flexible observation schemes. For instance, in the space-time context, the number of spatial observations can be very small compared to the number of temporal observations, which is a situation often found in real data. However, our results also cover more general observation schemes.

We show all methods at work in both simulation studies and in an analysis of extremal rainfall in Florida, USA, and confirm a very good finite sample behaviour. We propose and apply a variety of goodness-of-fit test procedures to assess the quality of the model fit and to perform model selection. In particular, we can reject the hypothesis that the Florida rainfall data are observations of a spatially isotropic Brown-Resnick space-time process.

Acknowledgements

First of all, I would like to thank Claudia Klüppelberg who, three years ago, offered me the opportunity to return from industry to university. She has been an outstanding supervisor. Working with her and conducting numerous fruitful and efficient discussions have been a pleasure.

Thank you very much also to Thomas Mikosch and Gernot Müller for completing my examination committee and for interesting discussions about my dissertation.

I am also exceptionally grateful to Anthony Davison and his research group, in particular Sebastian Engelke, Peiman Asadi, Raphaël de Fondeville and Léo Belzile, who hosted me at EPFL in Lausanne for three months in 2016, giving me valuable advice and support in many discussions and offering me the opportunity to take part in CUSO summer and winter schools in the Swiss mountains.

I must say that I consider myself extremely happy to have so many exceptional colleagues and friends here at our chair. In particular, I would like to mention Nadine. We were almost always the last to leave the office (like today) and she helped me out of a couple of desperate LaTeX situations (like today). It has also been a pleasure to be part of the Monday Mensa lunch group. Thanks to Andi, Patrick and Matthias for sharing many interesting culinary experiences with me.

My special thanks goes to my co-authors Richard Davis and Christina Kienlein (Steinkohl). Together with Claudia, Christina led me to the topic of max-stable space-time processes when they supervised my master's thesis four years ago. Besides, I would like to thank Erwin Bolthausen and Thomas Mikosch for enlightening discussions on central limit theorems.

I also appreciate fruitful work and discussions with the members of the Deutsches Geodätisches Forschungsinstitut Eva Börgens, Denise Dettmering, Gerhard Ressler and Florian Seitz, who are members or very much associated with my IGSSE research team REWAP.

I also thank Dennis Leber for comparing simulation methods of the Brown-Resnick process, and Ton Dieker and Marius Hofert for providing and improving the simulation programme.

Furthermore, I thank Chin Man Mok and the Southwest Florida Water Management District (SWFWMD) for providing the Florida rainfall data analysed in this thesis.

I acknowledge support from the Deutsche Forschungsgemeinschaft (DFG) through the TUM International Graduate School of Science and Engineering (IGSSE).

Besonders wichtig ist es mir auch, mich bei meiner Familie und meinen Freunden zu bedanken. Bina, Mama, Papa, Volker, Björn, Mimi, Nani, Elfriede, Jola, Stefan, Felix, Merlin, Jea, Manu, Andi, Basti und viele mehr... ohne euch wäre das hier alles nicht möglich gewesen! Vor allem dir, Bina, muss ich danken. Ich weiß, es ist nicht immer alles einfach mit mir und vor allem die letzten drei Jahre waren schwierig. Aber du hast mich immer unterstützt, wie du mich generell bei allem unterstützt. Dafür bin ich dir unendlich dankbar.

Contents

Chapter 1

Introduction

1.1 Motivation and objectives of extreme value theory

The central aim of extreme value theory and extreme value statistics is to quantify and analyse rare and extreme events and estimate tails rather than the median of underlying distributions. These events comprise natural catastrophes such as extreme wind gusts, hurricanes or flooding. They are typically very costly and thus of particular interest for society and industry. The theory developed in this thesis is applied to daily rainfall maxima taken over hourly accumulated measurements in inches at 144 spatial locations in the period 1999-2004. The locations lie on a grid of size 12×12 in a region in Florida, USA. The measurements are block maxima both in space and time and computed from raw data provided by the Southwest Florida Water Management District (SWFWMD). Their position in Florida is visualised in Figure 1.1. More details on the data are given in Sections 3.6 and 5.5.

From a historical point of view, there are basically three stages of evolution of extreme value theory. The most classical one is *univariate extreme value theory*, which dates back to Fisher and Tippett [36]. It deals with a number $n \in \mathbb{N}$ of univariate real observations $X_1, \ldots, X_n$. This can be for instance the (extreme) amount of rainfall at a particular fixed location or site on earth, such as the observed time series of daily maxima at a fixed location on the grid of size 12×12 in Florida. For an exemplary fixed location, the daily maxima are visualised for the wet seasons (June-September) of the years 1999-2004 in Figure 1.2. The observations are often assumed to be independent, which is reasonable for example if they are given on a yearly basis. But this assumption can be relaxed, which opens the door to a more general time series setting. A typical question or task in univariate extreme value theory is to compute high quantiles of an extreme value distribution underlying the observations. This can be interpreted as a value (e.g. rainfall amount) that is exceeded at the fixed location extremely rarely, for example on average only every 100 (or 200,...) years, and may not even have been observed yet. Profound introductions into univariate extreme value theory can be found in Coles [16] and Embrechts et al. [32].

The next stage is the generalisation from univariate to d-variate observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^d$ for $d \in \mathbb{N}$, summarised under the term *multivariate extreme value theory*. The $\boldsymbol{X}_i$ are vectors

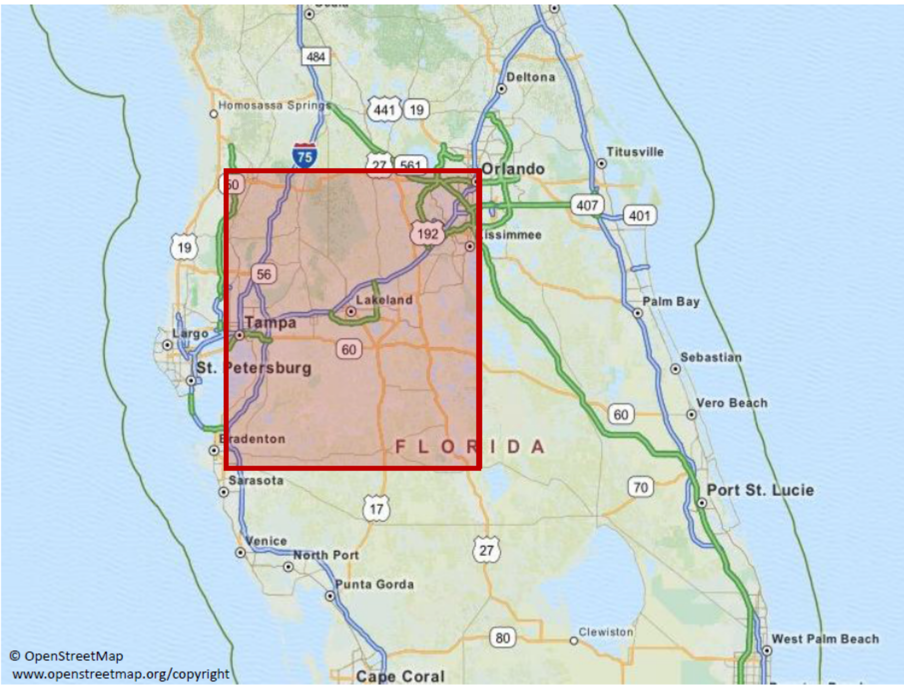

Figure 1.1: Rainfall observation area in Florida

whose components may be interpreted as different distinct locations, thus constituting d series of univariate observations. On top of treating each of those components (often called "univariate margins") with methods of univariate extreme value theory, the focus lies on modelling and determining the structure of (extreme) dependence between them. One is particularly interested in joint (i.e., simultaneous) exceedances of high levels or thresholds of the univariate margins. In the context of the rainfall example, an interesting question might be "given high rainfall at location A, how likely is high rainfall at a (nearby) location B?". Detailed introductions into multivariate extreme value theory are given in Beirlant et al. [3] and, with focus on the bivariate case, in [16], Chapter 8.

As the last natural step, multivariate extreme value theory has been generalised to infinite dimensions, which gives rise to stochastic extreme value processes $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ for some $d \in \mathbb{N}$. These processes are commonly interpreted as spatial processes but we stress that this includes processes $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^{d-1} \times [0, \infty)\}$, which we interpret as space-time processes, and time series $\{X(t) : t \in [0, \infty)\}$. In contrast to (finite-dimensional) multivariate extreme value theory, extreme events are not modelled on a finite number of distinct locations, but on a continuous spatial map and possibly in continuous time. A theoretical introduction into this setting can be found in de Haan and Ferreira [25]. Observations of the processes are often given on a grid in space and/or time, such as the daily rainfall maxima in Florida, which are visualised on the regular grid of size 12×12 at four consecutive time points in Figure 1.3. Central aims are to find and develop models for extreme value processes that are appropriate with respect to the area of application. Furthermore, these models have to be fitted to the observed data and the goodness of the fit needs to be assessed. This is the setting this thesis is mainly committed to.

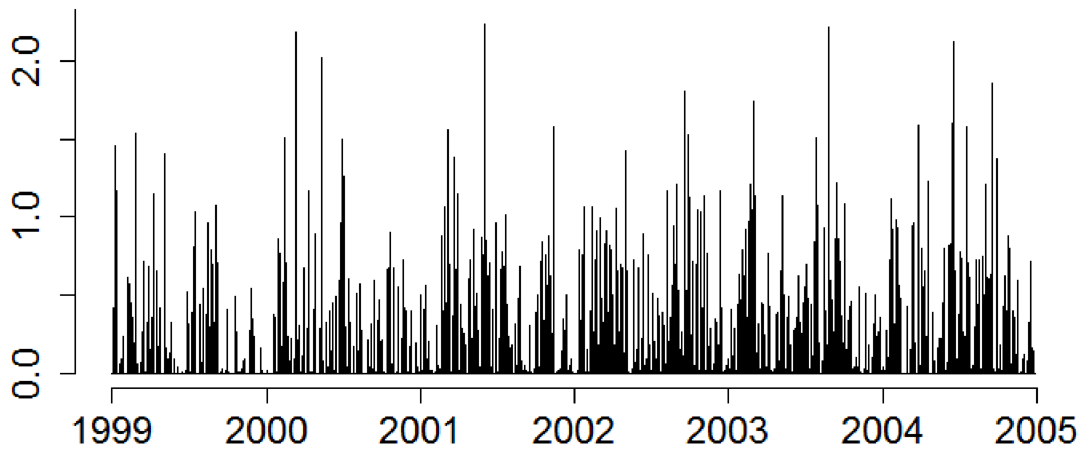

Figure 1.2: Daily rainfall maxima in inches taken over hourly accumulated measurements from 1999-2004 at some fixed location in Florida, USA.

1.2 Scope and goals of this thesis

This thesis pursues two central objectives. One aim is to develop new statistical models that can be applied in a variety of different areas of applications in extreme value statistics. These models need to capture extremal dependence properties in time series, spatial or space-time stochastic processes appropriately.

The second central goal is to provide suitable inference methods which enable us to fit the developed statistical extremal dependence models to real data such as the Florida rainfall data visualised in Figures 1.1-1.3. These inference methods comprise parametric, semiparametric, as well as non-parametric procedures.

Throughout this thesis we consider two classes of strictly stationary spatial (or space-time) processes $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ with dimension $d \in \mathbb{N}$, which are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and prominent in extreme value theory.

Regularly varying stochastic processes

The first class consists of *regularly varying* processes. Here and in what follows, for two positive functions f and g we write $f(n) \sim g(n)$ as $n \to \infty$ for $\lim_{n\to\infty} f(n)/g(n) = 1$. A stochastic process $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ is called regularly varying, if there exists some normalising sequence $(a_n)_{n\in\mathbb{N}}$ with $0 < a_n \to \infty$ such that $\mathbb{P}(|X(\mathbf{0})| > a_n) \sim n^{-d}$ as $n \to \infty$ and for every finite set $\mathcal{I} \subset \mathbb{R}^d$ with cardinality $|\mathcal{I}| < \infty$,

$$n^d \mathbb{P}\Big(\frac{X_{\mathcal{I}}}{a_n} \in \cdot\Big) =: \mu_{\mathcal{I},n}(\cdot) \xrightarrow{v} \mu_{\mathcal{I}}(\cdot), \quad n \to \infty, \tag{1.1}$$

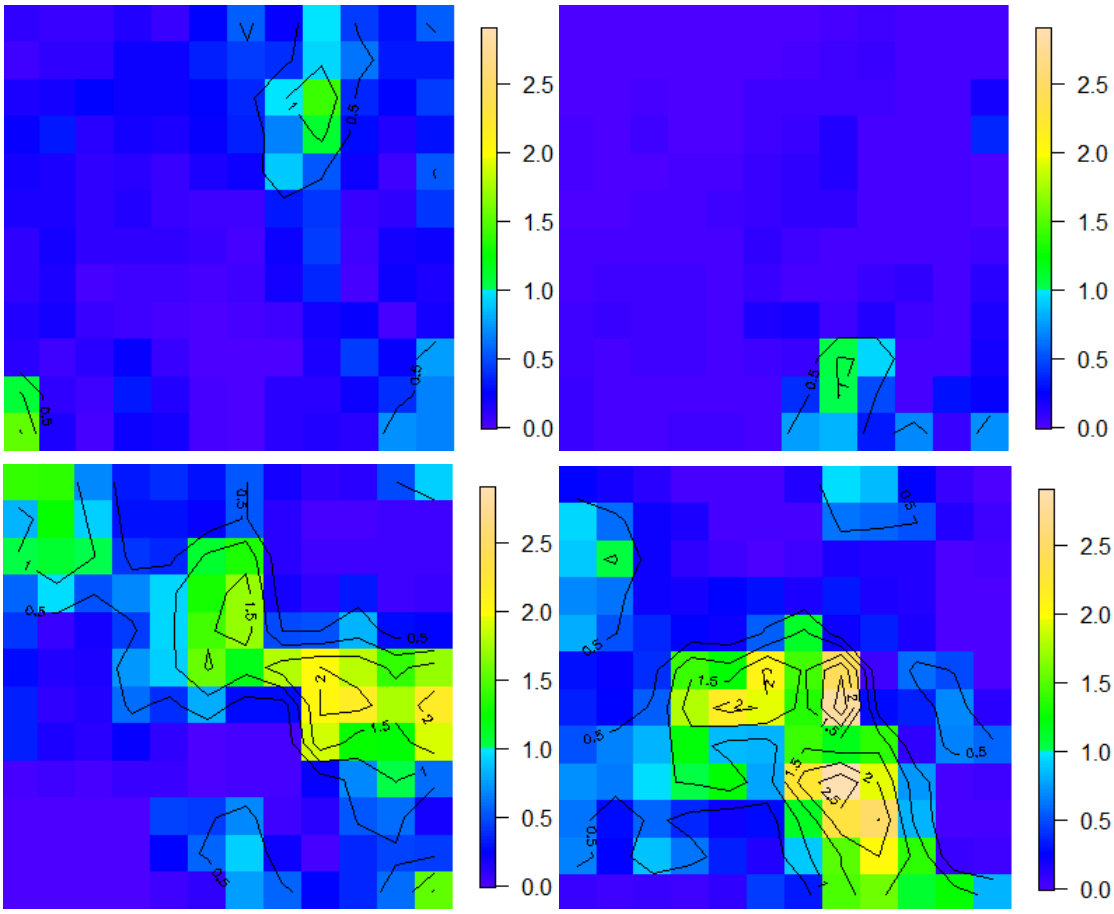

Figure 1.3: Daily rainfall maxima in inches taken over hourly accumulated measurements at 144 gridded locations in Florida, USA, and at four consecutive days (from left to right and top to bottom) in the period 1999-2004.

for some non-null *Radon* measure $\mu_{\mathcal{I}}$ on the Borel sets in $\overline{\mathbb{R}}^{|\mathcal{I}|}\backslash\{\mathbf{0}\}$; i.e., $\mu_{\mathcal{I}}$ is finite on compact sets. If $\mathcal{I}$ is a singleton; i.e., $\mathcal{I} = \{\boldsymbol{s}\}$ for some $\boldsymbol{s} \in \mathbb{R}^d$, we set

$$\mu_{\{\boldsymbol{s}\}}(\cdot) = \mu_{\{\mathbf{0}\}}(\cdot) =: \mu(\cdot), \tag{1.2}$$

which is justified by stationarity. We let $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ and we let $X_{\mathcal{I}}$ denote the vector $(X(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{I})$. The notation $\stackrel{v}{\to}$ stands for *vague convergence* meaning that

$$\int_{\overline{\mathbb{R}}^{|\mathcal{I}|}\backslash\{\mathbf{0}\}} f(\boldsymbol{x})\mu_{\mathcal{I},n}(d\boldsymbol{x}) \to \int_{\overline{\mathbb{R}}^{|\mathcal{I}|}\backslash\{\mathbf{0}\}} f(\boldsymbol{x})\mu_{\mathcal{I}}(d\boldsymbol{x}), \quad n \to \infty, \tag{1.3}$$

for all continuous nonnegative functions $f : \overline{\mathbb{R}}^{|\mathcal{I}|}\backslash\{\mathbf{0}\} \to (0, \infty)$ with compact support. The limit measure $\mu_{\mathcal{I}}$ is furthermore *homogeneous* of order $-\beta$ where $\beta > 0$ is called the *index of regular variation*:

$$\mu_{\mathcal{I}}(xC) = x^{-\beta}\mu_{\mathcal{I}}(C), \quad x > 0,$$

for every Borel set $C \subset \overline{\mathbb{R}}^{|\mathcal{I}|} \backslash \{\mathbf{0}\}$. For more background on regular variation for stochastic processes and vectors see Hult and Lindskog [40] and Resnick [57, 59].

Max-stable processes

The second class of processes we consider are heavy tailed *max-stable* processes, which constitute a subclass of regularly varying processes. A process $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ is called max-stable if there exist sequences $a_n(\boldsymbol{s}) > 0$ and $b_n(\boldsymbol{s})$ for $\boldsymbol{s} \in \mathbb{R}^d$ and $n \in \mathbb{N}$ such that

$$\Big\{a_n^{-1}(\boldsymbol{s})\Big(\bigvee_{j=1}^{n} X_j(\boldsymbol{s}) - b_n(\boldsymbol{s})\Big) : \boldsymbol{s} \in \mathbb{R}^d\Big\} \stackrel{d}{=} \{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}, \tag{1.4}$$

where $\{X_j(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ are independent replicates of $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ and the maximum is taken componentwise. The symbol $\stackrel{d}{=}$ stands for equality in distribution. Max-stable processes provide a useful framework for modelling extremal dependence in continuous time and/or space. A max-stable process is a limit process which possesses a *max-domain of attraction*. A stochastic process $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ is in the max-domain of attraction of a max-stable process $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ if there exist sequences $c_n(\boldsymbol{s}) > 0$ and $d_n(\boldsymbol{s})$ for $\boldsymbol{s} \in \mathbb{R}^d$ and $n \in \mathbb{N}$ such that

$$\Big\{c_n^{-1}(\boldsymbol{s})\Big(\bigvee_{j=1}^{n} Y_j(\boldsymbol{s}) - d_n(\boldsymbol{s})\Big) : \boldsymbol{s} \in \mathbb{R}^d\Big\} \stackrel{d}{\to} \{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}, \quad n \to \infty, \tag{1.5}$$

where $\{Y_j(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ are independent replicates of $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$. The symbol $\stackrel{d}{\to}$ stands for convergence in distribution.

For $D = \{\boldsymbol{s}^{(1)}, \ldots, \boldsymbol{s}^{(|D|)}\} \subset \mathbb{R}^d$ and $\boldsymbol{y} = (y_1, \ldots, y_{|D|}) > \mathbf{0}$ the finite-dimensional margins of $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ are given by

$$\mathbb{P}(X(\boldsymbol{s}^{(1)}) \le y_1, X(\boldsymbol{s}^{(2)}) \le y_2, \cdots, X(\boldsymbol{s}^{(|D|)}) \le y_{|D|}) = \exp\{-V_D(\boldsymbol{y})\}. \tag{1.6}$$

Here V_D denotes the *exponent measure* (cf. [3], Section 8.2.2).

The univariate margins of $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ follow a *generalised extreme value distribution* (GEV) (cf. [32], Definition 3.4.1) given for shape parameter $\xi \in \mathbb{R}$, location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$ by

$$\mathrm{GEV}_{(\xi,\mu,\sigma)}(x) = \begin{cases} \exp\Big\{ - \Big[1 + \xi\Big(\frac{x-\mu}{\sigma}\Big)\Big]^{-\frac{1}{\xi}}\Big\}, & \text{if } \xi \neq 0, \\ \exp\Big\{ - \exp\Big[- \Big(\frac{x-\mu}{\sigma}\Big)\Big]\Big\}, & \text{if } \xi = 0. \end{cases} \tag{1.7}$$

The GEV is the classical model for univariate block maxima such as yearly or daily rainfall maxima taken over blocks (consecutive values) of observations. Depending on the value of its shape parameter ξ, a GEV (and its max-domain of attraction) can be assigned to three different types (cf. [36]). The case $\xi < 0$ corresponds to the extremal Weibull class and $\xi = 0$ to the Gumbel class. Of particular importance in this thesis is the case $\xi > 0$ corresponding to the heavy tailed Fréchet class, which contains distributions with regularly varying right tail. Prominent examples are the Pareto and the loggamma distributions and they find applications in insurance and reinsurance industry. Hence max-stable processes with Fréchet marginal distributions are regularly varying, but the class of regularly varying processes includes also processes which lie in their max-domain of attraction. In this thesis the most prominent example of a max-stable process is the strictly stationary *Brown-Resnick process* $\{\eta(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ with Fréchet marginal distributions. It has representation

$$\eta(\boldsymbol{s}) = \bigvee_{j=1}^{\infty} \left\{ \xi_j \, e^{W_j(\boldsymbol{s}) - \delta(\boldsymbol{s})} \right\}, \quad \boldsymbol{s} \in \mathbb{R}^d, \tag{1.8}$$

where $\{\xi_j : j \in \mathbb{N}\}$ are points of a Poisson process on $[0, \infty)$ with intensity $\xi^{-2} d\xi$, the *dependence function* δ is nonnegative and conditionally negative definite, and $\{W_j(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ are independent replicates of a Gaussian process $\{W(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ with stationary increments, $W(\boldsymbol{0}) = 0$, $\mathbb{E}[W(\boldsymbol{s})] = 0$ and covariance function

$$\mathbb{C}ov[W(\boldsymbol{s}^{(1)}), W(\boldsymbol{s}^{(2)})] = \delta(\boldsymbol{s}^{(1)}) + \delta(\boldsymbol{s}^{(2)}) - \delta(\boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)}).$$

Representation (1.8) goes back to de Haan [24] and Giné et al. [39]. Brown-Resnick processes have been studied by Brown and Resnick [8] in a time series context, as a spatial model by Kabluchko et al. [47], and in a space-time setting by Huser and Davison [43], Davis et al. [19] and Steinkohl [62]. From a geostatistical point of view, the dependence function δ is called the *semivariogram* of the process $\{W(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ defined through $2\delta(\boldsymbol{h}) = \mathbb{V}ar[W(\boldsymbol{s}) - W(\boldsymbol{s} + \boldsymbol{h})]$ for $\boldsymbol{s}, \boldsymbol{h} \in \mathbb{R}^d$ and $\{\eta(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ is called the Brown-Resnick process associated to δ. For $D = \{\boldsymbol{s}, \boldsymbol{s} + \boldsymbol{h}\}$ where $\boldsymbol{s} \in \mathbb{R}^d$ and $\boldsymbol{h} \in \mathbb{R}^d$ is some fixed lag vector, we get for the bivariate exponent measure (1.6) of the Brown-Resnick process (cf. [19], Section 3),

$$V_D(y_1, y_2) = \frac{1}{y_1} \widetilde{\Phi}\Big(\frac{y_2}{y_1}\Big) + \frac{1}{y_2} \widetilde{\Phi}\Big(\frac{y_1}{y_2}\Big), \quad y_1, y_2 > 0, \tag{1.9}$$

with

$$\widetilde{\Phi}\Big(\frac{x}{y}\Big) = \widetilde{\Phi}\Big(\boldsymbol{h}; \frac{x}{y}\Big) := \Phi\Big(\frac{\log(x/y)}{\sqrt{2\delta(\boldsymbol{h})}} + \sqrt{\frac{\delta(\boldsymbol{h})}{2}}\Big), \quad x, y > 0. \tag{1.10}$$

The finite-dimensional distributions of $\{\eta(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ are characterised by δ; hence modelling the underlying extremal dependence structure is achieved by setting up parametric models for δ or related functions such as the *extremogram*, which can be defined for regularly varying processes as follows.

The extremogram as a correlogram for extreme events

The extremogram measures extremal dependence in a strictly stationary regularly varying stochastic process and can be seen as a correlogram for extreme events. Introduced for time series in Davis and Mikosch [17] and Fasen et al. [35], it has been generalised to spatial settings in Cho et al. [15] and to space-time settings in Buhl et al. [14] and in Steinkohl [62]. In the respective publications, asymptotic results like consistency and asymptotic normality of an empirical extremogram are proved under weak mixing conditions. Davis et al. [21] give a profound review of the estimation theory for time series with various examples. As stated in [17], the extremogram can be regarded as the covariance function of indicator functions of exceedance events in an asymptotic sense. It is defined for strictly stationary regularly varying processes $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$. Consider a sequence $0 < a_n \to \infty$ as in (1.1). For μ as in (1.2) and two μ-continuous Borel sets A and B in $\overline{\mathbb{R}}\backslash\{0\}$ (i.e., $\mu(\partial A) = \mu(\partial B) = 0$) such that $\mu(A) > 0$, the extremogram is defined as

$$\rho_{AB}(\boldsymbol{h}) = \lim_{n\to\infty} \frac{\mathbb{P}(X(\boldsymbol{0})/a_n \in A, X(\boldsymbol{h})/a_n \in B)}{\mathbb{P}(X(\boldsymbol{0})/a_n \in A)}, \quad \boldsymbol{h} \in \mathbb{R}^d. \tag{1.11}$$

For $A = B = (1, \infty)$, the extremogram $\rho_{AB}(\boldsymbol{h})$ is the *tail dependence coefficient* and characterises the extremal dependence structure between $X(\boldsymbol{0})$ and $X(\boldsymbol{h})$ (cf. [3], Section 9.5.1). If $\rho_{(1,\infty),(1,\infty)}(\boldsymbol{h}) = 0$ then $X(\boldsymbol{0})$ and $X(\boldsymbol{h})$ are called *asymptotically independent*, in case $0 < \rho_{(1,\infty),(1,\infty)}(\boldsymbol{h}) \leq 1$ they are said to be *asymptotically dependent*. This reveals the importance of developing appropriate inference methods in order to obtain extremogram estimates with good asymptotic properties. The methods can be non-parametric and based on empirical estimates, cf. [12, 15, 17] or Buhl and Klüppelberg [13]. They can also be parametric and rely on a parametric model set up for the extremogram. Also a semiparametric method; that is, a combination of both as carried out in [14], is possible. In this thesis we develop new, both parametric and semiparametric inference methods for the extremogram of strictly stationary heavy tailed max-stable processes and for more general regularly varying processes. In simulations studies and in real data analyses, we apply them to Brown-Resnick processes as defined in (1.8). A list of parametric models for their dependence function δ is given in the following.

Models for the dependence function of a Brown-Resnick process

As can be deduced from Lemma A.1 in Buhl and Klüppelberg [12], the extremogram ρ_{AB} of the max-stable Brown-Resnick process defined in (1.8) is characterised by its dependence function δ. Thus a parametric model for δ directly yields a model for the extremogram ρ_{AB} defined in (1.11).

A spatial model for δ commonly used in environmental applications is the fractional class

$$\big\{\delta(\boldsymbol{h}) = C\|\boldsymbol{h}\|^{\alpha} : \quad \boldsymbol{h} \in \mathbb{R}^d, \quad C \in (0,\infty), \quad \alpha \in (0,2]\big\}. \tag{1.12}$$

The special time series model, where $\delta(u) = |u|/2$ for $u \in \mathbb{R}$, which corresponds to the Gaussian processes $\{W_j(t) : t \in [0,\infty)\}$ in (1.8) being standard Wiener processes, was introduced in [8]. Another special case is the storm profile model introduced by Smith [61], which corresponds to $\alpha = 2$ in (1.12). It models a very smooth process, since the parameter α relates to the smoothness of the sample paths of the underlying Gaussian processes, with the boundary case $\alpha = 2$ corresponding to mean-square differentiable processes. It was shown in [47] that Brown-Resnick processes with dependence function δ as in (1.12) arise as limits of pointwise maxima of appropriately rescaled and normalised independent stationary Gaussian processes. These Gaussian processes and the resulting Brown-Resnick limit process are *spatially isotropic*, meaning that $\delta(\boldsymbol{h})$, and therefore extremal dependence between two process values $\eta(\mathbf{0})$ and $\eta(\boldsymbol{h})$, only depend on the norm of the lag vector $\boldsymbol{h}$, whereas directional influences are neglected.

For Brown-Resnick space-time processes $\{\eta(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^{d-1}, t \in [0,\infty)\}$, model (1.12) was generalised in [19] and [62] to the space-time fractional class

$$\big\{\delta(\boldsymbol{h},u) = C_1\|\boldsymbol{h}\|^{\alpha_1} + C_2|u|^{\alpha_2} : \quad (\boldsymbol{h},u) \in \mathbb{R}^d, \quad C \in (0,\infty), \quad \alpha \in (0,2]\big\} \tag{1.13}$$

and applied in [62] to the gridded space-time maxima of rainfall in Florida visualised in Figures 1.1-1.3. This model is again spatially isotropic, which is an assumption that is often unrealistic in environmental analyses, since for instance wind and rainfall have directional preferences.

In Chapters 4 and 5 of this thesis, which are based on the publications Buhl and Klüppelberg [13] and Buhl and Klüppelberg [11], we extend model (1.13) and allow for different rates of decay of extremal dependence along the axes of a d-dimensional spatial grid (which we call principal directions) by defining the dependence function $\delta(\boldsymbol{h},u)$ for spatial lag $\boldsymbol{h} = (h_1,\dots,h_d) \in \mathbb{R}^d$ and time lag $u \in \mathbb{R}$ as

$$\big\{\delta(\boldsymbol{h},u) = \sum_{j=1}^{d-1} C_j|h_j|^{\alpha_j} + C_d|u|^{\alpha_d} : \quad (\boldsymbol{h},u) \in \mathbb{R}^d, \quad C_j \in (0,\infty), \quad \alpha_j \in (0,2], \, j = 1,\dots,d\big\} \tag{1.14}$$

In Chapter 5, we also provide a variety of goodness-of-fit test procedures. Firstly, we present a procedure to test whether observed data originate from a max-stable process, which works well in the considered space-time setting. Such tests should be conducted before the actual fitting of a max-stable model in order to check whether this is an appropriate assumption. We furthermore provide a test procedure to assess the quality of the model fit, which relies on simulation diagnostics. Besides, based on the model fit, we present a method to test for spatial anisotropy versus isotropy, which is designed for the new Brown-Resnick model (1.14). We fit this model to the Florida rainfall data using pairwise likelihood estimation, which is described in more detail below in this section. Model (1.14) can be further generalised to allow for arbitrary principal orthogonal directions by considering a rotation matrix R and the dependence function $\delta_R(\boldsymbol{h},u) = \delta(R\boldsymbol{h},u)$ for $(\boldsymbol{h},u) \in \mathbb{R}^d$. This involves the rotation angle as a further model parameter. We examine this rotated model in more detail in Chapter 4.

Also in that chapter we introduce in a general Brown-Resnick space-time model as defined in (1.8) some influence of the spatial dependence from past values of the process. We do this by time-shifting the Gaussian process $\{W(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^{d-1}, t \in [0,\infty)\}$ and setting

$$W^{(\boldsymbol{\tau})}(\boldsymbol{s},t) := W(\boldsymbol{s} - t\boldsymbol{\tau}, t)$$

for some $\boldsymbol{\tau} \in \mathbb{R}^{d-1}$, called the direction of propagation. This time-shift then directly translates into the dependence function δ. The direction of propagation $\boldsymbol{\tau}$ is a model parameter and needs to be estimated from observed data. It reflects directional preferences of storms or other weather extremes. Figure 1.4 visualises a realisation simulated from the space-time model (1.14) for $d = 3$ extended by a time-shift with direction of propagation $\boldsymbol{\tau} = (1, -1)$.

Fitting of parametric extremal dependence and extremogram models requires appropriate inference methods. In the following we give an overview over existing methods and present new methods developed in this thesis.

Inference methods for extremal dependence models

A common approach to estimate the parameters of a model for max-stable processes such as the Brown-Resnick process (1.8) is composite likelihood. Pairwise likelihood estimation based on the bivariate density of the models has been proposed in Padoan et al. [54] for spatial processes and studied in [43] and Davis et al. [20] and [62] in a space-time context. Results like consistency and asymptotic normality for pairwise likelihood estimates of their proposed model (1.13) are derived in [20]. Genton et al. [37] examine triplewise likelihood estimation for the max-stable Smith storm profile model and report a gain in efficiency. This investigation is extended to the more general Brown-Resnick processes as defined in (1.8) by Huser and Davison [42] who show via simulations that the gain in efficiency is substantial only for very smooth processes, for example for those which are associated to a dependence function δ as in (1.12) with a parameter α close to 2. Composite likelihood methods have become widely used in parameter estimation for max-stable processes since due to the exponential form of their finite-dimensional distribution functions (cf. Eq. (1.6)), the number of terms in the corresponding densities and thus in the likelihood function explodes. Recently, however, methods have been proposed that open the door to full likelihood estimation in specific scenarios, see for instance Wadsworth [66] who suggests to incorporate information on the occurence times of maxima, which simplifies the likelihood, or Wadsworth and Tawn [67] and Engelke et al. [34], who suggest threshold-based approaches. When full likelihood estimation is feasible, also frequentist or Bayesian approaches are applicable, see for example Dombry et al. [29] and Thibaud et al. [65].

In this thesis, we propose the following inference methods involving non- and semiparametric as well as likelihood-based approaches. In Chapter 2, which is based on the publication Buhl and Klüppelberg [12] and in the first part of Chapter 4, which is based on Buhl and Klüppelberg [13], we deal with empirical estimation of the extremogram (1.11) of a general regularly varying process $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$. We assume to observe the process on some domain $\mathcal{D}_n \subset \mathbb{R}^d$ whose cardinality increases to infinity as $n \to \infty$. We choose further a set $\mathcal{H} \subset \mathbb{R}^d$ of lag vectors which the estimation is based on, and define for $\boldsymbol{h} \in \mathcal{H}$ the set $D_n(\boldsymbol{h}) = \{\boldsymbol{s} \in \mathcal{D}_n : \boldsymbol{s} + \boldsymbol{h} \in \mathcal{D}_n\}$ as the set of vectors $\boldsymbol{s} \in \mathcal{D}_n$ such that with $\boldsymbol{s}$ also the lagged vector $\boldsymbol{s} + \boldsymbol{h}$ belongs to $\mathcal{D}_n$. For μ-continuous sets A and B in $\overline{\mathbb{R}}\backslash\{0\}$ such that $\mu(A) > 0$ and a sequence $m = m_n \to \infty$ and $m_n = o(n)$ as $n \to \infty$, the empirical extremogram is defined for $\boldsymbol{h} \in \mathcal{H}$ as

$$\widehat{\rho}_{AB,m_n}(\boldsymbol{h}) := \frac{\frac{1}{|\mathcal{D}_n(\boldsymbol{h})|} \sum_{\boldsymbol{s} \in \mathcal{D}_n(\boldsymbol{h})} \mathbb{1}_{\{X(\boldsymbol{s})/a_m \in A, X(\boldsymbol{s}+\boldsymbol{h})/a_m \in B\}}}{\frac{1}{|\mathcal{D}_n|} \sum_{\boldsymbol{s} \in \mathcal{D}_n} \mathbb{1}_{\{X(\boldsymbol{s})/a_m \in A\}}}. \tag{1.15}$$

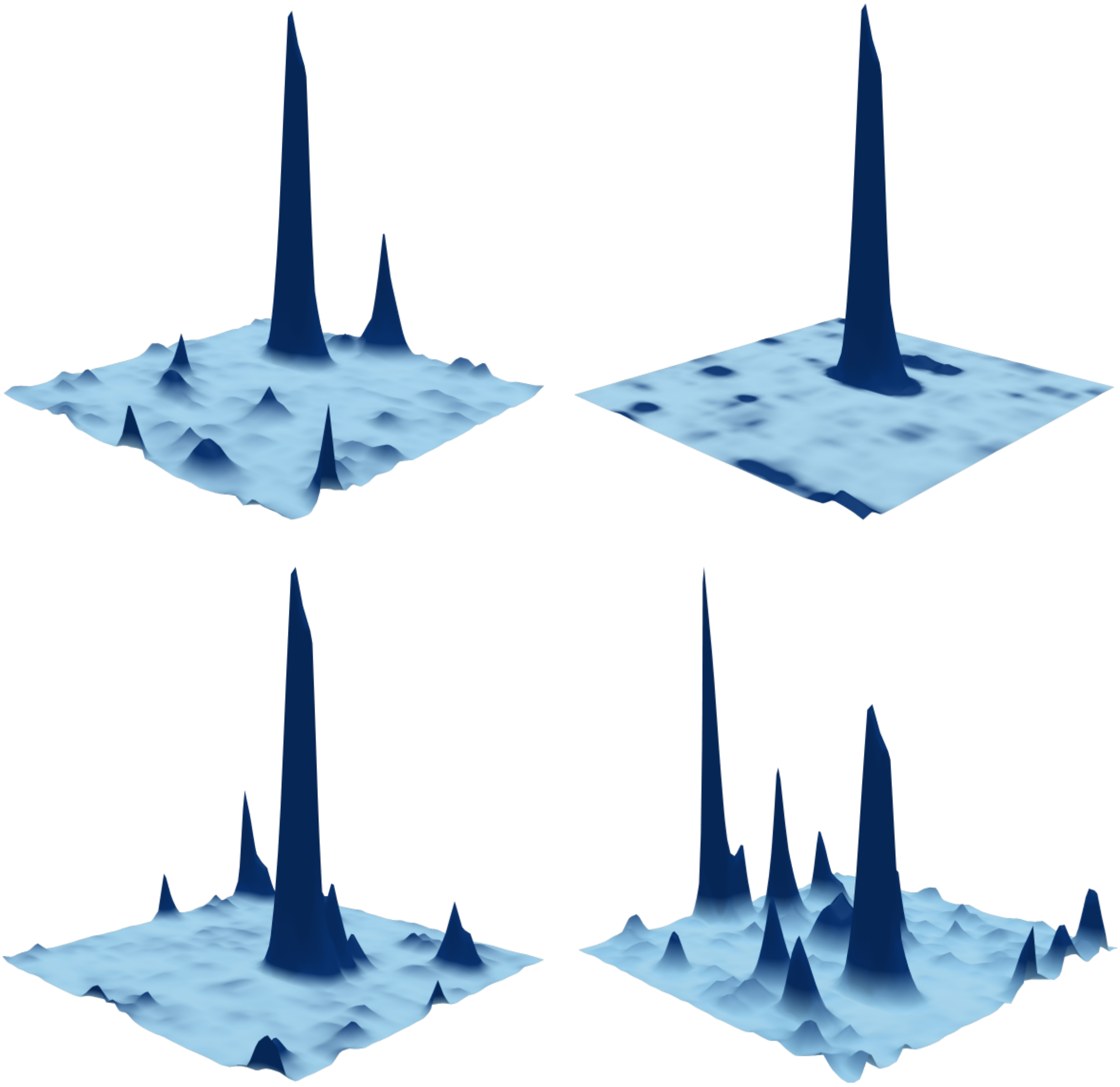

Figure 1.4: Simulation from a time-shifted Brown-Resnick space-time model shown at four consecutive time points (from left to right and top to bottom). The movement and scaling of the largest peak in the top left plot from the upper corner of the grid to the lower corner in the bottom right plot is clearly visible. Also smaller peaks can be observed to emerge and disappear.

A central limit theorem (CLT) for the empirical extremogram $(\widehat{\rho}_{AB,m_n}(\boldsymbol{h}) : \boldsymbol{h} \in \mathcal{H})$ centred by the so-called pre-asymptotic extremogram is proved in [15]. The proof is mainly based on the seminal paper [6] by Bolthausen.

The pre-asymptotic extremogram in the CLT can be replaced by the true one (1.11), if a certain bias condition is satisfied; in particular, the difference between the pre-asymptotic and the true extremogram must vanish with the same rate as the one given in the CLT. However, for many processes the assumptions required in [15] are too restrictive to satisfy this bias condition. In Chapter 2 we explain this in detail by means of two models which exactly fall into this class; the max-moving average process and the Brown-Resnick process (1.8). As the latter, the max-moving average process is max-stable with Fréchet margins.

We then prove a CLT for the empirical extremogram centred by the pre-asymptotic extremogram for strictly stationary regularly varying stochastic processes which relies on weaker conditions than the CLT stated in [15]. Our proof also partly relies on Bolthausen's CLT for spatial processes in [6]; however, we make important modifications so that the bias condition mentioned above can be satisfied, and thus CLTs for the empirical extremogram centred by the true one with optimal rates become possible for many more processes. The proof is based on a big block/small block argument, similarly as in the time series setting considered by [17].

Whether a CLT centred by the true extremogram is possible depends on the particular regularly varying process. If the process has finite-dimensional Fréchet distributions, we can state necessary and sufficient assumptions such that a CLT of that kind is possible, relying on weaker mixing conditions than given in [15]. Furthermore, under conditions such that a CLT centred by the true extremogram is not possible, a bias-corrected estimator can be defined, which we do in Chapter 4. This bias correction can also be introduced in order to improve the rate of convergence.

In Chapter 2 we consider as the observation area $\mathcal{D}_n = \{1, \ldots, n\}^d$ a regular grid that increases in all dimensions, which corresponds to the setting mostly considered so far, cf. [15, 17] or [62]. Irregularly observed data, possibly generated by a Poisson process, have been considered in [15] (also in [62] in the context of pairwise likelihood estimation). The choice of a regular grid $\mathcal{D}_n$ can be extended to arbitrary observation sets provided that they increase to $\mathbb{Z}^d$ and satisfy a central boundary condition required for Bolthausen's CLT in [6].

In the first part of Chapter 4 we extend the observation scheme to settings which are more realistic with regard to practical applications. In particular, instead of assuming that the observation area $\mathcal{D}_n$ increases in all dimensions, it is often more appropriate to assume $\mathcal{D}_n$ to increase in only some but not all dimensions. This means that $\mathcal{D}_n$ can (possibly after reordering) be decomposed into a fixed and an increasing part; i.e.,

$$\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n, \tag{1.16}$$

where for $q, w \in \mathbb{N}$ satisfying $w + q = d$:

(1) $\mathcal{F} \subset \mathbb{Z}^q$ is a fixed domain independent of n, and

(2) $\mathcal{I}_n = \{1, \ldots, n\}^w \subset \mathbb{N}^w$ is an increasing sequence of regular grids satisfying a boundary

condition.

This includes the special case where the observation area is given by

$$\mathcal{D}_n = \mathcal{F} \times \{1, \dots, n\} \tag{1.17}$$

for $\mathcal{F} \subset \mathbb{R}^{d-1}$, and we interpret the observations as generated by a space-time process $\{X(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathbb{R}^{d-1}, t \in [0, \infty)\}$ on a fixed spatial and an increasing temporal domain.

In Chapter 3, which is based on the publication Buhl et al. [14], we present a semiparametric method to estimate the parameters of a parametric model for the extremogram ρ_{AB} defined in (1.11) or the dependence function δ of Brown-Resnick space-time processes defined in (1.8). The model we assume for δ is the space-time fractional class given in (1.13). The method relies on the empirical extremogram (1.15) sampled at different lags $\boldsymbol{h} \in \mathcal{H}$ and on its asymptotic normality. Exploiting a specific closed-form expression of the theoretical extremogram of the Brown-Resnick process which depends on the chosen parametric model for δ, we use the empirical estimates to estimate the model parameters in an ordinary weighted linear regression setting. We separate space and time and estimate the corresponding parameters separately, which is possible due to the additive separability of the space-time model (1.13). We therefore require throughout Chapter 3 an observation scheme such that both the number of temporal and the number of spatial observations increase to infinity, which corresponds to $q = 0$ in (1.16). For the spatial empirical extremogram we apply the CLT with mixing conditions as provided in Chapter 2, and for the timewise estimate that of [17]. We then prove asymptotic normality of the weighted least squares parameter estimates, where constrained optimisation has to be applied, since one of the spatial and one of the time parameters in model (1.13) has bounded support. Also the limit laws differ depending on whether the parameter lies on the boundary or not.

In the second part of Chapter 4, which is based on the publication Buhl and Klüppelberg [13], we generalise the semiparametric estimation method introduced in Chapter 3 in various ways. First of all, the observation scheme (1.16) underlying the empirical extremogram (1.15) is allowed to be much more flexible, see the comments above. Secondly, the new method is not restricted to max-stable Brown-Resnick processes, but can be applied to all regularly varying spatial and space-time processes satisfying appropriate mixing conditions. Motivated by geostatistical least squares variogram estimation as done in Lahiri et al. [48], we work in a generalised least squares regression framework and can assume very general (identifiably parametrised) dependence models for the extremogram (1.11). These models need to satisfy certain regularity conditions which we state in detail. All model parameters (spatial and temporal) are estimated simultaneously; in particular, the method is not restricted to additively separable models as (1.13). We prove consistency and asymptotic normality of the least squares parameter estimates, which respectively rely on consistency and asymptotic normality of the empirical extremogram.

Finally, in Chapter 5, which is based on the publication [11], we propose a purely parametric approach and generalise the pairwise likelihood estimation method proposed in [20] for the space-time Brown-Resnick model (1.13) to the more flexible spatially anisotropic model (1.14). The estimation is based on observations on a regular spatial grid $\mathcal{S}_M = \{1, \dots, M\}^{d-1}$ and at equidistant time points $\mathcal{T} = \{1, \dots T\}$ for M and $T \in \mathbb{N}$; i.e., the observation area is given by

$$\mathcal{D} = \mathcal{D}_{M,T} = \mathcal{S}_M \times \mathcal{T}. \tag{1.18}$$

We show strong consistency and asymptotic normality of the estimates if both M and T tend to infinity, relying on mild regularity and mixing conditions that hold for the space-time Brown-Resnick process model (1.14). We furthermore allow for flexible observation schemes in the sense that, for instance, in (1.18) the number M and thus the number of spatial observations can remain fix such that only the number of observed time points T tends to infinity. This corresponds to Eq. (1.17) above and accounts for the setting of a rather small number of spatial observations compared to a large observed time series, which is a situation often found in real applications. Also in this context we verify the asymptotic properties of the estimates, which requires appropriate adaptations of the required regularity and mixing conditions.

Likelihood-based approaches often yield more accurate estimates than non- or semiparametric methods; however, they can be substantially more time-consuming, see Chapter 6 in [62]. A way to reduce computation time is to use the semiparametric estimates obtained in Chapters 3 and 4 as starting values in the likelihood optimisation routine.

What all estimation methods mentioned above have in common is that, applied to estimation of the parameters of the dependence function δ of a Brown-Resnick process, the asymptotic covariance matrices of the estimates are difficult to access. To produce reliable asymptotic confidence intervals, we therefore use subsampling methods as proposed for random fields in Chapter 5 of Politis et al. [56]. Subsampling requires only weak assumptions to work asymptotically correct, including the existence of continuous limit distributions of the estimates, which holds in all mentioned cases, and a particular mixing condition.

1.3 Outline of this thesis

This PhD thesis consists of the first four of the following six research papers. Each of Chapters 2-5 corresponds to one of the papers [P1]-[P4], respectively, and is self-contained with its own abstract and introduction. Notations and abbreviations might differ among the different chapters since different notations seem reasonable in different settings. There are four appendices at the end of this thesis, one for each chapter.

[P1] S. Buhl and C. Klüppelberg. Limit theory for the empirical extremogram of random fields. *Submitted for publication*, 2017.

[P2] S. Buhl, R. Davis, C. Klüppelberg, and C. Steinkohl. A semiparametric estimation procedure for max-stable space-time processes. *Submitted for publication*, 2017.

[P3] S. Buhl and C. Klüppelberg. Generalised least squares estimation of regularly varying space-time processes based on flexible observation schemes. *Submitted for publication*, 2017.

[P4] S. Buhl and C. Klüppelberg. Anisotropic Brown-Resnick space-time processes: estimation and model assessment. *Extremes*, 19(4): 627-660, 2016.

[P5] E. Boergens, S. Buhl, D. Dettmering, C. Klüppelberg, and F. Seitz. Combination of multi-mission altimetry data along the Mekong River with spatio-temporal kriging. *Journal of Geodesy*, 2016.

[P6] S. Buhl and A.C. Davison. Parametric change point detection in samples of random variables with adjustment for multiple testing under dependency. *In preparation*, 2017.

In Chapter 2 we define the extremogram of regularly varying spatial processes, its empirical estimator and the pre-asymptotic version. We state conditions under which the empirical extremogram centred by the pre-asymptotic version is asymptotically normal and compare our conditions to those of [15]. Under slightly stronger conditions we show a CLT centred by the true extremogram. As two specific examples we examine the extremogram and its estimator for the max-moving average process and the Brown-Resnick process.

In Chapter 3 we introduce the new two-step semiparametric method to estimate the parameters of the Brown-Resnick space-time model (1.13), which is based on a closed-form expression of its empirical extremogram, and define the ordinary weighted least squares estimator. We state and verify its asymptotic properties. We examine the finite sample behaviour of the method in a simulation study and conclude with an analysis of the Florida rainfall data visualised in Figures 1.1-1.3.

The first part of Chapter 4 generalises the results concerning estimation of the empirical extremogram of regularly varying processes obtained in Chapter 2 to very flexible observation schemes as described in (1.16). For processes with Fréchet marginal distributions we give precise conditions under which a CLT centred by the true extremogram can be obtained. In case those conditions are not satisfied we introduce a bias-corrected empirical extremogram.

In the second part of Chapter 4 we generalise the semiparametric estimation method introduced in Chapter 3 in the various ways described in Section 1.2. In particular, we define the generalised least squares estimator and verify its properties of consistency and asymptotic normality. We apply the method to Brown-Resnick space-time processes. We give a list of existing and new parametric models for its dependence function δ and estimate their parameters in the course of a simulation study.

Chapter 5 is dedicated to pairwise likelihood estimation of the new spatially anisotropic Brown-Resnick space-time model (1.14) from data that is observed within a flexible observation scheme. We state and verify the asymptotic properties of the pairwise likelihood estimates. We show the method at work for finite samples in an analysis of the Florida rainfall data also analysed in Chapter 3 and apply a variety of goodness-of-fit test procedures. In particular, we reject the hypothesis of spatial isotropy of the process underlying the observations, thus concluding that an anisotropic model is more appropriate than the isotropic model applied in Chapter 3. Based on the model fit, we furthermore produce conditional probability fields answering questions of the type "Conditional on extreme rainfall at location $\boldsymbol{s}_1$ at time t_1, what is the probability of extreme rainfall at a (nearby) location $\boldsymbol{s}_2$ at time t_2?", where "extreme rainfall" is specified by means of a large empirical quantile.

Chapter 2

Limit theory for the empirical extremogram of random fields

Abstract

Regularly varying stochastic processes are able to model extremal dependence between process values at locations in random fields. We investigate the empirical extremogram as an estimator of dependence in the extremes. We provide conditions to ensure asymptotic normality of the empirical extremogram centred by a pre-asymptotic version. The proof relies on a CLT for exceedance variables. For max-stable processes with Fréchet margins we provide conditions such that the empirical extremogram centred by its true version is asymptotically normal. The results of this chapter apply to a variety of spatial and space-time processes, and to time series models. We apply our results to max-moving average processes and Brown-Resnick processes.

AMS 2010 Subject Classifications: primary: 60F05, 60G70, 62G32; secondary: 37A25, 62M30

Keywords: Brown-Resnick process; empirical extremogram; extremogram; max-moving average process; max-stable process; random field; spatial CLT; spatial mixing

2.1 Introduction

The extremogram measures extremal dependence in a strictly stationary regularly varying stochastic process and can hence be seen as a correlogram for extreme events. It was introduced in Davis and Mikosch [17] for time series (also in Fasen et al. [35]), and they show consistency and asymptotic normality of an empirical extremogram under weak mixing conditions. Davis et al. [21] give a profound review of the estimation theory for time series with various examples. For a discussion of the role of the extremogram in dependence modelling of extremes we refer again to [17]. As it is spelt out there, it is the covariance function of indicator functions of exceedance events in an asymptotic sense. Also in that paper classical mixing conditions as presented in

Ibragimov and Linnik [45], on which we rely in our work, are compared to the extreme mixing conditions D and D' often used in extreme value theory (cf. Embrechts et al. [32], Section 4.4, and Leadbetter et al. [49], Sections 3.1 and 3.2).

The extremogram and its empirical estimate have been formulated for spatial d-dimensional stochastic processes by Cho et al. [15] and for space-time processes in Buhl et al. [14] and Steinkohl [62], when observed on a regular grid. The extremogram is defined for strictly stationary regularly varying stochastic processes, where all finite-dimensional distributions are in the maximum domain of attraction of some Fréchet distribution. Among other results, based on the seminal paper [6] by Bolthausen, [15] prove a CLT for the empirical extremogram sampled at different spatial lags, centred by the so-called pre-asymptotic extremogram. Such results also compare with a CLT for sample space-time covariance estimators derived in Li et al. [51], also based on [6].

The pre-asymptotic extremogram can be replaced in the CLT by the true one, if a certain bias condition is satisfied; in particular, the difference between the pre-asymptotic and the true extremogram must vanish with the same rate as the one given in the CLT. However, for many processes the assumptions required in [15] are too restrictive to satisfy this bias condition. We explain this in detail and present two models which exactly fall into this class; the max-moving average process and the Brown-Resnick process. These two processes are max-stable with Fréchet margins.

In this chapter, we prove a CLT for the empirical extremogram centred by the pre-asymptotic extremogram for strictly stationary regularly varying stochastic processes, which relies on weaker conditions than the CLT stated in [15]. Our proof also partly relies on Bolthausen's CLT for spatial processes in [6]; however, we make important modifications so that the bias condition mentioned above can be satisfied, and thus a CLT for the empirical extremogram centred by the true one for many more processes becomes possible. The proof is based on a big block/small block argument, similarly to [17].

Our interest is of course in a CLT centred by the true extremogram, and whether such a CLT is possible depends on the specific regularly varying process. If the process has finite-dimensional max-stable distributions, in our framework equivalent to having finite-dimensional Fréchet distributions, we can give conditions such that a CLT of that kind is possible. Here we need the weaker mixing conditions of our version of Bolthausen's CLT compared to [15]. Furthermore, under conditions such that a CLT centred by the true extremogram is not possible, a bias-corrected estimator can be defined, which we do in Chapter 3 for the Brown-Resnick process and in Chapter 4 for more general processes with Fréchet marginal distributions.

This chapter is organised as follows. In Section 2.2 we present the general model class of strictly stationary regularly varying processes in $\mathbb{R}^d$ for $d \in \mathbb{N}$. We also define here the extremogram for such processes. In Section 2.3 we define the empirical extremogram based on grid observations, and also the pre-asymptotic extremogram. Section 2.4 is devoted to the CLT for the empirical extremogram centred by the pre-asymptotic extremogram and to our examples of max-stable spatial processes; max-moving average processes and Brown-Resnick processes. We discuss in detail the problem of a CLT for the empirical extremogram and compare our new conditions for the CLT to hold with those in previous work, particularly with those given in Cho et al. [15]. For processes with Fréchet margins we prove a CLT for the empirical extremogram centred by the true extremogram. The proof of the CLT is given in Section 2.5.

2.2 Regularly varying spatial processes

As a natural model class in extreme value theory we consider strictly stationary regularly varying processes $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ for $d \in \mathbb{N}$, where all finite-dimensional distributions are regularly varying (cf. Hult and Lindskog [41] for definitions and results in a general framework and Resnick [59] for details about multivariate regular variation). As a prerequisite, we define for every finite set $\mathcal{I} \subset \mathbb{R}^d$ the vector

$$X_{\mathcal{I}} := (X(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{I})^{\intercal}.$$

Throughout we assume that $X_{\mathcal{I}}$ inherits the strict stationarity from $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$, which is guaranteed, if we consider lagged vectors of $X_{\mathcal{I}}$. Furthermore, $|\mathcal{I}|$ denotes the cardinality of $\mathcal{I}$. As usual, $f(n) \sim g(n)$ as $n \to \infty$ means that $f(n)/g(n) \to 1$ for two positive functions f and g.

Definition 2.1 (Regularly varying process)**.** *A strictly stationary stochastic process $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ is called* regularly varying*, if there exists some normalising sequence $0 < a_n \to \infty$ such that $\mathbb{P}(|X(\boldsymbol{0})| > a_n) \sim n^{-d}$ as $n \to \infty$ and for every finite set $\mathcal{I} \subset \mathbb{R}^d$*

$$n^d \mathbb{P}\Big(\frac{X_{\mathcal{I}}}{a_n} \in \cdot\Big) \overset{v}{\to} \mu_{\mathcal{I}}(\cdot), \quad n \to \infty, \tag{2.1}$$

for some non-null Radon measure $\mu_{\mathcal{I}}$ on the Borel sets in $\overline{\mathbb{R}}^{|\mathcal{I}|}\backslash\{\boldsymbol{0}\}$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. In that case,

$$\mu_{\mathcal{I}}(xC) = x^{-\beta}\mu_{\mathcal{I}}(C), \quad x > 0,$$

for every Borel set $C \subset \overline{\mathbb{R}}^{|\mathcal{I}|}\backslash\{\boldsymbol{0}\}$. The notation $\overset{v}{\to}$ stands for vague convergence, and $\beta > 0$ is called the index of regular variation*.*

For every $\boldsymbol{s} \in \mathbb{R}^d$ and $\mathcal{I} = \{\boldsymbol{s}\}$ we set $\mu_{\{\boldsymbol{s}\}}(\cdot) = \mu_{\{\boldsymbol{0}\}}(\cdot) =: \mu(\cdot)$, which is justified by stationarity. The focus of the present chapter is on the extremogram, defined for values in $\mathbb{R}^d$ as follows.

Definition 2.2 (Extremogram)**.** *Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ be a strictly stationary regularly varying process in $\mathbb{R}^d$. For two μ-continuous Borel sets A and B in $\overline{\mathbb{R}}\backslash\{0\}$ (i.e., $\mu(\partial A) = \mu(\partial B) = 0$) such that $\mu(A) > 0$, the* extremogram *is defined as*

$$\rho_{AB}(\boldsymbol{h}) = \lim_{n\to\infty} \frac{\mathbb{P}(X(\boldsymbol{0})/a_n \in A, X(\boldsymbol{h})/a_n \in B)}{\mathbb{P}(X(\boldsymbol{0})/a_n \in A)}, \quad \boldsymbol{h} \in \mathbb{R}^d. \tag{2.2}$$

Our goal is to estimate the extremogram for arbitrary strictly stationary regularly varying processes by its empirical version and prove asymptotic properties like consistency and asymptotic normality.

Such results also allow for semiparametric estimation in a parametric spatial or space-time model as presented in Chapters 3 and 4.

Analogous asymptotic results for the empirical extremogram of time series have been shown in Davis and Mikosch [17] and of d-dimensional random fields in Cho et al. [15]. However, in certain situations, for example in the case of the Brown-Resnick process (2.26), the rates obtained in [15] are too crude to allow for a CLT. We apply a small block/large block argument in space (similarly to [17] for time series), which leads to more precise rates in the CLT. Arguments of our proof are based on spatial mixing conditions, and rely on general results of Bolthausen [6] and Ibragimov and Linnik [45].

2.3 Large sample properties of the spatial empirical extremogram

The estimation of the extremogram is based on data observed on

$$\mathcal{S}_n = \{1,\dots,n\}^d = \{\boldsymbol{s}_j : j = 1,\dots,n^d\}, \text{ the regular grid of side length } n.$$

Let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^d$. Define the following quantities for $\gamma > 0$:

$$\begin{aligned} &B(\mathbf{0},\gamma) = \big\{\boldsymbol{s} \in \mathbb{Z}^d : \|\boldsymbol{s}\| \le \gamma\big\}, \\ &B(\boldsymbol{s},\gamma) = \big\{\boldsymbol{s'} \in \mathbb{Z}^d : \|\boldsymbol{s} - \boldsymbol{s'}\| \le \gamma\big\} = \boldsymbol{s} + B(\mathbf{0},\gamma), && (2.3) \\ &\mathcal{H} \subseteq \{\boldsymbol{h} = \boldsymbol{s} - \boldsymbol{s}' : \boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}_n\} \cap B(\mathbf{0},\gamma), \text{ a finite set of observed lags.} && (2.4) \end{aligned}$$

We further define the vectorised process $\{\boldsymbol{Y}(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ by

$$\boldsymbol{Y}(\boldsymbol{s}) := X_{B(\boldsymbol{s},\gamma)};$$

i.e., $\boldsymbol{Y}(\boldsymbol{s})$ is the vector of values of X with indices in $B(\boldsymbol{s},\gamma)$ as defined in (2.3). We introduce the balls $B(\mathbf{0},\gamma)$ in order to express events like $\{X(\boldsymbol{s}) \in A, X(\boldsymbol{s}+\boldsymbol{h}) \in B\}$ or $\{X(\boldsymbol{s}) \in A\}$ for $\boldsymbol{s} \in \mathbb{R}^d$ and $\boldsymbol{h} \in \mathcal{H} \subseteq B(\mathbf{0},\gamma)$ as well as Borel sets A, B in $\overline{\mathbb{R}}\backslash\{0\}$ through events $\{\boldsymbol{Y}(\boldsymbol{s}) \in C\}$ for appropriately chosen Borel sets C in $\overline{\mathbb{R}}^{|B(\mathbf{0},\gamma)|}\backslash\{\mathbf{0}\}$. This notation simplifies the presentation of the proofs of consistency and asymptotic normality considerably.

For $j \in \mathbb{N}$ let $\boldsymbol{e}_j$ be the j-th unit vector in $\mathbb{R}^d$. The choice of a regular grid $\mathcal{S}_n$ can be extended to arbitrary observation sets provided that they increase to $\mathbb{Z}^d$ and have boundaries $\partial\mathcal{S}_n := \{\boldsymbol{s} \in \mathcal{S}_n : \exists\, \boldsymbol{z} \in \mathbb{Z}^d \setminus \mathcal{S}_n \text{ and } j \in \mathbb{N} \text{ with } \|\boldsymbol{z} - \boldsymbol{s}\| = \|\boldsymbol{e}_j\|\}$ satisfying $\lim_{n\to\infty} |\partial\mathcal{S}_n|/|\mathcal{S}_n| = 0$. The natural extension to grids with different side lengths does not involve any additional mathematical difficulty, but notational complexity, since our proofs are based on big/small block arguments common in extreme value statistics, which are much simpler to formulate for a regular grid.

For fixed n and observations on the grid $\mathcal{S}_n$ there will be points $\boldsymbol{s} \in \mathcal{S}_n$ near the boundary, such that not all components of $\boldsymbol{Y}(\boldsymbol{s})$ can be observed. However, since we investigate asymptotic properties for $\mathcal{S}_n$ and the boundary points become negligible, this is irrelevant for our results and we suppress such technical details. As a consequence, our results apply to spatial as well as time series observations, thus include the frameworks considered in Cho et al. [15] and Davis and Mikosch [17].

We shall also need the following relations and definitions, where the limits exist by regular variation of $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$. Let C be a $\mu_{B(\mathbf{0},\gamma)}$-continuous Borel set in $\overline{\mathbb{R}}^{|B(\mathbf{0},\gamma)|} \backslash \{\mathbf{0}\}$ and $C \times D$ a $\tau_{B(\mathbf{0},\gamma)\times B(\boldsymbol{h},\gamma)}$-continuous Borel set in the product space, where we define

$$\mu_{B(\mathbf{0},\gamma)}(C) := \lim_{n\to\infty} n^d \mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_n} \in C\Big), \tag{2.5}$$

$$\tau_{B(\mathbf{0},\gamma)\times B(\boldsymbol{h},\gamma)}(C \times D) := \lim_{n\to\infty} n^d \mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_n} \in C, \frac{\boldsymbol{Y}(\boldsymbol{h})}{a_n} \in D\Big). \tag{2.6}$$

We enumerate the lags in $\mathcal{H}$ by $\mathcal{H} = \{\boldsymbol{h}_1, \dots, \boldsymbol{h}_p\}$. Following ideas of Davis and Mikosch [17] (also used in Cho et al. [15]) we define $\mu_{B(\mathbf{0},\gamma)}$-continuous Borel sets $D_1, \dots, D_p, D_{p+1}$ in $\overline{\mathbb{R}}^{|B(\mathbf{0},\gamma)|} \backslash \{\mathbf{0}\}$ by the property

$$\{\boldsymbol{Y}(\boldsymbol{s}) \in D_i\} = \{X(\boldsymbol{s}) \in A, X(\boldsymbol{s} + \boldsymbol{h}_i) \in B\} \tag{2.7}$$

for $i = 1, \dots, p$, and $\{\boldsymbol{Y}(\boldsymbol{s}) \in D_{p+1}\} = \{X(\boldsymbol{s}) \in A\}$. Note in particular that, by the relation between $\{\boldsymbol{Y}(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ and $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ and regular variation, for every μ-continuous Borel set A in $\overline{\mathbb{R}} \setminus \{0\}$,

$$\mu_{B(\mathbf{0},\gamma)}(D_{p+1}) = \lim_{n\to\infty} n^d \mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_n} \in D_{p+1}\Big) = \lim_{n\to\infty} n^d \mathbb{P}\Big(\frac{X(\mathbf{0})}{a_n} \in A\Big) = \mu(A).$$

The extremogram can be estimated from data by the following empirical version.

Definition 2.3 (Empirical extremogram)**.** *Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ be a strictly stationary regularly varying process in $\mathbb{R}^d$, observed on $\mathcal{S}_n$, and set $\mathcal{S}_n(\boldsymbol{h}) := \{\boldsymbol{s} \in \mathcal{S}_n : \boldsymbol{s} + \boldsymbol{h} \in \mathcal{S}_n\}$ for $\boldsymbol{h} \in \mathcal{H}$. Let A and B be μ-continuous Borel sets in $\overline{\mathbb{R}}\backslash\{0\}$ such that $\mu(A) > 0$. For a sequence $m = m_n \to \infty$ and $m_n = o(n)$ as $n \to \infty$, the* empirical extremogram *is defined for $\boldsymbol{h} \in \mathcal{H}$ as*

$$\widehat{\rho}_{AB,m_n}(\boldsymbol{h}) := \frac{\frac{1}{|\mathcal{S}_n(\boldsymbol{h})|} \sum_{\boldsymbol{s}\in\mathcal{S}_n(\boldsymbol{h})} \mathbb{1}_{\{X(\boldsymbol{s})/a_m \in A, X(\boldsymbol{s}+\boldsymbol{h})/a_m \in B\}}}{\frac{1}{|\mathcal{S}_n|} \sum_{\boldsymbol{s}\in\mathcal{S}_n} \mathbb{1}_{\{X(\boldsymbol{s})/a_m \in A\}}}. \tag{2.8}$$

The following pre-asymptotic extremogram plays an important role when proving asymptotic normality of the empirical extremogram (2.8).

Definition 2.4 (Pre-asymptotic extremogram)**.** *Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ be a strictly stationary regularly varying process in $\mathbb{R}^d$. Let A and B be μ-continuous Borel sets in $\overline{\mathbb{R}}\backslash\{0\}$ such that $\mu(A) > 0$. For a sequence $m = m_n \to \infty$ and $m_n = o(n)$ as $n \to \infty$, the* pre-asymptotic extremogram *is defined as*

$$\rho_{AB,m_n}(\boldsymbol{h}) = \frac{P\left(X(\mathbf{0})/a_m \in A, X(\boldsymbol{h})/a_m \in B\right)}{\mathbb{P}(X(\mathbf{0})/a_m \in A)}. \tag{2.9}$$

The next section is devoted to the asymptotic properties of the empirical extremogram and the inherent bias-variance problem with its solution.

2.4 Consistency and CLT for the empirical extremogram

In this section we derive relevant asymptotic properties of the empirical extremogram. First we establish large sample properties of the empirical estimator of $\mu_{B(\mathbf{0},\gamma)}(C)$. Based on these results, the asymptotic normality is established.

Throughout this section we assume that $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ is a strictly stationary regularly varying process in $\mathbb{R}^d$, observed on $\mathcal{S}_n$.

We need the concept of α-mixing for such processes; see e.g. Bolthausen [6] or Doukhan [30].

Definition 2.5 (α-mixing coefficients and α-mixing)**.** *Consider a strictly stationary random field* $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ *and let* $d(\cdot,\cdot)$ *be some metric induced by a norm* $\|\cdot\|$ *on* $\mathbb{R}^d$. *For* $\Lambda_1, \Lambda_2 \subset \mathbb{Z}^d$ *set*

$$d(\Lambda_1, \Lambda_2) := \inf\left\{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| : \ \boldsymbol{s}_1 \in \Lambda_1, \boldsymbol{s}_2 \in \Lambda_2\right\}.$$

Further, for $i = 1, 2$ *denote* $\mathcal{F}_{\Lambda_i} = \sigma\left\{X(\boldsymbol{s}) : \boldsymbol{s} \in \Lambda_i\right\}$ *the* σ*-algebra generated by* $\{X(\boldsymbol{s}) : \ \boldsymbol{s} \in \Lambda_i\}$.

(i) The α-mixing coefficients *are defined for* $k, \ell \in \mathbb{N} \cup \{\infty\}$ *and* $r \geq 0$ *by*

$$\begin{aligned}\alpha_{k,\ell}(r) := \sup\{&|\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| : \\ &A_i \in \mathcal{F}_{\Lambda_i}, |\Lambda_1| \leq k, |\Lambda_2| \leq \ell, d(\Lambda_1, \Lambda_2) \geq r\}.\end{aligned} \tag{2.10}$$

(ii) The random field is called α-mixing, *if* $\alpha_{k,l}(r) \to 0$ *as* $r \to \infty$ *for all* $k, \ell \in \mathbb{N}$.

In what follows we have to control the dependence between vector processes $\{\boldsymbol{Y}(\boldsymbol{s}) = X_{B(\boldsymbol{s},\gamma)} : s \in \Lambda_1'\}$ and $\{\boldsymbol{Y}(\boldsymbol{s}) = X_{B(\boldsymbol{s},\gamma)} : \boldsymbol{s} \in \Lambda_2'\}$ for subsets $\Lambda_i' \subset \mathbb{Z}^d$ with cardinalities $|\Lambda_1'| \leq k$ and $|\Lambda_2'| \leq \ell$. In the context of Definition 2.5, this means that the Λ_i in (2.10) are unions of balls $\Lambda_i = \cup_{\boldsymbol{s} \in \Lambda_i'} B(\boldsymbol{s}, \gamma)$. Since $\gamma > 0$ is some predetermined finite constant independent of n, we keep notation simple by redefining the α-mixing coefficients with respect to the vector processes as

$$\begin{aligned}\alpha_{k,\ell}(r) := \sup\{&|\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| : \\ &A_i \in \mathcal{F}_{\Lambda_i}, \ \Lambda_i = \cup_{\boldsymbol{s} \in \Lambda_i'} B(\boldsymbol{s}, \gamma), \ |\Lambda_1'| \leq k, |\Lambda_2'| \leq \ell, d(\Lambda_1', \Lambda_2') \geq r\}.\end{aligned} \tag{2.11}$$

Theorem 2.6. *Let* $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ *be a strictly stationary regularly varying process, observed on* $\mathcal{S}_n$ *and let* $\mathcal{H} = \{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_p\}$ *be a finite set of lags in* $\mathbb{Z}^d$ *satisfying* $\mathcal{H} \subseteq B(\mathbf{0}, \gamma)$ *for some* $\gamma > 0$. *Suppose that the following conditions are satisfied:*

(M1) $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ *is* α*-mixing with* α*-mixing coefficients* $\alpha_{k,\ell}(r)$ *defined in* (2.10).

There exist sequences $m = m_n, r = r_n \to \infty$ *with* $m_n/n \to 0$ *and* $r_n/m_n \to 0$ *as* $n \to \infty$ *such that the following hold:*

(M2) $m_n^2 r_n^2 / n \to 0$.

(M3) For all $\epsilon > 0$:

$$\lim_{k\to\infty}\limsup_{n\to\infty}\sum_{\boldsymbol{h}\in\mathbb{Z}^d:k<\|\boldsymbol{h}\|\le r_n} m_n^d \\ \mathbb{P}\Big(\max_{\boldsymbol{s}\in B(\boldsymbol{0},\gamma)}|X(\boldsymbol{s})|>\epsilon a_m,\ \max_{\boldsymbol{s}'\in B(\boldsymbol{h},\gamma)}|X(\boldsymbol{s}')|>\epsilon a_m\Big)=0.$$

(M4) *(i)* $\lim_{n\to\infty} m_n^d \sum_{\boldsymbol{h}\in\mathbb{Z}^d:\|\boldsymbol{h}\|>r_n}\alpha_{1,1}(\|\boldsymbol{h}\|)=0,$

(ii) $\sum_{\boldsymbol{h}\in\mathbb{Z}^d}\alpha_{p,q}(\|\boldsymbol{h}\|)<\infty$ *for* $2\le p+q\le 4,$

(iii) $\lim_{n\to\infty} m_n^{d/2}n^{d/2}\ \alpha_{1,n^d}(r_n)=0,$

Then the empirical extremogram $\widehat{\rho}_{AB,m_n}(\boldsymbol{h})$ for $\boldsymbol{h}\in\mathcal{H}$ as in (2.8)*, centred by the pre-asymptotic extremogram in* (2.9)*, is asymptotically normal; more precisely,*

$$\Big(\frac{n}{m_n}\Big)^{d/2}\big(\widehat{\rho}_{AB,m_n}(\boldsymbol{h})-\rho_{AB,m_n}(\boldsymbol{h})\big)_{\boldsymbol{h}\in\mathcal{H}}\xrightarrow{d}\mathcal{N}(\boldsymbol{0},\Pi),\quad n\to\infty, \tag{2.12}$$

where $\Pi=\mu(A)^{-4}F\Sigma F^\intercal\in\mathbb{R}^{p\times p}$, and the matrix $\Sigma\in\mathbb{R}^{(p+1)\times(p+1)}$ has for $1\le i,j\le p+1$ components

$$\Sigma_{ii}=\mu_{B(\boldsymbol{0},\gamma)}(D_i)+\sum_{\boldsymbol{h}\in\mathbb{Z}^d\setminus\{\boldsymbol{0}\}}\tau_{B(\boldsymbol{0},\gamma)\times B(\boldsymbol{h},\gamma)}(D_i\times D_i)=:\sigma^2_{B(\boldsymbol{0},\gamma)}(D_i), \tag{2.13}$$

$$\Sigma_{ij}=\mu_{B(\boldsymbol{0},\gamma)}(D_i\cap D_j)+\sum_{\boldsymbol{h}\in\mathbb{Z}^d\setminus\{\boldsymbol{0}\}}\tau_{B(\boldsymbol{0},\gamma)\times B(\boldsymbol{h},\gamma)}(D_i\times D_j),\quad i\ne j. \tag{2.14}$$

The matrix F consists of a diagonal matrix F_1 and a vector F_2 in the last column:

$$F=[F_1,F_2]\quad\textit{with} \\ F_1=\operatorname{diag}(\mu(A))\in\mathbb{R}^{p\times p},\quad F_2=(-\mu_{B(\boldsymbol{0},\gamma)}(D_1),\dots,-\mu_{B(\boldsymbol{0},\gamma)}(D_p))^\intercal. \tag{2.15}$$

Remark 2.7. (i) In the proof (given in Section 2.5) we use a big block/small block argument. For simplicity we assume that n/m_n is an integer and subdivide $\mathcal{S}_n$ into $(n/m_n)^d$ non-overlapping d-dimensional blocks with side length m_n. Theorem 1 of Cho et al. [15] divides n^d into n^d/m_n blocks; i.e., their sequence m_n corresponds to our m_n^d, so that their assumptions look slightly different. In our notation, they require that $m_n^{2(d+1)}/n\to 0$ as $n\to\infty$, which is more restrictive than condition (M2) of $m_n^2r_n^2/n\to 0$, and indeed too restrictive for processes such as the max-moving average process and the Brown-Resnick process discussed below.

(ii) If the choice $m_n=n^{\beta_1}$ and $r_n=n^{\beta_2}$ with $0<\beta_2<\beta_1<1$ satisfies conditions (M3) and (M4), then for $\beta_1\in(0,1/2)$ and $\beta_2\in(0,\min\{\beta_1;1/2-\beta_1\})$ the condition (M2) also holds and we obtain the CLT (2.12).

The pre-asymptotic extremogram (2.9) in the central limit theorem can be replaced by the theoretical one (2.2), if it converges to the theoretical extremogram with the same convergence rate as the empirical extremogram to the pre-asymptotic extremogram; i.e., if

$$\Big(\frac{n}{m_n}\Big)^{d/2}\,(\rho_{AB,m_n}(\boldsymbol{h})-\rho_{AB}(\boldsymbol{h}))_{\boldsymbol{h}\in\mathcal{H}}\to\mathbf{0},\quad n\to\infty. \tag{2.16}$$

Relation (2.16) does not hold for every strictly stationary regularly varying spatial process or time series for which (2.12) is satisfied. Theorem 2.8 states a necessary and sufficient condition for processes with Fréchet marginal distributions such that both (2.12) and (2.16) hold. For general regularly varying stochastic processes such a result would hold under second order conditions on the finite-dimensional regularly varying distributions of the process, but we do not pursue this topic further.

Theorem 2.8 (CLT for processes with Fréchet margins)**.** *Let $\{X(\boldsymbol{s}) : \boldsymbol{s}\in\mathbb{R}^d\}$ be a strictly stationary max-stable process in $\mathbb{R}^d$ with standard unit Fréchet margins. Let ρ_{AB} be its extremogram* (2.2) *and ρ_{AB,m_n} the corresponding pre-asymptotic version* (2.9) *for sets $A=(a_1,a_2)$ and $B=(b_1,b_2)$ with $0<a_1<a_2\le\infty$ and $0<b_1<b_2\le\infty$. Assume that the process is observed on $\mathcal{S}_n$ and let $\mathcal{H}=\{\boldsymbol{h}_1,\dots,\boldsymbol{h}_p\}$ be a finite set of lags in $\mathbb{Z}^d$ satisfying $\mathcal{H}\subseteq B(\mathbf{0},\gamma)$ for some $\gamma>0$. Furthermore, suppose that conditions (M1)–(M4) of Theorem 2.6 hold for appropriately chosen sequences $m_n, r_n\to\infty$. Then the limit relation* (2.16) *holds if and only if $n/m_n^3\to 0$ as $n\to\infty$. In this case we obtain*

$$\Big(\frac{n}{m_n}\Big)^{d/2}\big(\widehat{\rho}_{AB,m_n}(\boldsymbol{h})-\rho_{AB}(\boldsymbol{h})\big)_{\boldsymbol{h}\in\mathcal{H}}\stackrel{d}{\to}\mathcal{N}(\mathbf{0},\Pi),\quad n\to\infty, \tag{2.17}$$

where Π is given in Theorem 2.6.

Proof. First note that, since all finite-dimensional distributions are max-stable distributions with standard unit Fréchet margins, they are multivariate regularly varying. Let $V_2(\boldsymbol{h};\cdot,\cdot)$ be the bivariate exponent measure defined through $\mathbb{P}(X(\mathbf{0})\le x_1, X(\boldsymbol{h})\le x_2)=\exp\{-V_2(\boldsymbol{h};x_1,x_2)\}$ for $x_1,x_2>0$, cf. Beirlant et al. [3], Section 8.2.2. From Lemma A.1(b) we know that for $\boldsymbol{h}\in\mathcal{H}$

$$\rho_{AB,m_n}(\boldsymbol{h})=\Big[\rho_{AB}(\boldsymbol{h})+\frac{1}{2\,m_n^d}\overline{V}_2^2(\boldsymbol{h})\Big](1+o(1)),\quad n\to\infty,$$

where $\overline{V}_2^2(\boldsymbol{h}):=\frac{a_1a_2}{a_2-a_1}(V_2^2(\boldsymbol{h};a_2,b_2)+V_2^2(\boldsymbol{h};a_2,b_1)+V_2^2(\boldsymbol{h};a_1,b_2)+V_2^2(\boldsymbol{h};a_1,b_1))$ as given in (A.3) and appropriate adaptations for $a_2=\infty$ and/or $b_2=\infty$ given in (A.4). Hence, for $\boldsymbol{h}\in\mathcal{H}$,

$$\Big(\frac{n}{m_n}\Big)^{d/2}\big(\rho_{AB,m_n}(\boldsymbol{h})-\rho_{AB}(\boldsymbol{h})\big)\sim\Big(\frac{n}{m_n^3}\Big)^{d/2}\frac{\overline{V}_2^2(\boldsymbol{h})}{2},$$

which converges to 0 if and only if $n/m_n^3\to 0$. $\square$

Remark 2.9. The requirement $n/m_n^3\to 0$ as $n\to\infty$ needed in Theorem 2.8 contradicts the condition $m_n^{2(d+1)}/n\to 0$ required in Cho et al. [15]; thus, under the conditions stated in that paper, only the CLT (2.12) centred by the pre-asymptotic extremogram can be proved. However, $n/m_n^3\to 0$ as $n\to\infty$ does not contradict the assumptions of Theorem 2.6 above, in particular, the much weaker assumption (M2).

(ii) From Theorem 2.8 in relation to Remark 2.7 (ii) we conclude that we need to choose $\beta_1 > 1/3$ in order to satisfy the CLT (2.17). This is not a contradiction to the conditions of Theorem 2.6 and we conclude that for $\beta_1 \in (1/3, 1/2)$ and $\beta_2 \in (0, \min\{\beta_1; 1/2 - \beta_1\})$, we have

$$n^{\frac{d}{2}(1-\beta_1)}\big(\widehat{\rho}_{AB,m_n}(\boldsymbol{h}) - \rho_{AB}(\boldsymbol{h})\big)_{\boldsymbol{h}\in\mathcal{H}} \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \Pi), \quad n \to \infty. \tag{2.18}$$

We discuss our findings for two prominent examples.

Example 2.10. [Max-moving average (MMA) process]
We start with a model for the process $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{Z}^d\}$ corresponding to the discrete observation scheme. It has been suggested in Cho et al. [15] based on a time series model of Davis and Resnick [18]. Let $Z(\boldsymbol{s})$ for $\boldsymbol{s} \in \mathbb{Z}^d$ be i.i.d. standard unit Fréchet random variables and set

$$X^\star(\boldsymbol{s}) := \max_{\boldsymbol{z}\in\mathbb{Z}^d} \phi^{\|\boldsymbol{z}\|} Z(\boldsymbol{s} - \boldsymbol{z}), \quad \boldsymbol{s} \in \mathbb{Z}^d, \tag{2.19}$$

for some $0 < \phi < 1$. Then $\{X^\star(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{Z}^d\}$ is a stationary *max-moving average* (MMA) process, also considered in equation (25) of [15]. As in [15] we deduce the following marginal distributions. The number $N(j)$ of lag vectors $\boldsymbol{h} \in \mathbb{Z}^d$ with norm $j = \|\boldsymbol{h}\|$ is of order $\mathcal{O}(j^{d-1})$ and

$$V_1 := \sum_{\boldsymbol{h}\in\mathbb{Z}^d} \phi^{\|\boldsymbol{h}\|} = \sum_{0\le j<\infty} \phi^j N(j) < \infty.$$

The univariate margins are unit Fréchet with scale parameter V_1; i.e.,

$$\mathbb{P}(X^\star(\boldsymbol{0}) \le x) = \exp\Big\{-\frac{V_1}{x}\Big\}, \quad x > 0.$$

Define $Q_{\boldsymbol{h}}(j) := \big|\{\boldsymbol{s} \in \mathbb{Z}^d : \min\{\|\boldsymbol{s}\|, \|\boldsymbol{s}+\boldsymbol{h}\|\} = j\}\big| \le 2N(j)$. With $V_2(\boldsymbol{h}) := V_2(\boldsymbol{h}; 1, 1) = \sum_{0\le j<\infty} Q_{\boldsymbol{h}}(j)\,\phi^j$, we have for the bivariate margins at lag $\boldsymbol{h} \in \mathbb{Z}^d$,

$$\mathbb{P}(X^\star(\boldsymbol{0}) \le x, X^\star(\boldsymbol{h}) \le x) = \exp\Big\{-\frac{V_2(\boldsymbol{h})}{x}\Big\}, \quad x > 0.$$

We standardise the process (2.19) by setting

$$X(\boldsymbol{s}) := X^\star(\boldsymbol{s})/V_1, \quad \boldsymbol{s} \in \mathbb{Z}^d. \tag{2.20}$$

As a consequence we can choose $a_m = m_n^d$ in Definition 2.2. We further conclude that the extremal coefficient (cf. Beirlant et al. [3], Section 8.2.7) at lag $\boldsymbol{h} \in \mathbb{Z}^d$ for the process (2.20) is given by

$$\theta(\boldsymbol{h}) := \frac{V_2(\boldsymbol{h})}{V_1} = \frac{1}{V_1}\sum_{0\le j<\infty} Q_{\boldsymbol{h}}(j)\,\phi^j. \tag{2.21}$$

Note that $2 - \theta(\boldsymbol{h}) = \frac{1}{V_1}\sum_{\|\boldsymbol{h}\|/2\le j<\infty} \phi^j[2N(j) - Q_{\boldsymbol{h}}(j)]$, where we use that $Q_{\boldsymbol{h}}(j) = 2N(j)$ for $j < \|\boldsymbol{h}\|/2$; see [15], p. 8.

We now verify the conditions of Theorem 2.6 for the process $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{Z}^d\}$ as in (2.20) and a chosen set of lags $\mathcal{H} \subseteq B(\mathbf{0}, \gamma)$ for some $\gamma > 0$.

We start with an upper bound for the α-mixing coefficients $\alpha_{k,\ell}(r)$ for $k, \ell \in \mathbb{N}$ and $r > 0$ defined in (2.11). To simplify notation, we use C throughout to denote some positive constant, although the actual value of C may change from line to line. We use Corollary 2.2 of Dombry and Eyi-Minko [27] to deduce

$$\begin{aligned}\alpha_{k,\ell}(r) &\le Ck\ell \sup_{\|\boldsymbol{h}\|>r} 2-\theta(\boldsymbol{h})\\ &\le Ck\ell \sup_{\|\boldsymbol{h}\|>r} \sum_{\|\boldsymbol{h}\|/2\le j<\infty} \phi^j [2N(j) - Q_{\boldsymbol{h}}(j)]\\ &\le Ck\ell \sum_{r/2\le j<\infty} 2N(j)\,\phi^j \le Ck\ell \sum_{r/2\le j<\infty} j^{d-1}\,\phi^j, \end{aligned}\tag{2.22}$$

since $N(j)$ is of order $\mathcal{O}(j^{d-1})$ as mentioned above. An integral bound yields for fixed $k, \ell \in \mathbb{N}$ and sufficiently large r such that the sequence $j^{d-1}\phi^j$ is monotonously decreasing for $j \ge r/2$,

$$\begin{aligned}\alpha_{k,\ell}(r) &\le C\Big(r^{d-1}\phi^{r/2} + \int_{r/2}^{\infty} t^{d-1}\,\phi^t\,dt\Big) = C\Big(r^{d-1}\phi^{r/2} + |\log(\phi)|^{-d}\Gamma\big(d, \frac{r}{2}|\log(\phi)|\big)\Big)\\ &\le C\phi^{r/2}\Big(r^{d-1} + \sum_{k=0}^{d-1}\frac{r^k|\log(\phi)|^k}{2^k k!}\Big) = \mathcal{O}(r^{d-1}\,\phi^{r/2}), \end{aligned}\tag{2.23}$$

as $r \to \infty$. We denote by $\Gamma(s,y) = \int_y^\infty t^{s-1}e^{-t}dt = (s-1)!e^{-y}\sum_{i=0}^{s-1} y^i/i!$ for $s \in \mathbb{N}$ the incomplete gamma function. Since $r^{d-1}\phi^{r/2} \to 0$, this implies that $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{Z}^d\}$ is α-mixing; thus (M1) is satisfied.

Now we verify (M3). To this end we compute for $\boldsymbol{s}, \boldsymbol{s'} \in \mathbb{Z}^d$ and $x > 0$, using a Taylor expansion, the limit as $x \to \infty$:

$$\begin{aligned}\mathbb{P}(X(\boldsymbol{s}) > x, X(\boldsymbol{s'}) > x) &= 1 - 2\mathbb{P}(X(\mathbf{0}) \le x) + \mathbb{P}(X(\boldsymbol{s}) \le x, X(\boldsymbol{s'}) \le x)\\ &= 1 - 2\exp\Big\{-\frac{1}{x}\Big\} + \exp\Big\{-\frac{V_2(\boldsymbol{s}-\boldsymbol{s'})}{V_1 x}\Big\}\\ &= \frac{2}{x} - \frac{V_2(\boldsymbol{s}-\boldsymbol{s'})}{V_1 x} + \mathcal{O}\big(\frac{1}{x^2}\big)\\ &= \frac{1}{x}(2-\theta(\boldsymbol{s}-\boldsymbol{s'})) + \mathcal{O}\big(\frac{1}{x^2}\big)\\ &\le \frac{C}{x}\sum_{\|\boldsymbol{s}-\boldsymbol{s'}\|/2\le j<\infty} j^{d-1}\,\phi^j + \mathcal{O}\big(\frac{1}{x^2}\big) \end{aligned}$$

by (2.22). Hence, for $\epsilon > 0$, as $n \to \infty$,

$$\begin{aligned}&\mathbb{P}\Big(\max_{\boldsymbol{s}\in B(\mathbf{0},\gamma)} X(\boldsymbol{s}) > \epsilon m_n^d, \max_{\boldsymbol{s'}\in B(\boldsymbol{h},\gamma)} X(\boldsymbol{s'}) > \epsilon m_n^d\Big)\\ &\le \sum_{\boldsymbol{s}\in B(\mathbf{0},\gamma)}\sum_{\boldsymbol{s'}\in B(\boldsymbol{h},\gamma)} \mathbb{P}\big(X(\boldsymbol{s}) > \epsilon m_n^d, X(\boldsymbol{s'}) > \epsilon m_n^d\big)\\ &\le \sum_{\boldsymbol{s}\in B(\mathbf{0},\gamma)}\sum_{\boldsymbol{s'}\in B(\boldsymbol{h},\gamma)} \Big\{\frac{C}{\epsilon m_n^d}\sum_{\|\boldsymbol{s}-\boldsymbol{s'}\|/2\le j<\infty} j^{d-1}\,\phi^j + \mathcal{O}\big(\frac{1}{m_n^{2d}}\big)\Big\} \end{aligned}$$

$$\leq \sum_{\boldsymbol{s}\in B(\boldsymbol{0},\gamma)} \sum_{\boldsymbol{s}'\in B(\boldsymbol{h},\gamma)} \Big\{ \frac{C}{\epsilon m_n^d} \|\boldsymbol{s}-\boldsymbol{s}'\|^{d-1} \phi^{\frac{\|\boldsymbol{s}-\boldsymbol{s}'\|}{2}} + \mathcal{O}\big(\frac{1}{m_n^{2d}}\big) \Big\}$$
$$\leq \frac{C|B(\boldsymbol{0},\gamma)|^2}{\epsilon m_n^d} (\|\boldsymbol{h}\|-2\gamma)^{d-1} \phi^{\frac{\|\boldsymbol{h}\|-2\gamma}{2}} + \mathcal{O}\big(\frac{1}{m_n^{2d}}\big),$$

where in the second last step we use the same bound as in (2.23), and in the last step we use that $\|\boldsymbol{h}\|^{d-1}\phi^{\|\boldsymbol{h}\|/2}$ decreases for sufficiently large $\|\boldsymbol{h}\|$. Therefore, we conclude

$$\lim_{k\to\infty} \limsup_{n\to\infty} \sum_{\boldsymbol{h}\in\mathbb{Z}^d: k<\|\boldsymbol{h}\|\leq r_n} \Big\{ m_n^d \mathbb{P}\Big(\max_{\boldsymbol{s}\in B(\boldsymbol{0},\gamma)} X(\boldsymbol{s}) > \epsilon m_n^d,\ \max_{\boldsymbol{s}'\in B(\boldsymbol{h},\gamma)} X(\boldsymbol{s}') > \epsilon m_n^d \Big) \Big\}$$
$$\leq \lim_{k\to\infty} \limsup_{n\to\infty} C \sum_{\boldsymbol{h}\in\mathbb{Z}^d: k<\|\boldsymbol{h}\|\leq r_n} \Big\{ (\|\boldsymbol{h}\|-2\gamma)^{d-1} \phi^{\frac{\|\boldsymbol{h}\|-2\gamma}{2}} \Big\} + \limsup_{n\to\infty} \mathcal{O}\big(\frac{r_n^d}{m_n^d}\big) = 0,$$

since $r_n = o(m_n)$, where we use for the last inequality that $\big|\{\boldsymbol{h}\in\mathbb{Z}^d : \|\boldsymbol{h}\|\leq r_n\}\big| = \mathcal{O}(r_n^d)$.
Turning to condition (M4i), using (2.23), we have as $n\to\infty$,

$$m_n^d \sum_{\boldsymbol{h}\in\mathbb{Z}^d:\|\boldsymbol{h}\|>r_n} \alpha_{1,1}(\|\boldsymbol{h}\|) \leq C m_n^d \sum_{j>r_n} j^{d-1}\,\alpha_{1,1}(j)$$
$$\leq C m_n^d \sum_{j>r_n} j^{2(d-1)}\,\phi^{j/2} \leq C m_n^d r_n^{2(d-1)} \phi^{r_n/2},$$

which follows again from an integral bound. Since

$$m_n^d r_n^{2(d-1)} \phi^{r_n/2} = \exp\{d\log(m_n) - r_n|\log(\phi^{1/2})| + 2(d-1)\log(r_n)\},$$

if we choose m_n and r_n such that

$$\log(m_n) = o(r_n), \quad n\to\infty, \tag{2.24}$$

then condition (M4i) is satisfied.
Now observe that for $2\leq p+q\leq 4$, using again (2.23),

$$\sum_{\boldsymbol{h}\in\mathbb{Z}^d} \alpha_{p,q}(\|\boldsymbol{h}\|) \leq \alpha_{p,q}(0) + C\sum_{j>0} j^{d-1}\,\alpha_{p,q}(j) \leq \alpha_{p,q}(0) + C\sum_{j>0} j^{2(d-1)}\,\phi^{j/2} < \infty.$$

This shows (M4ii).
Finally, we turn to the condition (M4iii) and compute, using (2.22) and (2.23),

$$m_n^{d/2} n^{d/2} \alpha_{1,n^d}(r_n) \leq C m_n^{d/2} n^{(3d)/2} r_n^{d-1} \phi^{r_n/2}$$
$$= C\exp\big\{ \frac{3d}{2}\log(n) - r_n|\log(\phi)| + \frac{d}{2}\log(m_n) + (d-1)\log(r_n) \big\}.$$

Thus, we must choose r_n such that

$$\log(n) = o(r_n), \quad n\to\infty. \tag{2.25}$$

To satisfy both (2.24) and (2.25) and the conditions $r_n = o(m_n)$, $m_n = o(n)$, we can thus choose $m_n = n^{\beta_1}$ and $r_n = n^{\beta_2}$ with $0 < \beta_2 < \beta_1 < 1$. Hence, Remarks 2.7(ii) and 2.9(ii) apply such that (2.18) holds for $\beta_1 \in (1/3, 1/2)$ and $\beta_2 \in (0, \min\{\beta_1; 1/2 - \beta_1\})$.

Example 2.11. [Brown-Resnick process]
Consider a strictly stationary Brown-Resnick process, which has representation

$$X(\boldsymbol{s}) = \bigvee_{j=1}^{\infty} \left\{ \xi_j \, e^{W_j(\boldsymbol{s}) - \delta(\boldsymbol{s})} \right\}, \quad \boldsymbol{s} \in \mathbb{R}^d. \tag{2.26}$$

where $\{\xi_j : j \in \mathbb{N}\}$ are points of a Poisson process on $[0, \infty)$ with intensity $\xi^{-2} d\xi$, the *dependence function* δ is non-negative and conditionally negative definite and $\{W_j(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ are independent replicates of a Gaussian process $\{W(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ with stationary increments, $W(\boldsymbol{0}) = 0$, $\mathbb{E}[W(\boldsymbol{s})] = 0$, and covariance function

$$\mathbb{C}ov[W(\boldsymbol{s}^{(1)}), W(\boldsymbol{s}^{(2)})] = \delta(\boldsymbol{s}^{(1)}) + \delta(\boldsymbol{s}^{(2)}) - \delta(\boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)}).$$

All finite-dimensional distributions are multivariate extreme value distributions with standard unit Fréchet margins. Representation (2.26) goes back to de Haan [24] and Giné et al. [39]; for more details on Brown-Resnick processes we refer to Brown and Resnick [8], Davis et al. [19], and Kabluchko et al. [47]. Brown-Resnick processes have been successfully fitted to time series, spatial data and space-time data. Inference methods include both parametric and semi- or non-parametric approaches. Empirical studies can for example be found in Buhl and Klüppelberg [11], Davis et al. [20], Buhl et al. [14], Cho et al. [15], Huser and Davison [43], Padoan et al. [54] and Steinkohl [62]. This important model is treated in detail in Chapter 3. There it is proved that the mixing conditions of Theorem 2.6 hold for sequences $r_n = o(m_n)$, $m_n = o(n)$ and that we can choose $m_n = n^{\beta_1}$ and $r_n = n^{\beta_2}$ with $0 < \beta_2 < \beta_1 < 1$. Hence, Remarks 2.7(ii) and 2.9(ii) apply such that (2.18) holds for $\beta_1 \in (1/3, 1/2)$ and $\beta_2 \in (0, \min\{\beta_1; 1/2 - \beta_1\})$. Moreover, we prove there that for $\beta_1 \leq 1/3$, the empirical extremogram can be bias-corrected such that the resulting empirical estimator satisfies a CLT to the true extremogram. We further derive a semiparametric estimator for a parametrised extremogram based on a least squares procedure, investigate its behaviour in a simulation study, and apply it to space-time data.

2.5 Proof of Theorem 2.6

The empirical extremogram as defined in (2.8) can be viewed as a ratio of estimates of $\mu_{B(\boldsymbol{0},\gamma)}(C)$ and $\mu_{B(\boldsymbol{0},\gamma)}(D)$ for two suitably chosen sets C and D. Thus we first derive a LLN and a CLT for such estimates, formulated in the two Lemmas 2.12 and 2.13 below.

We consider estimates of $\mu_{B(\boldsymbol{0},\gamma)}(C)$, where C is a $\mu_{B(\boldsymbol{0},\gamma)}$-continuous Borel set in $\overline{\mathbb{R}}^{|B(\boldsymbol{0},\gamma)|} \backslash \{\boldsymbol{0}\}$ (i.e. $\mu_{B(\boldsymbol{0},\gamma)}(\partial C) = 0$). In particular, there exists some $\varepsilon > 0$ such that $C \subset \{\boldsymbol{x} \in \mathbb{R}^{|B(\boldsymbol{0},\gamma)|} :$

$\|\boldsymbol{x}\| > \varepsilon\}$. In view of (2.5) a natural estimator for $\mu_{B(\mathbf{0},\gamma)}(C)$ is

$$\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(C) := \Big(\frac{m_n}{n}\Big)^d \sum_{\boldsymbol{s}\in\mathcal{S}_n} \mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{s})}{a_m}\in C\}}. \tag{2.27}$$

The proof is based on a big block/small block argument as follows. We choose sequences m_n and r_n satisfying the conditions of Theorem 2.6, and divide the grid $\mathcal{S}_n$ into $(n/m_n)^d$ big d-dimensional blocks of side length m_n, where for simplicity we assume that n/m_n is an integer. From those blocks we then cut off smaller blocks, which consist of the first r_n elements in each of the d dimensions. The large blocks are then separated (by these small blocks) with at least the distance r_n in all dimensions and shown to be asymptotically independent. The construction is an extension of the corresponding time series construction; an interpretation of the big and small blocks in that framework can be found for example in Davis et al. [21] at the end of page 15.

Lemma 2.12. *Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ be a strictly stationary regularly varying process in $\mathbb{R}^d$. Let C be some $\mu_{B(\mathbf{0},\gamma)}$-continuous Borel set in $\overline{\mathbb{R}}^{|B(\mathbf{0},\gamma)|} \setminus \{\mathbf{0}\}$. Suppose that the following mixing conditions are satisfied.*

(1) $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ is α-mixing with α-mixing coefficients $\alpha_{k,l}(r)$ defined in (2.10).

(2) There exist sequences $m := m_n, r := r_n \to \infty$ with $m_n/n \to 0$ and $r_n/m_n \to 0$ as $n \to \infty$ such that (M3) and (M4i) hold.

Then, as $n \to \infty$,

$$\mathbb{E}\big[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(C)\big] \to \mu_{B(\mathbf{0},\gamma)}(C), \tag{2.28}$$

$$\begin{aligned}\mathbb{V}ar\big[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(C)\big] &\sim \Big(\frac{m_n}{n}\Big)^d \Big(\mu_{B(\mathbf{0},\gamma)}(C) + \sum_{\boldsymbol{h}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}} \tau_{B(\mathbf{0},\gamma)\times B(\boldsymbol{h},\gamma)}(C\times C)\Big)\\ &=: \Big(\frac{m_n}{n}\Big)^d \sigma^2_{B(\mathbf{0},\gamma)}(C).\end{aligned} \tag{2.29}$$

If $\mu_{B(\mathbf{0},\gamma)}(C) = 0$, (2.29) is interpreted as $\mathbb{V}ar\big[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(C)\big] = o(m_n/n)$. In particular, we have

$$\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(C) \xrightarrow{P} \mu_{B(\mathbf{0},\gamma)}(C), \quad n \to \infty. \tag{2.30}$$

Proof. Strict stationarity and relation (2.5) imply that

$$\begin{aligned}\mathbb{E}\big[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(C)\big] &= \Big(\frac{m_n}{n}\Big)^d \sum_{\boldsymbol{s}\in\mathcal{S}_n} \mathbb{P}\Big(\frac{\boldsymbol{Y}(\boldsymbol{s})}{a_m} \in C\Big) = m_n^d \mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_m} \in C\Big)\\ &\to \mu_{B(\mathbf{0},\gamma)}(C), \quad n \to \infty.\end{aligned}$$

Further observe that

$$\mathbb{V}ar\big[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(C)\big] = \Big(\frac{m_n}{n}\Big)^{2d} \mathbb{V}ar\big[\sum_{\boldsymbol{s}\in\mathcal{S}_n} \mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{s})}{a_m}\in C\}}\big]$$

$$= \Big(\frac{m_n}{n}\Big)^{2d} \Big(n^d \mathbb{V}ar\big[\mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{0})}{a_m}\in C\}}\big] + \sum_{\substack{\boldsymbol{s},\boldsymbol{s}'\in\mathcal{S}_n\\ \boldsymbol{s}\neq\boldsymbol{s}'}} \mathbb{C}ov\big[\mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{s})}{a_m}\in C\}}, \mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{s}')}{a_m}\in C\}}\big]\Big)$$

$$=: A_1 + A_2. \tag{2.31}$$

By (2.5) and since $\mathbb{P}(\boldsymbol{Y}(\boldsymbol{0})/a_m \in C) \to 0$ as $n \to \infty$,

$$A_1 = \Big(\frac{m_n^2}{n}\Big)^d \mathbb{P}\Big(\frac{\boldsymbol{Y}(\boldsymbol{0})}{a_m} \in C\Big)\Big(1 - \mathbb{P}\Big(\frac{\boldsymbol{Y}(\boldsymbol{0})}{a_m} \in C\Big)\Big) \sim \Big(\frac{m_n}{n}\Big)^d \mu_{B(\boldsymbol{0},\gamma)}(C) \to 0.$$

Let

$$L = L(n) := \{\boldsymbol{h} = (\boldsymbol{s} - \boldsymbol{s}') \in \mathbb{Z}^d : \boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}_n, \boldsymbol{s} \neq \boldsymbol{s}'\}$$

be the set of spatial lags on the observation grid. We divide the spatial lags in L into different sets. Observe that a spatial lag $\boldsymbol{h} = (h_1, \ldots, h_d)$ appears in L exactly $\prod_{j=1}^d (n - |h_j|)$ times. For fixed $k \in \mathbb{N}$ we therefore have by stationarity

$$\begin{aligned}
\Big(\frac{n}{m_n}\Big)^d A_2 &= m_n^d \sum_{\boldsymbol{h}\in L} \prod_{j=1}^d \Big(1 - \frac{|h_j|}{n}\Big) \mathbb{C}ov[\mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{0})}{a_m}\in C\}}, \mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{h})}{a_m}\in C\}}] \\
&= m_n^d \Big(\sum_{\substack{\boldsymbol{h}\in L\\ 0<\|\boldsymbol{h}\|\leq k}} + \sum_{\substack{\boldsymbol{h}\in L\\ k<\|\boldsymbol{h}\|\leq r_n}} + \sum_{\substack{\boldsymbol{h}\in L\\ \|\boldsymbol{h}\|>r_n}} \Big) \\
&\qquad\qquad \prod_{j=1}^d \Big(1 - \frac{|h_j|}{n}\Big) \mathbb{C}ov\big[\mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{0})}{a_m}\in C\}}, \mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{h})}{a_m}\in C\}}\big] \\
&=: A_{21} + A_{22} + A_{23}.
\end{aligned} \tag{2.32}$$

Concerning A_{21} we have,

$$A_{21} = m_n^d \sum_{\substack{\boldsymbol{h}\in L\\ 0<\|\boldsymbol{h}\|\leq k}} \prod_{j=1}^d \Big(1 - \frac{|h_j|}{n}\Big) \Big[\mathbb{P}\Big(\frac{\boldsymbol{Y}(\boldsymbol{0})}{a_m} \in C, \frac{\boldsymbol{Y}(\boldsymbol{h})}{a_m} \in C\Big) - \mathbb{P}\Big(\frac{\boldsymbol{Y}(\boldsymbol{0})}{a_m} \in C\Big)^2\Big].$$

We have by (2.5),

$$m_n^d \mathbb{P}\Big(\frac{\boldsymbol{Y}(\boldsymbol{0})}{a_m} \in C\Big)^2 \sim \mu_{B(\boldsymbol{0},\gamma)}(C)\mathbb{P}\Big(\frac{\boldsymbol{Y}(\boldsymbol{0})}{a_m} \in C\Big) \to 0, \quad n \to \infty.$$

Moreover, for $\boldsymbol{h} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\}$, by (2.6),

$$m_n^d \mathbb{P}\Big(\frac{\boldsymbol{Y}(\boldsymbol{0})}{a_m} \in C, \frac{\boldsymbol{Y}(\boldsymbol{h})}{a_m} \in C\Big) \to \tau_{B(\boldsymbol{0},\gamma)\times B(\boldsymbol{h},\gamma)}(C \times C), \quad n \to \infty. \tag{2.33}$$

Finally, by dominated convergence,

$$\lim_{k\to\infty} \limsup_{n\to\infty} A_{21} = \sum_{\boldsymbol{h}\in\mathbb{Z}^d\setminus\{\boldsymbol{0}\}} \tau_{B(\boldsymbol{0},\gamma)\times B(\boldsymbol{h},\gamma)}(C \times C). \tag{2.34}$$

As to A_{22}, observe that for all $n \geq 0$ we have $\prod_{j=1}^{d}(1-\frac{|h_j|}{n}) \leq 1$ for $\boldsymbol{h} \in L$. Furthermore, since C is bounded away from $\mathbf{0}$, there exists $\epsilon > 0$ such that $C \subset \{\boldsymbol{x} \in \overline{\mathbb{R}}^{|B(\mathbf{0},\gamma)|} : \|\boldsymbol{x}\| > \epsilon\}$. Hence, we obtain

$$
\begin{aligned}
|A_{22}| \leq & \sum_{\substack{\boldsymbol{h}\in L\\ k<\|\boldsymbol{h}\|\leq r_n}} m_n^d \mathbb{P}\big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_m} \in C, \frac{\boldsymbol{Y}(\boldsymbol{h})}{a_m} \in C\big) + m_n^d \mathbb{P}\big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_m} \in C\big)^2 \\
\leq & \sum_{\substack{\boldsymbol{h}\in \mathbb{Z}^d\\ k<\|\boldsymbol{h}\|\leq r_n}} \Big\{ m_n^d \mathbb{P}\big(\|\boldsymbol{Y}(\mathbf{0})\| > \epsilon a_m, \|\boldsymbol{Y}(\boldsymbol{h})\| > \epsilon a_m\big) + \frac{1}{m_n^d}\big(m_n^d \mathbb{P}\big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_m} \in C\big)\big)^2 \Big\}.
\end{aligned}
$$

From (2.5) we know that $m_n^d \mathbb{P}(\boldsymbol{Y}(\mathbf{0}) \in a_m C) \to \mu_{B(\mathbf{0},\gamma)}(C)$ and, hence, the second summand can be estimated by $(r_n/m_n)^d \to 0$ as $n \to \infty$. The first sum tends to 0 by (M3), exploiting the equivalence of norms on $\mathbb{R}^{|B(\mathbf{0},\gamma)|}$, and it follows that

$$
\lim_{k\to\infty} \limsup_{n\to\infty} A_{22} = 0
$$

Using the definition of α-mixing for $A_1 = \{\boldsymbol{Y}(\mathbf{0})/a_m \in C\}$ and $A_2 = \{\boldsymbol{Y}(\boldsymbol{h})/a_m \in C\}$, we obtain,

$$
\begin{aligned}
|A_{23}| &\leq m_n^d \sum_{\boldsymbol{h}\in L:\|\boldsymbol{h}\|>r_n} \Big|\mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_m} \in C, \frac{\boldsymbol{Y}(\boldsymbol{h})}{a_m} \in C\Big) - \mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_m} \in C\Big)^2\Big| \\
&\leq m_n^d \sum_{\boldsymbol{h}\in \mathbb{Z}^d:\|\boldsymbol{h}\|>r_n} \alpha_{1,1}(\|\boldsymbol{h}\|) \to 0, \quad n \to \infty, \qquad (2.35)
\end{aligned}
$$

by condition (M4i).

Summarising these computations, we obtain from (2.32) and (2.34) that

$$
A_2 \sim \Big(\frac{m_n}{n}\Big)^d \sum_{\boldsymbol{h}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}} \tau_{B(\mathbf{0},\gamma)\times B(\boldsymbol{h},\gamma)}(C \times C), \quad n \to \infty,
$$

and, therefore, (2.31) implies (2.29). Since $m_n/n \to 0$ as $n \to \infty$, equations (2.28) and (2.29) imply (2.30). ☐

For the proof of the next lemma, in contrast to Cho et al. [15], we proceed similarly as in the proofs of Lemma 2.12 above and of Theorem 3.2 of Davis and Mikosch [17] and keep the sequence r_n (instead of m_n in [15]) in (2.40) as the distance between the large blocks. This construction allows for the much weaker conditions (M2).

Lemma 2.13. *Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ be a strictly stationary regularly varying process in $\mathbb{R}^d$. Let the assumptions of Theorem 2.6 hold for some $\gamma \geq 0$. Let C be some $\mu_{B(\mathbf{0},\gamma)}$-continuous Borel set in $\overline{\mathbb{R}}^{|B(\mathbf{0},\gamma)|}\backslash\{\mathbf{0}\}$. Then*

$$
\widehat{S}_{B(\mathbf{0},\gamma),m_n} := \Big(\frac{m_n}{n}\Big)^{d/2} \sum_{\boldsymbol{s}\in\mathcal{S}_n} \Big[\mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{s})}{a_m}\in C\}} - \mathbb{P}\Big(\frac{\boldsymbol{Y}(\boldsymbol{s})}{a_m} \in C\Big)\Big]
$$

$$= \Big(\frac{n}{m_n}\Big)^{d/2} (\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(C) - \mu_{B(\mathbf{0},\gamma),m_n}(C)) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{B(\mathbf{0},\gamma)}(C)), \tag{2.36}$$

as $n \to \infty$ *with* $\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(C)$ *as in* (2.27), $\mu_{B(\mathbf{0},\gamma),m_n}(C)) := m_n^d \mathbb{P}(\boldsymbol{Y}(\mathbf{0})/a_m \in C)$ *and* $\sigma^2_{B(\mathbf{0},\gamma)}(C)$ *given in* (2.29).

Proof. Like Cho et al. [15] we follow Lemma 2 in Bolthausen [6] and define

$$I(\boldsymbol{s}) := \mathbb{1}_{\{\boldsymbol{Y}(\boldsymbol{s})/a_m \in C\}} - \mathbb{P}(\boldsymbol{Y}(\mathbf{0})/a_m \in C), \quad \boldsymbol{s} \in \mathcal{S}_n. \tag{2.37}$$

Note that by stationarity,

$$\widehat{S}_{B(\mathbf{0},\gamma),m_n} = \Big(\frac{m_n}{n}\Big)^{d/2} \sum_{\boldsymbol{s}\in\mathcal{S}_n} I(\boldsymbol{s}). \tag{2.38}$$

The boundary condition required in equation (1) in Bolthausen [6] is trivially satisfied for the regular grid $\mathcal{S}_n$.

We first observe that $0 \le \mathbb{E}[I(\boldsymbol{s})I(\boldsymbol{s}')] = \mathbb{C}ov[\mathbb{1}_{\{\boldsymbol{Y}(\boldsymbol{s})/a_m \in C\}}, \mathbb{1}_{\{\boldsymbol{Y}(\boldsymbol{s}')/a_m \in C\}}]$ and, using (2.29) for the asymptotic result, that

$$0 < \sigma^2_{B(\mathbf{0},\gamma)} \sim \mathbb{V}ar[\widehat{S}_{B(\mathbf{0},\gamma),m_n}],\ \mathbb{V}ar[\widehat{S}_{B(\mathbf{0},\gamma),m_n}] \le \Big(\frac{m_n}{n}\Big)^d \sum_{\boldsymbol{s},\boldsymbol{s}'\in\mathbb{Z}^d} |\mathbb{E}[I(\boldsymbol{s})I(\boldsymbol{s}')]| < \infty, \tag{2.39}$$

such that $\sum_{\boldsymbol{s},\boldsymbol{s}'\in\mathbb{Z}^d} \mathbb{E}[I(\boldsymbol{s})I(\boldsymbol{s}')] > 0$. Finiteness in (2.39) follows from a classic result found e.g. in Ibragimov and Linnik [45], Theorems 17.2.2 and 17.2.3, and the required summability conditions of the α-mixing coefficients.

Next, we define

$$v_n := \Big(\frac{m_n}{n}\Big)^d \sum_{\substack{\boldsymbol{s},\boldsymbol{s}'\in\mathcal{S}_n \\ \|\boldsymbol{s}-\boldsymbol{s}'\|\le r_n}} \mathbb{E}\big[I(\boldsymbol{s})I(\boldsymbol{s}')\big]. \tag{2.40}$$

Decompose

$$\begin{aligned} v_n &= \Big(\frac{m_n}{n}\Big)^d \sum_{\substack{\boldsymbol{s},\boldsymbol{s}'\in\mathcal{S}_n \\ \|\boldsymbol{s}-\boldsymbol{s}'\|\le r_n}} \mathbb{E}[I(\boldsymbol{s})I(\boldsymbol{s}')] \\ &= \Big(\frac{m_n}{n}\Big)^d \sum_{\boldsymbol{s},\boldsymbol{s}'\in\mathcal{S}_n} \mathbb{E}[I(\boldsymbol{s})I(\boldsymbol{s}')] - \Big(\frac{m_n}{n}\Big)^d \sum_{\substack{\boldsymbol{s},\boldsymbol{s}'\in\mathcal{S}_n \\ \|\boldsymbol{s}-\boldsymbol{s}'\|> r_n}} \mathbb{E}[I(\boldsymbol{s})I(\boldsymbol{s}')] \\ &= \mathbb{V}ar[\widehat{S}_{B(\mathbf{0},\gamma),m_n}] - \Big(\frac{m_n}{n}\Big)^d \sum_{\substack{\boldsymbol{s},\boldsymbol{s}'\in\mathcal{S}_n \\ \|\boldsymbol{s}-\boldsymbol{s}'\|> r_n}} \mathbb{E}[I(\boldsymbol{s})I(\boldsymbol{s}')]. \end{aligned} \tag{2.41}$$

Hence, using the asymptotic result in (2.39),

$$\frac{v_n}{\mathbb{V}ar[\widehat{S}_{B(\mathbf{0},\gamma),m_n}]} = 1 - \Big(\frac{m_n}{n}\Big)^d \frac{1}{\sigma^2_{B(\mathbf{0},\gamma)}(C)} \sum_{\substack{\boldsymbol{s},\boldsymbol{s}'\in\mathcal{S}_n \\ \|\boldsymbol{s}-\boldsymbol{s}'\|> r_n}} \mathbb{E}[I(\boldsymbol{s})I(\boldsymbol{s}')](1+o(1)).$$

Now note that

$$\begin{aligned}
&\Big(\frac{m_n}{n}\Big)^d \sum_{\substack{\boldsymbol{s},\boldsymbol{s}'\in\mathcal{S}_n\\ \|\boldsymbol{s}-\boldsymbol{s}'\|>r_n}} \mathbb{E}[I(\boldsymbol{s})I(\boldsymbol{s}')]\\
\leq& m_n^d \sum_{\boldsymbol{h}\in L:\|\boldsymbol{h}\|>r_n} \prod_{j=1}^d \Big(1-\frac{|h_j|}{n}\Big)\\
&\qquad \Big|\mathbb{P}\Big(\frac{\boldsymbol{Y}(\boldsymbol{0})}{a_m}\in C, \frac{\boldsymbol{Y}(\boldsymbol{h})}{a_m}\in C\Big) - \Big[\mathbb{P}\Big(\frac{\boldsymbol{Y}(\boldsymbol{0})}{a_m}\in C\Big)\Big]^2\Big|\\
\leq& m_n^d \sum_{\boldsymbol{h}\in\mathbb{Z}^d:\|\boldsymbol{h}\|>r_n} \alpha_{1,1}(\|\boldsymbol{h}\|) \to 0, \quad n\to\infty,
\end{aligned}$$

as in (2.35), with α-mixing coefficients defined in (2.11). Therefore,

$$v_n \sim \mathbb{V}ar[\widehat{S}_{B(\boldsymbol{0},\gamma),m_n}] \to \sigma^2_{B(\boldsymbol{0},\gamma)}(C), \quad n\to\infty. \tag{2.42}$$

Next we define the standardised quantities

$$\begin{aligned}
\overline{S}_n &:= v_n^{-1/2}\widehat{S}_{B(\boldsymbol{0},\gamma),m_n} = v_n^{-1/2}\Big(\frac{m_n}{n}\Big)^{d/2}\sum_{\boldsymbol{s}\in\mathcal{S}_n} I(\boldsymbol{s}),\\
\overline{S}_{\boldsymbol{s},n} &:= v_n^{-1/2}\Big(\frac{m_n}{n}\Big)^{d/2} \sum_{\substack{\boldsymbol{s}'\in\mathcal{S}_n\\ \|\boldsymbol{s}-\boldsymbol{s}'\|\leq r_n}} I(\boldsymbol{s}').
\end{aligned}$$

We now show condition (b) of Lemma 2 in Bolthausen [6]. To this end let $\mathrm{i}\in\mathbb{C}$ be the complex imaginary unit. If $\lim_{n\to\infty}\mathbb{E}\big[(\mathrm{i}\lambda-\overline{S}_n)\exp\{\mathrm{i}\lambda\overline{S}_n\}\big]=0$ for all $\lambda\in\mathbb{R}$, then (by Stein's Lemma) the law of $\overline{S}_n$ converges to the standard normal one and we obtain (2.36) by (2.38) and (2.42). First note that for arbitrary $\lambda\in\mathbb{R}$,

$$\begin{aligned}
&(\mathrm{i}\lambda-\overline{S}_n)\exp\{\mathrm{i}\lambda\overline{S}_n\}\\
=&\mathrm{i}\lambda\exp\{\mathrm{i}\lambda\overline{S}_n\}\Big(1-v_n^{-1/2}\sum_{\boldsymbol{s}\in\mathcal{S}_n}\Big(\frac{m_n}{n}\Big)^{d/2} I(\boldsymbol{s})\overline{S}_{\boldsymbol{s},n}\Big)\\
&-v_n^{-1/2}\exp\{\mathrm{i}\lambda\overline{S}_n\}\sum_{\boldsymbol{s}\in\mathcal{S}_n}\Big(\frac{m_n}{n}\Big)^{d/2} I(\boldsymbol{s})(1-\exp\{-\mathrm{i}\lambda\overline{S}_{\boldsymbol{s},n}\}-\mathrm{i}\lambda\overline{S}_{\boldsymbol{s},n})\\
&-v_n^{-1/2}\sum_{\boldsymbol{s}\in\mathcal{S}_n}\Big(\frac{m_n}{n}\Big)^{d/2} I(\boldsymbol{s})\exp\{-\mathrm{i}\lambda(\overline{S}_{\boldsymbol{s},n}-\overline{S}_n)\}\\
=:&B_1-B_2-B_3.
\end{aligned}$$

Since $|\exp\{ix\}|=1$ for all $x\in\mathbb{R}$, we compute

$$|B_1| = |\lambda|\Big|1-v_n^{-1}\Big(\frac{m_n}{n}\Big)^d \sum_{\substack{\boldsymbol{s},\boldsymbol{s}'\in\mathcal{S}_n\\ \|\boldsymbol{s}-\boldsymbol{s}'\|\leq r_n}} I(\boldsymbol{s})I(\boldsymbol{s}')\Big|$$

and, using (2.40),

$$\begin{aligned}|B_1| &= |\lambda| v_n^{-1}\Big(\frac{m_n}{n}\Big)^d \Big| \sum_{\|\boldsymbol{s}-\boldsymbol{s}'\|\le r_n} I(\boldsymbol{s})I(\boldsymbol{s}') - \sum_{\|\boldsymbol{s}-\boldsymbol{s}'\|\le r_n} \mathbb{E}\big[I(\boldsymbol{s})I(\boldsymbol{s}')\big]\Big| \\ &= |\lambda| v_n^{-1}\Big(\frac{m_n}{n}\Big)^d \Big| \sum_{\|\boldsymbol{s}-\boldsymbol{s}'\|\le r_n} \Big(I(\boldsymbol{s})I(\boldsymbol{s}') - \mathbb{E}\big[I(\boldsymbol{s})I(\boldsymbol{s}')\big]\Big)\Big|,\end{aligned}$$

such that

$$E[|B_1|^2] = \lambda^2 v_n^{-2}\Big(\frac{m_n}{n}\Big)^{2d} \sum_{\|\boldsymbol{s}-\boldsymbol{s}'\|\le r_n} \sum_{\|\boldsymbol{\ell}-\boldsymbol{\ell}'\|\le r_n} \mathbb{C}ov\big[I(\boldsymbol{s})I(\boldsymbol{s}'), I(\boldsymbol{\ell})I(\boldsymbol{\ell}')\big].$$

We use definition (2.11) of the α-mixing coefficients for

$$\Lambda_1' = \{\boldsymbol{s}, \boldsymbol{s}'\} \quad \text{and} \quad \Lambda_2' = \{\boldsymbol{\ell}, \boldsymbol{\ell}'\},$$

then $|\Lambda_1'|, |\Lambda_2'| \le 2$, and for $d(\Lambda_1', \Lambda_2')$ we consider the following two cases:

(1) $\|\boldsymbol{s}-\boldsymbol{\ell}\| \ge 3r_n$. Then $2r_n \le (2/3)\|\boldsymbol{s}-\boldsymbol{\ell}\|$ and $d(\Lambda_1', \Lambda_2') \ge \|\boldsymbol{s}-\boldsymbol{\ell}\| - 2r_n$. Since indicator variables are bounded, by Theorem 17.2.1 of Ibragimov and Linnik [45] we have

$$|\mathbb{C}ov\big[I(\boldsymbol{s})I(\boldsymbol{s}'), I(\boldsymbol{\ell})I(\boldsymbol{\ell}')\big]| \le 4\alpha_{2,2}\Big(\|\boldsymbol{s}-\boldsymbol{\ell}\| - 2r_n\Big) \le 4\alpha_{2,2}\Big(\frac{1}{3}\|\boldsymbol{s}-\boldsymbol{\ell}\|\Big).$$

The last inequality holds, since $\alpha_{2,2}$ is a decreasing function.

(2) $\|\boldsymbol{s}-\boldsymbol{\ell}\| < 3r_n$. Set $j := \min\{\|\boldsymbol{s}-\boldsymbol{\ell}\|, \|\boldsymbol{s}-\boldsymbol{\ell}'\|, \|\boldsymbol{s}'-\boldsymbol{\ell}\|, \|\boldsymbol{s}'-\boldsymbol{\ell}'\|\}$, then $d(\Lambda_1', \Lambda_2') \ge j$ and, again by Theorem 17.2.1 of [45],

$$\mathbb{C}ov\big[I(\boldsymbol{s})I(\boldsymbol{s}'), I(\boldsymbol{\ell})I(\boldsymbol{\ell}')\big] \le 4\alpha_{p,q}(j), \quad 2 \le p+q \le 4.$$

Therefore,

$$\begin{aligned}E[|B_1|^2] &\le \frac{4\lambda^2}{v_n^2}\Big(\frac{m_n}{n}\Big)^{2d} \\ &\quad \Big[\sum_{\|\boldsymbol{s}-\boldsymbol{\ell}\|\ge 3r_n} \sum_{\substack{\|\boldsymbol{s}-\boldsymbol{s}'\|\le r_n \\ \|\boldsymbol{\ell}-\boldsymbol{\ell}'\|\le r_n}} \alpha_{2,2}\Big(\frac{1}{3}\|\boldsymbol{s}-\boldsymbol{\ell}\|\Big) + \sum_{\|\boldsymbol{s}-\boldsymbol{\ell}\|< 3r_n} \sum_{\substack{\|\boldsymbol{s}-\boldsymbol{s}'\|\le r_n \\ \|\boldsymbol{\ell}-\boldsymbol{\ell}'\|\le r_n}} \alpha_{p,q}(j)\Big] \\ &\le \frac{4\lambda^2}{v_n^2}\Big(\frac{m_n}{n}\Big)^{2d} n^d {r_n}^{2d}\Big[\sum_{\boldsymbol{h}\in\mathbb{Z}^d:\|\boldsymbol{h}\|\ge 3r_n} \alpha_{2,2}\Big(\frac{1}{3}\|\boldsymbol{h}\|\Big) + \sum_{\boldsymbol{h}\in\mathbb{Z}^d:\|\boldsymbol{h}\|<3r_n} \alpha_{p,q}(\|\boldsymbol{h}\|)\Big].\end{aligned}$$

The last inequality unfolds by stationarity as follows: we obtain n^d by summation over all $\boldsymbol{s} \in \mathcal{S}_n$, whereas r_n^{2d} arises from summation over all $\boldsymbol{s}'$ and $\boldsymbol{\ell}'$ such that $\|\boldsymbol{s}-\boldsymbol{s}'\| \le r_n$ and $\|\boldsymbol{\ell}-\boldsymbol{\ell}'\| \le r_n$, respectively. By (M4ii) the sums in brackets are finite and thus

$$E[|B_1|^2] = \mathcal{O}\Big(\Big(\frac{m_n^2 r_n^2}{n}\Big)^d\Big),$$

which converges to 0 as $n \to \infty$ by (M2).

Now we show that $\mathbb{E}[|B_2|] \to 0$ as $n \to \infty$. Since $|1 - \exp\{-\mathrm{i}x\} - \mathrm{i}x| \le x^2/2$ for $x \in \mathbb{R}$ and $|I(\boldsymbol{s})| \le 1$ for $\boldsymbol{s} \in \mathcal{S}_n$, we find

$$|B_2| \le \frac{1}{2}\lambda^2 v_n^{-1/2}\Big(\frac{m_n}{n}\Big)^{d/2} \sum_{\boldsymbol{s}\in\mathcal{S}_n} \overline{S}_{\boldsymbol{s},n}^2.$$

By stationarity and (2.40) for the second equality below,

$$\begin{aligned}
\mathbb{E}[|B_2|] &\le \frac{1}{2}\lambda^2 v_n^{-1/2}\Big(\frac{m_n}{n}\Big)^{d/2}\ n^d \mathbb{E}[\overline{S}_{\boldsymbol{0},n}^2] \\
&= \frac{1}{2}\lambda^2 v_n^{-3/2}\Big(\frac{m_n}{n}\Big)^{d/2}\ m_n^d \sum_{\boldsymbol{s}\in\mathcal{S}_n:\|\boldsymbol{s}\|\le r_n}\ \sum_{\boldsymbol{s}'\in\mathcal{S}_n:\|\boldsymbol{s}'\|\le r_n} \mathbb{E}[I(\boldsymbol{s})I(\boldsymbol{s}')] \\
&\le \frac{1}{2}\lambda^2 v_n^{-3/2}\Big(\frac{m_n}{n}\Big)^{d/2}\ r_n^d\ m_n^d \sum_{\boldsymbol{s}\in\mathcal{S}_n} \mathbb{E}[I(\boldsymbol{0})I(\boldsymbol{s})] \\
&= \mathcal{O}\Big(\Big(\frac{m_n}{n}\Big)^{d/2} r_n^d\Big) = \mathcal{O}\Big(\Big(\frac{m_n r_n^2}{n}\Big)^{d/2}\Big),
\end{aligned}$$

where we used (2.39) to obtain $m_n^d \sum_{\boldsymbol{s}\in\mathcal{S}_n} \mathbb{E}[I(\boldsymbol{0})I(\boldsymbol{s})] = \mathcal{O}(1)$. Again by (M2) we find that $\mathbb{E}[|B_2|] \to 0$ as $n \to \infty$.

Next we estimate B_3:

$$\begin{aligned}
\mathbb{E}[B_3] &= v_n^{-\frac{1}{2}}\Big(\frac{m_n}{n}\Big)^{d/2} \sum_{\boldsymbol{s}\in\mathcal{S}_n} \mathbb{E}\Big[I(\boldsymbol{s})\exp\big\{-\mathrm{i}\lambda(\overline{S}_{\boldsymbol{s},n} - \overline{S}_n)\big\}\Big] \\
&= v_n^{-\frac{1}{2}}\Big(\frac{m_n}{n}\Big)^{d/2} \sum_{\boldsymbol{s}\in\mathcal{S}_n} \mathbb{E}\Big[I(\boldsymbol{s})\exp\Big\{\mathrm{i}\lambda v_n^{-\frac{1}{2}}\Big(\frac{m_n}{n}\Big)^{d/2} \sum_{\substack{\boldsymbol{s}'\in\mathcal{S}_n \\ \|\boldsymbol{s}-\boldsymbol{s}'\|>r_n}} I(\boldsymbol{s}')\Big\}\Big] \\
&= v_n^{-\frac{1}{2}} m_n^{d/2} n^{d/2} \mathbb{E}\Big[I(\boldsymbol{0})\exp\Big\{\mathrm{i}\lambda v_n^{-\frac{1}{2}}\Big(\frac{m_n}{n}\Big)^{d/2} \sum_{\|\boldsymbol{s}\|>r_n} I(\boldsymbol{s})\Big\}\Big],
\end{aligned}$$

where the last equality holds by stationarity. We use definition (2.11) of the α-mixing coefficients for

$$\Lambda_1' = \{\boldsymbol{0}\} \quad \text{and} \quad \Lambda_2' = \{\boldsymbol{s}\in\mathcal{S}_n : \|\boldsymbol{s}\| > r_n\},$$

then $|\Lambda_1'| = 1$, $|\Lambda_2'| \le n^d$ and $d(\Lambda_1', \Lambda_2') > r_n$. Abbreviate

$$\eta(r_n) := \exp\Big\{\mathrm{i}\lambda v_n^{-\frac{1}{2}}\Big(\frac{m_n}{n}\Big)^{d/2} \sum_{\|\boldsymbol{s}\|>r_n} I(\boldsymbol{s})\Big\},$$

then $I(\boldsymbol{0})$ and $\eta(r_n)$ are measurable with respect to $\mathcal{F}_{\Lambda_1}$ and $\mathcal{F}_{\Lambda_2}$, respectively, where $\Lambda_i = \cup_{\boldsymbol{s}\in\Lambda_i'} B(\boldsymbol{s},\gamma)$ for $i = 1, 2$. Now we apply Theorem 17.2.1 of Ibragimov and Linnik to obtain

$$|\mathbb{E}[B_3]| \le 4 v_n^{-1/2} m_n^{d/2} n^{d/2} \alpha_{1,n^d}(r_n) \to 0,$$

where convergence to 0 is guaranteed by condition (M4iii). □

The proof of Theorem 2.6 follows now similarly as that of Corollary 3.4 in Davis and Mikosch [17] (also used in Theorem 1 in Cho et al. [15]). In order to keep this chapter self-contained, we summarise the main ideas.

Sketch of the proof of Theorem 2.6. For Borel sets $C, D \subseteq \overline{\mathbb{R}}^{|B(0,\gamma)|} \backslash \{\mathbf{0}\}$ such that $\mu_{B(0,\gamma)}(D) > 0$, define the ratio

$$R_n(C,D) := \frac{\mathbb{P}(\boldsymbol{Y}(\mathbf{0})/a_m \in C)}{\mathbb{P}(\boldsymbol{Y}(\mathbf{0})/a_m \in D)}$$

and the correspondent empirical estimator

$$\widehat{R}_n(C,D) := \frac{\sum\limits_{\boldsymbol{s} \in \mathcal{S}_n} \mathbb{1}_{\{\boldsymbol{Y}(\boldsymbol{s})/a_m \in C\}}}{\sum\limits_{\boldsymbol{s} \in \mathcal{S}_n} \mathbb{1}_{\{\boldsymbol{Y}(\boldsymbol{s})/a_m \in D\}}}.$$

Recall the definition of $\mathcal{H} = \{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_p\}$. For $1 \le i \le p$ fix a lag $\boldsymbol{h}_i = (h_i^1, \ldots, h_i^d) \in \mathcal{H}$ and denote as before

$$\mathcal{S}_n(\boldsymbol{h}_i) = \{\boldsymbol{s} \in \mathcal{S}_n : \boldsymbol{s} + \boldsymbol{h}_i \in \mathcal{S}_n\} \text{ with } |\mathcal{S}_n(\boldsymbol{h}_i)| = \prod_{j=1}^{d} (n - |h_i^j|) \sim n^d, \; n \to \infty.$$

Then the empirical extremogram as defined in (2.8) for Borel sets A, B in $\overline{\mathbb{R}} \backslash \{0\}$ satisfies as $n \to \infty$,

$$\begin{aligned}
\widehat{\rho}_{AB,m_n}(\boldsymbol{h}_i) &\sim \frac{\sum\limits_{\boldsymbol{s} \in \mathcal{S}_n(\boldsymbol{h}_i)} \mathbb{1}_{\{X(\boldsymbol{s})/a_m \in A, X(\boldsymbol{s}+\boldsymbol{h}_i)/a_m \in B\}}}{\sum\limits_{\boldsymbol{s} \in \mathcal{S}_n} \mathbb{1}_{\{X(\boldsymbol{s})/a_m \in A\}}} \\
&\sim \frac{\sum\limits_{\boldsymbol{s} \in \mathcal{S}_n} \mathbb{1}_{\{\boldsymbol{Y}(\boldsymbol{s})/a_m \in D_i\}}}{\sum\limits_{\boldsymbol{s} \in \mathcal{S}_n} \mathbb{1}_{\{\boldsymbol{Y}(\boldsymbol{s})/a_m \in D_{p+1}\}}} = \widehat{R}_n(D_i, D_{p+1}),
\end{aligned}$$

by definition (2.7) of the sets D_i for $i = 1, \ldots, p$. Moreover, the pre-asymptotic extremogram defined in (2.9) can be written as

$$\begin{aligned}
\rho_{AB,m_n}(\boldsymbol{h}_i) &= \frac{\mathbb{P}(X(\mathbf{0})/a_m \in A, X(\boldsymbol{h}_i)/a_m \in B)}{\mathbb{P}(X(\mathbf{0})/a_m \in A)} = \frac{\mathbb{P}(\boldsymbol{Y}(\mathbf{0})/a_m \in D_i)}{\mathbb{P}(\boldsymbol{Y}(\mathbf{0})/a_m \in D_{p+1})} \\
&= R_n(D_i, D_{p+1}).
\end{aligned}$$

This implies that proving (2.12) requires a central limit theorem for the scaled vector of ratio differences

$$\Big(\frac{n}{m_n}\Big)^{d/2} \big[\widehat{R}_n(D_i, D_{p+1}) - R_n(D_i, D_{p+1})\big]_{i=1,\ldots,p}. \tag{2.43}$$

Now observe that for fixed $i \in \{1, \dots, p\}$,

$$\begin{aligned}
\widehat{R}_n(D_i, D_{p+1}) - R_n(D_i, D_{p+1}) &= \frac{\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i)}{\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_{p+1})} - \frac{\mu_{B(\mathbf{0},\gamma),m_n}(D_i)}{\mu_{B(\mathbf{0},\gamma),m_n}(D_{p+1})} \\
&= \frac{\mu_{B(\mathbf{0},\gamma),m_n}(D_{p+1})/\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_{p+1})}{(\mu_{B(\mathbf{0},\gamma),m_n}(D_{p+1}))^2} \\
&\quad \times \Big[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i)\mu_{B(\mathbf{0},\gamma),m_n}(D_{p+1}) - \widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_{p+1})\mu_{B(\mathbf{0},\gamma),m_n}(D_i)\Big] \\
&= \frac{\mu_{B(\mathbf{0},\gamma),m_n}(D_{p+1})/\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_{p+1})}{(\mu_{B(\mathbf{0},\gamma),m_n}(D_{p+1}))^2} \\
&\quad \times \Big[\big(\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i) - \mu_{B(\mathbf{0},\gamma),m_n}(D_i)\big)\mu_{B(\mathbf{0},\gamma),m_n}(D_{p+1}) \\
&\quad - \big(\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_{p+1}) - \mu_{B(\mathbf{0},\gamma),m_n}(D_{p+1})\big)\mu_{B(\mathbf{0},\gamma),m_n}(D_i)\Big] \\
&= \frac{1 + o_p(1)}{(\mu_{B(\mathbf{0},\gamma)}(D_{p+1}))^2} \\
&\quad \times \Big[\big(\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i) - \mu_{B(\mathbf{0},\gamma),m_n}(D_i)\big)\mu_{B(\mathbf{0},\gamma)}(D_{p+1}) \\
&\quad - \big(\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_{p+1}) - \mu_{B(\mathbf{0},\gamma),m_n}(D_{p+1})\big)\mu_{B(\mathbf{0},\gamma)}(D_i)\Big]
\end{aligned}$$

by (2.5), Lemma 2.12 and Slutzky's lemma. For the vector in (2.43), recalling that $\mu_{B(\mathbf{0},\gamma)}(D_{p+1}) = \mu(A)$, and $F \in \mathbb{R}^{(p+1)\times(p+1)}$ as given in (2.15), we find

$$\begin{aligned}
&\Big(\frac{n}{m_n}\Big)^{d/2}\big[\widehat{R}_n(D_i, D_{p+1}) - R_n(D_i, D_{p+1})\big]_{i=1,\dots,p} \\
=&\Big(\frac{n}{m_n}\Big)^{d/2} \frac{1 + o_p(1)}{(\mu_{B(\mathbf{0},\gamma)}(D_{p+1}))^2} F \begin{pmatrix} \widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_1) - \mu_{B(\mathbf{0},\gamma),m_n}(D_1) \\ \vdots \\ \widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_p) - \mu_{B(\mathbf{0},\gamma),m_n}(D_p) \\ \widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_{p+1}) - \mu_{B(\mathbf{0},\gamma),m_n}(D_{p+1}) \end{pmatrix} \\
=:&\Big(\frac{n}{m_n}\Big)^{d/2} \frac{1 + o_p(1)}{(\mu_{B(\mathbf{0},\gamma)}(D_{p+1}))^2} F \boldsymbol{\mu}_{m_n}.
\end{aligned}$$

Thus, it remains to prove that

$$\Big(\frac{n}{m_n}\Big)^{d/2} \boldsymbol{\mu}_{m_n} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma), \tag{2.44}$$

where Σ is given in the statement of the Theorem. This can be done as in Davis and Mikosch [17], Corollary 3.3 using the Cramér-Wold device and similar ideas as in the proofs of Lemmas 2.12 and 2.13. In particular, note that for all $i, j \in \{1, \dots, p+1\}$ as $n \to \infty$,

$$\begin{aligned}
&\mathbb{C}ov[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i), \widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_j)] \\
&\quad \sim \frac{m_n}{n^d}\Big(\mu_{B(\mathbf{0},\gamma)}(D_i \cap D_j) + \sum_{\boldsymbol{h}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}} \tau_{B(\mathbf{0},\gamma)\times B(\boldsymbol{h},\gamma)}(D_i \times D_j)\Big).
\end{aligned}$$

□

Chapter 3

Semiparametric estimation for isotropic max-stable space-time processes

Abstract

Max-stable space-time processes have been developed to study extremal dependence in space-time data. We propose a semiparametric estimation procedure based on a closed form expression of the extremogram to estimate the parameters in a max-stable space-time process. We establish the asymptotic properties of the resulting parameter estimates and propose subsampling procedures to obtain asymptotically correct confidence intervals. A simulation study shows that the proposed procedure works well for moderate sample sizes. Finally, we apply this estimation procedure to fitting a max-stable model to radar rainfall measurements in a region in Florida.

AMS 2010 Subject Classifications: primary: 60G70, 62F12, 62G32; secondary: 62M10, 62M30, 62P12

Keywords: Brown-Resnick process; extremogram; max-stable process; semiparametric estimation; space-time process; subsampling; mixing

3.1 Introduction

Max-stable processes are a natural extension of the generalised extreme value distributions to infinite dimensions and provide a useful framework for modelling extremal dependence in continuous time or space. In this chapter we focus on the max-stable Brown-Resnick process, which was introduced in a time series framework in Brown and Resnick [8], in a spatial setting in Kabluchko et al. [47], and extended to a space-time setting in Davis et al. [19].

In the literature, various max-stable models and estimation procedures have been proposed for extremal data. For the Brown-Resnick process with parametrised dependence structure, inference has been based on composite likelihood methods. In particular, pairwise likelihood estimation has been found useful to estimate parameters in a max-stable process. A description of this method can be found in Padoan et al. [54] for the spatial setting, and Huser and Davison [43] in a space-time setting. Asymptotic results for pairwise likelihood estimates and detailed analyses in the space-time setting for the model analysed in this chapter are given in Davis et al. [20]. Unfortunately, parameter estimation using composite likelihood methods can be laborious, since the computation and subsequent optimisation of the objective function is time-consuming. Also the choice of good initial values for the optimisation of the composite likelihood is essential.

In this chapter we introduce a new semiparametric estimation procedure as an alternative to or as a prerequisite for composite likelihood methods. It is based on the extremogram as a natural extremal analog of the correlation function for stationary processes. It was introduced in Davis and Mikosch [17] for time series (also in Fasen et al. [35]), and they show consistency and asymptotic normality of an empirical extremogram estimate under weak mixing conditions. The empirical extremogram and its asymptotic properties in a spatial setting have been investigated in Cho et al. [15] and Chapter 2 of this thesis, which is based on the publication Buhl and Klüppelberg [12].

Assuming the same dependence structure for the Brown-Resnick space-time process as in [19, 20], we obtain a closed form expression of the extremogram containing the parameters of interest. We first estimate the extremogram nonparametrically by its empirical version, where we separate space and time. Weighted linear regression is then applied in order to produce parameter estimates.

Asymptotic normality of these semiparametric estimates requires asymptotic normality of the extremogram. For the spatial estimate we apply the CLT with mixing conditions as provided in Chapter 2, and for the timewise estimate that of [17]. The rate of convergence can be improved by a bias correction term, which we explain in detail for space and time. In a second step we prove then asymptotic normality of the weighted least squares parameter estimates, where constrained optimisation has to be applied, since one of the space and one of the time parameters has bounded support. Also the limit laws differ depending whether the parameter lies on the boundary or not. Since the asymptotic covariance matrices in the normal limits are difficult to access, we apply subsampling procedures to obtain pointwise confidence intervals for the parameters, also taking care of the different normal limits.

This chapter is organised as follows. Section 3.2 defines the isotropic Brown-Resnick process with its choice of dependence function used throughout for modelling extremes observed in space and time. The extremogram is introduced and its parametric form for our model is given. Based on gridded data, the nonparametric extremogram estimation is derived. Asymptotic normality of the parameter estimates is established in Section 3.3. Section 3.3.1 is dedicated to the asymptotic normality of the empirical spatial extremogram and its bias correction; Section 3.3.2 deals with the asymptotic properties of the spatial parameter estimates. Sections 3.3.3 and 3.3.4 present the analogues for the time parameters. In Section 3.4 we explain the subsampling procedure. We test our new semiparametric estimation procedure in a simulation study in Section 3.5. The chapter concludes with an analysis of daily rainfall maxima in a region in Florida in Section 3.6, where we also compare the semiparametric estimates with the previously obtained pairwise likelihood estimates. Some auxiliary results are summarised in an appendix.

3.2 Model description and semiparametric estimates

Throughout the chapter we consider a strictly stationary Brown-Resnick process in space and time with representation

$$\eta(\boldsymbol{s},t) = \bigvee_{j=1}^{\infty} \left\{ \xi_j \, e^{W_j(\boldsymbol{s},t)-\delta(\|\boldsymbol{s}\|,t)} \right\}, \quad (\boldsymbol{s},t) \in \mathbb{R}^2 \times [0,\infty), \tag{3.1}$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^2$, $\{\xi_j : j \in \mathbb{N}\}$ are points of a Poisson process on $[0,\infty)$ with intensity $\xi^{-2}d\xi$ and the dependence function δ is *nonnegative and conditionally negative definite*; i.e., for every $m \in \mathbb{N}$ and every $(\boldsymbol{s}^{(1)},t^{(1)}),\dots,(\boldsymbol{s}^{(m)},t^{(m)}) \in \mathbb{R}^2 \times [0,\infty)$, it holds that

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j \delta(\|\boldsymbol{s}^{(i)}-\boldsymbol{s}^{(j)}\|, |t^{(i)}-t^{(j)}|) \le 0$$

for all $a_1,\dots,a_m \in \mathbb{R}$ summing up to 0. The processes $\{W_j(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^2, t \in [0,\infty)\}$ are independent replicates of a Gaussian process $\{W(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^2, t \in [0,\infty)\}$ with stationary increments, $W(\mathbf{0},0)=0$, $\mathbb{E}[W(\boldsymbol{s},t)]=0$ and covariance function

$$\begin{aligned}\mathbb{C}ov[W(\boldsymbol{s}^{(1)},t^{(1)}),&W(\boldsymbol{s}^{(2)},t^{(2)})] \\ &= \delta(\|\boldsymbol{s}^{(1)}\|,t^{(1)}) + \delta(\|\boldsymbol{s}^{(2)}\|,t^{(2)}) - \delta(\|\boldsymbol{s}^{(1)}-\boldsymbol{s}^{(2)}\|, |t^{(1)}-t^{(2)}|).\end{aligned}$$

Representation (3.1) goes back to de Haan [24], Giné et al. [39] and Kabluchko et al. [47]. All finite-dimensional distributions are multivariate extreme value distributions with standard unit Fréchet margins, hence they are in particular multivariate regularly varying. Furthermore, they are perfectly characterised by the dependence function δ, which is termed the *semivariogram* of the process $\{W(\boldsymbol{s},t)\}$ in geostatistics: For $(\boldsymbol{s}^{(1)},t^{(1)}),(\boldsymbol{s}^{(2)},t^{(2)}) \in \mathbb{R}^2 \times [0,\infty)$, it is given by

$$\mathbb{V}ar[W(\boldsymbol{s}^{(1)},t^{(1)}) - W(\boldsymbol{s}^{(2)},t^{(2)})] = 2\delta(\|\boldsymbol{s}^{(1)}-\boldsymbol{s}^{(2)}\|, |t^{(1)}-t^{(2)}|).$$

Since we assume δ to depend only on the norm of $\boldsymbol{s}^{(1)}-\boldsymbol{s}^{(2)}$, the associated process is *(spatially) isotropic.*

In this chapter we assume the dependence function δ to be given for $v,u \ge 0$ by

$$\delta(v,u) = 2\theta_1 v^{\alpha_1} + 2\theta_2 u^{\alpha_2}, \tag{3.2}$$

where $0 < \alpha_1,\alpha_2 \le 2$ and $\theta_1,\theta_2 > 0$. This is the fractional class frequently used for dependence modelling, and here defined with respect to space and time.

The bivariate distribution function of $(\eta(\mathbf{0},0),\eta(\boldsymbol{h},u))$ is given for $x_1,x_2>0$ by

$$F(x_1,x_2)=\exp\Bigg\{-\frac{1}{x_1}\Phi\left(\frac{\log(x_2/x_1)}{\sqrt{2\delta(\|\boldsymbol{h}\|,|u|)}}+\sqrt{\frac{\delta(\|\boldsymbol{h}\|,|u|)}{2}}\right)\\ -\frac{1}{x_2}\Phi\left(\frac{\log(x_1/x_2)}{\sqrt{2\delta(\|\boldsymbol{h}\|,|u|)}}+\sqrt{\frac{\delta(\|\boldsymbol{h}\|,|u|)}{2}}\right)\Bigg\}, \tag{3.3}$$

where Φ denotes the standard normal distribution function (cf. Davis et al. [19]).

The parameters of interest are contained in the dependence function δ. We refer to (θ_1,α_1) as the *spatial parameter* and to (θ_2,α_2) as the *temporal parameter*. From the bivariate distribution function in (3.3), the pairwise density can be derived and pairwise likelihood methods can be used to estimate the parameters; cf. Davis et al. [20], Huser and Davison [43] and Padoan et al. [54]. Full likelihood inference is typically hardly tractable in a general multidimensional setting, as the number of terms occurring in the likelihood explode. More recently, however, parametric inference methods based on higher-dimensional margins have been proposed that work in specific scenarios, see for instance Genton et al. [37], who use triplewise instead of pairwise likelihood, Engelke et al. [34], who propose a threshold-based approach, or Thibaud and Opitz [64] and Wadsworth and Tawn [67], who use a censoring scheme for bias reduction.

In the following we introduce an alternative estimation approach, which is based on a closed form expression of the *extremogram*. The latter was introduced for time series by Davis and Mikosch [17] and for spatial and space-time processes by Cho et al. [15] and Steinkohl [62], respectively, and can be regarded as a correlogram for extreme events.

In this chapter we consider an isotropic Brown-Resnick process as a regularly varying stochastic processes $\{\eta(\boldsymbol{s}):\boldsymbol{s}\in\mathbb{R}^d\}$ where $d=1$ corresponds to a time series and $d=2$ to a spatial process, such that $d=3$ holds for the space-time process.

More precisely, we consider strictly stationary regularly varying processes $\{\eta(\boldsymbol{s}):\boldsymbol{s}\in\mathbb{R}^d\}$ for $d\in\mathbb{N}$, where all finite-dimensional distributions are regularly varying (cf. Hult and Lindskog [41] for definitions and results in a general framework and Resnick [59] for details about multivariate regular variation). As a prerequisite, we define for every finite set $\mathcal{I}\subset\mathbb{R}^d$ with cardinality $|\mathcal{I}|$ the vector

$$\eta_{\mathcal{I}}:=(\eta(\boldsymbol{s}):\boldsymbol{s}\in\mathcal{I})^\intercal.$$

Throughout, we abbreviate $\overline{\mathbb{R}}:=\mathbb{R}\cup\{-\infty,\infty\}$. For two positive functions f and g, we define the relation "$\sim$" as usually by $f(n)\sim g(n)$ as $n\to\infty$ if $\frac{f(n)}{g(n)}\to 1$ as $n\to\infty$.

Definition 3.1 (Regularly varying stochastic process)**.** *A strictly stationary stochastic process $\{\eta(\boldsymbol{s}):\boldsymbol{s}\in\mathbb{R}^d\}$ is called* regularly varying, *if there exists some normalizing sequence $0<a_n\to\infty$ such that $\mathbb{P}(|\eta(\mathbf{0})|>a_n)\sim n^{-d}$ as $n\to\infty$, and if for every finite set $\mathcal{I}\subset\mathbb{R}^d$,*

$$n^d\mathbb{P}\Big(\frac{\eta_{\mathcal{I}}}{a_n}\in\cdot\Big)\xrightarrow{v}\mu_{\mathcal{I}}(\cdot),\quad n\to\infty, \tag{3.4}$$

for some non-null Radon measure $\mu_{\mathcal{I}}$ *on the Borel sets in* $\overline{\mathbb{R}}^{|\mathcal{I}|}\backslash\{\mathbf{0}\}$*. In that case,*

$$\mu_{\mathcal{I}}(xC) = x^{-\beta}\mu_{\mathcal{I}}(C), \quad x > 0,$$

for every Borel set C *in* $\overline{\mathbb{R}}^{|\mathcal{I}|}\backslash\{\mathbf{0}\}$*. The notation* $\stackrel{v}{\to}$ *stands for vague convergence, and* $\beta > 0$ *is called the* index of regular variation*.*

For every $\boldsymbol{s} \in \mathbb{R}^d$ and $\mathcal{I} = \{\boldsymbol{s}\}$ we set $\mu_{\{\boldsymbol{s}\}}(\cdot) = \mu_{\{\mathbf{0}\}}(\cdot) =: \mu(\cdot)$, which is justified by stationarity.

Assuming strict stationarity and spatial isotropy of a regularly varying space-time process $\{\eta(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathbb{R}^2, t \in [0, \infty)\}$ we can define its extremogram at two points $(\boldsymbol{s}_1, t_1)$ and $(\boldsymbol{s}_2, t_2)$ only in terms of the spatial and temporal lags $v := \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$ and $u := |t_1 - t_2|$.

Definition 3.2 (The extremogram)**.** *For a regularly varying strictly stationary isotropic space-time process* $\{\eta(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in \mathbb{R}^2 \times [0, \infty)\}$ *we define the* space-time extremogram *for two* μ*-continuous Borel sets* A *and* B *in* $\overline{\mathbb{R}}\backslash\{0\}$ *(i.e.* $\mu(\partial A) = \mu(\partial B) = 0$*) such that* $\mu(A) > 0$ *by*

$$\rho_{AB}(v, u) = \lim_{n\to\infty} \frac{P\left(\eta(\boldsymbol{s}_1, t_1)/a_n \in A, \eta(\boldsymbol{s}_2, t_2)/a_n \in B\right)}{P\left(\eta(\boldsymbol{s}_1, t_1)/a_n \in A\right)}, \tag{3.5}$$

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

Setting $A = B = (1, \infty)$, we rediscover the *tail dependence coefficient* $\chi(v, u) = \rho_{(1,\infty)(1,\infty)}(v, u)$. For the isotropic Brown-Resnick process there is a closed form expression for $\chi(v, u)$, which is the basis for our estimation procedure.

Lemma 3.3 (Davis et al. [19], equation (3.1))**.** *Let* η *be the strictly stationary isotropic Brown-Resnick process in* $\mathbb{R}^2 \times [0, \infty)$ *as defined in* (3.1) *with dependence function given in* (3.2)*. Then for* $A = B = (1, \infty)$ *the extremogram of* η *is given by*

$$\chi(v, u) = 2\Big(1 - \Phi\Big(\sqrt{\frac{1}{2}\delta(v, u)}\Big)\Big) = 2\big(1 - \Phi(\sqrt{\theta_1 v^{\alpha_1} + \theta_2 u^{\alpha_2}})\big), \quad v, u \geq 0. \tag{3.6}$$

Solving equation (3.6) for $\delta(v, u)$ leads to

$$\frac{\delta(v, u)}{2} = \theta_1 v^{\alpha_1} + \theta_2 u^{\alpha_2} = \Big(\Phi^{-1}\big(1 - \frac{1}{2}\chi(v, u)\big)\Big)^2. \tag{3.7}$$

For temporal lag 0 and taking the logarithm on both sides we have

$$2\log\Big(\Phi^{-1}\big(1 - \frac{1}{2}\chi(v, 0)\big)\Big) = \log(\theta_1) + \alpha_1 \log v.$$

In the same way, we obtain

$$2\log\Big(\Phi^{-1}\big(1 - \frac{1}{2}\chi(0, u)\big)\Big) = \log(\theta_2) + \alpha_2 \log u.$$

These equations are the basis for parameter estimates. We replace the extremogram on the left hand side in both of these equations by nonparametric estimates at different lags. Then we use constrained weighted least squares estimation in a linear regression framework to obtain parameter estimates.

The estimation procedure is based on the following observation scheme for the space-time data.

Condition 3.4. *(1) The locations lie on a regular 2-dimensional grid*

$$\mathcal{S}_n = \big\{(i_1, i_2) : i_1, i_2 \in \{1, \dots, n\}\big\} = \big\{\boldsymbol{s}_i : i = 1, \dots, n^2\big\}.$$

(2) The time points are equidistant, given by the set $\{t_1, \dots, t_T\}$.

Remark 3.5. The assumption of a regular grid can be relaxed in various ways. A simple, but notationally more involved extension is the generalisation to rectangular grids, cf. Section 2.3. Furthermore, it is possible to assume that the observation area consists of random locations given by points of a Poisson process, see for instance Cho et al. [15], Section 2.3, or Steinkohl [62], Section 4.5.2. Also deterministic, but irregularly spaced locations, could be considered as treated in [62] in Section 4.5.1 in the context of pairwise likelihood estimation. In order to make our method transparent we focus on observations on a regular grid.

The following scheme provides the semiparametric estimation procedure in detail.
Denote by $\mathcal{V}$ and $\mathcal{U}$ finite sets of spatial and temporal lags, on which the estimation is based. We denote by "lag" the norm or absolute value of the difference of two spatial locations or two time points, respectively. Concerning the choice of $\mathcal{V}$ and $\mathcal{U}$, we generally include those lags which show clear extremal dependence between locations or time points. Larger lags should not be considered, since they may introduce a bias in the least squares estimates, similarly as in pairwise likelihood estimation; cf. Section 5.5.3. One way to determine the range of clear extremal dependence are permutation tests, which we describe at the end of Section 3.6.

(1) Nonparametric estimates for the extremogram:
Summarise all pairs of $\mathcal{S}_n$ which give rise to the same spatial lag $v \in \mathcal{V}$ into

$$N(v) = \{(i,j) \in \{1, \dots, n^2\}^2 : \ \|\boldsymbol{s}_i - \boldsymbol{s}_j\| = v\}.$$

For all $t \in \{t_1, \dots, t_T\}$ estimate the *spatial extremogram* by

$$\widehat{\chi}^{(t)}(v, 0) = \frac{\dfrac{1}{|N(v)|} \displaystyle\sum_{i=1}^{n^2} \sum_{\substack{j=1 \\ \|\boldsymbol{s}_i - \boldsymbol{s}_j\| = v}}^{n^2} \mathbb{1}_{\{\eta(\boldsymbol{s}_i, t) > q, \eta(\boldsymbol{s}_j, t) > q\}}}{\dfrac{1}{n^2} \displaystyle\sum_{i=1}^{n^2} \mathbb{1}_{\{\eta(\boldsymbol{s}_i, t) > q\}}}, \quad v \in \mathcal{V}, \tag{3.8}$$

where q is a large quantile (to be specified) of the standard unit Frechét distribution.

For all $\boldsymbol{s} \in \mathcal{S}_n$ estimate the *temporal extremogram* by

$$\widehat{\chi}^{(\boldsymbol{s})}(0,u) = \frac{\frac{1}{T-u}\sum_{k=1}^{T-u} \mathbb{1}_{\{\eta(\boldsymbol{s},t_k)>q,\eta(\boldsymbol{s},t_k+u)>q\}}}{\frac{1}{T}\sum_{k=1}^{T} \mathbb{1}_{\{\eta(\boldsymbol{s},t_k)>q\}}}, \quad u \in \mathcal{U}\,, \tag{3.9}$$

where again q is a large (possibly different) quantile of the standard unit Frechét distribution (2) The overall "spatial" and "temporal" extremogram estimates are defined as averages over the temporal and spatial locations, respectively; i.e.,

$$\widehat{\chi}(v,0) = \frac{1}{T}\sum_{k=1}^{T} \widehat{\chi}^{(t_k)}(v,0), \quad v \in \mathcal{V}, \tag{3.10}$$

$$\widehat{\chi}(0,u) = \frac{1}{n^2}\sum_{i=1}^{n^2} \widehat{\chi}^{(\boldsymbol{s}_i)}(0,u), \quad u \in \mathcal{U}. \tag{3.11}$$

Parameter estimates for $\theta_1, \alpha_1, \theta_2$ and α_2 are found by using constrained weighted least squares estimation:

$$\begin{pmatrix}\widehat{\theta}_1\\ \widehat{\alpha}_1\end{pmatrix} = \underset{\substack{\theta_1,\alpha_1>0\\ \alpha_1\in(0,2]}}{\arg\min} \sum_{v\in\mathcal{V}} w_v\Big(2\log\Big(\Phi^{-1}\big(1-\frac{1}{2}\widehat{\chi}(v,0)\big)\Big) - \big(\log(\theta_1)+\alpha_1\log(v)\big)\Big)^2, \tag{3.12}$$

$$\begin{pmatrix}\widehat{\theta}_2\\ \widehat{\alpha}_2\end{pmatrix} = \underset{\substack{\theta_2,\alpha_2>0\\ \alpha_2\in(0,2]}}{\arg\min} \sum_{u\in\mathcal{U}} w_u\Big(2\log\Big(\Phi^{-1}(1-\frac{1}{2}\widehat{\chi}(0,u))\Big) - \big(\log(\theta_2)+\alpha_2\log(u)\big)\Big)^2, \tag{3.13}$$

with weights $w_u > 0$ and $w_v > 0$.

We call the estimates $(\widehat{\theta}_1, \widehat{\alpha}_1)$ and $(\widehat{\theta}_2, \widehat{\alpha}_2)$ *weighted least squares estimates* (WLSE).

3.3 Estimation of the isotropic Brown-Resnick process

In this section we investigate asymptotic properties of the WLSE $(\widehat{\theta}_1, \widehat{\alpha}_1)$ and $(\widehat{\theta}_2, \widehat{\alpha}_2)$.

For a central limit theorem of the extremogram we need a sufficiently precise estimate for the extremogram (3.6), which we give now.

Lemma 3.6. *Let $\boldsymbol{s}, \boldsymbol{h} \in \mathbb{R}^2$ and $t \in [0,\infty)$. For every sequence $a_n \to \infty$ we have*

$$\begin{aligned}&\frac{\mathbb{P}(\eta(\boldsymbol{s},t)>a_n, \eta(\boldsymbol{s}+\boldsymbol{h},t)>a_n)}{\mathbb{P}(\eta(\boldsymbol{s},t)>a_n)}\\ =&\Big[\chi(\|\boldsymbol{h}\|,0) + \frac{1}{2a_n}\big(\chi(\|\boldsymbol{h}\|,0)-2\big)\big(\chi(\|\boldsymbol{h}\|,0)-1\big)\Big](1+o(1)).\end{aligned}$$

Lemma 3.6 is a direct application of Lemma A.1(b) for $A = B = (1,\infty)$ and equation (A.4). This applies since $\{\eta(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^2\}$ has finite-dimensional standard unit Fréchet marginal distributions. Also note that $a_n \sim n^2$ as $n \to \infty$ according to Definition 3.1.

Since the WLSE are functions of the spatial and temporal extremograms, we first derive the asymptotic properties of $\widehat{\chi}^{(t)}$ and $\widehat{\chi}^{(\boldsymbol{s})}$ for a fixed time point t and a fixed location $\boldsymbol{s}$, respectively. Sections 3.3.1 and 3.3.2 focus on the spatial parameters, whereas Sections 3.3.3 and 3.3.4 handle the temporal parameters. We use several results for the extremogram provided in Appendix B.1 and in Chapter 2.

3.3.1 Asymptotics of the empirical spatial extremogram

We prove a central limit theorem for the empirical spatial extremogram of the Brown-Resnick process (3.1) based on a finite set of spatial lags

$$\mathcal{V} = \{v_1, \dots, v_p\},$$

which show clear extremal dependence as explained in Section 3.2. First we show that the empirical extremogram centred by the pre-asymptotic version is asymptotically normal.

Theorem 3.7. *For a fixed time point $t \in \{t_1, \dots, t_T\}$, consider the spatial Brown-Resnick process $\left\{\eta(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathbb{R}^2\right\}$ as defined in* (3.1) *with dependence function given in* (3.2)*. Set $m_n = n^{\beta_1}$ for $\beta_1 \in (0, 1/2)$. Then the empirical spatial extremogram $\widehat{\chi}^{(t)}(v, 0)$ defined in* (3.8) *with the quantile $q = m_n^2$ satisfies*

$$\frac{n}{m_n}\left(\widehat{\chi}^{(t)}(v,0) - \chi_n(v,0)\right)_{v\in\mathcal{V}} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Pi_1^{(\text{iso})}), \quad n \to \infty,$$

where the covariance matrix $\Pi_1^{(\text{iso})}$ is specified in equation (3.19) *below, and χ_n is the pre-asymptotic spatial extremogram,*

$$\chi_n(v,0) = \frac{\mathbb{P}(\eta(\mathbf{0},0) > m_n^2, \eta(\boldsymbol{h},0) > m_n^2)}{\mathbb{P}(\eta(\mathbf{0},0) > m_n^2)}, \quad v = \|\boldsymbol{h}\| \in \mathcal{V}\,. \tag{3.14}$$

Proof. As $\eta(\mathbf{0}, t)$ has standard unit Fréchet marginal distributions, we can choose $a_{m_n} = m_n^2$ by Definition 3.1 of regular variation.

We apply Theorem 2.6 by verifying conditions (M1)-(M4) of that theorem for $\{\eta(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathbb{R}^2\}$, $d = 2$, and $A = B = (1, \infty)$. Condition (M1) is satisfied by equation (B.2).

To show conditions (M2)-(M4) we choose sequences $m_n = n^{\beta_1}$ and $r_n = n^{\beta_2}$ for $0 < \beta_1 < 1/2$ and $0 < \beta_2 < \beta_1$. For this choice m_n and r_n increase to infinity with $m_n = o(n)$ and $r_n = o(m_n)$ as required.

Condition (M2); i.e., $m_n^2 r_n^2/n = n^{2(\beta_1+\beta_2)-1} \to 0$ holds if and only if $\beta_2 \in (0, \min\{\beta_1, (1/2-\beta_1)\})$.

We now show condition (M3). Choose $\gamma > 0$, such that all lags in $\mathcal{V}$ lie in $B(\mathbf{0}, \gamma) := \{\boldsymbol{s} \in \mathbb{Z}^2 : \|\boldsymbol{s}\| \leq \gamma\}$. Denote by $B(\boldsymbol{h}, \gamma) := \{\boldsymbol{s} \in \mathbb{Z}^2 : \|\boldsymbol{s} - \boldsymbol{h}\| \leq \gamma\} = \boldsymbol{h} + B(\mathbf{0}, \gamma)$ for $\boldsymbol{h} \in \mathbb{R}^2$. For $\varepsilon > 0$, like in Example 2.10, we have for $\boldsymbol{s}, \boldsymbol{s}' \in \mathbb{R}^2$ by a Taylor expansion,

$$\begin{aligned}&\mathbb{P}(\eta(\boldsymbol{s},t) > \varepsilon m_n^2, \eta(\boldsymbol{s}',t) > \varepsilon m_n^2)\\ &\qquad = 1 - 2\mathbb{P}(\eta(\mathbf{0},0) \leq \varepsilon m_n^2) + \mathbb{P}(\eta(\boldsymbol{s},t) \leq \varepsilon m_n^2, \eta(\boldsymbol{s}',t) \leq \varepsilon m_n^2)\end{aligned}$$

$$
\begin{aligned}
&= 1 - 2\exp\Big\{ -\frac{1}{x}\Big\} + \exp\Big\{ -\frac{2-\chi(\|\boldsymbol{s}-\boldsymbol{s}'\|,0)}{\varepsilon m_n^2}\Big\} \\
&= \frac{1}{\varepsilon m_n^2}\chi(\|\boldsymbol{s}-\boldsymbol{s}'\|,0) + \mathcal{O}\Big(\frac{1}{m_n^4}\Big), \quad n\to\infty.
\end{aligned}
$$

Therefore, for $\|\boldsymbol{h}\| \geq 2\gamma$,

$$
\begin{aligned}
&\mathbb{P}(\max_{\boldsymbol{s}\in B(\boldsymbol{0},\gamma)} \eta(\boldsymbol{s},t) > \varepsilon m_n^2, \max_{\boldsymbol{s}'\in B(\boldsymbol{h},\gamma)} \eta(\boldsymbol{s}',t) > \varepsilon m_n^2) \\
\leq &\sum_{\boldsymbol{s}\in B(\boldsymbol{0},\gamma)} \sum_{\boldsymbol{s}'\in B(\boldsymbol{h},\gamma)} \mathbb{P}(\eta(\boldsymbol{s},t) > \varepsilon m_n^2, \eta(\boldsymbol{s}',t) > \varepsilon m_n^2) \\
= &\sum_{\boldsymbol{s}\in B(\boldsymbol{0},\gamma)} \sum_{\boldsymbol{s}'\in B(\boldsymbol{h},\gamma)} \Big\{\frac{1}{\varepsilon m_n^2}\chi(\|\boldsymbol{s}-\boldsymbol{s}'\|,0) + \mathcal{O}\Big(\frac{1}{m_n^4}\Big)\Big\} \\
\leq &\frac{2|B(\boldsymbol{0},\gamma)|^2}{\varepsilon m_n^2}\big(1-\Phi(\sqrt{\theta_1(\|\boldsymbol{h}\|-2\gamma)^{\alpha_1}}\big) + \mathcal{O}\Big(\frac{1}{m_n^4}\Big),
\end{aligned}
\tag{3.15}
$$

as $n\to\infty$, where we have used (3.6). Summarise $V := \{v = \|\boldsymbol{h}\| : \boldsymbol{h}\in\mathbb{Z}^2\}$ and note that $|\{\boldsymbol{h}\in\mathbb{Z}^2 : \|\boldsymbol{h}\| = v\}| = \mathcal{O}(v)$. Therefore, for $k\geq 2\gamma$,

$$
\begin{aligned}
L_{m_n} &:= \limsup_{n\to\infty} m_n^2 \sum_{\substack{\boldsymbol{h}\in\mathbb{Z}^2\\ k<\|\boldsymbol{h}\|\leq r_n}} \mathbb{P}\Big(\max_{\boldsymbol{s}\in B(\boldsymbol{0},\gamma)} \eta(\boldsymbol{s},t) > \varepsilon m_n^2, \max_{\boldsymbol{s}'\in B(\boldsymbol{h},\gamma)} \eta(\boldsymbol{s}',t) > \varepsilon m_n^2\Big) \\
&\leq 2|B(\boldsymbol{0},\gamma)|^2 \limsup_{n\to\infty}\Big\{ \sum_{\substack{\boldsymbol{h}\in\mathbb{Z}^2\\ k<\|\boldsymbol{h}\|\leq r_n}} \Big\{\frac{1}{\varepsilon}(1-\Phi(\sqrt{\theta_1(\|\boldsymbol{h}\|-2\gamma)^{\alpha_1}}))\Big\} + \mathcal{O}\Big(\Big(\frac{r_n}{m_n}\Big)^2\Big)\Big\} \\
&\leq C_1 \limsup_{n\to\infty} \sum_{\substack{v\in V:\\ k<v\leq r_n}} \Big\{\frac{v}{\varepsilon}2\big(1-\Phi(\sqrt{\theta_1(v-2\gamma)^{\alpha_1}})\big)\Big\},
\end{aligned}
$$

for some constant $C_1 > 0$. For the term $\mathcal{O}((r_n/m_n)^2)$ we use that $r_n/m_n \to 0$. From Lemma B.3 and the fact that $1-\Phi(x) \leq \exp\{-x^2/2\}$ for $x>0$, we find for $C_2>0$,

$$
L_{m_n} \leq C_2 k^2 \exp\Big\{-\frac{1}{2}\theta_1(k-2\gamma)^{\alpha_1}\Big\}.
$$

Since $\alpha_1 > 0$, the right hand side converges to 0 as $k\to\infty$ ensuring condition (M3).

Now we turn to the mixing conditions (M4).
We start with (M4i). With V as before, and with equation (B.2), we estimate, recalling from above that the number of lags $\|\boldsymbol{h}\| = v$ is of oder $\mathcal{O}(v)$,

$$
m_n^2 \sum_{\boldsymbol{h}\in\mathbb{Z}^2:\|\boldsymbol{h}\|>r_n} \alpha_{1,1}(\|\boldsymbol{h}\|) \leq C_1 m_n^2 \sum_{v\in V:v>r_n} v\,\alpha_{1,1}(v) \leq 4C_1 m_n^2 \sum_{v\in V:v>r_n} v\,e^{-\theta_1 v^{\alpha_1}/2}.
$$

By Lemma B.3 we find

$$
m_n^2 \sum_{v\in V:v>r_n} v\,e^{-\theta_1 v^{\alpha_1}/2} \leq c m_n^2 r_n^2\; e^{-\theta_1 r_n^{\alpha_1}/2} = c m_n^2 r_n^2\; e^{-\theta_1 n^{\alpha_1\beta_2}/2} \to 0, \quad n\to\infty.
$$

By the same arguments condition (M4ii) is satisfied.
Condition (M4iii) holds by equation (B.2), since

$$m_n\, n\, \alpha_{1,n^2}(r_n) \leq 4n^3 m_n\, e^{-\theta_1 r_n^{\alpha_1}/2} \to 0, \quad n \to \infty.$$

For the specification of the asymptotic covariance matrix we apply Theorem 2.6 for the isotropic case, where each spatial lag v_i arises from a set of different vectors $\boldsymbol{h}$, all with same Euclidean norm v_i. For $i \in \{1, \ldots, p\}$ such that $v_i \in \mathcal{V}$, we summarise these into

$$L(v_i) := \{\boldsymbol{h} \in \mathbb{Z}^2 : \|\boldsymbol{h}\| = v_i\} = \{\boldsymbol{h}_1^{(i)}, \ldots, \boldsymbol{h}_{\ell_i}^{(i)}\},$$

where $\ell_i := |L(v_i)|$. Based on the preceding steps of the proof, we conclude from that theorem that

$$\frac{n}{m_n}\big(\widehat{\chi}^{(t)}(\boldsymbol{h}_1^{(i)}, 0) - \chi_n(\boldsymbol{h}_1^{(i)}, 0), \ldots, \widehat{\chi}^{(t)}(\boldsymbol{h}_{\ell_i}^{(i)}, 0) - \chi_n(\boldsymbol{h}_{\ell_i}^{(i)}, 0)\big)^{\intercal}_{i=1,\ldots,p} \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \Pi_1^{(\text{space})}),$$

where $\Pi_1^{(\text{space})}$ is specified in equations (2.13)-(2.15). Furthermore, $\widehat{\chi}^{(t)}(\boldsymbol{h}, 0)$ is the empirical extremogram for each vector $\boldsymbol{h}$ as specified above.

Define $N(\boldsymbol{h}) := \{(i,j) \in \{1, \ldots, n^2\} : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}\}$, then the numerator in (3.8) normalises by $|N(\boldsymbol{h})|$ (instead of $|N(v)|$) and the sum runs over $\boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}$ vector-wise (instead of equality in norm). Hence, $|N(v_i)| = \sum_{\boldsymbol{h} \in L(v_i)} |N(\boldsymbol{h})|$. Isotropy implies for the pre-asymptotic extremogram that $\chi_n(v_i, 0) = \chi_n(\boldsymbol{h}, 0)$ for all $\boldsymbol{h} \in L(v_i)$, such that

$$\chi_n(v_i, 0) = \sum_{\boldsymbol{h} \in L(v_i)} \frac{|N(\boldsymbol{h})|}{|N(v_i)|} \chi_n(v_i, 0) = \sum_{\boldsymbol{h} \in L(v_i)} \frac{|N(\boldsymbol{h})|}{|N(v_i)|} \chi_n(\boldsymbol{h}, 0) \tag{3.16}$$

as well as, by the definition of the estimator in (3.8),

$$\widehat{\chi}^{(t)}(v_i, 0) = \sum_{\boldsymbol{h} \in L(v_i)} \frac{|N(\boldsymbol{h})|}{|N(v_i)|} \widehat{\chi}^{(t)}(v_i, 0) = \sum_{\boldsymbol{h} \in L(v_i)} \frac{|N(\boldsymbol{h})|}{|N(v_i)|} \widehat{\chi}^{(t)}(\boldsymbol{h}, 0). \tag{3.17}$$

We conclude by (3.16) and (3.17) that

$$\widehat{\chi}^{(t)}(v_i, 0) - \chi_n(v_i, 0) = \sum_{\boldsymbol{h} \in L(v_i)} \frac{|N(\boldsymbol{h})|}{|N(v_i)|} \big(\widehat{\chi}^{(t)}(\boldsymbol{h}, 0) - \chi_n(\boldsymbol{h}, 0)\big).$$

To obtain a concise representation of the asymptotic normal law for the isotropic extremogram, we define row vectors $(|N(\boldsymbol{h})|/|N(v_i)| : \boldsymbol{h} \in L(v_i))$ for $i = 1, \ldots, p$. Set $L := \sum_{i=1}^p \ell_i$ and define the $p \times L-$matrix

$$N := \begin{pmatrix} \left(\frac{|N(\boldsymbol{h})|}{|N(v_1)|} : \boldsymbol{h} \in L(v_1)\right) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \left(\frac{|N(\boldsymbol{h})|}{|N(v_2)|} : \boldsymbol{h} \in L(v_2)\right) & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \left(\frac{|N(\boldsymbol{h})|}{|N(v_p)|} : \boldsymbol{h} \in L(v_p)\right) \end{pmatrix}. \tag{3.18}$$

Then we find

$$\begin{aligned} & \frac{n}{m_n}\left(\widehat{\chi}^{(t)}(v_i,0) - \chi_n(v_i,0)\right)^\intercal_{i=1,\dots,p} \\ = \; & \frac{n}{m_n} N\left(\widehat{\chi}^{(t)}(\boldsymbol{h}_1^{(i)},0) - \chi_n(\boldsymbol{h}_1^{(i)},0), \dots, \widehat{\chi}^{(t)}(\boldsymbol{h}_{\ell_i}^{(i)},0) - \chi_n(\boldsymbol{h}_{\ell_i}^{(i)},0)\right)^\intercal_{i=1,\dots,p} \\ \xrightarrow{d} \; & \mathcal{N}(\mathbf{0}, N\Pi_1^{(\text{space})}N^\intercal), \quad n \to \infty, \end{aligned}$$

such that

$$\Pi_1^{(\text{iso})} := N\Pi_1^{(\text{space})}N^\intercal. \tag{3.19}$$

□

Corollary 3.8. *Under the conditions of Theorem 3.7 the averaged spatial extremogram in* (3.10) *satisfies*

$$\frac{n}{m_n}\Big(\frac{1}{T}\sum_{k=1}^{T}\widehat{\chi}^{(t_k)}(v,0) - \chi_n(v,0)\Big)_{v\in\mathcal{V}} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Pi_2^{(\text{iso})}), \quad n \to \infty,$$

with covariance matrix $\Pi_2^{(\text{iso})}$ *specified in* (3.23) *below.*

Proof. For the first part of the proof, we neglect spatial isotropy. This part is similar to the proof of Theorem 2.6 and Corollary 3.4 of Davis and Mikosch [17]. We use the notation of the proof of Theorem 3.7. Enumerate the set of spatial lag vectors inherent in the estimation of the extremogram as $\{\boldsymbol{h}_1^{(i)}, \dots, \boldsymbol{h}_{\ell_i}^{(i)} : i = 1, \dots, p\}$ and let $\gamma \geq \max\{v_1, \dots, v_p\}$. Define the vector process

$$\{\boldsymbol{Y}(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^2\} = \{(\eta(\boldsymbol{s}+\boldsymbol{h}, t_k) : \boldsymbol{h} \in B(\mathbf{0},\gamma))^\intercal_{k=1,\dots,T} : \boldsymbol{s} \in \mathbb{R}^2\}.$$

Let $A = B = (1,\infty)$. Consider $i = 1, \dots, p$, $j = 1, \dots, \ell_i$, and $k = 1, \dots, T$. Define sets $D_{j,k}^{(i)}$ by

$$\{\boldsymbol{Y}(\boldsymbol{s}) \in D_{j,k}^{(i)}\} = \{\eta(\boldsymbol{s}, t_k) \in A, \eta(\boldsymbol{s}', t_k) \in B : \boldsymbol{s} - \boldsymbol{s}' = \boldsymbol{h}_j^{(i)}\},$$

and the sets D_k by

$$\{\boldsymbol{Y}(\boldsymbol{s}) \in D_k\} = \{\eta(\boldsymbol{s}, t_k) \in A\}.$$

For $\boldsymbol{h} \in \mathbb{R}^2$ let $B_T(\boldsymbol{h},\gamma) := B(\boldsymbol{h},\gamma) \times \{t_1, \dots, t_T\}$. For $\mu_{B_T(\mathbf{0},\gamma)}$-continuous Borel sets C and D

in $\overline{\mathbb{R}}^{T|B(\mathbf{0},\gamma)|}\backslash\{\mathbf{0}\}$, regular variation yields the existence of the limit measures

$$\mu_{B_T(\mathbf{0},\gamma)}(C) := \lim_{n\to\infty} m_n^2 \mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{m_n^2} \in C\Big)$$
$$\tau_{B_T(\mathbf{0},\gamma)\times B_T(\boldsymbol{h},\gamma)}(C\times D) := \lim_{n\to\infty} m_n^2 \mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{m_n^2} \in C, \frac{\boldsymbol{Y}(\boldsymbol{h})}{m_n^2} \in D\Big).$$

By time stationarity we have $\mu_{B_T(\mathbf{0},\gamma)}(D_k) = \mu(A)$,

$$\widehat{\chi}^{(t_k)}(\boldsymbol{h}_j^{(i)},0) \sim \widehat{R}_{m_n}(D_{j,k}^{(i)}, D_k) := \widehat{\mu}_{B_T(\mathbf{0},\gamma),m_n}(D_{j,k}^{(i)})/\widehat{\mu}_{B_T(\mathbf{0},\gamma),m_n}(D_k), \quad n\to\infty, \tag{3.20}$$

where the $\widehat{\mu}_{B_T(\mathbf{0},\gamma),m_n}(\cdot)$ are empirical estimators of $\mu_{B_T(\mathbf{0},\gamma)}(\cdot)$ defined as

$$\widehat{\mu}_{B_T(\mathbf{0},\gamma),m_n}(\cdot) := \Big(\frac{m_n}{n}\Big)^2 \sum_{\boldsymbol{s}\in\mathcal{S}_n} \mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{s})}{m_n^2}\in\cdot\}}. \tag{3.21}$$

Likewise we have for the pre-asymptotic quantities

$$\chi_n(\boldsymbol{h}_j^{(i)},0) = R_{m_n}(D_{j,k}^{(i)}, D_k) := \frac{\mathbb{P}(\boldsymbol{Y}(\mathbf{0})/m_n^2 \in D_{j,k}^{(i)})}{\mathbb{P}(\boldsymbol{Y}(\mathbf{0})/m_n^2 \in D_k)} =: \frac{\mu_{B_T(\mathbf{0},\gamma),m_n}(D_{j,k}^{(i)})}{\mu_{B_T(\mathbf{0},\gamma),m_n}(D_k)}, \tag{3.22}$$

which are independent of time t_k by stationarity. For notational ease we abbreviate in the following

$$\mu_{B_T(\mathbf{0},\gamma)}(\cdot) = \mu_\gamma(\cdot), \quad \mu_{B_T(\mathbf{0},\gamma),m_n}(\cdot) = \mu_{\gamma,m_n}(\cdot), \quad \text{and} \quad \widehat{\mu}_{B_T(\mathbf{0},\gamma),m_n}(\cdot) = \widehat{\mu}_{\gamma,m_n}(\cdot)$$

For each $k\in\{1,\dots,T\}$ we now define the matrices

$$F^{(k)} = [F_1, F_2^{(k)}]$$

with $F_1 \in \mathbb{R}^{L\times L}$ and $F_2^{(k)} \in \mathbb{R}^L$ given by

$$F_1 = \operatorname{diag}(\mu(A)) \quad \text{and} \quad F_2^{(k)} := (-\mu_\gamma(D_{1,k}^{(1)}),\dots,-\mu_\gamma(D_{\ell_1,k}^{(1)}),\dots,-\mu_\gamma(D_{\ell_p,k}^{(p)}))^\top.$$

Although $F_2^{(k)}$ is constant over $k\in\{1,\dots,T\}$ by time stationarity, we keep the index to clarify the notation. Define the $TL\times T(L+1)$-matrix $\boldsymbol{F}$ and the column vector $\widehat{\boldsymbol{\chi}}-\boldsymbol{\chi}_n$ with TL components as

$$\boldsymbol{F} := \begin{pmatrix} F^{(1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & F^{(2)} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & F^{(T)} \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{\chi}}-\boldsymbol{\chi}_n := \begin{pmatrix} \widehat{\chi}^{(t_1)}(h_1^{(1)},0) - \chi_n(h_1^{(1)},0) \\ \vdots \\ \widehat{\chi}^{(t_1)}(h_{\ell_1}^{(1)},0) - \chi_n(h_{\ell_1}^{(1)},0) \\ \vdots \\ \widehat{\chi}^{(t_1)}(h_{\ell_p}^{(p)},0) - \chi_n(h_{\ell_p}^{(p)},0) \\ \vdots \\ \widehat{\chi}^{(t_T)}(h_{\ell_p}^{(p)},0) - \chi_n(h_{\ell_p}^{(p)},0) \end{pmatrix}.$$

Define the vector $(\widehat{\boldsymbol{R}}_{m_n} - \boldsymbol{R}_{m_n})$ with the quantities from (3.20) and the corresponding pre-asymptotic quantities from (3.22) exactly in the same way. Furthermore, define for $k = 1, \ldots, T$ the vectors in $\mathbb{R}^{L+1}$

$$\boldsymbol{\mu}^{(k)}_{\gamma,m_n} = \\ \big(\mu_{\gamma,m_n}(D^{(1)}_{1,k}), \ldots, \mu_{\gamma,m_n}(D^{(1)}_{\ell_1,k}), \ldots\ldots, \mu_{\gamma,m_n}(D^{(p)}_{1,k}), \ldots, \mu_{\gamma,m_n}(D^{(p)}_{\ell_p,k}), \mu_{\gamma,m_n}(D_k)\big)^{\intercal},$$

which we stack one on top of the other giving a vector $\boldsymbol{\mu}_{\gamma,m_n}$ in $\mathbb{R}^{T(L+1)}$, and $\widehat{\boldsymbol{\mu}}_{\gamma,m_n}$ analogously. Then we obtain

$$\widehat{\boldsymbol{\chi}} - \boldsymbol{\chi}_n = (1+o(1))(\widehat{\boldsymbol{R}}_{m_n} - \boldsymbol{R}_{m_n}) = \frac{1+o_p(1)}{\mu(A)^2} \boldsymbol{F} \, (\widehat{\boldsymbol{\mu}}_{\gamma,m_n} - \boldsymbol{\mu}_{\gamma,m_n}),$$

where the last step follows as in the proof of Theorem 2.6 and involves Slutzky's theorem. Using ideas of the proof of Lemma (2.12), we observe that as $n \to \infty$,

$$\mathbb{C}ov\Big[\widehat{\mu}_{B_T(\boldsymbol{0},\gamma),m_n}(C), \widehat{\mu}_{B_T(\boldsymbol{0},\gamma),m_n}(D)\Big] \\ \sim \Big(\frac{m_n}{n}\Big)^2 \Big(\mu_{B_T(\boldsymbol{0},\gamma)}(C \cap D) + \sum_{\boldsymbol{0} \neq \boldsymbol{h} \in \mathbb{Z}^2} \tau_{B_T(\boldsymbol{0},\gamma) \times B_T(\boldsymbol{h},\gamma)}(C \times D)\Big) =: \Big(\frac{m_n}{n}\Big)^2 c_{C,D}.$$

With $\Sigma \in \mathbb{R}^{T(L+1) \times T(L+1)}$ defined as

$$\Sigma = \begin{pmatrix} c_{D^{(1)}_{1,1}, D^{(1)}_{1,1}} & \cdots & c_{D^{(1)}_{1,1}, D_1} & \cdots & c_{D^{(1)}_{1,1}, D^{(p)}_{1,T}} & \cdots & c_{D^{(1)}_{1,1}, D_T} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{D_T, D^{(1)}_{1,1}} & \cdots & c_{D_T, D_1} & \cdots & c_{D_T, D^{(p)}_{1,T}} & \cdots & c_{D_T, D_T} \end{pmatrix},$$

we thus conclude that

$$\frac{n}{m_n} \begin{pmatrix} \widehat{\chi}^{(t_1)}(\boldsymbol{h}^{(1)}_1, 0) - \chi_n(\boldsymbol{h}^{(1)}_1, 0) \\ \vdots \\ \widehat{\chi}^{(t_T)}(\boldsymbol{h}^{(p)}_{\ell_p}, 0) - \chi_n(\boldsymbol{h}^{(p)}_{\ell_p}, 0) \end{pmatrix} \stackrel{d}{\to} \mathcal{N}(\boldsymbol{0}, \mu(A)^{-4} \boldsymbol{F} \Sigma (\boldsymbol{F})^{\top}).$$

To obtain the asymptotic covariance matrix in the spatially isotropic case, we proceed as in the proof of Theorem 3.7. We define the $Tp \times TL$-matrix

$$\boldsymbol{N} := \begin{pmatrix} N & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & N & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & N \end{pmatrix}$$

with N given in equation (3.18). Then we have

$$\frac{n}{m_n}\begin{pmatrix}\widehat{\chi}^{(t_1)}(v_1,0)-\chi_n(v_1,0)\\ \vdots \\ \widehat{\chi}^{(t_T)}(v_p,0)-\chi_n(v_p,0)\end{pmatrix} = \frac{n}{m_n}\boldsymbol{N}\begin{pmatrix}\widehat{\chi}^{(t_1)}(\boldsymbol{h}_1^{(1)},0)-\chi_n(\boldsymbol{h}_1^{(1)},0)\\ \vdots \\ \widehat{\chi}^{(t_T)}(\boldsymbol{h}_{\ell_p}^{(p)},0)-\chi_n(\boldsymbol{h}_{\ell_p}^{(p)},0)\end{pmatrix}$$
$$\stackrel{d}{\to}\quad \mathcal{N}(\boldsymbol{0},\mu(A)^{-4}\boldsymbol{N}\boldsymbol{F}\Sigma(\boldsymbol{N}\boldsymbol{F})^\top),\quad n\to\infty,$$

and we conclude that for the averaged spatial extremogram the statement holds with

$$\Pi_2^{(\text{iso})} = \mu(A)^{-4}T^{-2}\begin{pmatrix}1 & 0\cdots0 & 1 & 0\cdots0 & \cdots & 1 & 0\cdots0\\ 0 & 1\cdots0 & 0 & 1\cdots0 & \cdots & 0 & 1\cdots0\\ & & & \ddots & & & \\ 0 & 0\cdots1 & 0 & 0\cdots1 & \cdots & 0 & 0\cdots1\end{pmatrix}\boldsymbol{N}\boldsymbol{F}\Sigma(\boldsymbol{N}\boldsymbol{F})^\top$$
$$\begin{pmatrix}1 & 0\cdots0 & 1 & 0\cdots0 & \cdots & 1 & 0\cdots0\\ 0 & 1\cdots0 & 0 & 1\cdots0 & \cdots & 0 & 1\cdots0\\ & & & \ddots & & & \\ 0 & 0\cdots1 & 0 & 0\cdots1 & \cdots & 0 & 0\cdots1\end{pmatrix}^\top. \tag{3.23}$$

□

Remark 3.9. In the central limit theorem the pre-asymptotic extremogram (3.14) can be replaced by the theoretical one (3.6), provided that

$$\frac{n}{m_n}\left(\chi_n(v,0)-\chi(v,0)\right)\to 0,\quad n\to\infty, \tag{3.24}$$

is satisfied for all spatial lags $v\in\mathcal{V}$. For the Brown-Resnick process (3.1) we obtain from Lemma 3.6,

$$\begin{aligned}&\frac{n}{m_n}\left(\chi_n(v,0)-\chi(v,0)\right)\\ &=\frac{n}{m_n}\left(\frac{P(\eta(\boldsymbol{s},t)>m_n^2,\eta(\boldsymbol{s}+\boldsymbol{h},t)>m_n^2)}{\mathbb{P}(\eta(\boldsymbol{s},t)>m_n^2)}-\chi(v,0)\right)\\ &\sim\frac{n}{2m_n^3}\big(\chi(v,0)-2\big)\big(\chi(v,0)-1\big)\\ &=n^{1-3\beta_1}\frac{1}{2}\big(\chi(v,0)-2\big)\big(\chi(v,0)-1\big)\quad\to\quad 0\quad\text{ if and only if }\beta_1>1/3;\end{aligned}$$

cf. Theorem 2.8. Thus we have to distinguish two cases:

(I) For $\beta_1\le 1/3$ we cannot replace the pre-asymptotic extremogram by the theoretical version, but can resort to a bias correction, which is decribed in (3.27) below.

(II) For $1/3<\beta_1<1/2$ we obtain indeed

$$n^{1-\beta_1}\big(\widehat{\chi}^{(t)}(v,0)-\chi(v,0)\big)_{v\in\mathcal{V}}\stackrel{d}{\to}\mathcal{N}(\boldsymbol{0},\Pi_1^{(\text{iso})}),\quad n\to\infty. \tag{3.25}$$

We now turn to the bias correction needed in case (I). By Lemma 3.6 the pre-asymptotic extremogram has representation

$$\begin{aligned}\chi_n(v,0) &= \Big[\chi(v,0) + \frac{1}{2m_n^2}\big(\chi(v,0)-2\big)\big(\chi(v,0)-1\big)\Big](1+o(1)) \\ &= \Big[\chi(v,0) + \frac{1}{2m_n^2}\nu(v,0)\Big](1+o(1)), \quad n\to\infty, \end{aligned} \tag{3.26}$$

where $\nu(v,0) := \big(\chi(v,0)-2\big)\big(\chi(v,0)-1\big)$.

Consequently, we propose for fixed $t \in \{t_1,\ldots,t_T\}$ and all $v \in \mathcal{V}$ the *bias corrected empirical spatial extremogram*

$$\widehat{\chi}^{(t)}(v,0) - \frac{1}{2m_n^2}\big(\widehat{\chi}^{(t)}(v,0)-2\big)\big(\widehat{\chi}^{(t)}(v,0)-1\big) =: \widehat{\chi}^{(t)}(v,0) - \frac{1}{2m_n^2}\widehat{\nu}^{(t)}(v,0),$$

and set

$$\widetilde{\chi}^{(t)}(v,0) := \begin{cases} \widehat{\chi}^{(t)}(v,0) - \dfrac{1}{2m_n^2}\widehat{\nu}^{(t)}(v,0) & \text{if } m_n = n^{\beta_1} \text{ with } \beta_1 \in (\frac{1}{5},\frac{1}{3}], \\ \widehat{\chi}^{(t)}(v,0) & \text{if } m_n = n^{\beta_1} \text{ with } \beta_1 \in (\frac{1}{3},\frac{1}{2}). \end{cases} \tag{3.27}$$

Theorem 3.10 below shows asymptotic normality of the bias corrected extremogram centred by the true one and, in particular, why β_1 has to be larger than 1/5.

Theorem 3.10. *For a fixed time point $t \in \{t_1,\ldots,t_T\}$ consider the spatial Brown-Resnick process $\{\eta(\boldsymbol{s},t), \boldsymbol{s}\in\mathbb{R}^2\}$ defined in* (3.1) *with dependence function given in* (3.2). *Set $m_n = n^{\beta_1}$ for $\beta_1 \in \big(\frac{1}{5},\frac{1}{3}\big]$. Then the bias corrected empirical spatial extremogram* (3.27) *satisfies*

$$\frac{n}{m_n}\big(\widetilde{\chi}^{(t)}(v,0) - \chi(v,0)\big)_{v\in\mathcal{V}} \stackrel{d}{\to} \mathcal{N}(\mathbf{0},\Pi_1^{(\text{iso})}), \quad n\to\infty, \tag{3.28}$$

where $\Pi_1^{(\text{iso})}$ is the covariance matrix as given in equation (3.19). *Furthermore, the corresponding bias corrected version of* (3.10) *satisfies*

$$\frac{n}{m_n}\Big(\frac{1}{T}\sum_{k=1}^{T}\widetilde{\chi}^{(t_k)}(v,0) - \chi(v,0)\Big)_{v\in\mathcal{V}} \stackrel{d}{\to} \mathcal{N}(\mathbf{0},\Pi_2^{(\text{iso})}), \quad n\to\infty,$$

with covariance matrix $\Pi_2^{(\text{iso})}$ specified in (3.23).

Proof. For simplicity we suppress the time point t in the notation. By (3.26) and (3.27) we have as $n\to\infty$,

$$\frac{n}{m_n}(\widetilde{\chi}(v,0)-\chi(v,0)) \sim \frac{n}{m_n}(\widehat{\chi}(v,0)-\chi_n(v,0)) - \frac{n}{2m_n^3}(\widehat{\nu}(v,0)-\nu(v,0)).$$

By Theorem 3.7 it suffices to show that $(n/(2m_n^3))(\widehat{\nu}(v,0)-\nu(v,0)) \stackrel{P}{\to} 0$. Setting $\nu_n(v,0) :=$

$\big(\chi_n(v,0)-2\big)\big(\chi_n(v,0)-1\big)$ we have

$$\frac{n}{2m_n^3}(\widehat{\nu}(v,0)-\nu(v,0))=\frac{n}{2m_n^3}(\widehat{\nu}(v,0)-\nu_n(v,0))+\frac{n}{2m_n^3}(\nu_n(v,0)-\nu(v,0))=:A_1+A_2.$$

We calculate

$$\begin{aligned}
&\frac{n}{m_n(2\chi(v,0)-3)}\Big(\widehat{\nu}(v,0)-\nu_n(v,0)\Big)\\
&=\frac{n}{m_n(2\chi(v,0)-3)}\Big(\widehat{\chi}^2(v,0)-3\widehat{\chi}(v,0)-(\chi_n^2(v,0)-3\chi_n(v,0))\Big)\\
&=\frac{n}{m_n(2\chi(v,0)-3)}\Big((\widehat{\chi}(v,0)-\chi_n(v,0))(\widehat{\chi}(v,0)+\chi_n(v,0))-3(\widehat{\chi}(v,0)-\chi_n(v,0))\Big)\\
&=\frac{n}{m_n}\Big(\widehat{\chi}(v,0)-\chi_n(v,0)\Big)\frac{\widehat{\chi}(v,0)+\chi_n(v,0)-3}{2\chi(v,0)-3}.
\end{aligned}$$

The first term converges by Theorem 3.7 weakly to a normal distribution, and the second term, together with the fact that $\widehat{\chi}(v,0)\xrightarrow{P}\chi(v,0)$ and $\chi_n(v,0)\xrightarrow{P}\chi(v,0)$, converges to 1 in probability. Hence, it follows from Slutzky's theorem that $A_1\xrightarrow{P}0$. Now we turn to A_2 and calculate

$$\begin{aligned}
\nu_n(v,0)&=\chi_n^2(v,0)-3\chi_n(v,0)+2\\
&\sim\Big(\chi(v,0)+\frac{1}{2m_n^2}\nu(v,0)\Big)^2-3\Big(\chi(v,0)+\frac{1}{2m_n^2}\nu(v,0)\Big)+2\\
&=\chi^2(v,0)-3\chi(v,0)+2+\frac{1}{m_n^2}\chi(v,0)\nu(v,0)+\frac{1}{4m_n^4}\nu(v,0)^2-\frac{3}{2m_n^2}\nu(v,0)\\
&=\big(\chi(v,0)-2\big)\big(\chi(v,0)-1\big)+\frac{1}{m_n^2}\chi(v,0)\nu(v,0)+\frac{1}{4m_n^4}\nu(v,0)^2-\frac{3}{2m_n^2}\nu(v,0)\\
&=\nu(v,0)+\frac{\nu(v,0)}{m_n^2}\Big(\chi(v,0)+\frac{1}{4m_n^2}\nu(v,0)-\frac{3}{2}\Big),
\end{aligned}$$

where we have used (3.26). Therefore, A_2 converges to 0, if $n/m_n^5\to 0$ as $n\to\infty$. With $m_n=n^{\beta_1}$ it follows that $\beta_1>\frac{1}{5}$. Finally, the last statement follows from Corollary 3.8. □

Remark 3.11. Note that in (3.25) the rate of convergence is of the order n^a for $a\in(1/2,2/3)$. On the other hand, after bias correction in (3.28) we obtain convergence of the order n^a for $a\in[2/3,4/5)$; i.e. a better rate.

Example 3.12. We generate 100 realisations of the Brown-Resnick process in (3.1) using the `R`-package `RandomFields` [60] and the exact method via extremal functions proposed in Dombry et al. [28], Section 2. We then compare the empirical estimates of the spatial extremogram $\widehat{\chi}(v,0)$ in (3.8) and the bias corrected ones $\widetilde{\chi}(v,0)$ in (3.27) with the true theoretical extremogram $\chi(v,0)$ for lags $v\in\{1,\sqrt{2},2,\sqrt{5},\sqrt{8},3,\sqrt{10},\sqrt{13},4,\sqrt{17}\}$. We choose the parameters $\theta_1=0.4$ and $\alpha_1=1.5$. The grid size and the number of time points are given by $n=70$ and $T=10$. The results are summarised in Figure 3.1. We see that the bias corrected extremogram is closer to the true one.

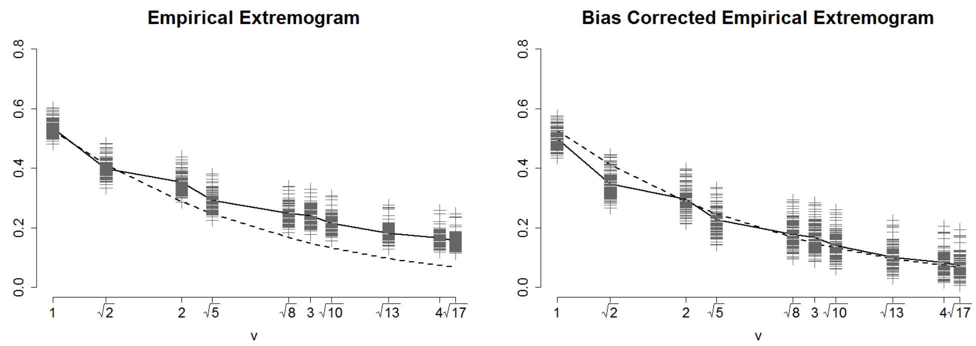

Figure 3.1: Empirical spatial extremogram (left) and its bias corrected version (right) for one hundred simulated max-stable random fields in (3.1) with $\delta(v,0) = 2 \cdot 0.4 v^{1.5}$. The dotted line represents the theoretical spatial extremogram and the solid line is the mean over all estimates.

3.3.2 Asymptotic properties of spatial parameter estimates

In this section we prove asymptotic normality of the WLSE $(\widehat{\theta}_1, \widehat{\alpha}_1)$ of Section 3.2. We use the following notation:

$$y_v := 2 \log\left(\Phi^{-1}\big(1 - \frac{1}{2}\widetilde{\chi}(v,0)\big)\right) \quad \text{and} \quad x_v := \log(v), \quad v \in \mathcal{V},$$

with $\widetilde{\chi}(v,0) = \frac{1}{T}\sum_{k=1}^{T} \widetilde{\chi}^{(t_k)}(v,0)$ as in (3.10), possibly after a bias correction, which depends on the two cases described in Remark 3.9. Then (3.12) reads as

$$\begin{pmatrix} \widehat{\theta}_1 \\ \widehat{\alpha}_1 \end{pmatrix} = \underset{\substack{\theta_1, \alpha_1 > 0 \\ \alpha_1 \in (0,2]}}{\arg\min} \sum_{v \in \mathcal{V}} w_v \big(y_v - \big(\log(\theta_1) + \alpha_1 x_v\big)\big)^2 \tag{3.29}$$

and we are in the setting of weighted linear regression. To show asymptotic normality of the WLSE as in (3.29), we define the design matrix X and weight matrix W as

$$X = [\mathbf{1}, (x_v)^{\intercal}_{v \in \mathcal{V}}] \in \mathbb{R}^{p \times 2} \quad \text{and} \quad W = \operatorname{diag}\{w_v : v \in \mathcal{V}\} \in \mathbb{R}^{p \times p},$$

respectively, where $\mathbf{1} = (1, \dots, 1)^{\intercal} \in \mathbb{R}^p$. Let $\boldsymbol{\psi}_1 = (\log(\theta_1), \alpha_1)^{\intercal}$ be the parameter vector with parameter space $\Psi = \mathbb{R} \times (0,2]$. Then the WLSE; i.e., the solution to (3.29) is given by

$$\widehat{\boldsymbol{\psi}}_1 := \begin{pmatrix} \log(\widehat{\theta}_1) \\ \widehat{\alpha}_1 \end{pmatrix} = (X^{\intercal} W X)^{-1} X^{\intercal} W (y_v)^{\intercal}_{v \in \mathcal{V}}.$$

Without any constraints $\widehat{\boldsymbol{\psi}}_1$ may produce estimates of α_1 outside its parameter space $(0,2]$. In such cases we set the parameter estimate equal to 2, and we denote the resulting estimate by $\widehat{\boldsymbol{\psi}}_1^c = (\log(\widehat{\theta}_1^c), \widehat{\alpha}_1^c)^{\intercal}$.

Theorem 3.13. *Let $\widehat{\boldsymbol{\psi}}_1^c = (\log(\widehat{\theta}_1^c), \widehat{\alpha}_1^c)^{\intercal}$ denote the WLSE resulting from the constrained minimisation problem* (3.29) *and $\boldsymbol{\psi}_1^* = (\log(\theta_1^*), \alpha_1^*)^{\intercal} \in \Psi$ the true parameter vector. Set $m_n = n^{\beta_1}$*

for $\beta_1 \in (1/5, 1/2)$. Then as $n \to \infty$,

$$\frac{n}{m_n}\Big(\widehat{\boldsymbol{\psi}}_1^c - \boldsymbol{\psi}_1^*\Big) \xrightarrow{d} \begin{cases} \boldsymbol{Z}_1 & \text{if } \alpha_1^* < 2, \\ \boldsymbol{Z}_2 & \text{if } \alpha_1^* = 2, \end{cases} \tag{3.30}$$

where $\boldsymbol{Z}_1 \sim \mathcal{N}(\mathbf{0}, \Pi_3^{(\text{iso})})$, and the distribution of $\boldsymbol{Z}_2$ is given by

$$\begin{aligned} \mathbb{P}\left(\boldsymbol{Z}_2 \in B\right) = &\int_{B \cap \{(b_1,b_2)\in\mathbb{R}^2 : b_2 < 0\}} \varphi_{\mathbf{0},\Pi_3^{(\text{iso})}}(z_1, z_2) dz_1 dz_2 \\ &+ \int_0^\infty \int_{\{b_1 \in \mathbb{R} : (b_1, 0) \in B\}} \varphi_{\mathbf{0},\Pi_3^{(\text{iso})}}\Big(z_1 - \frac{1}{\sum_{v\in\mathcal{V}} w_v} \sum_{v\in\mathcal{V}} (w_v x_v)\; z_2,\; z_2\Big) dz_1 dz_2 \end{aligned} \tag{3.31}$$

for every Borel set B in $\mathbb{R}^2$, and $\varphi_{\mathbf{0},\Sigma}$ denotes the bivariate normal density with mean vector $\mathbf{0}$ and covariance matrix Σ. In particular, the joint distribution function of $\boldsymbol{Z}_2$ is given for $(p_1, p_2)^\intercal \in \mathbb{R}^2$ by

$$\begin{aligned} \mathbb{P}\left(\boldsymbol{Z}_2 \le (p_1, p_2)^\intercal\right) = &\int_{-\infty}^{\min\{0,p_2\}} \int_{-\infty}^{p_1} \varphi_{\mathbf{0},\Pi_3^{(\text{iso})}}(z_1, z_2) dz_1 dz_2 \\ &+ \mathbb{1}_{\{p_2 \ge 0\}} \int_0^\infty \int_{-\infty}^{p_1} \varphi_{\mathbf{0},\Pi_3^{(\text{iso})}}\Big(z_1 - \frac{1}{\sum_{v\in\mathcal{V}} w_v} \sum_{v\in\mathcal{V}} (w_v x_v)\; z_2,\; z_2\Big) dz_1 dz_2. \end{aligned} \tag{3.32}$$

The covariance matrix of $\boldsymbol{Z}_1$ has representation

$$\Pi_3^{(\text{iso})} = Q_x^{(w)} G \Pi_2^{(\text{iso})} G Q_x^{(w)\intercal}, \tag{3.33}$$

where $\Pi_2^{(\text{iso})}$ is the covariance matrix given in (3.23),

$$Q_x^{(w)} = (X^\intercal W X)^{-1} X^\intercal W \quad \text{and} \tag{3.34}$$

$$G = \operatorname{diag}\left\{ \sqrt{\frac{2\pi}{\theta_1^* v^{\alpha_1^*}}} \exp\left\{\frac{1}{2}\theta_1^* v^{\alpha_1^*}\right\} : \; v \in \mathcal{V} \right\}. \tag{3.35}$$

Proof. For $v \in \mathcal{V}$ we have $y_v = g(\widetilde{\chi}(v, 0))$ with $g(x) = 2\log(\Phi^{-1}(1 - x/2))$. The derivative of g is given by

$$g'(x) = -\Big(\Phi^{-1}(1 - \frac{x}{2})\, \varphi(\Phi^{-1}(1 - \frac{x}{2}))\Big)^{-1},$$

where φ is the univariate standard normal density. Thus,

$$g'(\chi(v, 0)) = -\Big(\sqrt{\theta_1^* v^{\alpha_1^*}}\, \varphi\big(\sqrt{\theta_1^* v^{\alpha_1^*}}\big)\Big)^{-1} = -\sqrt{\frac{2\pi}{\theta_1^* v^{\alpha_1^*}}} \exp\left\{\frac{1}{2}\theta_1^* v^{\alpha_1^*}\right\}.$$

Using the multivariate delta method together with Theorems 3.7 and 3.10 it follows that

$$\frac{n}{m_n}\big(y_v - g(\chi(v, 0))\big)_{v\in\mathcal{V}} \xrightarrow{d} \mathcal{N}(\mathbf{0}, G\Pi_2^{(\text{iso})} G), \quad n \to \infty,$$

where G is defined in (3.35). Since

$$\min_{\substack{\theta_1,\alpha_1\\ \alpha_1\in(0,2]}} \sum_{v\in\mathcal{V}} w_v\big(g(\chi(v,0)) - \big(\log(\theta_1)+\alpha_1 x_v\big)\big)^2 = \sum_{v\in\mathcal{V}} w_v\big(g(\chi(v,0)) - \big(\log(\theta_1^*)+\alpha_1^* x_v\big)\big)^2,$$

we find the well-known property of unbiasedness of the WLSE,

$$Q_x^{(w)}(g(\chi(v,0)))_{v\in\mathcal{V}}^\intercal = \arg\min_{\substack{\theta_1,\alpha_1\\ \alpha_1\in(0,2]}} \sum_{v\in\mathcal{V}} w_v\big(g(\chi(v,0)) - \big(\log(\theta_1)+\alpha_1 x_v\big)\big)^2 = \boldsymbol{\psi}_1^*.$$

It follows that, as $n\to\infty$,

$$\frac{n}{m_n}\left(\widehat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_1^*\right) = \frac{n}{m_n} Q_x^{(w)}\big(y_v - g(\chi(v,0))\big)_{v\in\mathcal{V}} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, Q_x^{(w)} G \Pi_2^{(\mathrm{iso})} G Q_x^{(w)\intercal}\right).$$

We now turn to the constraints on α_1. Since the objective function is quadratic, if the unconstrained estimate exceeds two, the constraint $\alpha_1 \in (0,2]$ results in an estimate $\widehat{\alpha}_1^c = 2$. We consider separately the cases $\alpha_1^* < 2$ and $\alpha_1^* = 2$; i.e., the true parameter lies either in the interior or on the boundary of the parameter space. The constrained estimator $\widehat{\boldsymbol{\psi}}_1^c$ can be written as

$$\widehat{\boldsymbol{\psi}}_1^c = \widehat{\boldsymbol{\psi}}_1 \mathbb{1}_{\{\widehat{\alpha}_1\le 2\}} + (\widehat{\theta}_1, 2)^\intercal \mathbb{1}_{\{\widehat{\alpha}_1 > 2\}}.$$

We calculate the asymptotic probabilities for the events $\{\widehat{\alpha}_1 \le 2\}$ and $\{\widehat{\alpha}_1 > 2\}$,

$$\mathbb{P}(\widehat{\alpha}_1 \le 2) = P\Big(\frac{n}{m_n}(\widehat{\alpha}_1 - \alpha_1^*) \le \frac{n}{m_n}(2-\alpha_1^*)\Big).$$

Since for $\alpha_1^* < 2$ as $n\to\infty$

$$\frac{n}{m_n}\big(\widehat{\alpha}_1 - \alpha_1^*\big) \xrightarrow{d} \mathcal{N}\left(0, (0,1)\Pi_3^{(\mathrm{iso})}(0,1)^\intercal\right) \quad \text{and} \quad \frac{n}{m_n}(2-\alpha_1^*) \to \infty,$$

it follows that

$$\mathbb{P}(\widehat{\alpha}_1 \le 2) \to 1 \quad \text{and} \quad \mathbb{P}(\widehat{\alpha}_1 > 2) \to 0, \quad n\to\infty. \tag{3.36}$$

Therefore, for $\alpha_1^* < 2$,

$$\frac{n}{m_n}\big(\widehat{\boldsymbol{\psi}}_1^c - \boldsymbol{\psi}_1^*\big) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Pi_3^{(\mathrm{iso})}), \quad n\to\infty.$$

We now consider the case $\alpha_1^* = 2$ and $\widehat{\alpha}_1 > 2$ (the unconstrained estimate exceeds 2). In this case (3.29) leads to the constrained optimisation problem

$$\begin{aligned}&\min_{\boldsymbol{\psi}_1}\{[W^{1/2}((y_v)_{v\in\mathcal{V}}^\intercal - X\boldsymbol{\psi}_1)]^\intercal [W^{1/2}((y_v)_{v\in\mathcal{V}}^\intercal - X\boldsymbol{\psi}_1]\},\\ &\text{s.t.} \quad (0,1)\boldsymbol{\psi}_1 = 2.\end{aligned}$$

To obtain asymptotic results for $\widehat{\boldsymbol{\psi}}_1^c - \boldsymbol{\psi}_1^*$, the vector $\widehat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_1^*$ is projected onto the line $\Lambda = \{\boldsymbol{\psi}\in\mathbb{R}^2, (0,1)\boldsymbol{\psi} = 0\}$, i.e., denoting by I_2 the 2×2-identity matrix, the projection matrix with respect to the induced norm $\boldsymbol{\psi} \mapsto (\boldsymbol{\psi}^\intercal X^\intercal W X \boldsymbol{\psi})^{1/2}$ is given by (cf. Andrews [1], page 1365)

$$P_\Lambda = I_2 - (X^\intercal W X)^{-1}(0,1)^\intercal((0,1)(X^\intercal W X)^{-1}(0,1)^\intercal)^{-1}(0,1).$$

For simplicity we use the abbreviation $p_{wx} = \sum_{v\in\mathcal{V}} w_v x_v / \sum_{v\in\mathcal{V}} w_v$. We calculate

$$\begin{aligned}
&(\widehat{\boldsymbol{\psi}}_1^c - \boldsymbol{\psi}_1^*)\mathbb{1}_{\{\widehat{\alpha}_1>2\}} = P_\Lambda(\widehat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_1^*)\mathbb{1}_{\{\widehat{\alpha}_1>2\}} \\
&= (\widehat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_1^*)\mathbb{1}_{\{\widehat{\alpha}_1>2\}} - (X^\intercal W X)^{-1}(0,1)^\intercal \left((0,1)(X^\intercal W X)^{-1}(0,1)^\intercal\right)^{-1}(\widehat{\alpha}_1 - 2)\mathbb{1}_{\{\widehat{\alpha}_1>2\}} \\
&= (\widehat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_1^*)\mathbb{1}_{\{\widehat{\alpha}_1>2\}} + \begin{pmatrix} p_{wx} \\ -1 \end{pmatrix}(\widehat{\alpha}_1 - 2)\mathbb{1}_{\{\widehat{\alpha}_1>2\}}.
\end{aligned}$$

For the joint constrained estimator $\boldsymbol{\psi}_1^c$ we obtain

$$\begin{aligned}
\widehat{\boldsymbol{\psi}}_1^c - \boldsymbol{\psi}_1^* &= (\widehat{\boldsymbol{\psi}}_1^c - \boldsymbol{\psi}_1^*)\mathbb{1}_{\{\widehat{\alpha}_1\le 2\}} + (\widehat{\boldsymbol{\psi}}_1^c - \boldsymbol{\psi}_1^*)\mathbb{1}_{\{\widehat{\alpha}_1>2\}} \\
&= (\widehat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_1^*)\mathbb{1}_{\{\widehat{\alpha}_1\le 2\}} + (\widehat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_1^*)\mathbb{1}_{\{\widehat{\alpha}_1>2\}} + \begin{pmatrix} p_{wx} \\ -1 \end{pmatrix}(\widehat{\alpha}_1 - 2)\mathbb{1}_{\{\widehat{\alpha}_1>2\}} \\
&= (\widehat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_1^*) + \begin{pmatrix} p_{wx} \\ -1 \end{pmatrix}(\widehat{\alpha}_1 - 2)\mathbb{1}_{\{\widehat{\alpha}_1>2\}}.
\end{aligned}$$

This implies

$$\frac{n}{m_n}(\widehat{\boldsymbol{\psi}}_1^c - \boldsymbol{\psi}_1^*) = \frac{n}{m_n}\begin{pmatrix} (\log(\widehat{\theta}_1) - \log(\theta_1^*)) + p_{wx}(\widehat{\alpha}_1 - 2)\mathbb{1}_{\{\widehat{\alpha}_1>2\}} \\ (\widehat{\alpha}_1 - 2) - (\widehat{\alpha}_1 - 2)\mathbb{1}_{\{\widehat{\alpha}_1>2\}} \end{pmatrix}.$$

Let $f(x_1,x_2) = (x_1 + p_{wx}x_2\mathbb{1}_{\{x_2>0\}}, x_2 - x_2\mathbb{1}_{\{x_2>0\}})^\intercal$ and observe that $f(c(x_1,x_2)) = cf(x_1,x_2)$ for $c \ge 0$. For the asymptotic distribution we calculate

$$\begin{aligned}
&\mathbb{P}\Big(\frac{n}{m_n}(\widehat{\boldsymbol{\psi}}_1^c - \boldsymbol{\psi}_1^*) \in B\Big) \\
&= \mathbb{P}\Big(\frac{n}{m_n}f(\widehat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_1^*) \in B\Big) = \mathbb{P}\Big(f\big(\frac{n}{m_n}(\widehat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_1^*)\big) \in B\Big) \\
&= \mathbb{P}\Big(\frac{n}{m_n}(\widehat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_1^*) \in f^{-1}(B \cap \{(b_1,b_2)\in\mathbb{R}^2 : b_2 < 0\}) \cup f^{-1}(B \cap \{(b_1,0) : b_1 \in \mathbb{R}\})\Big) \\
&= \mathbb{P}\Big(\frac{n}{m_n}(\widehat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_1^*) \in (B \cap \{(b_1,b_2)\in\mathbb{R}^2 : b_2 < 0\}) \\
&\quad \cup (\{(b_1 - p_{wx}b_2, b_2), b_2 \ge 0, (b_1,0) \in B)\Big) \\
&\to \int_{B\cap\{(b_1,b_2)\in\mathbb{R}^2, b_2<0\}} \varphi_{\mathbf{0},\Pi_3^{(\mathrm{iso})}}(z_1,z_2)dz_1dz_2 \\
&\qquad + \int_0^\infty \int_{\{b_1\in\mathbb{R},(b_1,0)\in B\}} \varphi_{\mathbf{0},\Pi_3^{(\mathrm{iso})}}(z_1 - p_{wx}z_2, z_2)dz_1dz_2, \quad n\to\infty.
\end{aligned}$$

Plugging in $B = (-\infty, p_1] \times (-\infty, p_2]$ and using the Fubini-Tonelli theorem yields (3.32). □

Remark 3.14. The derivation of the asymptotic properties for the constrained estimate is in fact a special case of Corollary 1 in Andrews [1], who shows asymptotic properties of parameter estimates in a very general setting, when the true parameter is on the boundary of the parameter space. The asymptotic distribution of the estimates in the case $\alpha_1^* = 2$ results from the fact that approximately half of the estimates lie above the true value and are therefore equal to two, which is reflected by the second term in the asymptotic distribution of the estimates.

3.3.3 Asymptotic properties of the empirical temporal extremogram

The results for the temporal parameter (θ_2, α_2) are analogous to those for the spatial parameter as presented in Sections 3.3.1 and 3.3.2. The finite set of temporal lags

$$\mathcal{U} = \{1, \dots, \overline{p}\},$$

used for the estimation are those which show clear extremal dependence as explained in Section 3.2.

Theorem 3.15. *For fixed location $\boldsymbol{s} \in \mathcal{S}_n$, consider the Brown-Resnick time series $\{\eta(\boldsymbol{s}, t) : t \in [0, \infty)\}$ as defined in* (3.1) *with dependence function given in* (3.2)*. Set $m_T = T^{\beta_1}$ for $\beta_1 \in (0, 1)$. Then the empirical temporal extremogram $\widehat{\chi}^{(\boldsymbol{s})}(0, u)$ defined in* (3.9) *with the quantile $q = m_T$ satisfies*

$$\left(\frac{T}{m_T}\right)^{1/2} \left(\widehat{\chi}^{(\boldsymbol{s})}(0, u) - \chi_T(0, u)\right)_{u \in \mathcal{U}} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Pi_1^{(\text{time})}), \quad T \to \infty,$$

where the covariance matrix $\Pi_1^{(\text{time})}$ is specified in Corollary 3.4 of Davis and Mikosch [17], and χ_T is the pre-asymptotic extremogram

$$\chi_T(0, u) \quad = \quad \frac{\mathbb{P}(\eta(\mathbf{0}, 0) > m_T, \eta(\mathbf{0}, u) > m_T)}{\mathbb{P}(\eta(\mathbf{0}, 0) > m_T)}, \quad u \in \mathcal{U}. \tag{3.37}$$

Proof. We verify the mixing conditions for the central limit theorem of the temporal extremogram in Davis and Mikosch [17], Corollary 3.4.

Define sequences $m_T = T^{\beta_1}$ for $\beta_1 \in (0, 1)$ and $r_T = T^{\beta_2}$ for $0 < \beta_2 < \beta_1$, which both tend to infinity as $T \to \infty$ as well as $m_T/T \to 0$ and $r_T/m_T \to 0$. From equation (B.3) and Lemma B.3 the time series $\{\eta(\boldsymbol{s}, t), t \in [0, \infty)\}$ is α-mixing with mixing coefficients $\alpha(u) \leq Cu^2 \exp\{-\theta_2 u^\alpha / 2\}$ for some positive constant C. Hence, by Lemma B.3, and temporal parameters (θ_2, α_2),

$$\begin{aligned} m_T \sum_{u=r_T}^{\infty} \alpha(u) \quad &\leq \quad C m_T \sum_{u=r_T}^{\infty} u^2 e^{-\theta_2 u^{\alpha_2}/2} \leq C m_T \, r_T^3 \, e^{-\theta_2 r_T^{\alpha_2}/2} \\ &= \quad C T^{\beta_1 + 3\beta_2} \exp\Big\{-\frac{1}{2}\theta_2 T^{\alpha_2 \beta_2}\Big\} \to 0, \quad T \to \infty. \end{aligned} \tag{3.38}$$

In addition, for arbitrary $\varepsilon > 0$,

$$\begin{aligned} & m_T \sum_{u=k}^{r_T} P\left(\|(\eta(\boldsymbol{s}, t_{1+u}), \dots, \eta(\boldsymbol{s}, t_{1+u+p}))\| > \varepsilon m_T, \|(\eta(\boldsymbol{s}, t_1), \dots, \eta(\boldsymbol{s}, t_{1+p}))\| > \varepsilon m_T\right) \\ & \leq m_T \sum_{u=k}^{\infty} \sum_{i=u}^{u+p} \sum_{j=0}^{p} P\left(\eta(\boldsymbol{s}, t_{1+i}) > \varepsilon m_T, \eta(\boldsymbol{s}, t_{1+j}) > \varepsilon m_T\right). \end{aligned}$$

By a time-wise version of (3.15) and the fact that $\mathbb{P}(\eta(\boldsymbol{s}, t_k) > m_T) \sim m_T^{-1}$, it suffices to show that the following sum is finite, which we estimate by

$$\begin{aligned}&\sum_{u=k}^{\infty}\sum_{i=u}^{u+p}\sum_{j=0}^{p} 2\Big(1-\Phi\big(\sqrt{\theta_2|t_{i+1}-t_{j+1}|^{\alpha_2}}\big)\Big)\\ \leq& 2\sum_{u=k}^{\infty}\sum_{i=0}^{p}\sum_{j=0}^{p}\exp\left\{-\frac{\theta_2}{2}|t_{i+1}-t_{j+1}|^{\alpha_2}\right\}<\infty,\end{aligned}$$

where we use that $1-\Phi(x) \leq \exp\{-x^2/2\}$ for $x > 0$. This establishes Condition (M) in Davis and Mikosch [17]. As in equation (3.38), we get for $m_T = T^{\beta_1}$,

$$\frac{T}{m_T}\alpha(r_T) \leq CT^{1+2\beta_2-\beta_1}\exp\Big\{-\frac{1}{2}\theta_2 T^{\alpha_2\beta_2}\Big\} \to 0, \quad T \to \infty.$$

Furthermore, we need one of the following conditions:

(I) $m_T = o(T^{1/3})$, which is satisfied if and only if $\beta_1 < 1/3$; or

(II) $m_T r_T^3/T = T^{\beta_1+3\beta_2-1} \to 0$ as $T \to \infty$,
which in particular holds for $\beta_1 \in [\frac{1}{3}, 1)$ and $\beta_2 \in (0, \min\{\beta_1, \frac{1}{3}(1-\beta_1)\})$, and
$m_T^4 T^{-1}\sum_{u=r_T}^{m_T}\alpha(u) \leq CT^{4\beta_1+3\beta_2-1}\exp\big\{-\theta_2 T^{\alpha_2\beta_2}/2\big\} \to 0$ as $T \to \infty$,
which is satisfied for $\beta_1 \in [1/3, 1)$ and $\beta_2 < \beta_1$.

□

Remark 3.16. The following is the time-wise analogue of Remark 3.9, and gives the convergence rate of the pre-asymptotic extremogram (3.37) to the extremogram. First note that the pre-asymptotic extremogram has for $u \in \mathcal{U}$ the representation

$$\chi_T(0,u) = \Big[\chi(0,u) + \frac{1}{2m_T}\big(\chi(0,u)-2\big)\big(\chi(0,u)-1\big)\Big](1+o(1)), \quad T \to \infty, \tag{3.39}$$

which can, similarly as Lemma 3.6, be deduced from equation (A.4). Hence we have for $u \in \mathcal{U}$,

$$\begin{aligned}&\Big(\frac{T}{m_T}\Big)^{1/2}\Big(\chi_T(0,u)-\chi(0,u)\Big)\\ \sim&\Big(\frac{T}{m_T}\Big)^{1/2}\frac{1}{2m_T}\Big[\big(\chi(0,u)-2\big)\big(\chi(0,u)-1\big)\Big] \to 0, \quad T\to\infty\end{aligned}$$

for $m_T = T^{\beta_1}$ if and only if $\beta_1 > 1/3$. Thus we have the two cases:

(I) For $\beta_1 \leq 1/3$ we cannot replace the pre-asymptotic extremogram with the theoretical version, but can resort to a bias correction, which is described in (3.41) below.

(II) For $1/3 < \beta_1 < 1$ we obtain indeed

$$\Big(\frac{T}{m_T}\Big)^{1/2}\big(\widehat{\chi}^{(\boldsymbol{s})}(0,u)-\chi(0,u)\big)_{u\in\mathcal{U}} \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \Pi_1^{(\text{time})}), \quad T \to \infty. \tag{3.40}$$

We now turn to the bias correction needed in case (I) for asymptotic normality. Motivated by equation (3.39), we propose for fixed $\boldsymbol{s} \in \mathcal{S}_n$ and all $u \in \mathcal{U}$ the *bias corrected empirical temporal extremogram*

$$\widehat{\chi}^{(\boldsymbol{s})}(0,u) - \frac{1}{2m_T}\big(\widehat{\chi}^{(\boldsymbol{s})}(0,u)-2\big)\big(\widehat{\chi}^{(\boldsymbol{s})}(0,u)-1\big), \quad \boldsymbol{s} \in \mathcal{S}_n,$$

and set

$$\widetilde{\chi}^{(\boldsymbol{s})}(0,u) := \tag{3.41}$$

$$\begin{cases} \widehat{\chi}^{(\boldsymbol{s})}(0,u) - \frac{1}{2m_T}\big(\widehat{\chi}^{(\boldsymbol{s})}(0,u)-2\big)\big(\widehat{\chi}^{(\boldsymbol{s})}(0,u)-1\big) & \text{if } m_T = T^{\beta_1} \text{ with } \beta_1 \in (\frac{1}{5},\frac{1}{3}], \\ \widehat{\chi}^{(\boldsymbol{s})}(0,u) & \text{if } m_T = T^{\beta_1} \text{ with } \beta_1 \in (\frac{1}{3},1). \end{cases} \tag{3.42}$$

We conclude this section by proving asymptotic normality of the bias corrected temporal extremogram centred by the true one. The proof is analogous to that of Theorem 3.10 and shows in particular why β_1 needs to be larger than $1/5$. The extension of the statement to spatial means of extremograms follows in the same way as in Corollary 3.8 by using the vectorised process

$$\{\boldsymbol{Y}(t) : t \in [0,\infty)\} = \{(\eta(\boldsymbol{s},t),\ldots,\eta(\boldsymbol{s},t+\overline{p}))^\intercal_{\boldsymbol{s}\in\mathcal{S}_n} : t \in [0,\infty)\}$$

and defining sets $D_{u,k}$ and D_k for $u = 1,\ldots,\overline{p}$ and $k = 1,\ldots,n^2$ properly to extend the covariance matrix. This leads to the statement in (3.44), where

$$\begin{aligned}\Pi_2^{(\text{time})} =& (n\,\mu((1,\infty)))^{-4} \begin{pmatrix} 1 & 0\cdots0 & 1 & 0\cdots0 & \cdots & 1 & 0\cdots0 \\ 0 & 1\cdots0 & 0 & 1\cdots0 & \cdots & 0 & 1\cdots0 \\ & & & \ddots & & & \\ 0 & 0\cdots1 & 0 & 0\cdots1 & \cdots & 0 & 0\cdots1 \end{pmatrix} \boldsymbol{F}'\Sigma'(\boldsymbol{F}')^\intercal \\ & \begin{pmatrix} 1 & 0\cdots0 & 1 & 0\cdots0 & \cdots & 1 & 0\cdots0 \\ 0 & 1\cdots0 & 0 & 1\cdots0 & \cdots & 0 & 1\cdots0 \\ & & & \ddots & & & \\ 0 & 0\cdots1 & 0 & 0\cdots1 & \cdots & 0 & 0\cdots1 \end{pmatrix}^\intercal, \end{aligned} \tag{3.43}$$

and $\boldsymbol{F}'$ and Σ' are defined in a similar fashion as the matrices $\boldsymbol{F}$ and Σ in Corollary 3.8.

Theorem 3.17. *For a fixed location $\boldsymbol{s} \in \mathcal{S}_n$ consider the Brown-Resnick time series $\{\eta(\boldsymbol{s},t), t \in [0,\infty)\}$ as defined in* (3.1) *with dependence function given in* (3.2)*. Set $m_T = T_1^{\beta}$ for $\beta_1 \in \big(1/5, 1/3\big]$. Then the bias corrected empirical temporal extremogram* (3.41) *satisfies*

$$\left(\frac{T}{m_T}\right)^{1/2}\Big(\widetilde{\chi}^{(\boldsymbol{s})}(0,u) - \chi(0,u)\Big)_{u\in\mathcal{U}} \xrightarrow{d} \mathcal{N}(\boldsymbol{0},\Pi_1^{(\text{time})}), \quad T \to \infty,$$

with covariance matrix $\Pi_1^{(\text{time})}$ as in Theorem 3.15. Furthermore, the corresponding bias corrected *version* $\widetilde{\chi}(0,u) = n^{-2}\sum_{i=1}^{n^2}\widetilde{\chi}^{(\boldsymbol{s}_i)}(0,u)$ *of* (3.11) *satisfies*

$$\Big(\frac{T}{m_T}\Big)^{1/2}\Big(\frac{1}{n^2}\sum_{i=1}^{n^2}\widetilde{\chi}^{(\boldsymbol{s}_i)}(0,u)-\chi(0,u)\Big)_{u\in\mathcal{U}} \xrightarrow{d} \mathcal{N}(\mathbf{0},\Pi_2^{(\text{time})}), \quad T\to\infty, \tag{3.44}$$

with covariance matrix $\Pi_2^{(\text{time})}$ *specified in equation* (3.43).

Remark 3.18. Note that in (3.40) the rate of convergence is of the order n^a for $a\in(0,1/3)$. On the other hand, after bias correction in (3.44) we obtain convergence of the order n^a for $a\in[1/3,2/5)$. Thus, the bias correction leads to better rates compared to those in Davis and Mikosch [17], where no bias correction is applied.

3.3.4 Asymptotic properties of temporal parameter estimates

The asymptotic normality of the WLSE $(\widehat{\theta}_2,\widehat{\alpha}_2)$ of Section 3.2 can be derived in exactly the same way as for the spatial parameter estimates $(\widehat{\theta}_1,\widehat{\alpha}_1)$. Accordingly, we define

$$y_u := 2\log\Big(\Phi^{-1}\big(1-\frac{1}{2}\widetilde{\chi}(0,u)\big)\Big) \quad \text{and} \quad x_u := \log(u), \quad u\in\mathcal{U},$$

where $\widetilde{\chi}(0,u)=\frac{1}{n^2}\sum_{i=1}^{n^2}\widetilde{\chi}^{(\boldsymbol{s}_i)}(0,u)$ as in (3.11), possibly after a bias correction, which depends on the two cases described in Remark 3.16. Then (3.13) reads as

$$\begin{pmatrix}\widehat{\theta}_2\\ \widehat{\alpha}_2\end{pmatrix} = \underset{\substack{\theta_2,\alpha_2>0\\ \alpha_2\in(0,2]}}{\arg\min}\sum_{v\in\mathcal{V}} w_v\big(y_v-\big(\log(\theta_2)+\alpha_2 x_v\big)\big)^2 \tag{3.45}$$

We also define the design matrix X and weight matrix W as

$$X=[\mathbf{1},(x_u)_{u\in\mathcal{U}}^\intercal]\in\mathbb{R}^{\overline{p}\times 2} \quad \text{and} \quad W=\operatorname{diag}\{w_u : u\in\mathcal{U}\}\in\mathbb{R}^{\overline{p}\times\overline{p}},$$

where $\mathbf{1}=(1,\ldots,1)^\intercal\in\mathbb{R}^{\overline{p}}$. We state asymptotic normality of the WLSE of the time parameters.

Theorem 3.19. *Let* $\widehat{\boldsymbol{\psi}}_2^c = (\log(\widehat{\theta}_2^c),\widehat{\alpha}_2^c)^\intercal$ *denote the WLSE resulting from the constrained minimisation problem in* (3.45) *and* $\boldsymbol{\psi}_2^* = (\log(\theta_2^*),\alpha_2^*)^\intercal \in \Psi$ *the true parameter vector. Set* $m_T = T^{\beta_1}$ *for* $\beta_1\in(1/5,1)$. *Then as* $T\to\infty$,

$$\Big(\frac{T}{m_T}\Big)^{1/2}\Big(\widehat{\boldsymbol{\psi}}_2^c-\boldsymbol{\psi}_2^*\Big) \xrightarrow{d} \begin{cases}\boldsymbol{Z}_1 & \text{if } \alpha_2^*<2,\\ \boldsymbol{Z}_2 & \text{if } \alpha_2^*=2,\end{cases} \tag{3.46}$$

where $\boldsymbol{Z}_1\sim\mathcal{N}(\mathbf{0},\Pi_3^{(\text{time})})$, *and the distribution of* $\boldsymbol{Z}_2$ *is given by*

$$\begin{aligned} P(\boldsymbol{Z}_2\in B) &= \int_{B\cap\{(b_1,b_2)\in\mathbb{R}^2 : b_2<0\}} \varphi_{\mathbf{0},\Pi_3^{(\text{time})}}(z_1,z_2)dz_1dz_2\\ &\quad + \int_0^\infty\int_{\{b_1\in\mathbb{R}:(b_1,0)\in B\}} \varphi_{\mathbf{0},\Pi_3^{(\text{time})}}\Big(z_1-\frac{1}{\sum_{u\in\mathcal{U}}w_u}\sum_{u\in\mathcal{U}}w_u x_u\ z_2,\ z_2\Big)dz_1dz_2 \end{aligned} \tag{3.47}$$

for every Borel set $B \subset \mathbb{R}$, and $\varphi_{\mathbf{0},\Sigma}$ denotes the bivariate normal density with mean vector $\mathbf{0}$ *and covariance matrix Σ. In particular, the joint distribution function of $\boldsymbol{Z}_2$ is given for $(p_1, p_2)^\intercal \in \mathbb{R}^2$ by*

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{Z}_2 \leq (p_1,p_2)^\intercal\right) = &\int_{-\infty}^{\min\{0,p_2\}} \int_{-\infty}^{p_1} \varphi_{\mathbf{0},\Pi_3^{(\text{time})}}(z_1,z_2)dz_1dz_2 \\ &+ \mathbb{1}_{\{p_2\geq 0\}} \int_0^{\infty}\int_{-\infty}^{p_1} \varphi_{\mathbf{0},\Pi_3^{(\text{time})}}\Big(z_1 - \frac{1}{\sum_{u\in\mathcal{U}} w_u}\sum_{u\in\mathcal{U}}(w_u x_u)\ z_2,\ z_2\Big)dz_1dz_2. \end{aligned} \quad (3.48)$$

The covariance matrix of $\boldsymbol{Z}_1$ has representation

$$\Pi_3^{(\text{time})} = Q_x^{(w)} G \Pi_2^{(\text{time})} G Q_x^{(w)\intercal}, \quad (3.49)$$

where $\Pi_2^{(\text{time})}$ is the covariance matrix given in (3.43),

$$Q_x^{(w)} = (X^\intercal W X)^{-1} X^\intercal W \quad \textit{and} \quad (3.50)$$

$$G = \operatorname{diag}\left\{ \sqrt{\frac{2\pi}{\theta_2^* u^{\alpha_2^*}}} \exp\left\{\theta_2^* u^{\alpha_2^*}/2\right\},\ u \in \mathcal{U} \right\}. \quad (3.51)$$

3.4 Subsampling for confidence regions

Let $\{\eta(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^2, t \in [0,\infty)\}$ be the isotropic Brown-Resnick process as in (3.1) with dependence function δ given in (3.2); i.e.,

$$\delta(v,u) = 2\theta_1 v^{\alpha_1} + 2\theta_2 u^{\alpha_2}, \quad v,u \geq 0,$$

for $\theta_1, \theta_2 > 0$ and $\alpha_1, \alpha_2 \in (0,2]$. We assume to observe the process on a regular grid $\mathcal{S}_n = \{(i_1,i_2) : i_1, i_2 \in \{1,\dots,n\}\}$ and at time points $\{t_1,\dots,t_T\} = \{1,\dots,T\}$ as specified in Condition 3.4. The estimation methods based on the spatial and temporal extremograms described in Sections 3.2 and 3.3 yield consistent estimators $\widehat{\boldsymbol{\psi}}_1^c = (\log(\widehat{\theta}_1^c), \widehat{\alpha}_1^c)^\intercal$ and $\widehat{\boldsymbol{\psi}}_2^c = (\log(\widehat{\theta}_2^c), \widehat{\alpha}_2^c)^\intercal$ of the true spatial and temporal parameters $\boldsymbol{\psi}_1^\star = (\log(\theta_1^\star), \alpha_1^\star)^\intercal$ and $\boldsymbol{\psi}_2^\star = (\log(\theta_2^\star), \alpha_2^\star)^\intercal$, respectively. Furthermore, we have the limit theorems

$$\tau_n\Big(\widehat{\boldsymbol{\psi}}_1^c - \boldsymbol{\psi}_1^\star\Big) \xrightarrow{d} \boldsymbol{Z}^{(1)}, \quad n\to\infty, \quad \text{and} \quad \tau_T\Big(\widehat{\boldsymbol{\psi}}_2^c - \boldsymbol{\psi}_2^\star\Big) \xrightarrow{d} \boldsymbol{Z}^{(2)}, \quad T\to\infty,$$

where the bivariate distributions of $\boldsymbol{Z}^{(1)}$ and $\boldsymbol{Z}^{(2)}$ depend on the true parameter values $\alpha_1^\star$ and $\alpha_2^\star$, respectively. The rates of convergence are given by $\tau_n := \frac{n}{m_n}$ and $\tau_T := \sqrt{\frac{T}{m_T}}$, where m_n and m_T are appropriately chosen scaling sequences.

Due to the complicated forms of the covariance matrices of $\boldsymbol{Z}^{(1)}$ and $\boldsymbol{Z}^{(2)}$ (cf. Theorems 3.13 and 3.19) we use resampling methods to construct asymptotic confidence regions of the parameter vectors $\boldsymbol{\psi}_1^\star$ and $\boldsymbol{\psi}_2^\star$. One appealing method is subsampling (see Politis et al. [56], Chapter 5), since it works under weak regularity conditions asymptotically correct. The central assumption is the existence of weak limit laws, which is guaranteed by Theorems 3.13 and 3.19. In Section 3.4.1 we consider the spatial case, whereas Section 3.4.2 deals with the temporal case.

Subsampling is also successfully applied for confidence bounds of pairwise likelihood estimates of the space-time Brown-Resnick process in Section 5.4. The procedure is as follows: understanding inequalities between vectors componentwise, we choose block lengths $(1,1,1) \leq \boldsymbol{b} = (b_s, b_s, b_t) \leq (n, n, T)$ and the degree of overlap $(1,1,1) \leq \boldsymbol{e} = (e_s, e_s, e_t) \leq (b_s, b_s, b_t)$, where $\boldsymbol{e} = (1,1,1)$ corresponds to maximum overlap and $\boldsymbol{e} = \boldsymbol{b}$ to no overlap. The blocks are indexed by $\boldsymbol{i} = (i_1, i_2, i_3) \in \mathbb{N}^3$ with $i_j \leq q_s$ for $q_s := \lfloor \frac{n-b_s}{e_s} \rfloor + 1$ and $j = 1,2$ and $i_3 \leq q_t := \lfloor \frac{T-b_t}{e_t} \rfloor + 1$. This results in a total number of $q = q_s^2 q_t$ blocks, which we summarise in the sets

$$E_{\boldsymbol{i},\boldsymbol{b},\boldsymbol{e}} = \big\{(s_1, s_2, t) \in \mathcal{S}_n \times \{1, \ldots, T\} : (i_j - 1)e_s + 1 \leq s_j \leq (i_j - 1)e_s + b_s \text{ for } j = 1,2; \\ (i_3 - 1)e_t + 1 \leq t \leq (i_3 - 1)e_t + b_t\big\}.$$

We estimate $\theta_1, \alpha_1, \theta_2, \alpha_2$ based on the observations in each block as described in the previous sections. This yields different estimates for each spatial and temporal parameter, which we denote by $\widehat{\boldsymbol{\psi}}^c_{1,\boldsymbol{i}}$ and $\widehat{\boldsymbol{\psi}}^c_{2,\boldsymbol{i}}$, respectively.

3.4.1 Subsampling: the spatial parameters

Our first theorem below provides a basis for constructing asymptotically valid confidence intervals for the true spatial parameters $\theta_1^\star$ and $\alpha_1^\star$. We define τ_{b_s} as the analogue of $\tau_n = n/m_n = n^{1-\beta}$ where $\beta \in (1/5, 1/2)$; i.e., $\tau_{b_s} := b_s^{1-\beta}$ (cf. Remark 3.9 and Theorem 3.10)

Theorem 3.20. *Assume that the conditions of Theorem 3.13 hold and, as $n \to \infty$,*

(i) $b_s \to \infty$ such that $b_s = o(n)$ (hence, $\tau_{b_s}/\tau_n = (b_s/n)^{1-\beta} \to 0$),

(ii) $\boldsymbol{e}$ does not depend on n.

Define the empirical distribution function $L_{b_s,\boldsymbol{s}}$

$$L_{b_s,s}(x) := \frac{1}{q} \sum_{i_1=1}^{q_s} \sum_{i_2=1}^{q_s} \sum_{i_3=1}^{q_t} \mathbb{1}_{\{\tau_{b_s} \|\widehat{\boldsymbol{\psi}}^c_{1,\boldsymbol{i}} - \widehat{\boldsymbol{\psi}}^c_1\| \leq x\}}, \quad x \in \mathbb{R},$$

and the empirical quantile function

$$c_{b_s,s}(1-\alpha) := \inf\{x \in \mathbb{R} : L_{b_s,s}(x) \geq 1 - \alpha\}, \quad \alpha \in (0,1).$$

Then

$$\mathbb{P}\Big(\tau_n \|\widehat{\boldsymbol{\psi}}^c_1 - \boldsymbol{\psi}_1^\star\| \leq c_{b_s,s}(1-\alpha)\Big) \to 1 - \alpha, \quad n \to \infty. \tag{3.52}$$

Proof. We apply Corollary 5.3.4 of Politis et al. [56]. Their main Assumption 5.3.4 is the existence of a weak limit distribution of $\tau_n \|\widehat{\boldsymbol{\psi}}^c_1 - \boldsymbol{\psi}_1^\star\|$. By Theorem 3.13, the continuous mapping theorem and the Fubini-Tonelli theorem we have for $\gamma \geq 0$, as $n \to \infty$,

$$\mathbb{P}(\tau_n \|\widehat{\boldsymbol{\psi}}_1^c - \boldsymbol{\psi}_1^\star\| \leq \gamma) \to \mathbb{P}(\|\boldsymbol{Z}_1\| \leq \gamma) = \mathbb{P}(\boldsymbol{Z}_1 \in B(\mathbf{0}, \gamma)) = 2 \int\limits_{-\gamma}^{\gamma} \int\limits_{0}^{\sqrt{\gamma^2 - r^2}} \varphi_{\mathbf{0}, \Pi_3^{(\mathrm{iso})}}(r, s) ds dr$$

if $\alpha_1^\star < 2$. For $\alpha_1^\star = 2$ we obtain

$$\begin{aligned}
&\mathbb{P}(\tau_n \|\widehat{\boldsymbol{\psi}}_1^c - \boldsymbol{\psi}_1^\star\| \leq \gamma) \;\to\; \mathbb{P}(\|\boldsymbol{Z}_2\| \leq \gamma) = \mathbb{P}(\boldsymbol{Z}_2 \in B(\mathbf{0}, \gamma)) \\
&= \int\limits_{-\gamma}^{\gamma} \int\limits_{-\sqrt{\gamma^2 - r^2}}^{0} \varphi_{\mathbf{0}, \Pi_3^{(\mathrm{iso})}}(r, s) ds dr \\
&\quad + \int\limits_{-\gamma}^{\gamma} \int\limits_{0}^{\infty} \varphi_{\mathbf{0}, \Pi_3^{(\mathrm{iso})}}(r - \frac{1}{\sum_{v \in \mathcal{V}} w_v} \sum_{v \in \mathcal{V}} (w_v x_v) s, s) ds dr \\
&= \int\limits_{-\gamma}^{\gamma} \Big\{ \int\limits_{-\sqrt{\gamma^2 - r^2}}^{0} \varphi_{\mathbf{0}, \Pi_3^{(\mathrm{iso})}}(r, s) ds + \int\limits_{0}^{\infty} \varphi_{\mathbf{0}, \Pi_3^{(\mathrm{iso})}}\Big(r - \frac{1}{\sum_{v \in \mathcal{V}} w_v} \sum_{v \in \mathcal{V}} (w_v x_v) s, s\Big) ds \Big\} dr.
\end{aligned}$$

In particular, the limiting distribution function of the scaled norm $\tau_n \|\widehat{\boldsymbol{\psi}}_1^c - \boldsymbol{\psi}_1^\star\|$ is continuous in γ both for $\alpha_1^\star < 2$ and $\alpha_1^\star = 2$. Assumptions (i) and (ii) are also presumed in Politis et al. [56]. The required condition on the α-mixing coefficients is satisfied, similarly as in the proof of Theorem 3.7, by equation (B.2) and Lemma B.3, and the result follows. □

As a consequence of equation (3.52), for n large enough, an approximate $(1-\alpha)$-confidence region for the true parameter vector $\boldsymbol{\psi}_1^\star = (\log(\theta_1^\star), \alpha_1^\star)$ is given by

$$\{\boldsymbol{\psi} \in \mathbb{R} \times (0, 2] : \|\boldsymbol{\psi} - \widehat{\boldsymbol{\psi}}_1^c\| \leq c_{b_s, s}(1-\alpha)/\tau_n\}.$$

The one-dimensional approximate $(1-\alpha)$-confidence intervals for the parameters $\theta_1^\star$ and $\alpha_1^\star$ can be read off from this as

$$\Big[\widehat{\theta}_1^c \exp\Big\{ - \frac{c_{b_s,s}(1-\alpha)}{\tau_n}\Big\}, \widehat{\theta}_1^c \exp\Big\{\frac{c_{b_s,s}(1-\alpha)}{\tau_n}\Big\}\Big] \text{ and }$$
$$\Big[\widehat{\alpha}_1^c - \frac{c_{b_s,s}(1-\alpha)}{\tau_n}, \widehat{\alpha}_1^c + \frac{c_{b_s,s}(1-\alpha)}{\tau_n}\Big] \cap (0, 2].$$

3.4.2 Subsampling: the temporal parameters

The theorem below provides a basis for constructing asymptotically valid confidence intervals for the true temporal parameters $\theta_2^\star$ and $\alpha_2^\star$. We define τ_{b_t} as the analogue of $\tau_T = \sqrt{\frac{T}{m_T}} = T^{(1-\beta)/2}$ where $\beta \in (1/5, 1)$; i.e., $\tau_{b_t} := b_t^{(1-\beta)/2}$ (cf. Remark 3.16 and Theorem 3.17). The proof is completely analogous to that of Theorem 3.20 for the spatial parameters.

Theorem 3.21. *Assume that the conditions of Theorem 3.19 hold and, as $T \to \infty$,* *(i)* $b_t \to \infty$ *such that* $b_t = o(T)$ *(hence,* $\tau_{b_t}/\tau_T = (b_t/T)^{(1-\beta)/2} \to 0$*),*

(ii) $\boldsymbol{e}$ *does not depend on* T*.*

Define the empirical distribution function $L_{b_t,t}$

$$L_{b_t,t}(x) := \frac{1}{q}\sum_{i_1=1}^{q_s}\sum_{i_2=1}^{q_s}\sum_{i_3=1}^{q_t} \mathbb{1}_{\{\tau_{b_t}\|\widehat{\psi}^c_{2,i}-\widehat{\psi}^c_2\|\le x\}}, \quad x \in \mathbb{R},$$

and the empirical quantile function

$$c_{b_t,t}(1-\alpha) := \inf\left\{x \in \mathbb{R} : L_{b_t,t}(x) \ge 1-\alpha\right\}, \quad \alpha \in (0,1).$$

Then

$$\mathbb{P}\left(\tau_T\|\widehat{\boldsymbol{\psi}}^c_2 - \boldsymbol{\psi}^\star_2\| \le c_{b_t,t}(1-\alpha)\right) \to 1-\alpha, \quad n \to \infty. \tag{3.53}$$

As a consequence of equation (3.53), for T large enough, an approximate $(1-\alpha)$-confidence region for the true parameter vector $\boldsymbol{\psi}^\star_2 = (\log(\theta^\star_2), \alpha^\star_2)$ is given by

$$\{\boldsymbol{\psi} \in \mathbb{R}\times(0,2] : \|\boldsymbol{\psi} - \widehat{\boldsymbol{\psi}}^c_2\| \le c_{b_t,t}(1-\alpha)/\tau_T\}.$$

The one-dimensional approximate $(1-\alpha)$-confidence intervals for the parameters $\theta^\star_2$ and $\alpha^\star_2$ can be read off from this as

$$\left[\widehat{\theta}^c_2 \exp\left\{-\frac{c_{b_t,t}(1-\alpha)}{\tau_T}\right\}, \widehat{\theta}^c_2 \exp\left\{\frac{c_{b_t,t}(1-\alpha)}{\tau_T}\right\}\right] \text{ and}$$
$$\left[\widehat{\alpha}^c_2 - \frac{c_{b_t,t}(1-\alpha)}{\tau_T}, \widehat{\alpha}^c_2 + \frac{c_{b_t,t}(1-\alpha)}{\tau_T}\right] \cap (0,2].$$

3.5 Simulation study

We examine the performance of the WLSEs by a simulation study. The estimation of the spatial parameters relies on a rather large number of spatial observations and the estimation of the temporal parameters on a rather large number of observed time points. However, simulation of Brown-Resnick space-time processes based on the exact method proposed by Dombry et al. [28] can be time consuming, if both a large number of spatial locations and of time points is taken. For a time-saving method we generate the process on two different space-time observation areas, one for examining the performance of the spatial estimates and one for the temporal ones, which we call $\mathcal{S}^{(1)} \times \mathcal{T}^{(1)}$ and $\mathcal{S}^{(2)} \times \mathcal{T}^{(2)}$, respectively. The design for the simulation experiment is given in more details as follows:

1. We choose two space-time observation areas

$$\mathcal{S}^{(1)} \times \mathcal{T}^{(1)} = \{(i_1,i_2) : i_1, i_2 \in \{1,\dots,70\}\} \times \{1,\dots,10\}$$
$$\mathcal{S}^{(2)} \times \mathcal{T}^{(2)} = \{(i_1,i_2) : i_1, i_2 \in \{1,\dots,5\}\} \times \{1,\dots,300\}$$

and the sets $\mathcal{V} = \{1, \sqrt{2}, 2, \sqrt{5}, \sqrt{8}, 3, \sqrt{10}, \sqrt{13}, 4, \sqrt{17}\}$ and $\mathcal{U} = \{1, \ldots, 10\}$.

2. We simulate the Brown-Resnick space-time process (3.1) based on the exact method proposed in Dombry et al. [28], using the R-package `RandomFields` [60]. The dependence function δ is modelled as in (3.2); i.e.,

$$\delta(v, u) = 2\theta_1 v^{\alpha_1} + 2\theta_2 u^{\alpha_2}, \quad v, u \geq 0,$$

with parameters

$$\theta_1 = 0.4, \; \alpha_1 = 1.5, \quad \theta_2 = 0.2, \; \alpha_2 = 1.$$

3. The parameters $\theta_1, \alpha_1, \theta_2$ and α_2 are estimated.
   - For the estimation of the empirical extremograms (cf. equations 3.8-3.11) we have to choose high empirical quantiles q. In practice, q is chosen from an interval of high quantiles for which the empirical extremogram is robust, see the remarks of Davis et al. [21] after Theorem 2.1. We choose the 90%−empirical quantile for the estimation of the spatial parameters and the 70%−quantile for the temporal part. The quantile for the temporal part is lower to ensure reliable estimation of the extremogram, because the number of time points (300) used for the estimation of the time parameters is much smaller than the number of spatial locations ($70 \cdot 70 = 4900$) used for the estimation of the spatial parameters.
   - The weights in the constrained weighted linear regression problem (see 3.29 and 3.45) are chosen such that locations and time points which are further apart of each other have less influence on the estimation. More precisely, we choose

$$w_u = \exp\{-u^2\} \text{ for } u \in \mathcal{U} \quad \text{and} \quad w_v = \exp\{-v^2\} \text{ for } v \in \mathcal{V}.$$

   This choice of weights reflects the exponential decay of $\chi(v, 0)$ and $\chi(0, u)$ defined in (3.6), which are tail probabilities of the standard normal distribution Φ.

4. Pointwise confidence bounds are computed by subsampling as described in Sections 3.4.1 and 3.4.2. We choose block lengths $\boldsymbol{b} = (60, 60, 10)$ and overlap $\boldsymbol{e} = (2, 2, 2)$ for the space-time process with observation area $\mathcal{S}^{(1)} \times \mathcal{T}^{(1)}$ and $\boldsymbol{b} = (5, 5, 200)$, $\boldsymbol{e} = (1, 1, 1)$ for the process with observation area $\mathcal{S}^{(2)} \times \mathcal{T}^{(2)}$.

5. Steps 1 - 5 are repeated 100 times.

Figures 3.2 and 3.3 show the estimates of the spatial parameters (θ_1, α_1) and temporal parameters (θ_2, α_2) for each of the 100 realisations of the Brown-Resnick space-time process. The dotted lines above and below the dots are pointwise confidence intervals based on subsampling. Table 3.1 shows the mean, RMSE and MAE of the 100 simulations. Altogether, we observe that the estimates are close to the true values. Moreover, the spatial confidence intervals are more accurate than the temporal ones, which is due to the larger number of observations in space than in time.

We carried out the same simulations with $\alpha_1 = 2$ and $\alpha_2 = 2$ and obtained equally satisfying results. The WLSEs are again very accurate and the upper bounds of the subsampling confidence intervals are equal to 2, thus again containing the true value.

	MEAN	RMSE	MAE
θ_1	0.4033	0.0678	0.0559
α_1	1.4984	0.0521	0.0400
θ_2	0.2249	0.0649	0.0526
α_2	0.9563	0.0939	0.0767

Table 3.1: Mean, root mean squared error and mean absolute error of the WLSE.

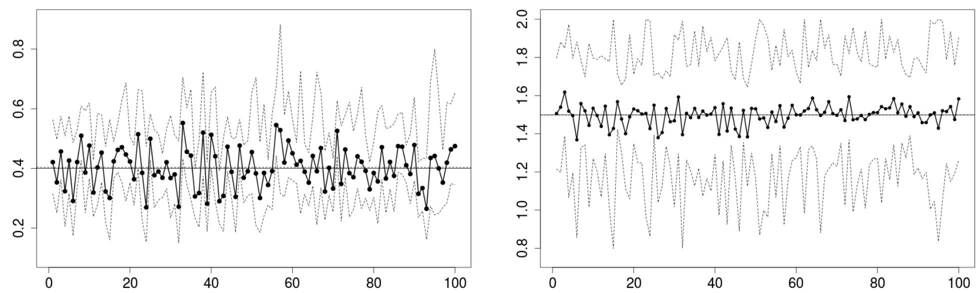

Figure 3.2: WLSEs of θ_1 (left) and α_1 (right) for 100 simulated Brown-Resnick space-time processes together with pointwise 95%−subsampling confidence intervals (dotted). The middle solid line is the true value and the middle dotted line represents the mean over all estimates.

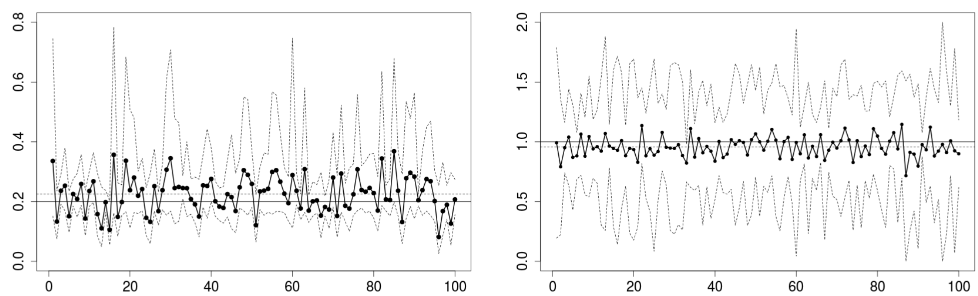

Figure 3.3: WLSEs of θ_2 (left) and α_2 (right) for 100 simulated Brown-Resnick space-time processes together with pointwise 95%−subsampling confidence intervals (dotted). The middle solid line is the true value and the middle dotted line represents the mean over all estimates.

In her Chapter 6, Steinkohl [62] carries out a detailed simulation study with the goal to compare the performance of the semiparametric estimation method with the pairwise likelihood approach in finite samples. To summarise her findings, the WLSE is slightly more biased than the pairwise likelihood estimator. This is due to the fact that the bias correction in the semiparametric estimation depends crucially on the chosen threshold as it applies only asymptotically. A big advantage of the semiparametric method is the substantial reduction of computation time by about a factor 15 compared to the pairwise likelihood estimation.

3.6 Analysis of radar rainfall measurements

Finally, we apply the Brown-Resnick space-time model in (3.1) and the WLSE to radar rainfall data. The data were collected by the Southwest Florida Water Management District (SWFWMD)*. Our objective is to quantify the extremal behaviour of radar rainfall data in a region in Florida by using spatial and temporal block maxima and fitting a Brown-Resnick space-time model to the block maxima.

The data base consists of radar values in inches measured on a 120×120km region containing 3600 grid locations. We calculate the spatial and temporal maxima over subregions of size 10×10km and over 24 subsequent measurements of the corresponding hourly accumulated time series in the wet season (June to September) from the years 1999-2004 for further analysis. In this way we obtain 12×12 locations during 732 days containing space-time block maxima of rainfall observations.

We denote the set of locations by $\mathcal{S} = \{(i_1, i_2), i_1, i_2 \in \{1, \ldots, 12\}\}$ and the space-time observations by $\{\eta(\boldsymbol{s}, t), \boldsymbol{s} \in \mathcal{S}, t \in \{t_1, \ldots, t_{732}\}\}$. This setup is also considered in Section 5.5 and Steinkohl [62], Chapter 7. To make the results obtained there comparable with the results here, we use the data preprocessed as there and after the same marginal modelling steps; for a precise description cf. Section 5.5.1. Since the data do not fail the max-stability check described in Section 5.5.2, we assume that $\{\eta(\boldsymbol{s}, t), \boldsymbol{s} \in \mathcal{S}, t \in \{t_1, \ldots, t_{732}\}\}$ are realisations of a max-stable space-time process with standard unit Fréchet margins.

We then fit the Brown-Resnick model (3.1) by estimating the extremal dependence structure (3.2) as follows:

1. We estimate the parameters θ_1, α_1, θ_2 and α_2 by WLSE as described in Section 3.2 based on the sets $\mathcal{V} = \{1, \sqrt{2}, 2, \sqrt{5}, \sqrt{8}, 3, \sqrt{10}, \sqrt{13}, 4, \sqrt{17}\}$ and $\mathcal{U} = \{1, \ldots, 10\}$ for the linear regression. Permutation tests as described below and visualised in Figure 3.6 indicate that these lags are sufficient to cover the relevant extremal dependence structure. Since the true extremogram χ is unknown, we choose as weights for the different spatial and temporal lags $v \in \mathcal{V}$ and $u \in \mathcal{U}$ the corresponding estimated averaged extremogram values; i.e., $w_v = T^{-1} \sum_{k=1}^{T} \widetilde{\chi}^{(t_k)}(v, 0)$ and $w_u = n^{-2} \sum_{i=1}^{n^2} \widetilde{\chi}^{(\boldsymbol{s}_i)}(0, u)$, respectively. Since the so defined weights are random, what follows is conditional on the realisations of these weights.

   As the number of spatial points in the analysis is rather small, we cannot choose a very high empirical quantile q, since this would in turn result in a too small number of exceedances to get a reliable estimate of the extremogram. Hence, we choose q as the empirical $60\%-$quantile, relying on the fact that the block maxima generate a max-stable process.

   For the temporal estimation, we choose the empirical $90\%-$quantile for q.

*http://www.swfwmd.state.fl.us/

2. We perform subsampling (see Section 3.4) to construct 95%-confidence intervals for each parameter estimate. As subsample block sizes we choose $b_s = 12$ (due to the small number of spatial locations) and $b_t = 300$. We further choose $e_s = e_t = 1$, which corresponds to the maximum degree of overlap.

The results are shown in Figures 3.5, 3.6 and Table 3.2. Figure 3.4 visualises the daily rainfall maxima for the two grid locations $(1,1)$ and $(5,6)$. The semiparametric estimates together with subsampling confidence intervals are given in Table 3.2.

For comparison we present the parameter estimates from the pairwise likelihood estimation (for details for the isotropic Brown-Resnick model see Davis et al. [19] and [62], Chapter 7), where we obtained $\widetilde{\theta}_1 = 0.3485$, $\widetilde{\alpha}_1 = 0.8858$, $\widetilde{\theta}_2 = 2.4190$ and $\widetilde{\alpha}_2 = 0.1973$. From Table 3.2 we recognize that these estimates are close to the semiparametric estimates and even lie in most cases in the 95%-subsampling confidence intervals.

Figure 3.5 shows the spatial and temporal mean of empirical temporal (left) and spatial (right) extremograms as described in (3.10) and (3.11) together with 95% subsampling confidence intervals. We perform a permutation test to test the presence of extremal independence. To this end we randomly permute the space-time data and calculate empirical extremograms as before. More precisely, we compute the empirical temporal extremogram as before and repeat the procedure 1000 times. From the resulting temporal extremogram sample we determine nonparametric 97.5% and 2.5% empirical quantiles, which gives a 95%−confidence region for temporal extremal independence. The analogue procedure is performed for the spatial extremogram.

The results are shown in Figure 3.6 together with the extremogram fit based on the WLSE. The plots indicate that for time lags larger than 3 there is no temporal extremal dependence, and for spatial lags larger than 4 no spatial extremal dependence.

Estimate	$\widehat{\theta}_1$	0.3611	$\widehat{\alpha}_1$	0.9876
Subsampling-CI		[0.3472,0.3755]		[0.9482,1.0267]
Estimate	$\widehat{\theta}_2$	2.3650	$\widehat{\alpha}_2$	0.0818
Subsampling-CI		[1.9110,2.7381]		[0.0000,0.2680]

Table 3.2: Semiparametric estimates for the spatial parameters θ_1 and α_1 and the temporal parameters θ_2 and α_2 of the Brown-Resnick process in (3.1) together with 95% subsampling confidence intervals.

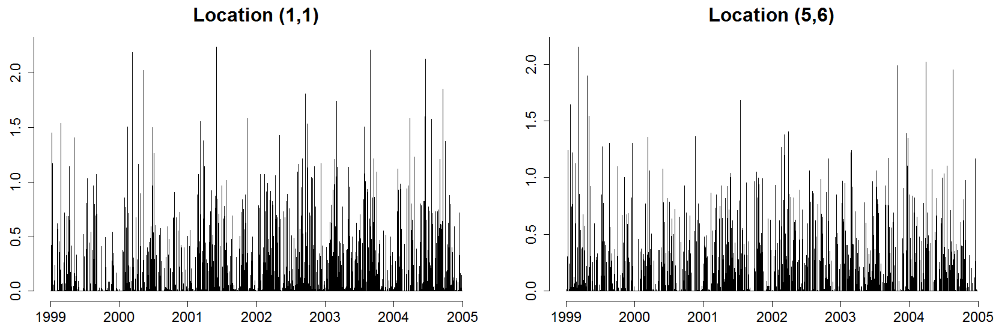

Figure 3.4: Daily rainfall maxima over hourly accumulated measurements from 1999-2004 in inches for two grid locations.

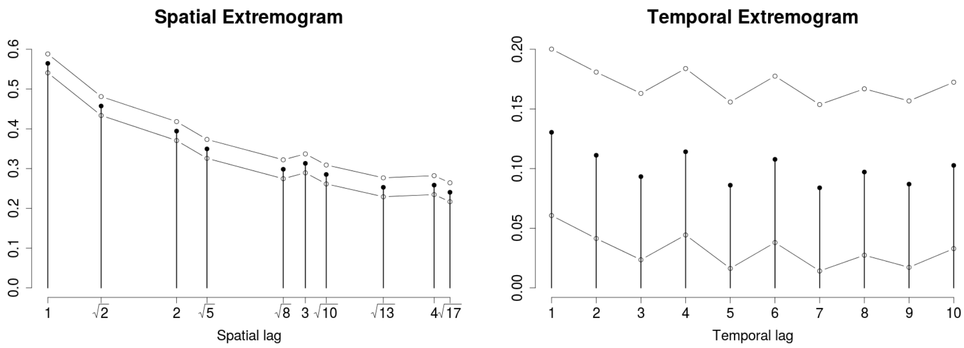

Figure 3.5: Empirical spatial (left) and temporal (right) extremogram based on spatial and temporal means for the space-time observations as given in (3.10) and (3.11) together with 95%−subsampling confidence intervals.

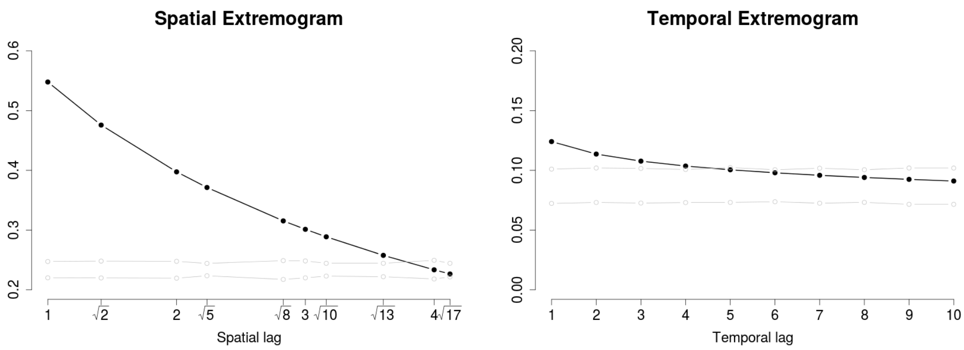

Figure 3.6: Permutation test for extremal independence: The gray lines show the 97.5%− and 2.5%−quantiles of the extremogram estimates for 1000 random space-time permutations for the empirical spatial (left) and the temporal (right) extremogram estimates.

3.7 Conclusions and Outlook

For the isotropic Brown-Resnick space-time process with flexible dependence structure we have suggested a new semiparametric estimation method, which works remarkably well in an extreme value setting. The method results in quite reliable estimates, much faster than the composite likelihood methods used so far. These estimates can also be used as initial values for a composite likelihood optimisation routine to obtain more accurate estimates.

Future work will be dedicated to generalisations of the semiparametric method based on extremogram estimation. At present we work on three topics:

1. Generalise the dependence function (3.2) to anisotropic and appropriate mixed models.

2. Generalise the sampling scheme to allow for a fixed (small) number of spatial observations and consider limit results for the number of temporal observations to tend to infinity.

3. Generalise the least squares estimation to estimate spatial and temporal parameters simultaneously, also in the situation described in topic 2.

Another interesting question concerns the optimal choice of the weight matrix W, such that the asymptotic variance of the WLSE is minimal. Some ideas can be found in the geostatistics literature in the context of least squares estimation of the variogram parameters; see e.g. Lahiri et al. [48], Section 4. They describe the situation, where the optimal choice of the weight matrix is given by the inverse of the asymptotic covariance matrix of the nonparametric estimates; i.e., of $\big(T^{-1}\sum_{k=1}^{T}\widetilde{\chi}^{(t_k)}(v,0)\big)^{\intercal}_{v\in\mathcal{V}}$ in the spatial case and of $\big(n^{-2}\sum_{i=1}^{n^2}\widetilde{\chi}^{(\boldsymbol{s}_i)}(0,u)-\chi(0,u)\big)^{\intercal}_{u\in\mathcal{U}}$ in the temporal case. In our case, however, this involves the matrices $\Pi_2^{(\mathrm{iso})}$ and $\Pi_2^{(\mathrm{time})}$ (given in (2.13)-(2.15)), whose components are infinite sums.

Chapter 4

Generalised least squares estimation of regularly varying space-time processes based on flexible observation schemes

Abstract

Regularly varying stochastic processes model extreme dependence between process values at different locations and/or time points. For such processes we propose a two-step parameter estimation of the extremogram, when some part of the domain of interest is fixed and another increasing. We provide conditions for consistency and asymptotic normality of the empirical extremogram centred by a pre-asymptotic version for such observation schemes. For max-stable processes with Fréchet margins we provide conditions, such that the empirical extremogram (or a bias-corrected version) centred by its true version is asymptotically normal. In a second step, for a parametric extremogram model, we fit the parameters by generalised least squares estimation and prove consistency and asymptotic normality of the estimates. We propose subsampling procedures to obtain asymptotically correct confidence intervals. Finally, we apply our results to a variety of Brown-Resnick processes. A simulation study shows that the procedure works well also for moderate sample sizes.

AMS 2010 Subject Classifications: primary: 60F05, 60G70, 62F12, 62G32
secondary: 37A25, 62M30 62P12,

Keywords: Brown-Resnick process; extremogram; generalised least squares estimation; max-stable process; observations schemes; regularly varying process; semiparametric estimation; space-time process.

4.1 Introduction

Max-stable processes and regularly varying processes have in recent years attracted attention as time series models, spatial processes and space-time processes. Regularly varying processes have been investigated in Hult and Lindskog [41, 40] and basic results on max-stable processes can be found in de Haan and Ferreira [25]. Such processes provide a useful framework for modelling and estimation of extremal events in their different settings.

Among the various regularly varying models considered in the literature, max-stable Brown-Resnick processes play a prominent role allowing for flexible fractional variogram models as often observed in environmental data. They have been introduced for time series in Brown and Resnick [8], for spatial processes in Kabluchko et al. [47], and in a space-time setting in Davis et al. [19].

For max-stable processes with parametrised dependence structure, various estimation procedures have been proposed for extremal data. Composite likelihood methods have been described in Padoan et al. [54] and Huser and Davison [43]. Threshold-based likelihood methods have been proposed in Engelke et al. [34] and Wadsworth and Tawn [67]. For the max-stable Brown-Resnick process asymptotic results of composite likelihood estimators are derived in Huser and Davison [42], Davis et al. [20] and Chapter 5 of this thesis, which is based on the publication Buhl and Klüppelberg [11].

Parameter estimation based on likelihood methods can be laborious and time consuming, and also the choice of good initial values for the optimisation procedure is essential. As a consequence, a semiparametric estimation procedure can be an alternative or a prerequisite for a subsequent likelihood method. Such an estimation method has been suggested and analysed for Brown-Resnick processes in Steinkohl [62] and Chapter 3 based on the extremogram, which is a natural extremal analogue of the correlation function for stationary processes. The extremogram was introduced for time series in Davis and Mikosch [17] and Fasen et al. [35], and extended to a spatial setting in Cho et al. [15] and space-time setting in [62]. A semiparametric estimation requires a parametric extremogram model. The parameter estimation is then based on the empirical extremogram, and a subsequent least squares estimation of the parameters.

The processes considered in [15, 62] and in Chapters 2 and 3, which are based on the publications Buhl and Klüppelberg [12] and Buhl et al. [14], are isotropic Brown-Resnick spatial or space-time processes associated to the class of fractional variogram models with additively separable dependence function in the space-time case, cf. model (I) in Section 4.5.3 below.

The central goal of this chapter is to generalise the semiparametric method developed for the spatially anisotropic Brown-Resnick process in Chapter 3 in various aspects. Firstly, we allow for general regularly varying processes, thus leaving the max-stable models for those in their domains of attraction. Secondly, whereas in Chapter 3 we carried out least squares estimation of the spatial and temporal dependence parameters separately, we allow for a much larger class of dependence models provided they satisfy certain regularity conditions. Thirdly, we develop a generalised least squares estimation, which estimates all dependence parameter in one go. Fourthly, we focus on extremogram estimation based on gridded data, but extend the observation scheme to a more realistic setting. In practice one often observes data on a d-dimensional grid ($d \in \mathbb{N}$) which is small with respect to some of its dimensions (e.g. the spatial dimensions) and large with respect to others (e.g. the temporal dimensions). Hence, with regard to such cases, instead of assuming that the grid increases in all dimensions, it is appropriate to assume for example a number of observed time points which tends to infinity, but a fixed and rather small number of observed spatial data. The extension to such observation schemes makes it necessary to split up every point and every lag in its components corresponding to the fixed and increasing domain. For a parametric extremogram model we derive asymptotic results of its generalised least squares estimators which differ considerably from those obtained when the grid increases in all dimensions. As a general result and not surprisingly, the fixed observation terms are still part of the limits.

This chapter is organised as follows. In Section 4.2 we introduce the theoretical framework of strictly stationary regularly varying processes. We define the extremogram, the observation scheme with its fixed and increasing dimensions as well as assumptions and asymptotic second order properties following from regular variation. Section 4.3 presents the empirical extremogram and its pre-asymptotic version. Here we prove a CLT for the empirical extremogram centred by the pre-asymptotic extremogram. We also specify the asymptotic covariance matrix. We prove a CLT for the empirical extremogram centred by the true extremogram under more restrictive assumptions. To formally state the asymptotic properties of the empirical extremogram, we need to quantify the dependence in a stochastic process, taking into account the different types of observation areas. For processes with Fréchet margins we prove asymptotic normality of the empirical extremogram centred by the true one. In case the required conditions are not satisfied, we provide weaker assumptions under which a CLT for a bias corrected version of the empirical extremogram can be obtained. Section 4.4 is dedicated to the parameter estimation by a generalised least squares method. Under appropriate regularity conditions we prove consistency and asymptotic normality, where the rate of convergence depends on the observation scheme. We also present the covariance matrix in a semi-explicit form. In Section 4.5 we show our method at work for Brown-Resnick space-time processes. We state conditions for Brown-Resnick processes that imply the mixing conditions from Section 4.3 and are hence sufficient to obtain the corresponding CLTs for the empirical extremogram. These conditions depend highly on the model for the associated variogram. Finally, in Section 4.5.3 we apply these results to three different dependence models of the Brown-Resnick process, and prove the mixing conditions, which guarantee the asymptotic normality of the empirical extremogram, as well as the regularity conditions of the generalised least squares estimates. In Section 4.6 we examine the finite sample properties of the GLSEs in a simulation study, fitting the parametric models described in Section 4.5.3 to simulated Brown-Resnick processes. We apply subsampling methods to obtain asymptotically valid confidence bounds of the parameters. Many proofs are rather technical and postponed to an Appendix.

4.2 Model description and the observation scheme

We consider the same theoretical framework as in Chapters 2 and 3 of *a strictly stationary regularly varying stochastic process* $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ for $d \in \mathbb{N}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This implies that there exists some normalizing sequence $0 < a_n \to \infty$ such that $\mathbb{P}(|X(\mathbf{0})| > a_n) \sim n^{-d}$ as $n \to \infty$ and that for every finite set $\mathcal{I} \subset \mathbb{R}^d$ with cardinality $|\mathcal{I}| < \infty$,

$$n^d \mathbb{P}\Big(\frac{X_{\mathcal{I}}}{a_n} \in \cdot\Big) \xrightarrow{v} \mu_{\mathcal{I}}(\cdot), \quad n \to \infty, \tag{4.1}$$

for some non-null Radon measure $\mu_{\mathcal{I}}$ on the Borel sets in $\overline{\mathbb{R}}^{|\mathcal{I}|} \backslash \{\mathbf{0}\}$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ and $X_{\mathcal{I}}$ denotes the vector $(X(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{I})$. The limit measure is homogeneous:

$$\mu_{\mathcal{I}}(xC) = x^{-\beta} \mu_{\mathcal{I}}(C), \quad x > 0,$$

for every Borel set $C \subset \overline{\mathbb{R}}^{|\mathcal{I}|} \backslash \{\mathbf{0}\}$. The notation $\xrightarrow{v}$ stands for vague convergence, and $\beta > 0$ is called the *index of regular variation*. Furthermore, $f(n) \sim g(n)$ as $n \to \infty$ means that $\lim_{n \to \infty} f(n)/g(n) = 1$. If $\mathcal{I}$ is a singleton; i.e., $\mathcal{I} = \{\boldsymbol{s}\}$ for some $\boldsymbol{s} \in \mathbb{R}^d$, we set

$$\mu_{\{\boldsymbol{s}\}}(\cdot) = \mu_{\{\mathbf{0}\}}(\cdot) =: \mu(\cdot), \tag{4.2}$$

which is justified by stationarity. For more details see Chapter 2. For background on regular variation for stochastic processes and vectors see Hult and Lindskog [40, 41] and Resnick [57, 59].

The extremogram for values in $\mathbb{R}^d$ is defined as follows.

Definition 4.1 (Extremogram)**.** *Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ be a strictly stationary regularly varying process and $a_n \to \infty$ a sequence satisfying* (4.1)*. For μ as in* (4.2) *and two μ-continuous Borel sets A and B in $\overline{\mathbb{R}} \backslash \{0\}$ (i.e., $\mu(\partial A) = \mu(\partial B) = 0$) such that $\mu(A) > 0$, the* extremogram *is defined as*

$$\rho_{AB}(\boldsymbol{h}) = \lim_{n \to \infty} \frac{\mathbb{P}(X(\mathbf{0})/a_n \in A, X(\boldsymbol{h})/a_n \in B)}{\mathbb{P}(X(\mathbf{0})/a_n \in A)}, \quad \boldsymbol{h} \in \mathbb{R}^d. \tag{4.3}$$

For $A = B = (1, \infty)$, the extremogram $\rho_{AB}(\boldsymbol{h})$ is the tail dependence coefficient *between $X(\mathbf{0})$ and $X(\boldsymbol{h})$ (cf. Beirlant et al. [3], Section 9.5.1).*

For the data we allow for realistic observation schemes described in the following.

Assumption 4.2. *The data are given in an observation area $\mathcal{D}_n \subset \mathbb{Z}^d$ that can (possibly after reordering) be decomposed into*

$$\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n, \tag{4.4}$$

where for $q, w \in \mathbb{N}$ satisfying $w + q = d$:

(1) $\mathcal{F} \subset \mathbb{Z}^q$ is a fixed domain independent of n, and

(2) $\mathcal{I}_n = \{1, \dots, n\}^w$ is an increasing sequence of regular grids.

This setting is similar to that used in Li et al. [51], where asymptotic properties of space-time covariance estimators are derived. The natural extension of the regular grid $\mathcal{I}_n$ to grids with different side lengths only increases notational complexity, which we avoid here. Our focus is on observations schemes, which are partially fixed and partially tend to infinity.

Example 4.3. In the special case where the observation area is given by

$$\mathcal{D}_n = \mathcal{F} \times \{1, \dots, n\}$$

for $\mathcal{F} \subset \mathbb{R}^{d-1}$, we interpret the observations as generated by a space-time process $\{X(\boldsymbol{s}, t) : \boldsymbol{s} \in \mathbb{R}^{d-1}, t \in [0, \infty)\}$ on a fixed spatial and an increasing temporal domain.

We shall need some definitions and assumptions, which we summarise as follows.

Assumption 4.4.
(1) *For some fixed $\gamma > 0$ and $\mathbf{0}, \boldsymbol{\ell} \in \mathbb{R}^d$ we define the balls*

$$B(\mathbf{0}, \gamma) = \big\{\boldsymbol{s} \in \mathbb{Z}^d : \|\boldsymbol{s}\| \le \gamma\big\} \text{ and } B(\boldsymbol{\ell}, \gamma) = \big\{\boldsymbol{s} \in \mathbb{Z}^d : \|\boldsymbol{\ell} - \boldsymbol{s}\| \le \gamma\big\} = \boldsymbol{\ell} + B(\mathbf{0}, \gamma).$$

(2) *The estimation of the extremogram is based on a set $\mathcal{H} = \{\boldsymbol{h}^{(1)}, \dots, \boldsymbol{h}^{(p)}\} \subset \mathbb{Z}^d \cap B(\mathbf{0}, \gamma)$ of observed lag vectors.*

(3) *We decompose points $\boldsymbol{s} \in \mathbb{R}^d$ with respect to the fixed and increasing domains into $\boldsymbol{s} = (\boldsymbol{f}, \boldsymbol{i}) \in \mathbb{R}^q \times \mathbb{R}^w$.*

(4) *Similarly, we decompose lag vectors $\boldsymbol{h} = \boldsymbol{s} - \boldsymbol{s}'$ or $\boldsymbol{\ell} = \boldsymbol{s} - \boldsymbol{s}'$ for some $\boldsymbol{s}, \boldsymbol{s}' \in \mathbb{R}^d$ into $\boldsymbol{h} = (\boldsymbol{h}_{\mathcal{F}}, \boldsymbol{h}_{\mathcal{I}})$ or $\boldsymbol{\ell} = (\boldsymbol{\ell}_{\mathcal{F}}, \boldsymbol{\ell}_{\mathcal{I}})$ in $\mathbb{R}^q \times \mathbb{R}^w$. The letter $\boldsymbol{h}$ is used throughout as argument of the extremogram or its estimators.*

(5) *We define the vectorised process $\{\boldsymbol{Y}(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ by*

$$\boldsymbol{Y}(\boldsymbol{s}) := X_{B(\boldsymbol{s}, \gamma)};$$

i.e., $\boldsymbol{Y}(\boldsymbol{s})$ is the vector of values of X with indices in the ball $B(\boldsymbol{s}, \gamma)$.

(6) *We shall also need the following relations, also stated in Chapter 2. For $a_n \to \infty$ as in (4.1), the following limits exist by regular variation of $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$. For $\boldsymbol{\ell} \in \mathbb{R}^d$ and $\gamma > 0$,*

$$\mu_{B(\mathbf{0},\gamma)}(C) := \lim_{n\to\infty} n^d \mathbb{P}(\boldsymbol{Y}(\mathbf{0})/a_n \in C), \tag{4.5}$$

$$\tau_{B(\mathbf{0},\gamma)\times B(\boldsymbol{\ell},\gamma)}(C \times D) := \lim_{n\to\infty} n^d \mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_n} \in C, \frac{\boldsymbol{Y}(\boldsymbol{\ell})}{a_n} \in D\Big), \tag{4.6}$$

for a $\mu_{B(\mathbf{0},\gamma)}$-continuous Borel set C in $\overline{\mathbb{R}}^{|B(\mathbf{0},\gamma)|} \backslash \{\mathbf{0}\}$ and a $\tau_{B(\mathbf{0},\gamma)\times B(\boldsymbol{\ell},\gamma)}$-continuous Borel set $C \times D$ in the product space.

(7) *We define sets $D_1, \dots, D_p, D_{p+1}$ by the identity*

$$\{\boldsymbol{Y}(\boldsymbol{s}) \in D_i\} = \{X(\boldsymbol{s}) \in A, X(\boldsymbol{s} + \boldsymbol{h}^{(i)}) \in B\} \tag{4.7}$$

for $i = 1, \dots, p$, and $\{\boldsymbol{Y}(\boldsymbol{s}) \in D_{p+1}\} = \{X(\boldsymbol{s}) \in A\}$. Note in particular that, by the relation between $\{\boldsymbol{Y}(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ and $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ and regular variation, for every μ-continuous Borel set A in $\overline{\mathbb{R}} \setminus \{0\}$,

$$\mu_{B(\mathbf{0},\gamma)}(D_{p+1}) = \lim_{n\to\infty} n^d \mathbb{P}(\boldsymbol{Y}(\mathbf{0})/a_n \in D_{p+1}) = \lim_{n\to\infty} n^d \mathbb{P}(X(\mathbf{0})/a_n \in A) = \mu(A).$$

□

4.3 Limit theory for the empirical extremogram

We suppose that a strictly stationary regularly varying process $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ is observed as in Assumption 4.2 and derive asymptotic properties of the empirical extremogram. We do this by formulating appropriate mixing conditions, generalising the results obtained in Chapter 2 to the more realistic setting of this paper. The proofs are based on spatial mixing conditions, which have to be adapted to the decomposition into a fixed and an increasing observation domain. In principle, our proofs rely on general results of Ibragimov and Linnik [45] and Bolthausen [6].

The main theorem of this section states asymptotic normality of the empirical extremogram sampled at lag vectors $\boldsymbol{h} \in \mathcal{H}$ and centred by its pre-asymptotic counterpart. The empirical and the pre-asymptotic extremograms are defined in Eq. (4.9) and (4.10).

For the definition of the empirical extremogram we need the following notation: for $k \in \mathbb{N}$, an arbitrary set $\mathcal{Z} \subset \mathbb{Z}^k$ and a fixed vector $\boldsymbol{h} \in \mathbb{Z}^k$, define the sets

$$\mathcal{Z}(\boldsymbol{h}) := \{\boldsymbol{z} \in \mathcal{Z} : \boldsymbol{z} + \boldsymbol{h} \in \mathcal{Z}\}, \tag{4.8}$$

which is the set of vectors $\boldsymbol{z} \in \mathcal{Z}$ such that with $\boldsymbol{z}$ also the lagged vector $\boldsymbol{z} + \boldsymbol{h}$ belongs to $\mathcal{Z}$.

Definition 4.5. *Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ be a strictly stationary regularly varying process, which is observed on $\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n$ as in (4.4). Let A and B be μ-continuous Borel sets in $\overline{\mathbb{R}}\setminus\{0\}$ such that $\mu(A) > 0$. For a sequence $m = m_n \to \infty$ and $m_n = o(n)$ as $n \to \infty$ define the following quantities:*

(1) The empirical extremogram

$$\widehat{\rho}_{AB,m_n}(\boldsymbol{h}) := \frac{\frac{1}{|\mathcal{D}_n(\boldsymbol{h})|} \sum\limits_{\boldsymbol{s}\in\mathcal{D}_n(\boldsymbol{h})} \mathbb{1}_{\{X(\boldsymbol{s})/a_m \in A, X(\boldsymbol{s}+\boldsymbol{h})/a_m \in B\}}}{\frac{1}{|\mathcal{D}_n|} \sum\limits_{\boldsymbol{s}\in\mathcal{D}_n} \mathbb{1}_{\{X(\boldsymbol{s})/a_m \in A\}}}, \quad \boldsymbol{h} \in \mathcal{H}. \tag{4.9}$$

For a fixed data set the value $a_m = a_{m_n}$ has to be specified as a large empirical quantile.

(2) The pre-asymptotic extremogram

$$\rho_{AB,m_n}(\boldsymbol{h}) = \frac{\mathbb{P}\left(X(\mathbf{0})/a_m \in A, X(\boldsymbol{h})/a_m \in B\right)}{\mathbb{P}(X(\mathbf{0})/a_m \in A)}, \quad \boldsymbol{h} \in \mathbb{R}^d. \tag{4.10}$$

Key of the proofs of consistency and asymptotic normality of the empirical extremogram below is the fact that $\widehat{\rho}_{AB,m_n}(\boldsymbol{h})$ is the empirical version of the pre-asymptotic extremogram $\rho_{AB,m_n}(\boldsymbol{h})$, which can in turn be viewed as a ratio of pre-asymptotic versions of $\mu_{B(\mathbf{0},\gamma)}(C(\boldsymbol{h}))$ (cf. Eq. (4.5)) for suitably chosen sets $C(\boldsymbol{h})$ that depend on A and B. In particular, by (4.7), for $\boldsymbol{h} \in B(\mathbf{0},\gamma)$,

$$\mathbb{P}\Big(\frac{X(\mathbf{0})}{a_m} \in A, \frac{X(\boldsymbol{h})}{a_m} \in B\Big) = \mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_m} \in C(\boldsymbol{h})\Big)$$

with $C(\boldsymbol{h})$ implicitly defined by $\{\boldsymbol{Y}(\boldsymbol{s}) \in C(\boldsymbol{h})\} = \{X(\boldsymbol{s}) \in A, X(\boldsymbol{s}+\boldsymbol{h}) \in B\}$ for $\boldsymbol{s} \in \mathbb{R}^d$. Note that if $\boldsymbol{h} = \boldsymbol{h}^{(i)} \in \mathcal{H}$, then $C(\boldsymbol{h}) = D_i$, and if $\boldsymbol{h} = \mathbf{0}$ and $A = B$ then $C(\boldsymbol{h}) = D_{p+1}$.

In view of (4.5), $\mu_{B(\mathbf{0},\gamma)}(C(\boldsymbol{h}))$ can be estimated by an empirical mean, where the estimator has to cope with Assumption 4.2 of an observation area with fixed and increasing domain.

Definition 4.6. Assume the situation of Definition 4.5. Based on observations on $\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n$ as in (4.4) decompose the observations $\boldsymbol{s} = (\boldsymbol{f}, \boldsymbol{i}) \in \mathcal{F} \times \mathcal{I}_n$ and the lags $\boldsymbol{h} = (\boldsymbol{h}_{\mathcal{F}}, \boldsymbol{h}_{\mathcal{I}}) \in \mathcal{H}$ as in Assumption 4.4(3) and (4). For $\boldsymbol{h}_{\mathcal{F}} \in \mathcal{H}$ define $\mathcal{F}(\boldsymbol{h}_{\mathcal{F}})$ as in (4.8). Then an empirical version of $\mu_{B(\mathbf{0},\gamma)}(C(\boldsymbol{h}))$ is for $\boldsymbol{h} \in \mathcal{H}$ given by

$$\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(C(\boldsymbol{h})) := \frac{m_n^d}{n^w} \sum_{\boldsymbol{i} \in \mathcal{I}_n} \frac{1}{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}})|} \sum_{\boldsymbol{f} \in \mathcal{F}(\boldsymbol{h}_{\mathcal{F}})} \mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})}{a_m} \in C(\boldsymbol{h})\}}. \tag{4.11}$$

□

Observe that for fixed $\boldsymbol{h}_{\mathcal{F}} \in \mathbb{Z}^q$ and observations on $\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n$ there will be points $\boldsymbol{s} = (\boldsymbol{f}, \boldsymbol{i}) \in \mathcal{F}(\boldsymbol{h}_{\mathcal{F}}) \times \mathcal{I}_n$ with $\boldsymbol{i}$ near the boundary of $\mathcal{I}_n$, such that not all components of the vector $\boldsymbol{Y}(\boldsymbol{s}) = \boldsymbol{Y}(\boldsymbol{f}, \boldsymbol{i})$ are observed. However, since we investigate asymptotic properties of $\mathcal{I}_n$ whose boundary points are negligible, we can ignore such technical details. As will be seen in the proofs below, for every $\boldsymbol{h} \in \mathcal{H}$, the empirical extremogram $\widehat{\rho}_{AB,m_n}(\boldsymbol{h})$ is asymptotically equivalent to the ratio of estimates $\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(C(\boldsymbol{h}))/\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_{p+1})$.

Limit results for the empirical extremogram (4.9) involve the calculation of mean and variance of $\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(C(\boldsymbol{h}^{(i)})) = \widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i)$ for $\boldsymbol{h}^{(i)} \in \mathcal{H}$. Strict stationarity and Assumption 4.4(6) yields immediately by a law of large numbers that $\mathbb{E}[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i)] \to \mu_{B(\mathbf{0},\gamma)}(D_i)$ as $n \to \infty$. Calculation of the variance involves the covariance structure and we decompose as in Assumption 4.4(4) $\boldsymbol{h}^{(i)}$ into $\boldsymbol{h}^{(i)} = (\boldsymbol{h}_{\mathcal{F}}^{(i)}, \boldsymbol{h}_{\mathcal{I}}^{(i)}) \in \mathbb{R}^q \times \mathbb{R}^w$. We have to calculate for $\boldsymbol{f}, \boldsymbol{f}' \in \mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})$ and $\boldsymbol{i}, \boldsymbol{i}' \in \mathcal{I}_n$,

$$\mathbb{C}ov\Big[\mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})}{a_m} \in D_i\}}, \mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{f}',\boldsymbol{i}')}{a_m} \in D_i\}}\Big] = \mathbb{C}ov\Big[\mathbb{1}_{\{\frac{\boldsymbol{Y}(\mathbf{0})}{a_m} \in D_i\}}, \mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}})}{a_m} \in D_i\}}\Big]$$

with $\boldsymbol{\ell}_{\mathcal{F}} = \boldsymbol{f} - \boldsymbol{f}'$ and $\boldsymbol{\ell}_{\mathcal{I}} = \boldsymbol{i} - \boldsymbol{i}'$, where the equality holds by stationarity. The lag vectors $\boldsymbol{\ell}_{\mathcal{F}}$ and $\boldsymbol{\ell}_{\mathcal{I}}$ are contained in

$$L_{\mathcal{F}}^{(i)} = \{\boldsymbol{f} - \boldsymbol{f}' : \boldsymbol{f} \in \mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)}), \boldsymbol{f}' \in \mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})\} \quad \text{and} \quad L_n := \{\boldsymbol{i} - \boldsymbol{i}' : \boldsymbol{i}, \boldsymbol{i}' \in \mathcal{I}_n\}, \tag{4.12}$$

respectively. The number of appearances of the lag $\boldsymbol{\ell}_{\mathcal{F}}$ we denote by

$$\mathrm{N}^{(i)}_{\mathcal{F}}(\boldsymbol{\ell}_{\mathcal{F}}) := \sum_{\boldsymbol{f},\boldsymbol{f}'\in\mathcal{F}(\boldsymbol{h}^{(i)}_{\mathcal{F}})} \mathbb{1}_{\{\boldsymbol{f}-\boldsymbol{f}'=\boldsymbol{\ell}_{\mathcal{F}}\}}. \tag{4.13}$$

Observe that a spatial lag $(\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}})$ with $\boldsymbol{\ell}_{\mathcal{I}} = (\ell^{(1)}_{\mathcal{I}},\ldots,\ell^{(w)}_{\mathcal{I}})$ appears in $L^{(i)}_{\mathcal{F}} \times L_n$ exactly $\mathrm{N}^{(i)}_{\mathcal{F}}(\boldsymbol{\ell}_{\mathcal{F}})\prod_{j=1}^{w}(n-|\ell^{(j)}_{\mathcal{I}}|)$ times. We show in Lemma C.2 that

$$\begin{aligned}
\mathbb{V}ar\big[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i)\big] &= \frac{m_n^{2d}}{n^{2w}|\mathcal{F}(\boldsymbol{h}^{(i)}_{\mathcal{F}})|^2}\mathbb{V}ar\Big[\sum_{\boldsymbol{f}\in\mathcal{F}(\boldsymbol{h}^{(i)}_{\mathcal{F}})}\sum_{\boldsymbol{i}\in\mathcal{I}_n}\mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})}{a_m}\in D_i\}}\Big] \\
&= \frac{m_n^{2d}}{n^{2w}|\mathcal{F}(\boldsymbol{h}^{(i)}_{\mathcal{F}})|^2}\Big(|\mathcal{F}(\boldsymbol{h}^{(i)}_{\mathcal{F}})|n^w\mathbb{V}ar\big[\mathbb{1}_{\{\frac{\boldsymbol{Y}(\mathbf{0})}{a_m}\in D_i\}}\big] \qquad (4.14)\\
&\quad + \sum_{\boldsymbol{f},\boldsymbol{f}'\in\mathcal{F}(\boldsymbol{h}^{(i)}_{\mathcal{F}})}\sum_{\substack{\boldsymbol{i},\boldsymbol{i}'\in\mathcal{I}_n\\(\boldsymbol{f},\boldsymbol{i})\neq(\boldsymbol{f}',\boldsymbol{i}')}}\mathbb{C}ov\big[\mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})}{a_m}\in D_i\}},\mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{f}',\boldsymbol{i}')}{a_m}\in D_i\}}\big]\Big) \\
&\sim \frac{m_n^d}{n^w}\frac{1}{|\mathcal{F}(\boldsymbol{h}^{(i)}_{\mathcal{F}})|}\Big(\mu_{B(\mathbf{0},\gamma)}(D_i) + \sum_{\boldsymbol{\ell}_{\mathcal{I}}\in\mathbb{Z}^w}\frac{1}{|\mathcal{F}(\boldsymbol{h}^{(i)}_{\mathcal{F}})|}\sum_{\substack{\boldsymbol{\ell}_{\mathcal{F}}\in L^{(i)}_{\mathcal{F}}\\(\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}})\neq\mathbf{0}}}\mathrm{N}^{(i)}_{\mathcal{F}}(\boldsymbol{\ell}_{\mathcal{F}})\,\tau_{B(\mathbf{0},\gamma)\times B((\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}}),\gamma)}(D_i\times D_i)\Big) \\
&=: \frac{m_n^d}{n^w}\sigma^2_{B(\mathbf{0},\gamma)}(D_i), \quad n\to\infty. \qquad (4.15)
\end{aligned}$$

Remark 4.7. For comparison we recall the expression in the corresponding Lemma 2.12, where $\mathcal{F}$ is not fixed, but part of the increasing regular grid. Then $|\mathcal{F}(\boldsymbol{h}^{(i)}_{\mathcal{F}})| \sim \mathrm{N}^{(i)}_{\mathcal{F}}(\boldsymbol{\ell}_{\mathcal{F}}) \sim n^q$ as $n\to\infty$, such that (4.14) can be approximated as follows:

$$\begin{aligned}
\mathbb{V}ar\big[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i)\big] &\sim \frac{m_n^d}{n^w n^q}\Big(\mu_{B(\mathbf{0},\gamma)}(D_i) + \sum_{\boldsymbol{\ell}_{\mathcal{I}}\in\mathbb{Z}^w}\sum_{\substack{\boldsymbol{\ell}_{\mathcal{F}}\in\mathbb{Z}^q\\(\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}})\neq\mathbf{0}}}\tau_{B(\mathbf{0},\gamma)\times B((\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}}),\gamma)}(D_i\times D_i)\Big) \\
&= \Big(\frac{m_n}{n}\Big)^d\Big(\mu_{B(\mathbf{0},\gamma)}(D_i) + \sum_{\boldsymbol{\ell}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}}\tau_{B(\mathbf{0},\gamma)\times B(\boldsymbol{\ell},\gamma)}(D_i\times D_i)\Big), \quad n\to\infty.
\end{aligned}$$

Thus, the difference from the setting of a partly fixed observation area $\mathcal{F}\subset\mathcal{D}_n$ is that the fixed observation terms do not disappear asymptotically, but remain as constants in the limit expression.

4.3.1 The extremogram for regularly varying processes

For proving asymptotic normality of the empirical extremogram we have to require appropriate mixing conditions and make use of a large/small block argument as in Chapter 2. For simplicity we assume that n^w/m_n^d is an integer and subdivide $\mathcal{D}_n$ into n^w/m_n^d non-overlapping d-dimensional large blocks $\mathcal{F}\times\mathcal{B}_i$ for $i=1,\ldots,n^w/m_n^d$, where the $\mathcal{B}_i$ are w-dimensional cubes with side lengths $m_n^{d/w}$. From those large blocks we then cut off smaller blocks, which consist of the first r_n elements in each of the w increasing dimensions. The large blocks are then separated (by these small blocks) with at least the distance r_n in all w increasing dimensions and shown to be asymptotically independent.

In order to formulate the CLT below, in particular, the asymptotic covariance matrix, we need to compute $\mathbb{C}ov[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i), \widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_j)]$ for possibly different $i,j \in \{1,\ldots,p\}$. To this end we extend the notation (4.12) and (4.13) as follows. The lag vectors $\boldsymbol{\ell}_{\mathcal{F}}$ are contained in

$$L_{\mathcal{F}}^{(i,j)} := \{\boldsymbol{f} - \boldsymbol{f}' : \boldsymbol{f} \in \mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)}), \boldsymbol{f}' \in \mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(j)})\}, \tag{4.16}$$

and we denote the number of appearances of the lag vector $\boldsymbol{\ell}_{\mathcal{F}}$ by

$$\mathrm{N}_{\mathcal{F}}^{(i,j)}(\boldsymbol{\ell}_{\mathcal{F}}) := \sum_{\boldsymbol{f} \in \mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})} \sum_{\boldsymbol{f}' \in \mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(j)})} \mathbb{1}_{\{\boldsymbol{f} - \boldsymbol{f}' = \boldsymbol{\ell}_{\mathcal{F}}\}} \tag{4.17}$$

If $i = j$, then we obtain again (4.12) and (4.13).

The asymptotic results stated in Theorem 4.8 below extend those in Theorem 2.6, where the observation area $\mathcal{D}_n$ is assumed to increase with n in all dimensions. The decomposition (4.4) into a fixed domain $\mathcal{F}$ and an increasing domain $\mathcal{I}_n$ results in mixing conditions which focus on properties for $\mathcal{I}_n$ increasing to $\mathbb{Z}^w$, while $\mathcal{F}$ remains fix and appears in the limit, similarly as in Eq. (4.14).

Theorem 4.8. *Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ be a strictly stationary regularly varying process, which is observed on $\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n$ as in* (4.4). *Let $\mathcal{H} = \{\boldsymbol{h}^{(1)}, \ldots, \boldsymbol{h}^{(p)}\} \subset \mathbb{Z}^d \cap B(\mathbf{0},\gamma)$ for some $\gamma > 0$ be a set of observed lag vectors. Suppose that the following conditions are satisfied.*

(M1) $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ is α-mixing with respect to $\mathbb{R}^w$ with mixing coefficients $\alpha_{k_1,k_2}(\cdot)$ defined in (C.1).

There exist sequences $m_n, r_n \to \infty$ with $m_n^d/n^w \to 0$ and $r_n^w/m_n^d \to 0$ as $n \to \infty$ such that:

(M2) $m_n^{2d} r_n^{2w}/n^w \to 0$.

(M3) For all $\epsilon > 0$, and for all fixed $\boldsymbol{\ell}_{\mathcal{F}} \in \mathbb{R}^q$ with $a_m = a_{m_n} \to \infty$ as in (4.1),

$$\lim_{k\to\infty} \limsup_{n\to\infty} \sum_{\substack{\boldsymbol{\ell}_{\mathcal{I}} \in \mathbb{Z}^w \\ k < \|\boldsymbol{\ell}_{\mathcal{I}}\| \le r_n}} m_n^d \, \mathbb{P}(\max_{\boldsymbol{s} \in B(\mathbf{0},\gamma)} |X(\boldsymbol{s})| > \epsilon a_m, \max_{\boldsymbol{s}' \in B((\boldsymbol{\ell}_{\mathcal{F}}, \boldsymbol{\ell}_{\mathcal{I}}),\gamma)} |X(\boldsymbol{s}')| > \epsilon a_m) = 0.$$

(M4) *(i)* $\lim_{n\to\infty} m_n^d \sum_{\boldsymbol{\ell} \in \mathbb{Z}^w : \|\boldsymbol{\ell}\| > r_n} \alpha_{1,1}(\|\boldsymbol{\ell}\|) = 0,$

(ii) $\sum_{\boldsymbol{\ell} \in \mathbb{Z}^w} \alpha_{k_1,k_2}(\|\boldsymbol{\ell}\|) < \infty$ *for* $2 \le k_1 + k_2 \le 4,$

(iii) $\lim_{n\to\infty} m_n^{d/2} n^{w/2} \, \alpha_{1,n^w}(r_n) = 0.$

Then the empirical extremogram $\widehat{\rho}_{AB,m_n}$ defined in (4.9)*, sampled at lags in $\mathcal{H}$ and centred by the pre-asymptotic extremogram ρ_{AB,m_n} given in* (4.10)*, is asymptotically normal; i.e.,*

$$\sqrt{\frac{n^w}{m_n^d}} \Big[\widehat{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)}) - \rho_{AB,m_n}(\boldsymbol{h}^{(i)})\Big]_{i=1,\ldots,p} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Pi), \quad n \to \infty, \tag{4.18}$$

where $\Pi = \mu(A)^{-4} F \Sigma F^{\intercal} \in \mathbb{R}^{p\times p}$. *Writing* $\boldsymbol{h}^{(i)} = (\boldsymbol{h}_{\mathcal{F}}^{(i)}, \boldsymbol{h}_{\mathcal{I}}^{(i)})$ *for* $1 \le i \le p+1$, *with the convention that* $(\boldsymbol{h}_{\mathcal{F}}^{(p+1)}, \boldsymbol{h}_{\mathcal{I}}^{(p+1)}) = \mathbf{0}$, *the matrix* $\Sigma \in \mathbb{R}^{(p+1)\times(p+1)}$ *has components*

$$\begin{aligned}\Sigma_{ii} =& \frac{1}{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})|} \mu_{B(\mathbf{0},\gamma)}(D_i) \\ &+ \sum_{\boldsymbol{\ell}_{\mathcal{I}} \in \mathbb{Z}^w} \frac{1}{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})|^2} \sum_{\substack{\boldsymbol{\ell}_{\mathcal{F}} \in L_{\mathcal{F}}^{(i)} \\ (\boldsymbol{\ell}_{\mathcal{F}}, \boldsymbol{\ell}_{\mathcal{I}}) \neq \mathbf{0}}} \mathrm{N}_{\mathcal{F}}^{(i)}(\boldsymbol{\ell}_{\mathcal{F}})\, \tau_{B(\mathbf{0},\gamma)\times B((\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}}),\gamma)}(D_i \times D_i) \\ =&: \sigma_{B(\mathbf{0},\gamma)}(D_i)^2, \quad 1 \le i \le p+1, \end{aligned} \tag{4.19}$$

$$\begin{aligned}\Sigma_{ij} =& \frac{1}{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})||\mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(j)})|} \Big(|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)}) \cap \mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(j)})| \mu_{B(\mathbf{0},\gamma)}(D_i \cap D_j) \\ &+ \sum_{\boldsymbol{\ell}_{\mathcal{I}} \in \mathbb{Z}^w} \sum_{\substack{\boldsymbol{\ell}_{\mathcal{F}} \in L_{\mathcal{F}}^{(i,j)} \\ (\boldsymbol{\ell}_{\mathcal{F}}, \boldsymbol{\ell}_{\mathcal{I}}) \neq \mathbf{0}}} \mathrm{N}_{\mathcal{F}}^{(i,j)}(\boldsymbol{\ell}_{\mathcal{F}})\, \tau_{B(\mathbf{0},\gamma)\times B((\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}}),\gamma)}(D_i \times D_j) \Big), \quad 1 \le i \neq j \le p+1. \end{aligned} \tag{4.20}$$

The matrix F *consists of a diagonal matrix* F_1 *and a vector* F_2 *in the last column:*

$$F = [F_1, F_2] \quad \textit{with} \quad F_1 = \mathrm{diag}(\mu(A)) \in \mathbb{R}^{p\times p}, \quad F_2 = (-\mu_{B(\mathbf{0},\gamma)}(D_1), \dots, -\mu_{B(\mathbf{0},\gamma)}(D_p))^{\intercal}.$$

Corollary 4.9. *Assume the situation as in Theorem 4.8. Suppose that the following conditions are satisfied.*

(1) $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ *is* α*-mixing with respect to* $\mathbb{R}^w$ *with mixing coefficients* $\alpha_{k_1,k_2}(z)$ *defined in* (C.1).

(2) There exist sequences $m := m_n, r := r_n \to \infty$ *with* $m_n^d/n^w \to 0$ *and* $r_n^w/m_n^d \to 0$ *as* $n \to \infty$ *such that (M3) and (M4i) hold.*

Then, as $n \to \infty$,

$$\widehat{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)}) \xrightarrow{P} \rho_{AB}(\boldsymbol{h}^{(i)}), \quad i = 1, \dots, p,$$

Proof. As in part II of the proof of Theorem 4.8 (cf. Appendix C.2), we find that for $i = 1, \dots, p$, as $n \to \infty$,

$$\widehat{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)}) \sim \frac{\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i)}{\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_{p+1})} \xrightarrow{P} \frac{\mu_{B(\mathbf{0},\gamma)}(D_i)}{\mu_{B(\mathbf{0},\gamma)}(D_{p+1})} = \rho_{AB}(\boldsymbol{h}^{(i)}),$$

where the sets D_i and D_{p+1} are defined in (4.7). Convergence in probability follows by Lemma C.2 and Slutzky's theorem. The last identity holds by definitions (4.3) and (4.5), recalling that $\mu_{B(\mathbf{0},\gamma)}(D_{p+1}) = \mu(A) > 0$. □

Remark 4.10. (i) If the choice $m_n = n^{\beta_1}$ and $r_n = n^{\beta_2}$ with $0 < \beta_2 < \beta_1 d/w < 1$ satisfies conditions (M3) and (M4), then for $\beta_1 \in (0, w/(2d))$ and $\beta_2 \in (0, \min\{\beta_1 d/w; 1/2 - \beta_1 d/w\})$ the condition (M2) also holds and we obtain the CLT (4.18).

(ii) The pre-asymptotic extremogram (4.10) in the CLT (4.18) can be replaced by the true one (4.3), if the pre-asymptotic extremogram converges to the true extremogram with the same convergence rate; i.e., if

$$\sqrt{\frac{n^w}{m_n^d}}\Big[\rho_{AB,m_n}(\boldsymbol{h}^{(i)})-\rho_{AB}(\boldsymbol{h}^{(i)})\Big]_{i=1,\ldots,p}\to\mathbf{0},\quad n\to\infty. \tag{4.21}$$

4.3.2 The extremogram of processes with Fréchet marginal distributions

There are strictly stationary regularly varying processes for which (4.18) is satisfied, but (4.21) does not hold. Theorem 4.11 below states a necessary and sufficient condition for max-stable processes with Fréchet marginal distributions such that both (4.18) and (4.21) hold, yielding the CLT (4.28) for the empirical extremogram (4.9) centred by the the true one (4.3). In case this condition is not satisfied, Theorem 4.12 states conditions such that (4.28) holds for a bias corrected version of the empirical extremogram.

Theorem 4.11 (CLT for processes with Fréchet margins)**.** *Let $\{X(\boldsymbol{s}):\boldsymbol{s}\in\mathbb{R}^d\}$ be a strictly stationary max-stable process with standard unit Fréchet margins, which is observed on $\mathcal{D}_n=\mathcal{F}\times\mathcal{I}_n$ as in* (4.4)*. Let $\mathcal{H}=\{\boldsymbol{h}^{(1)},\ldots,\boldsymbol{h}^{(p)}\}\subset\mathbb{Z}^d\cap B(\mathbf{0},\gamma)$ for some $\gamma>0$ be a set of observed lag vectors. Suppose that conditions (M1)–(M4) of Theorem 4.8 hold for appropriately chosen sequences $m_n,r_n\to\infty$. Let ρ_{AB} be the extremogram* (4.3) *and ρ_{AB,m_n} the pre-asymptotic version* (4.10) *for sets $A=(a_1,a_2)$ and $B=(b_1,b_2)$ with $0<a_1<a_2\leq\infty$ and $0<b_1<b_2\leq\infty$. Then the limit relation* (4.21) *holds if and only if $n^w/m_n^{3d}\to 0$ as $n\to\infty$. In this case we obtain*

$$\sqrt{\frac{n^w}{m_n^d}}\Big[\widehat{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)})-\rho_{AB}(\boldsymbol{h}^{(i)})\Big]_{i=1,\ldots,p}\xrightarrow{d}\mathcal{N}(\mathbf{0},\Pi),\quad n\to\infty, \tag{4.22}$$

with Π specified in Theorem 4.8.

Proof. All finite-dimensional distributions are max-stable distributions with standard unit Fréchet margins, hence they are multivariate regularly varying. Furthermore we can choose $a_m=m_n^d$ in Definition 4.1. Let $V_2(\boldsymbol{h};\cdot,\cdot)$ be the bivariate exponent measure defined by $\mathbb{P}(X(\mathbf{0})\leq x_1,X(\boldsymbol{h})\leq x_2)=\exp\{-V_2(\boldsymbol{h};x_1,x_2)\}$ for $x_1,x_2>0$, cf. Beirlant et al. [3], Section 8.2.2. From Lemma A.1(b) we know that for $\boldsymbol{h}\in\mathcal{H}$ and with $\overline{V}_2^2(\boldsymbol{h}):=a_1a_2/(a_2-a_1)(V_2^2(\boldsymbol{h};a_2,b_2)+V_2^2(\boldsymbol{h};a_2,b_1)+V_2^2(\boldsymbol{h};a_1,b_2)+V_2^2(\boldsymbol{h};a_1,b_1))$,

$$\rho_{AB,m_n}(\boldsymbol{h})=(1+o(1))\Big[\rho_{AB}(\boldsymbol{h})+\frac{1}{2\,m_n^d}\overline{V}_2^2(\boldsymbol{h})\Big],\quad n\to\infty. \tag{4.23}$$

If $a_2=\infty$ and/or $b_2=\infty$, appropriate adaptations need to be taken, which are described in Lemma A.1. Hence, for $\boldsymbol{h}\in\mathcal{H}$,

$$\sqrt{\frac{n^w}{m_n^d}}\big(\rho_{AB,m_n}(\boldsymbol{h})-\rho_{AB}(\boldsymbol{h})\big)=(1+o(1))\sqrt{\frac{n^w}{m_n^{3d}}}\frac{\overline{V}_2^2(\boldsymbol{h})}{2},\quad n\to\infty,$$

which converges to 0 if and only if $n^w/m_n^{3d}\to 0$. □

If $n^w/m_n^{3d}\not\to 0$ in Theorem 4.11, a CLT centred by the true extremogram can still be obtained for a bias corrected empirical estimator. Eq. (4.23) is the basis for such a bias correction if the sets A and B are given by $A = (a, \infty)$ and $B = (b, \infty)$ with $a, b > 0$. In that case we have

$$\rho_{AB,m_n}(\boldsymbol{h}) = (1 + o(1))\Big[\rho_{AB}(\boldsymbol{h}) + \frac{1}{2\,m_n^d a}\big(\rho_{AB}(\boldsymbol{h}) - 2\frac{a}{b}\big)\big(\rho_{AB}(\boldsymbol{h}) - 1\big)\Big], \quad n \to \infty; \tag{4.24}$$

see Eq. (A.4).

An asymptotically bias corrected estimator is given by

$$\widehat{\rho}_{AB,m_n}(\boldsymbol{h}) - \frac{1}{2m_n^d a}\big(\widehat{\rho}_{AB,m_n}(\boldsymbol{h}) - 2\frac{a}{b}\big)\big(\widehat{\rho}_{AB,m_n}(\boldsymbol{h}) - 1\big)$$

and we set, covering both cases,

$$\widetilde{\rho}_{AB,m_n}(\boldsymbol{h}) := \begin{cases} \widehat{\rho}_{AB,m_n}(\boldsymbol{h}) - \frac{1}{2m_n^d a}\big(\widehat{\rho}_{AB,m_n}(\boldsymbol{h}) - 2\frac{a}{b}\big)\big(\widehat{\rho}_{AB,m_n}(\boldsymbol{h}) - 1\big) & \text{if } n^w/m_n^{3d} \not\to 0 \text{ but } n^w/m_n^{5d} \to 0, \\ \widehat{\rho}_{AB,m_n}(\boldsymbol{h}) & \text{if } n^w/m_n^{3d} \to 0. \end{cases} \tag{4.25}$$

Theorem 4.12 below guarantees asymptotic normality of the bias corrected extremogram for an—according to Theorem 4.8—valid sequence m_n satisfying $n^w/m_n^{5d} \to 0$. The proof, which is given in Appendix C.3, generalises that of Theorem 3.10, which covers the special case $a = b = 1$ for Brown-Resnick processes.

Theorem 4.12 (CLT for the bias corrected extremogram for processes with Fréchet margins)**.** *Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ be a strictly stationary max-stable process with standard unit Fréchet margins. Assume the situation of Theorem 4.11 for sets $A = (a, \infty)$ and $B = (b, \infty)$ with $a, b > 0$. Then if and only if $n^w/m_n^{5d} \to 0$, the bias corrected extremogram* (4.25) *is asymptotically normal; i.e.,*

$$\sqrt{\frac{n^w}{m_n^d}}\Big[\widetilde{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)}) - \rho_{AB}(\boldsymbol{h}^{(i)})\Big]_{i=1,\dots,p} \stackrel{d}{\to} \mathcal{N}(\boldsymbol{0}, \Pi), \tag{4.26}$$

where Π *is specified in Theorem 4.8.*

Remark 4.13. From Theorems 4.11 and 4.12 in relation to Remark 4.10 (i) we deduce two cases:

(I) For $w/(5d) < \beta_1 \le w/(3d)$ we cannot replace the pre-asymptotic extremogram by the theoretical version in (4.22), but can resort to a bias correction as described in (4.25) to obtain

$$n^{(w-\beta_1 d)/2}\Big[\widetilde{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)}) - \rho_{AB}(\boldsymbol{h}^{(i)})\Big]_{i=1,\dots,p} \stackrel{d}{\to} \mathcal{N}(\boldsymbol{0}, \Pi), \quad n \to \infty, \tag{4.27}$$

for sets $A = (a, \infty)$ and $B = (b, \infty)$ with covariance matrix Π specified in Theorem 4.8.

(II) For $w/(3d) < \beta_1 < w/(2d)$ we obtain indeed

$$n^{(w-\beta_1 d)/2}\Big[\widehat{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)}) - \rho_{AB}(\boldsymbol{h}^{(i)})\Big]_{i=1,\dots,p} \stackrel{d}{\to} \mathcal{N}(\boldsymbol{0}, \Pi), \quad n \to \infty, \tag{4.28}$$

with covariance matrix Π specified in Theorem 4.8.

Observe that Remark 4.13 generalises Remark 3.9.

4.4 Generalised least squares extremogram estimates

Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ be a strictly stationary regularly varying process observed on $\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n$ as in (4.4). In this section we fit parametric models to the empirical extremogram using least squares techniques for the parameter estimation. Our approach and extremogram models extend the weighted least squares estimation developed in Chapter 3 and Steinkohl [62] considerably. In that work the isotropic space-time Brown-Resnick model (I) of Section 4.5.3 below has been estimated by separation of space and time, which is possible for that model, but not for all models of interest. In what follows we present generalised least squares approaches to fit general parametric extremogram models taking the observation scheme $\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n$ of a fixed and an increasing domain into account. The approach bears some similarity to the semiparametric variogram estimation in Lahiri et al. [48].

Our setting is as follows. Let $\{\rho_{AB,\boldsymbol{\theta}}(\boldsymbol{h}) : \boldsymbol{h} \in \mathbb{R}^d, \boldsymbol{\theta} \in \Theta\}$ be some parametric extremogram model with parameter space Θ and continuous in $\boldsymbol{h} \in \mathbb{R}^d$. Assume that $\rho_{AB}(\cdot) = \rho_{AB,\boldsymbol{\theta}^\star}(\cdot)$ with true parameter vector $\boldsymbol{\theta}^\star \in \Theta$. Denote by $\widehat{\rho}_{AB,m_n}(\boldsymbol{h})$ any of the estimators of Theorem 4.8, Theorem 4.11, or Theorem 4.12 for the appropriately chosen μ-continuous Borel sets A and B such that $\mu(A) > 0$ and lags $\boldsymbol{h} \in \mathcal{H} = \{\boldsymbol{h}^{(1)}, \ldots, \boldsymbol{h}^{(p)}\}$.

First note that under the much weaker conditions of Corollary 4.9 the empirical extremogram is a consistent estimator of the extremogram such that as $n \to \infty$,

$$\widehat{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)}) \xrightarrow{P} \rho_{AB,\boldsymbol{\theta}^\star}(\boldsymbol{h}^{(i)}), \quad i = 1, \ldots, p, \tag{4.29}$$

Under more restrictive conditions given in the three CLTs above,

$$\sqrt{\frac{n^w}{m_n^d}} \Big[\widehat{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)}) - \rho_{AB,\boldsymbol{\theta}^\star}(\boldsymbol{h}^{(i)})\Big]_{i=1,\ldots,p} \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \Pi), \tag{4.30}$$

where Π is the covariance matrix specified in Theorem 4.8.

As we shall prove below, consistency of the empirical extremogram entails consistent generalised least squares parameter estimates, whereas asymptotic normality of the empirical extremogram entails asymptotically normal generalised least squares parameter estimates.

Definition 4.14 (Generalised least squares extremogram estimator (GLSE))**.** *Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ be a strictly stationary regularly varying process, which is observed on $\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n$ as in (4.4). Let A and B be μ-continuous Borel sets in $\overline{\mathbb{R}}\backslash\{0\}$. For a sequence $m = m_n \to \infty$ and $m_n = o(n)$ as $n \to \infty$ define for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ the column vector*

$$\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}) := \big[\widehat{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)}) - \rho_{AB,\boldsymbol{\theta}}(\boldsymbol{h}^{(i)})\big]^{\intercal}_{i=1,\ldots,p}. \tag{4.31}$$

For some positive definite weight matrix $V(\boldsymbol{\theta}) \in \mathbb{R}^{p\times p}$, *the* GLSE *is defined as*

$$\widehat{\boldsymbol{\theta}}_{n,V} := \underset{\boldsymbol{\theta}\in\Theta}{\arg\min}\{\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta})^\intercal V(\boldsymbol{\theta})\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta})\}. \tag{4.32}$$

Assumption 4.15 presents a set of conditions, which imply consistency and asymptotic normality of the GLSE.

Assumption 4.15. *Assume the situation of Definition 4.14. We shall require the following conditions.*

(G1) Consistency: $\widehat{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)}) \xrightarrow{P} \rho_{AB,\boldsymbol{\theta}^\star}(\boldsymbol{h}^{(i)})$ *as* $n\to\infty$ *for* $i=1,\dots,p$.

(G2) Asymptotic normality: $\sqrt{\dfrac{n^w}{m_n^d}}\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star) \xrightarrow{d} \mathcal{N}(\mathbf{0},\Pi)$ *as* $n\to\infty$.

(G3) *(i) Identifiability condition: For all* $\epsilon>0$ *there exists some* $\delta>0$ *such that*
$\inf\Big\{\sum_{i=1}^{p}(\rho_{AB,\boldsymbol{\theta}_1}(\boldsymbol{h}^{(i)})-\rho_{AB,\boldsymbol{\theta}_2}(\boldsymbol{h}^{(i)}))^2 : \boldsymbol{\theta}^{(1)},\boldsymbol{\theta}^{(2)}\in\Theta, \|\boldsymbol{\theta}^{(1)}-\boldsymbol{\theta}^{(2)}\|\geq\epsilon\Big\}>\delta.$

(ii) $\sum_{i=1}^{p}(\rho_{AB,\boldsymbol{\theta}_1}(\boldsymbol{h}^{(i)})-\rho_{AB,\boldsymbol{\theta}_2}(\boldsymbol{h}^{(i)}))^2>0, \quad \boldsymbol{\theta}^{(1)}\neq\boldsymbol{\theta}^{(2)}\in\Theta.$

Note that (i) implies (ii).

(G4) Smoothness condition 1: For all $i=1,\dots,p$:

(i) $\sup_{\boldsymbol{\theta}\in\Theta}\{\rho_{AB,\boldsymbol{\theta}}(\boldsymbol{h}^{(i)})\}<\infty.$

(ii) $\rho_{AB,\boldsymbol{\theta}}(\boldsymbol{h}^{(i)})$ *has continuous partial derivatives of order* $z_1\geq 0$ *w.r.t.* $\boldsymbol{\theta}$, *where* $z_1=0$ *corresponds to* $\rho_{AB,\boldsymbol{\theta}}(\boldsymbol{h}^{(i)})$ *being continuous in* $\boldsymbol{\theta}$.

(G5) Smoothness condition 2:

(i) $\sup_{\boldsymbol{\theta}\in\Theta}\{\|V(\boldsymbol{\theta})\|_M+\|V(\boldsymbol{\theta})^{-1}\|_M\}<\infty$, *where* $\|\cdot\|_M$ *is some arbitrary matrix norm.*

(ii) The matrix valued function $V(\boldsymbol{\theta})$ *has continuous derivatives of order* $z_2\geq 0$ *w.r.t.* $\boldsymbol{\theta}$, *where* $z_2=0$ *corresponds to* $V(\boldsymbol{\theta})$ *being continuous in* $\boldsymbol{\theta}$.

(G6) Rank condition: For $\boldsymbol{\theta}=(\theta_1,\dots,\theta_k)\in\Theta\subset\mathbb{R}^k$ *we set*

- $\rho^{(\ell)}_{AB,\boldsymbol{\theta}}(\boldsymbol{h}^{(i)}) := \frac{\partial}{\partial\theta_\ell}\rho_{AB,\boldsymbol{\theta}}(\boldsymbol{h}^{(i)})$ *for* $1\leq i\leq p, 1\leq\ell\leq k$.
- $\boldsymbol{\rho}^{(\ell)}_{AB}(\boldsymbol{\theta}) := (\rho^{(\ell)}_{AB,\boldsymbol{\theta}}(\boldsymbol{h}^{(i)}) : i=1,\dots,p)^\intercal$ *for* $1\leq\ell\leq k$.
- *Denote by* $\mathrm{P}_{AB}(\boldsymbol{\theta})$ *the Jacobian matrix of* $(-\rho_{AB,\boldsymbol{\theta}}(\boldsymbol{h}^{(1)}),\dots,-\rho_{AB,\boldsymbol{\theta}}(\boldsymbol{h}^{(p)}))^\intercal$; *i.e.,*

$$\mathrm{P}_{AB}(\boldsymbol{\theta}) = (-\boldsymbol{\rho}^{(1)}_{AB}(\boldsymbol{\theta}),\dots,-\boldsymbol{\rho}^{(k)}_{AB}(\boldsymbol{\theta}))\in\mathbb{R}^{p\times k}. \tag{4.33}$$

The Jacobi matrix has full rank: $\mathrm{rank}(\mathrm{P}_{AB}(\boldsymbol{\theta}^\star))=k$. □

The proof of the next theorem can be found in Appendix C.4.

Theorem 4.16 (Consistency and asymptotic normality of the GLSE)**.** *Assume the situation of Definition 4.14. If Assumption 4.15(G1) and (G3) hold as well as (G4) and (G5) for* $z_1 = z_2 = 0$*, respectively, then the GLSE is consistent; i.e.,*

$$\widehat{\boldsymbol{\theta}}_{n,V} \xrightarrow{P} \boldsymbol{\theta}^\star, \quad n \to \infty. \tag{4.34}$$

If Assumption 4.15(G2) and (G3) hold as well as (G4) and (G5) for $z_1 = z_2 = 1$*, respectively, and the rank condition (G6) holds, then the GLSE is asymptotically normal; i.e.,*

$$\sqrt{\frac{n^w}{m_n^d}}(\widehat{\boldsymbol{\theta}}_{n,V} - \boldsymbol{\theta}^\star) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Pi_V), \quad n \to \infty, \tag{4.35}$$

with asymptotic covariance matrix

$$\Pi_V = B(\boldsymbol{\theta}^\star)\mathrm{P}_{AB}(\boldsymbol{\theta}^\star)^\intercal [V(\boldsymbol{\theta}^\star) + V(\boldsymbol{\theta}^\star)^\intercal]\, \Pi \,[V(\boldsymbol{\theta}^\star) + V(\boldsymbol{\theta}^\star)^\intercal]\mathrm{P}_{AB}(\boldsymbol{\theta}^\star)B(\boldsymbol{\theta}^\star),$$

where $B(\boldsymbol{\theta}^\star) := \big(\mathrm{P}_{AB}(\boldsymbol{\theta}^\star)^\intercal [V(\boldsymbol{\theta}^\star) + V(\boldsymbol{\theta}^\star)^\intercal]\mathrm{P}_{AB}(\boldsymbol{\theta}^\star)\big)^{-1}$ *and* Π *is the asymptotic covariance matrix in Eq.* (4.30)*.*

Remark 4.17. The quality of the GLSE depends on the matrix $V(\boldsymbol{\theta})$. Simple choices for the matrix $V(\boldsymbol{\theta})$ in (4.32) are the identity matrix, leading to the ordinary least squares estimator, or some general weight matrix, leading to weighted least squares estimators.

An asymptotically optimal matrix $V(\boldsymbol{\theta})$ can be obtained as follows. Let $\Pi = \Pi(\boldsymbol{\theta}^\star)$ be the asymptotic covariance matrix of the empirical extremogram in Eq. (4.30). Assume that $\Pi(\boldsymbol{\theta}^\star)$ can be extended to a matrix function $\Pi(\boldsymbol{\theta})$ on the whole parameter space $\boldsymbol{\Theta}$ and that $\Pi(\boldsymbol{\theta})$ is invertible for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Assume also that $V(\boldsymbol{\theta}) = \Pi^{-1}(\boldsymbol{\theta})$ satisfies the Assumption 4.15(G5) for $z_2 = 1$. Then, as pointed out in Lahiri et al. [48], Theorem 4.1, for spatial variogram estimators and in Einmahl et al. [31], Corollary 2.3, for extreme parameter estimation based on iid random vector observations, the resulting asymptotic covariance matrix $\Pi_V = \Pi_V(\boldsymbol{\theta}^\star)$ of the GLSE in (4.35) is asymptotically optimal among all valid matrices $V' = V'(\boldsymbol{\theta})$. This means that Π_V is minimal in the sense that for all valid matrices V', the difference $\Pi_{V'} - \Pi_V$ is positive semidefinite.

4.5 Estimation of Brown-Resnick space-time processes

4.5.1 Brown-Resnick processes

We consider a *strictly stationary Brown-Resnick process* with representation

$$\eta(\boldsymbol{s}) = \bigvee_{j=1}^{\infty} \left\{\xi_j \, e^{W_j(\boldsymbol{s}) - \delta(\boldsymbol{s})}\right\}, \quad \boldsymbol{s} \in \mathbb{R}^d, \tag{4.36}$$

where $\{\xi_j : j \in \mathbb{N}\}$ are points of a Poisson process on $[0, \infty)$ with intensity $\xi^{-2} d\xi$, the *dependence function* δ is nonnegative and conditionally negative definite, and $\{W_j(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ are independent replicates of a Gaussian process $\{W(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ with stationary increments, $W(\mathbf{0}) = 0$,

$\mathbb{E}[W(\boldsymbol{s})] = 0$ and covariance function

$$\mathbb{C}ov[W(\boldsymbol{s}^{(1)}), W(\boldsymbol{s}^{(2)})] = \delta(\boldsymbol{s}^{(1)}) + \delta(\boldsymbol{s}^{(2)}) - \delta(\boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)}).$$

Representation (4.36) goes back to de Haan [24] and Giné, Hahn, and Vatan [39]. The univariate margins of the process η follow standard unit Fréchet distributions. Non-stationary Brown-Resnick models have recently been discussed and fitted to data by Engelke et al. [34] and Asadi et al. [2].

There are various quantities to describe the dependence in (4.36), where explicit expressions can be derived:

- In geostatistics, the dependence function δ is termed the *semivariogram* of the process $\{W(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ based on the fact that for $\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)} \in \mathbb{R}^d$,

$$\mathbb{V}ar[W(\boldsymbol{s}^{(1)}) - W(\boldsymbol{s}^{(2)})] = 2\delta(\boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)}).$$

- For $\boldsymbol{h} \in \mathbb{R}^d$, the *tail dependence coefficient* is given by (see e.g. Davis, Klüppelberg, and Steinkohl [19], Section 3)

$$\rho_{(1,\infty)(1,\infty)}(\boldsymbol{h}) = \lim_{n\to\infty} \mathbb{P}\Big(\eta(\boldsymbol{h}) > n \,\Big|\, \eta(\boldsymbol{0}) > n\Big) = 2\Big(1 - \Phi\Big(\sqrt{\frac{\delta(\boldsymbol{h})}{2}}\Big)\Big), \tag{4.37}$$

  where Φ denotes the standard normal distribution function.

- For $D = \{\boldsymbol{s}^{(1)}, \ldots, \boldsymbol{s}^{(|D|)}\}$ and $\boldsymbol{y} = (y_1, \ldots, y_{|D|}) > \boldsymbol{0}$ the finite-dimensional margins are given by

$$\mathbb{P}(\eta(\boldsymbol{s}^{(1)}) \le y_1, \eta(\boldsymbol{s}^{(2)}) \le y_2, \cdots, \eta(\boldsymbol{s}^{(|D|)}) \le y_{|D|}) = \exp\{-V_D(\boldsymbol{y})\}. \tag{4.38}$$

  Here V_D denotes the *exponent measure* (cf. Beirlant et al. [3], Section 8.2.2), which is homogeneous of order -1 and depends solely on the dependence function δ. For $D = \{\boldsymbol{s}, \boldsymbol{s}+\boldsymbol{h}\}$ where $\boldsymbol{s} \in \mathbb{R}^d$ and $\boldsymbol{h} \in \mathbb{R}^d$ is some fixed lag vector, we get (cf. Davis et al. [19], Section 3)

$$V_2(y_1, y_2) = V_2(\boldsymbol{h}; y_1, y_2) = V_D(y_1, y_2) = \frac{1}{y_1}\widetilde{\Phi}\Big(\frac{y_2}{y_1}\Big) + \frac{1}{y_2}\widetilde{\Phi}\Big(\frac{y_1}{y_2}\Big), \quad y_1, y_2 > 0, \tag{4.39}$$

  with

$$\widetilde{\Phi}\Big(\frac{x}{y}\Big) = \widetilde{\Phi}\Big(\boldsymbol{h}; \frac{x}{y}\Big) := \Phi\Big(\frac{\log(x/y)}{\sqrt{2\delta(\boldsymbol{h})}} + \sqrt{\frac{\delta(\boldsymbol{h})}{2}}\Big), \quad x, y > 0. \tag{4.40}$$

- For $\boldsymbol{h} \in \mathbb{R}^d$ and sets $A = (a_1, a_2)$ and $B = (b_1, b_2)$ with $0 < a_1 < a_2 \le \infty$ and $0 < b_1 < b_2 \le \infty$, the extremogram (4.3) is given by (see Eq. (A.1))

$$\rho_{AB}(\boldsymbol{h}) = \frac{a_1 a_2}{a_2 - a_1}\Big(- V_2(a_2, b_2) + V_2(a_2, b_1) + V_2(a_1, b_2) - V_2(a_1, b_1)\Big) \tag{4.41}$$

for V_2 as in (4.39). For $A = (a, \infty)$ and $B = (b, \infty)$ we get formula (31) of Cho et al. [15]:

$$\rho_{AB}(\boldsymbol{h}) = a\Big\{\frac{1}{a}\Big(1 - \widetilde{\Phi}\Big(\frac{b}{a}\Big)\Big) + \frac{1}{b}\Big(1 - \widetilde{\Phi}\Big(\frac{a}{b}\Big)\Big)\Big\}. \tag{4.42}$$

- The *extremal coefficient* ξ_D (see [3], Section 8.2.7) for any finite set $D \subset \mathbb{R}^d$ is defined as

$$\mathbb{P}(\eta(\boldsymbol{s}^{(1)}) \leq y, \eta(\boldsymbol{s}^{(2)}) \leq y, \cdots, \eta(\boldsymbol{s}^{(|D|)}) \leq y) = \exp\{-\xi_D/y\}, \quad y > 0;$$

i.e., $\xi_D = V_D(1, \ldots, 1)$. If $|D| = 2$ and $\boldsymbol{h} = \boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)}$, then

$$\xi_D = 2 - \rho_{(1,\infty)(1,\infty)}(\boldsymbol{h}) = 2\Phi\Big(\sqrt{\frac{\delta(\boldsymbol{h})}{2}}\Big), \tag{4.43}$$

where the first identity holds in general (cf. Beirlant et al. [3], Section 9.5.1), and the last one by (4.37).

Our aim is to fit a parametric extremogram model of a Brown-Resnick process (4.36) based on observations given in $\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n$ as in (4.4). This approach is semiparametric in the sense that we first compute (possibly bias corrected) empirical estimates (4.25) of the extremogram $\rho_{AB}(\boldsymbol{h})$ for different $\boldsymbol{h} \in \mathcal{H}$, and fit a parametric model $\rho_{AB,\boldsymbol{\theta}}(\boldsymbol{h})$ by GLSE to the empirical extremogram. For sets $A = B = (a, \infty)$ with $a > 0$, this yields an estimator of the dependence function, since by (4.40) and (4.42) there is a one-to-one relation between extremogram and dependence function.

4.5.2 Asymptotic properties of the empirical extremogram of a Brown-Resnick process

Let $\{\eta(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ be a strictly stationary Brown-Resnick process as in (4.36) with some valid (i.e., nonnegative and conditionally negative definite) dependence function δ. Before investigating the asymptotic properties of the GLSE, we state sufficient conditions for δ so that the regularity conditions of Theorem 4.8 are satisfied.

Theorem 4.18. *Let $\{\eta(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ be a strictly stationary Brown-Resnick process as in* (4.36)*, observed on $\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n$ as in* (4.4)*. Let $\mathcal{H} = \{\boldsymbol{h}^{(1)}, \ldots, \boldsymbol{h}^{(p)}\} \subset \mathbb{Z}^d \cap B(\boldsymbol{0}, \gamma)$ for some $\gamma > 0$ be a set of observed lag vectors. Assume sequences*

$$m_n, r_n \to \infty, \quad m_n^d/n^w \to 0, \quad r_n^w/m_n^d \to 0, \quad m_n^{2d} r_n^{2w}/n^w \to 0, \quad n \to \infty. \tag{4.44}$$

Writing $\boldsymbol{u} = (\boldsymbol{u}_{\mathcal{F}}, \boldsymbol{u}_{\mathcal{I}}) \in \mathbb{R}^q \times \mathbb{R}^w$ according to the fixed and increasing domains, assume that the dependence function δ satisfies for arbitrary fixed finite set $L \subset \mathbb{Z}^q$:

(A) $m_n^d \sum_{z > r_n} z^{w-1} \exp\Big\{-\frac{1}{4} \inf_{\boldsymbol{u} \in L \times \mathbb{Z}^w : \|\boldsymbol{u}_{\mathcal{I}}\| \geq z} \delta(\boldsymbol{u})\Big\} \to 0$ *as* $n \to \infty$.

(B) $m_n^{d/2} n^{(3w)/2} \exp\Big\{-\frac{1}{4} \inf_{\boldsymbol{u} \in L \times \mathbb{Z}^w : \|\boldsymbol{u}_{\mathcal{I}}\| > r_n} \delta(\boldsymbol{u})\Big\} \to 0$ *as* $n \to \infty$.

Then conditions (M1)-(M4) of Theorem 4.8 are satisfied, and the empirical extremogram $\widehat{\rho}_{AB,m_n}$ *defined in* (4.9) *sampled at lags in* $\mathcal{H}$ *and centred by the pre-asymptotic extremogram* ρ_{AB,m_n} *given in* (4.10)*, is asymptotically normal; i.e.,*

$$\sqrt{\frac{n^w}{m_n^d}}\Big[\widehat{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)}) - \rho_{AB,m_n}(\boldsymbol{h}^{(i)})\Big]_{i=1,\ldots,p} \xrightarrow{d} \mathcal{N}(\boldsymbol{0},\Pi), \quad n\to\infty, \tag{4.45}$$

where the covariance matrix Π *is specified in Theorem 4.8.*

Proof. First note that, since all finite-dimensional distributions are max-stable distributions with standard unit Fréchet margins, they are multivariate regularly varying. We first show (M3). Let $\epsilon > 0$ and fix $\boldsymbol{\ell}_{\mathcal{F}} \in \mathbb{R}^q$. For $\gamma > 0$ define the set

$$L_\gamma(\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}}) := \{\boldsymbol{s}_1 - \boldsymbol{s}_2 : \boldsymbol{s}_1 \in B(\boldsymbol{0},\gamma), \boldsymbol{s}_2 \in B((\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}}),\gamma)\}.$$

Note that, writing $\boldsymbol{s}_1 = (\boldsymbol{f}_1,\boldsymbol{i}_1)$ and $\boldsymbol{s}_2 = (\boldsymbol{f}_2,\boldsymbol{i}_2) \in \mathbb{R}^q \times \mathbb{R}^w$ according to the fixed and increasing domains as before, it can be decomposed into $L_\gamma(\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}}) = L_\gamma^{(1)} \times L_\gamma^{(2)}(\boldsymbol{\ell}_{\mathcal{I}})$ where $L_\gamma^{(1)} := \{\boldsymbol{f}_1 - \boldsymbol{f}_2 : \boldsymbol{s}_1 \in B((\boldsymbol{0},\boldsymbol{0}),\gamma), \boldsymbol{s}_2 \in B((\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{0}),\gamma)\}$, which is independent of $\boldsymbol{\ell}_{\mathcal{I}}$, and $L_\gamma^{(2)} := \{\boldsymbol{i}_1 - \boldsymbol{i}_2 : \boldsymbol{s}_1 \in B((\boldsymbol{0},\boldsymbol{0}),\gamma), \boldsymbol{s}_2 \in B((\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}}),\gamma)\}$. Then, recalling that $a_m = m_n^d$, and using a second order Taylor expansion as in the proof of Theorem 3.7, we have as $n\to\infty$,

$$\begin{aligned}
&\mathbb{P}\big(\max_{\boldsymbol{s}\in B(\boldsymbol{0},\gamma)} \eta(\boldsymbol{s}) > \epsilon a_m, \max_{\boldsymbol{s}'\in B((\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}}),\gamma)} \eta(\boldsymbol{s}') > \epsilon a_m\big)\\
&\le \sum_{\boldsymbol{s}\in B(\boldsymbol{0},\gamma)} \sum_{\boldsymbol{s}'\in B((\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}}),\gamma)} \mathbb{P}(\eta(\boldsymbol{s}) > \epsilon m_n^d, \eta(\boldsymbol{s}') > \epsilon m_n^d)\\
&= \sum_{\boldsymbol{s}\in B(\boldsymbol{0},\gamma)} \sum_{\boldsymbol{s}'\in B((\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}}),\gamma)} \Big(1 - 2\exp\Big\{-\frac{1}{\epsilon m_n^d}\Big\} + \exp\Big\{-\frac{2}{\epsilon m_n^d}\Phi\Big(\sqrt{\frac{\delta(\boldsymbol{s}-\boldsymbol{s}')}{2}}\Big)\Big\}\Big)\\
&\le \frac{2|B(\boldsymbol{0},\gamma)|^2}{\epsilon m_n^d}\Big(1 - \Phi\Big(\Big(\frac{1}{2}\inf_{\boldsymbol{u}\in L_\gamma(\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}})} \delta(\boldsymbol{u})\Big)^{1/2}\Big)\Big) + \mathcal{O}\Big(\frac{1}{m_n^{2d}}\Big).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\limsup_{n\to\infty} \sum_{\substack{\boldsymbol{\ell}_{\mathcal{I}}\in\mathbb{Z}^w\\ k<\|\boldsymbol{\ell}_{\mathcal{I}}\|\le r_n}} m_n^d \mathbb{P}\big(\max_{\boldsymbol{s}\in B(\boldsymbol{0},\gamma)} \eta(\boldsymbol{s}) > \epsilon a_m, \max_{\boldsymbol{s}'\in B((\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}}),\gamma)} \eta(\boldsymbol{s}') > \epsilon a_m\big)\\
&\le 2|B(\boldsymbol{0},\gamma)|^2 \limsup_{n\to\infty} \sum_{\substack{\boldsymbol{\ell}_{\mathcal{I}}\in\mathbb{Z}^w\\ k<\|\boldsymbol{\ell}_{\mathcal{I}}\|\le r_n}} \Big\{\frac{1}{\epsilon}\Big(1 - \Phi\Big(\Big(\frac{1}{2}\inf_{\boldsymbol{u}\in L_\gamma(\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}})} \delta(\boldsymbol{u})\Big)^{1/2}\Big)\Big) + \mathcal{O}\Big(\frac{1}{m_n^d}\Big)\Big\}.
\end{aligned}$$

Since the number of grid points in $\mathbb{Z}^w$ with norm $z = \|\boldsymbol{\ell}_{\mathcal{I}}\|$ is of order $\mathcal{O}(z^{w-1})$, there exists a positive constant C such that the right hand side can be bounded from above by

$$\begin{aligned}
&2C|B(\boldsymbol{0},\gamma)|^2 \limsup_{n\to\infty} \sum_{k<z\le r_n} \Big\{\frac{z^{w-1}}{\epsilon}\Big(1 - \Phi\Big(\Big(\frac{1}{2}\inf_{\boldsymbol{u}\in L_\gamma(\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}}):\boldsymbol{\ell}_{\mathcal{I}}\in\mathbb{Z}^w,\|\boldsymbol{\ell}_{\mathcal{I}}\|=z} \delta(\boldsymbol{u})\Big)^{1/2}\Big)\Big) + \mathcal{O}\Big(\frac{z^{w-1}}{m_n^d}\Big)\Big\}\\
&\le \frac{2C|B(\boldsymbol{0},\gamma)|^2}{\epsilon} \limsup_{n\to\infty} \sum_{k<z<\infty} \Big\{z^{w-1}\Big(\exp\Big\{-\frac{1}{4}\inf_{\boldsymbol{u}\in L_\gamma(\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}}):\boldsymbol{\ell}_{\mathcal{I}}\in\mathbb{Z}^w,\|\boldsymbol{\ell}_{\mathcal{I}}\|=z} \delta(\boldsymbol{u})\Big\}\Big)\Big\} + \mathcal{O}\Big(\frac{r_n^w}{m_n^d}\Big)
\end{aligned}$$

$$\leq \frac{2C|B(\mathbf{0},\gamma)|^2}{\epsilon} \limsup_{n\to\infty} \sum_{k<z<\infty} \Big\{ z^{w-1} \Big(\exp\Big\{ -\frac{1}{4} \inf_{\boldsymbol{u}\in L_\gamma^{(1)}\times\mathbb{Z}^w:\|\boldsymbol{u}_\mathcal{I}\|\geq z-\gamma} \delta(\boldsymbol{u}) \Big\} \Big) \Big\} + \mathcal{O}\Big(\frac{r_n^w}{m_n^d}\Big),$$

where we have used in the second last step that $1-\Phi(x)\leq \exp\{-x^2/2\}$ for $x>0$ and in the last step the decomposition $L_\gamma(\boldsymbol{\ell}_\mathcal{F},\boldsymbol{\ell}_\mathcal{I}) = L_\gamma^{(1)} \times L_\gamma^{(2)}(\boldsymbol{\ell}_\mathcal{I})$. By condition (A), since we can neglect the constant γ, we have

$$\lim_{k\to\infty} \sum_{k<z<\infty} z^{w-1} \exp\Big\{ -\frac{1}{4} \inf_{\boldsymbol{u}\in L_\gamma^{(1)}\times\mathbb{Z}^w:\|\boldsymbol{u}_\mathcal{I}\|\geq z-\gamma} \delta(\boldsymbol{u}) \Big\} = 0.$$

Together with $r_n^w = o(m_n^d)$ as $n\to\infty$, this implies that

$$\lim_{k\to\infty} \limsup_{n\to\infty} \sum_{k<z\leq r_n} \Big\{ z^{w-1} \Big(\exp\Big\{ -\frac{1}{4} \inf_{\boldsymbol{u}\in L_\gamma^{(1)}\times\mathbb{Z}^w:\|\boldsymbol{u}_\mathcal{I}\|\geq z-\gamma} \delta(\boldsymbol{u}) \Big\} \Big) \Big\} + \mathcal{O}\Big(\frac{r_n^w}{m_n^d}\Big) = 0.$$

Next we prove (M1) and (M4i)-(M4iii). To this end we bound the α-mixing coefficients $\alpha_{k_1,k_2}(\cdot)$ for $k_1,k_2\in\mathbb{N}$ of $\{\eta(\boldsymbol{s}) : \boldsymbol{s}\in\mathbb{R}^d\}$ with respect to $\mathbb{R}^w$, which are defined in (C.2). Observe that $d(\Lambda_1,\Lambda_2)$ for sets $\Lambda_i\subset\mathbb{Z}^w$ as in Definition C.1 can only get large within the increasing domain. Define the set

$$L_\mathcal{F} := \{\boldsymbol{s}_1-\boldsymbol{s}_2 : \boldsymbol{s}_1,\boldsymbol{s}_2\in\mathcal{F}\}.$$

We use Eq. (4.43), as well as Dombry and Eyi-Minko [27], Eq. (3) and Corollary 2.2 to obtain

$$\begin{aligned}
\alpha_{k_1,k_2}(z) &\leq 2 \sup_{d(\Lambda_1,\Lambda_2)\geq z} \sum_{\boldsymbol{s}_1\in\mathcal{F}\times\Lambda_1} \sum_{\boldsymbol{s}_2\in\mathcal{F}\times\Lambda_2} \rho_{(1,\infty)(1,\infty)}(\boldsymbol{s}_1-\boldsymbol{s}_2) \\
&\leq 2k_1k_2|\mathcal{F}|^2 \sup_{\boldsymbol{u}\in L_\mathcal{F}\times\mathbb{Z}^w:\|\boldsymbol{u}_\mathcal{I}\|\geq z} \rho_{(1,\infty)(1,\infty)}(\boldsymbol{u}) \\
&= 4k_1k_2|\mathcal{F}|^2 \Big(1-\Phi\Big(\Big(\frac{1}{2}\inf_{\boldsymbol{u}\in L_\mathcal{F}\times\mathbb{Z}^w:\|\boldsymbol{u}_\mathcal{I}\|\geq z} \delta(\boldsymbol{u})\Big)^{\frac{1}{2}}\Big)\Big) \\
&\leq 4k_1k_2|\mathcal{F}|^2 \exp\Big\{ -\frac{1}{4} \inf_{\boldsymbol{u}\in L_\mathcal{F}\times\mathbb{Z}^w:\|\boldsymbol{u}_\mathcal{I}\|\geq z} \delta(\boldsymbol{u}) \Big\}. \qquad (4.46)
\end{aligned}$$

By condition (A) we have $\alpha_{k_1,k_2}(z)\to 0$, since necessarily $\inf_{\boldsymbol{u}\in L_\mathcal{F}\times\mathbb{Z}^w:\|\boldsymbol{u}_\mathcal{I}\|\geq z} \delta(\boldsymbol{u}) \to\infty$ as $z\to\infty$ and, therefore, the process $\{\eta(\boldsymbol{s}) : \boldsymbol{s}\in\mathbb{R}^d\}$ is α-mixing; i.e., (M1) holds. We continue by estimating

$$\begin{aligned}
&m_n^d \sum_{\boldsymbol{\ell}\in\mathbb{Z}^w:\|\boldsymbol{\ell}\|>r_n} \alpha_{1,1}(\|\boldsymbol{\ell}\|) \;\leq\; Cm_n^d \sum_{z>r_n} z^{w-1}\alpha_{1,1}(z) \\
&\leq 4C|\mathcal{F}|^2 m_n^d \sum_{z>r_n} z^{w-1} \exp\Big\{ -\frac{1}{4} \inf_{\boldsymbol{u}\in L_\mathcal{F}\times\mathbb{Z}^w:\|\boldsymbol{u}_\mathcal{I}\|\geq z} \delta(\boldsymbol{u}) \Big\} \to 0, \quad n\to\infty,
\end{aligned}$$

by condition (A). This shows (M4i). Similarly, it can be shown that (M4ii) holds, if (A) is satisfied. Finally, we show (M4iii). Using Eq. (4.46), we find

$$m_n^{d/2} n^{w/2} \alpha_{1,n^w}(r_n) \leq 4m_n^{d/2} n^{(3w)/2} \exp\Big\{ -\frac{1}{4} \inf_{\boldsymbol{u}\in L_\mathcal{F}\times\mathbb{Z}^w:\|\boldsymbol{u}_\mathcal{I}\|\geq r_n} \delta(\boldsymbol{u}) \Big\} \to 0$$

as $n \to \infty$ because of condition (B). $\square$

The following is an immediate corollary of Theorem 4.18.

Corollary 4.19. *Assume the situation as in Theorem 4.18. Suppose that the dependence function δ satisfies for some positive constants C and α and for an arbitrary norm $\|\cdot\|$ on $\mathbb{R}^w$ (that possibly differs from that considered in Theorem 4.18),*

$$\delta(\boldsymbol{u}) \geq C\|\boldsymbol{u}_{\mathcal{I}}\|^{\alpha} \tag{4.47}$$

for every $\boldsymbol{u} = (\boldsymbol{u}_{\mathcal{F}}, \boldsymbol{u}_{\mathcal{I}}) \in L \times \mathbb{Z}^w$, where $L \subset \mathbb{Z}^q$ is arbitrary, but fixed. In particular, $\delta(\boldsymbol{u}) \to \infty$ if $\|\boldsymbol{u}_{\mathcal{I}}\| \to \infty$. With $m_n = n^{\beta_1}$ and $r_n = n^{\beta_2}$ with $\beta_1 \in (0, w/(2d))$ and $\beta_2 \in \min\{\beta_1 d/w; 1/2 - \beta_1 d/w\}$, the conditions of Theorem 4.18 are satisfied for $\{\eta(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ and we conclude

$$n^{(w-d\beta_1)/2}\Big[\widehat{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)}) - \rho_{AB,m_n}(\boldsymbol{h}^{(i)})\Big]_{i=1,\dots,p} \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \Pi), \quad n \to \infty. \tag{4.48}$$

Proof. Due to equivalence of norms on $\mathbb{R}^w$ we will make no difference between the norm in (4.47) and the one used in Theorem 4.18. Clearly the sequences m_n and r_n satisfy the requirements $m_n, r_n \to \infty$, $m_n^d/n^w \to 0$, $r_n^w/m_n^d \to 0$ and $m_n^{2d} r_n^{2w}/n^w \to 0$ as $n \to \infty$. We have for $z > 0$,

$$\exp\Big\{-\frac{1}{4}\inf_{\boldsymbol{u}\in L\times\mathbb{Z}^w:\|\boldsymbol{u}_{\mathcal{I}}\|>z}\delta(\boldsymbol{u})\Big\} \leq \exp\Big\{-\frac{1}{4}\inf_{\boldsymbol{u}\in L\times\mathbb{Z}^w:\|\boldsymbol{u}_{\mathcal{I}}\|>z} C\|\boldsymbol{u}_{\mathcal{I}}\|^{\alpha}\Big\} \leq \exp\Big\{-\frac{Cz^{\alpha}}{4}\Big\}.$$

Condition (B) of Theorem 4.18 is satisfied since

$$\begin{aligned} n^{(\beta_1 d)/2} n^{(3w)/2} \exp\Big\{-\frac{Cr_n^{\alpha}}{4}\Big\} &= n^{(\beta_1 d)/2} n^{(3w)/2} \exp\Big\{-\frac{Cn^{\beta_2\alpha}}{4}\Big\} \\ &= \exp\Big\{-\frac{Cn^{\beta_2\alpha}}{4} + \frac{\beta_1 d + 3w}{2}\log(n)\Big\} \to 0 \end{aligned}$$

as $n \to \infty$. Condition (A) holds since by Lemma B.3, there is a positive constant K such that for sufficiently large n such that the sequence $z^{w-1}\exp\{-Cz^{\alpha}/4\}$ is monotonously decreasing for $z \geq r_n$,

$$\begin{aligned} m_n^d \sum_{z>r_n} z^{w-1}\exp\Big\{-\frac{Cz^{\alpha}}{4}\Big\} &\leq K m_n^d r_n^w \exp\Big\{-\frac{Cr_n^{\alpha}}{4}\Big\} \\ &= K\exp\Big\{-\frac{Cn^{\beta_2\alpha}}{4} + (\beta_1 d + \beta_2 w)\log(n)\Big\} \to 0. \end{aligned}$$

$\square$

With the particular choice of sequences $m_n = n^{\beta_1}$ and $r_n = n^{\beta_2}$ given in Corollary 4.19, we are in the setting of Remark 4.13. Hence, in addition to the CLT (4.48), we obtain the CLT (4.28) of the empirical extremogram centred by the true one and the CLT (4.27) corresponding to the bias corrected estimator.

4.5.3 Space-time Brown-Resnick processes: different models for the extremogram

We explore the semiparametric estimation for strictly stationary Brown-Resnick processes in their space-time form $\{\eta(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^{d-1}, t \in [0,\infty)\}$. For three classes of parametric models for the dependence function $\delta_{\boldsymbol{\theta}}$ we prove that the GLSE is consistent and asymptotically normal.

Note that by Eq. (4.42) every model $\{\delta_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ for the dependence function yields a model $\{\rho_{AB,\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ for its space-time extremogram. Moreover, the extremogram (4.42) is always of the same form, and only $\widetilde{\Phi}$ in (4.40) changes with the model. We consider three different model classes, which together cover a large field of environmental applications such as the modelling of extreme precipitation (cf. [11], [14], [19], [23] and Chapters 3 and 5), extreme wind speed (cf. [34]) or extremes on river networks (cf. [2]), provided they are valid (i.e., nonnegative and conditionally negative definite) dependence functions in the considered metric.

(I) Fractional space-time model.
Davis et al. [19] introduce the spatially isotropic model

$$\delta_{\boldsymbol{\theta}}(\boldsymbol{h},u) = C_1\|\boldsymbol{h}\|^{\alpha_1} + C_2|u|^{\alpha_2}, \quad (\boldsymbol{h},u) \in \mathbb{R}^d, \tag{4.49}$$

with parameter vector

$$\boldsymbol{\theta} \in \{(C_1, C_2, \alpha_1, \alpha_2) : C_1, C_2 \in (0,\infty), \alpha_1, \alpha_2 \in (0,2]\}.$$

The isotropy assumption, where (4.49) depends on the norm of the spatial lag $\boldsymbol{h}$, can be relaxed in a natural way by introducing *geometric anisotropy*. We only discuss the case $d-1=2$, but the approach is easily transferable to higher dimensions. Let $\varphi \in [0,\pi/2)$ be a rotation angle and $R = R(\varphi)$ a rotation matrix, and T a dilution matrix with $c > 0$; more precisely,

$$R = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix}.$$

The geometrically anisotropic model is then given by

$$\widetilde{\delta}_{\widetilde{\boldsymbol{\theta}}}(\boldsymbol{h},u) = \delta_{\boldsymbol{\theta}}(A\boldsymbol{h},u), \quad (\boldsymbol{h},u) \in \mathbb{R}^d \tag{4.50}$$

where $A = TR$ is the transformation matrix. The parameter vector of the transformed model is

$$\widetilde{\boldsymbol{\theta}} \in \{(C_1, C_2, \alpha_1, \alpha_2, c, \varphi) : C_1, C_2 \in (0,\infty), \alpha_1, \alpha_2 \in (0,2], c > 0, \varphi \in [0,\pi/2)\}.$$

For more details about geometric anisotropy see [19], Section 4.2, Blanchet and Davison [4], Section 4.2, or Engelke et al. [34], Section 5.2.

(II) Spatial anisotropy along orthogonal spatial directions

In Chapter 5 we generalise the fractional isotropic model (4.49) to

$$\delta_{\boldsymbol{\theta}}(\boldsymbol{h}, u) = \sum_{j=1}^{d-1} C_j|h_j|^{\alpha_j} + C_d|u|^{\alpha_d}, \quad (\boldsymbol{h}, u) \in \mathbb{R}^d \tag{4.51}$$

with parameter vector

$$\boldsymbol{\theta} \in \{(C_j, \alpha_j, j = 1, \ldots, d) : C_j \in (0, \infty), \alpha_j \in (0, 2], j = 1, \ldots, d\}.$$

It is more flexible than the isotropic model (I) as it allows for different rates of decay of extreme dependence along the axes of a d-dimensional spatial grid. Arbitrary principal orthogonal directions can be introduced by a rotation matrix R as introduced for the isotropic model in (I), here described for the case $d-1=2$:

$$\widetilde{\delta}_{\widetilde{\boldsymbol{\theta}}}(\boldsymbol{h}, u) = C_1|h_1 \cos\varphi - h_2 \sin\varphi|^{\alpha_1} + C_2|h_1 \sin\varphi + h_2 \cos\varphi|^{\alpha_2} + C_3|u|^{\alpha_3}, \quad (\boldsymbol{h}, u) \in \mathbb{R}^3. \tag{4.52}$$

The new parameter vector is

$$\widetilde{\boldsymbol{\theta}} \in \{(C_1, C_2, C_3, \alpha_1, \alpha_2, \alpha_3, \varphi) : C_j \in (0, \infty), \alpha_j \in (0, 2], j = 1, 2, 3, \varphi \in [0, \pi/2)\}.$$

In Chapter 5 this model is applied to extreme precipitation in Florida and, according to a specifically developed goodness-of-fit method, performs extremely well.

(III) Time-shifted Brown-Resnick processes

With the goal to allow for some influence of the spatial dependence from previous values of the process we time-shift the Gaussian processes in the definition of the Brown-Resnick model (4.36). For $\boldsymbol{\tau} = (\tau_1, \tau_2) \in \mathbb{R}^{d-1}$ define

$$W^{(\boldsymbol{\tau})}(\boldsymbol{s}, t) := W(\boldsymbol{s} - t\boldsymbol{\tau}, t).$$

Then $\{W^{(\boldsymbol{\tau})}(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in \mathbb{R}^{d-1}, t \in [0, \infty)\}$ is also a centred Gaussian process starting in 0 with stationary increments: for $(\boldsymbol{s}^{(1)}, t^{(1)}), (\boldsymbol{s}^{(2)}, t^{(2)}) \in \mathbb{R}^{d-1} \times [0, \infty)$, because of the stationary increments of $\{W(\boldsymbol{s}, t)\}$, where $\stackrel{d}{=}$ stands for equality in distribution,

$$\begin{aligned} W^{(\boldsymbol{\tau})}(\boldsymbol{s}^{(1)}, t^{(1)}) - W^{(\boldsymbol{\tau})}(\boldsymbol{s}^{(1)}, t^{(1)}) &\stackrel{d}{=} W(\boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)} - (t^{(1)} - t^{(2)})\boldsymbol{\tau}, t^{(1)} - t^{(2)}) \\ &= W^{(\boldsymbol{\tau})}(\boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)}, t^{(1)} - t^{(2)}), \end{aligned}$$

The corresponding time-shifted dependence function is given by

$$\delta^{(\boldsymbol{\tau})}(\boldsymbol{s}, t) := \frac{1}{2}\mathbb{V}ar[W^{(\boldsymbol{\tau})}(\boldsymbol{s}, t) - W^{(\boldsymbol{\tau})}(\boldsymbol{0}, 0)] = \frac{1}{2}\mathbb{V}ar[W(\boldsymbol{s} - t\boldsymbol{\tau}, t) - W(\boldsymbol{0}, 0)] = \delta(\boldsymbol{s} - t\tau, t),$$

which yields the covariance function

$$\mathbb{C}ov[W^{(\boldsymbol{\tau})}(\boldsymbol{s}^{(1)}, t^{(1)}), W^{(\boldsymbol{\tau})}(\boldsymbol{s}^{(2)}, t^{(2)})] =$$

$$\delta^{(\boldsymbol{\tau})}(\boldsymbol{s}^{(1)}, t^{(1)}) + \delta^{(\boldsymbol{\tau})}(\boldsymbol{s}^{(2)}, t^{(2)}) - \delta^{(\boldsymbol{\tau})}(\boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)}, t^{(1)} - t^{(2)}).$$

By Theorem 10 of Kabluchko et al. [47] the process

$$\eta^{(\boldsymbol{\tau})}(\boldsymbol{s}, t) := \bigvee_{i=1}^{\infty} \xi_i \mathrm{e}^{W_i^{(\boldsymbol{\tau})}(\boldsymbol{s},t) - \delta^{(\boldsymbol{\tau})}(\boldsymbol{s},t)} = \eta(\boldsymbol{s} - t\boldsymbol{\tau}, t), \quad (\boldsymbol{s}, t) \in \mathbb{R}^{d-1} \times [0, \infty), \tag{4.53}$$

defines a strictly stationary space-time Brown-Resnick process.

This method does not depend on the specific dependence function: every Brown-Resnick process $\{\eta(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in \mathbb{R}^{d-1}, t \in [0, \infty)\}$ with dependence function $\{\delta_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$ results in a time-shifted Brown-Resnick process with dependence function $\{\delta_{\boldsymbol{\theta}}^{(\boldsymbol{\tau})}, \boldsymbol{\theta} \in \Theta, \boldsymbol{\tau} \in \mathbb{R}^{d-1}\}$. To give an example, for the Brown-Resnick process (II) without rotation, the parametrised time-shifted dependence function is given by

$$\delta_{\boldsymbol{\theta}}^{(\boldsymbol{\tau})}(\boldsymbol{h}, u) = \sum_{i=1}^{d-1} C_i |h_i - u\tau_i|^{\alpha_i} + C_d |u|^{\alpha_d}, \quad (\boldsymbol{h}, u) \in \mathbb{R}^d \tag{4.54}$$

with parameter vector

$$(\boldsymbol{\theta}, \boldsymbol{\tau}) \in \{(C_j, \alpha_j, j = 1, \ldots, d) : C_j \in (0, \infty), \alpha_j \in (0, 2], j = 1, \ldots, d\} \times \mathbb{R}^{d-1}.$$

This model is somewhat motivated by the time-shifted moving maxima Brown-Resnick process introduced by Embrechts et al. [33], it is however much simpler to analyse and to estimate.

In the following we show that models (I)-(III) satisfy Assumption 4.15 and the conditions of Theorem 4.16 and Corollary 4.19.

Asymptotic properties of models (I)-(III)

As before, we assume space-time observations on $\mathcal{D}_n = \mathcal{S} \times \mathcal{T} = (\mathcal{S} \times \mathcal{T})(n)$, where $\mathcal{S} \subset \mathbb{Z}^{d-1}$ are the spatial and $\mathcal{T} \subset \mathbb{Z}$ the time series observations. Moreover, we assume that they decompose into $\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n$, where $\mathcal{F} \subset \mathbb{Z}^q$ is some fixed domain and $\mathcal{I}_n = \{1, \ldots, n\}^w$ is a sequence of regular grids, and $q + w = d$.

For two points $(\boldsymbol{s}^{(1)}, t^{(1)})$ and $(\boldsymbol{s}^{(2)}, t^{(2)}) \in \mathbb{R}^{d-1} \times [0, \infty)$, we denote by $(\boldsymbol{h}, u) = (\boldsymbol{s}^{(1)}, t^{(1)}) - (\boldsymbol{s}^{(2)}, t^{(2)}) \in \mathbb{R}^d$ their space-time lag vector. Furthermore, we choose Borel sets $A = B = (a, \infty)$ for some $a > 0$. We denote by $\widehat{\rho}_{AB,m_n}(\boldsymbol{h}, u)$ the (possibly bias-corrected) empirical space-time extremogram (4.25), sampled at lags in $\mathcal{H} \subset \mathbb{R}^d$, and by $\widehat{\boldsymbol{\theta}}_{n,V}$ the GLSE (4.32), referring to some positive definite weight matrix V.

To show consistency and asymptotic normality of the corresponding GLSE, we need to verify the assumptions required in Theorem 4.16; i.e. the relevant parts of Assumption 4.15. Note that Corollary 4.19 applies for all models, since they all satisfy $\delta_{\boldsymbol{\theta}}(\boldsymbol{h}, u) \geq C|u|^{\alpha}$ for $C > 0$ and $\alpha \in (0, 2]$. Thus we obtain the CLTs of the empirical extremogram centred by the pre-asymptotic extremogram (4.48), centred by the true one (4.22) and of the bias corrected empirical extremogram centred by the true one (4.27). Hence (G1) and (G2) hold for the empirical extremogram. Furthermore, we assume that the parameter space $\Theta \subset \mathbb{R}^k$, which contains the true parameter $\boldsymbol{\theta}^\star$, is a compact subset of the spaces introduced above for the corresponding models.

The following requirements concern the model-independent assumptions.

- In order to determine the GLSE we need to choose a matrix $V(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Theta$, and we take one, which satisfies condition (G5ii) with $z_2 = 1$. Due to compactness of the parameter space Θ, condition (G5i) is therefore automatically satisfied.
- We require that $|\mathcal{H}| \geq k$, such that the rank condition (G6) can be satisfied.

Next we discuss the model-dependent assumptions. First note that the smoothness condition (G4ii) is satisfied for $z_1 = 0$ for all models $\{\rho_{AB,\boldsymbol{\theta}}(\cdot)\}$ (equivalently $\{\delta_{\boldsymbol{\theta}}(\cdot)\}$). Due to compactness of the parameter space, condition (G4i) is therefore automatically satisfied. Besides it suffices to show condition (G3ii) in order to verify identifiability of the models. Condition (G3ii) is satisfied for models (I)-(III) if for two distinct parameter vectors $\boldsymbol{\theta}^{(1)} \neq \boldsymbol{\theta}^{(2)}$ there is at least one $(\boldsymbol{h}, u) \in \mathcal{H}$ such that $\rho_{AB,\boldsymbol{\theta}^{(1)}}(\boldsymbol{h}, u) \neq \rho_{AB,\boldsymbol{\theta}^{(2)}}(\boldsymbol{h}, u)$. This holds due to the power function structure of the models. For the geometric anisotropic model in (I) we need to exclude $c = 1$ to ensure identifiability of the angle φ; however, if $c = 1$ then φ has no influence on the dependence function and can be neglected. Thus, the GLSEs are consistent according to Theorem 4.16.

We now turn to the CLT (4.35), where it remains to show (G4ii) for $z_1 = 1$. Difficulties arise due to norms and absolute values of certain parameters in the model equations:

- In their basic forms without rotation or dilution, models (I) and (II) are infinitely often continuously partially differentiable in the model parameters. Hence asymptotic normality of the GLSEs follows by Theorem 4.16.
- If rotation and/or dilution parameters are included, continuous partial differentiability still holds under the following restrictions: Let α_1 (for model (I)) or $\alpha_1, \dots, \alpha_{d-1}$ (for model (II)) be the spatial smoothness parameters. Since they are the powers of some norm or absolute value, restricting them to values in $[1, 2]$ makes the models continuously partially differentiable. As to model (II), in the case $d - 1 = 2$, one of the parameters α_1 and α_2 being larger than 1 is already sufficient. To see this, recall that the spatial part of the dependence function is given by
$$C_1|h_1 \cos\varphi - h_2 \sin\varphi|^{\alpha_1} + C_2|h_1 \sin\varphi + h_2 \cos\varphi|^{\alpha_2}, \quad (h_1, h_2) \in \mathbb{R}^2.$$
Assume w.l.o.g that $\alpha_2 > 1$. Then critical values of $\varphi \in [0, \pi/2)$ are the roots of $h_1 \cos\varphi - h_2 \sin\varphi$. Given a value $h_2 \in \mathbb{R}$ we need to choose $h_1 \in \mathbb{R}$ such that $h_1 \neq h_2 \tan\varphi$ for all $\varphi \in [0, \pi/2)$. Since $\tan\varphi > 0$ for $\varphi \in [0, \pi_2)$, we can choose h_1 such that $\mathrm{sgn}(h_1) = -\mathrm{sgn}(h_2)$. If all lags $(h_1, h_2, u) \in \mathcal{H}$ are chosen such that (h_1, h_2) have opposite signs (or, trivially, are equal to $(0, 0)$) and if $\mathrm{rank}(\mathrm{P}_{AB}(\boldsymbol{\theta}^\star)) = k$, then the GLSE is asymptotically normal.
- Model (III) is continuous partially differentiable, if the spatial smoothness parameters α_i for $i = 1, \dots, d-1$ are all larger than 1. If $\alpha_i \leq 1$ for some i, then the term $C_i|h_i - u\tau_i|^{\alpha_i}$ is, as a function of τ_i, not differentiable at $\tau_i = h_i/u \in \mathbb{R}$. However, it is possible to restrict the parameter space such that such equalities do not occur.

4.6 Simulation study

Specifications

Consider the framework of Section 4.5.3. In particular, let $\{\eta(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^2, t \in [0,\infty)\}$ be a strictly stationary space-time Brown-Resnick process (4.36) observed on $\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n$. Denote by $\widehat{\rho}_{AB,m_n}(\boldsymbol{h},u)$ the space-time version of the (possibly bias corrected) empirical extremogram given in (4.25), sampled at lags in $\mathcal{H} \subset \mathbb{R}^d$, where $\mathcal{H}$ is specified below and we choose the sets $A = B = (1,\infty)$. As already indicated in its Definition 4.5(1), the computation involves the practical issue of choosing the value $a_{m_n} = m_n =: q$ as a large quantile, where the first equality is due to the standard unit Fréchet distribution of the marginals of the Brown-Resnick model, so that q should be chosen as a large quantile of the standard unit Fréchet distribution. In a data example it should be chosen from a set Q of large empirical quantiles of $\{\eta(\boldsymbol{s},t) : (\boldsymbol{s},t) \in \mathcal{D}_n\}$ for which the empirical extremograms $\widehat{\rho}_{AB,q}(\boldsymbol{h},u)$, are robust; cf. also Davis et al. [21] after their Theorem 2.1

In order to test the small sample performance of the GLSE $\widehat{\boldsymbol{\theta}}_{n,V}$ defined in (4.32), we consider some of the models (I)-(III) for the dependence function $\delta_{\boldsymbol{\theta}}$. For the simulations we use the R-package `RandomFields` [60] and the exact method via extremal functions proposed in Dombry et al. [28], Section 2.

(i) Spatially isotropic fractional space-time model
We generate 100 realisations from the model (4.49) on a grid of size 15x15x300. This corresponds to the situation of a fixed spatial and an increasing temporal observation area; i.e., it is given by $\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n$ with $\mathcal{F} = \{1,\dots,15\}^2$ and $\mathcal{I}_n = \{1,\dots,300\}$. We simulate the model with the true parameter vector

$$\boldsymbol{\theta}_1^\star = (0.8, 0.4, 1.5, 1),$$

which we assume to lie in a compact subset of

$$\Theta_1 = \{(C_1, C_2, \alpha_1, \alpha_2) : C_1, C_2 \in (0,\infty), \alpha_1, \alpha_2 \in (0,2]\}\,.$$

As the large empirical quantile q we take the 96%-quantile of $\{\eta(\boldsymbol{s},t) : (\boldsymbol{s},t) \in \mathcal{D}_n\}$.

(ii) Geometrically anisotropic fractional space-time model
We generate 100 realisations from model (4.50) on a grid of size 15x15x300. This corresponds to the same situation as in (i). We simulate the model with the true parameter vector

$$\boldsymbol{\theta}_2^\star = (0.8, 0.4, 1.5, 0.5, 3, \pi/4),$$

which we assume to lie in a compact subset of

$$\Theta_2 = \{(C_1, C_2, \alpha_1, \alpha_2, c, \varphi) : C_1, C_2 \in (0,\infty), \alpha_1 \in [1,2], \alpha_2 \in (0,2], c > 0, \varphi \in [0,\pi/2)\}\,,$$

where we choose $\alpha_1 \geq 1$ to ensure differentiability of the model, cf. the discussion in Section 4.5.3. As the large empirical quantile q we take the 97%-quantile of $\{\eta(\boldsymbol{s},t) : (\boldsymbol{s},t) \in \mathcal{D}_n\}$.

(iii) Spatially anisotropic time-shifted model
We generate 100 realisations from model (4.54) on a grid of size 40x40x40, and consider this as a situation where the observation area increases in all dimensions; i.e., it is given by $\mathcal{D}_n = \mathcal{I}_n$ with $\mathcal{I}_n = \{1, \ldots, 40\}^3$. We simulate the model with the true parameter vector

$$\boldsymbol{\theta}_3^\star = (0.4, 0.8, 0.5, 1.5, 1.5, 1, 1, 1),$$

which we assume to lie in a compact subset of

$$\Theta_3 = \left\{(C_1, C_2, C_3, \alpha_1, \alpha_2, \alpha_3, \tau_1, \tau_2) : C_j \in (0, \infty), \alpha_1, \alpha_2 \in [1, 2], \alpha_3 \in (0, 2], \tau_j \in \mathbb{R}\right\},$$

where we choose $\alpha_1, \alpha_2 \geq 1$ to ensure differentiability of the model, cf. the discussion in Section 4.5.3. As the large empirical quantile q we take the 95%-quantile of $\{\eta(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in \mathcal{D}_n\}$.
□

In all three settings we base the estimation on the set $\mathcal{H}$ of lags given by

$$\begin{aligned}\mathcal{H} = \{&(0,0,1),(0,0,2),(0,0,3),(0,0,4),(1,0,0),(2,0,0),(3,0,0),(4,0,0),(2,1,0),(4,2,0),\\ &(1,2,0),(2,4,0),(1,1,1),(2,2,2),(1,3,2)\}.\end{aligned}$$

With this choice we ensure that the lag vectors vary in all three dimensions so that we obtain reliable estimates. Generally one should choose $\mathcal{H}$ such that the whole range of clear extremal dependence is covered. However, beyond that, no lags should be included for the estimation, since independence effects can introduce a bias in the least squares estimates, similarly as in pairwise likelihood estimation; cf. Section 5.5.3. One way to determine the range of extremal dependence are permutation tests, which are described in Section 3.6. From those tests we know that our choice of lags satisfies this requirement for all three models.

For the weight matrix V in (4.32) we propose two choices, which yield equally good results in our statistical analysis. The first choice is $V_1 = \operatorname{diag}\{\exp(-\|(\boldsymbol{h}, u)\|^2) : (\boldsymbol{h}, u) \in \mathcal{H}\}$, which reflects the exponential decay of the tail dependence coefficients $\rho_{(1,\infty)(1,\infty)}(\boldsymbol{h}, u)$ of Brown-Resnick processes given by tail probabilities of the standard normal distribution. The second choice is to include the (possibly bias corrected) empirical extremogram estimates as in $V_2 = \operatorname{diag}\{\widehat{\rho}_{(1,\infty)(1,\infty),q}(\boldsymbol{h}, u) : (\boldsymbol{h}, u) \in \mathcal{H}\}$. Since the so defined weight matrix is random, what follows is conditional on its realisation. It is in practice not possible to incorporate the asymptotic covariance matrix Π of the empirical extremogram estimates $(\widehat{\rho}_{(1,\infty)(1,\infty),q}(\boldsymbol{h}, u) : (\boldsymbol{h}, u) \in \mathcal{H})$ (cf. Remark 4.17) to obtain a weight matrix that is optimal in theory. As can be seen from its specification in Theorem 4.8, it contains infinite sums and is, hence, numerically hardly tractable.

Results

For each of the scenarios (i)-(iii) we report the mean, the root mean squared error (RMSE) and the mean absolute error (MAE) of the resulting GLSEs for the 100 simulations. The results are summarised in Tables 4.1-4.3. Furthermore, in Figures 4.1-4.3 we plot the parameter estimates and add 95%-confidence bounds found by subsampling; cf. Politis et al. [56], Chapter 5. We use subsampling methods, since the asymptotic covariance matrix Π_V specified in Theorem 4.16 contains the matrix Π as specified in Theorem 4.8, which is, as explained above, hardly tractable. The fact that subsampling yields asymptotically valid confidence intervals for the true parameter vectors $\boldsymbol{\theta}_i^\star$ for $i = 1, 2, 3$ can be proved analogously to the proof of Theorem 3.20 based on Corollary 5.3.4 of [56]. It requires mainly the existence of continuous limit distributions of $\sqrt{n^w/m_n^d}\|(\widehat{\boldsymbol{\theta}}_{n,V} - \boldsymbol{\theta}_i^\star)\|$, which are guaranteed by Theorem 4.32, and some conditions on the α-mixing coefficients, which can be shown similarly as those required in Theorem 4.8.

	TRUE	MEAN	RMSE	MAE
$\widehat{C}_1$	0.8	0.7856	0.1763	0.1353
$\widehat{C}_2$	0.4	0.3987	0.0995	0.0785
$\widehat{\alpha}_1$	1.5	1.4830	0.1131	0.0897
$\widehat{\alpha}_2$	1	0.9916	0.0820	0.0625

Table 4.1: True parameter values (first column) and mean, RMSE and MAE of the estimates of the parameters of model (i).

	TRUE	MEAN	RMSE	MAE
$\widehat{C}_1$	0.8	0.7270	0.335	0.2750
$\widehat{C}_2$	0.4	0.3708	0.1377	0.1097
$\widehat{\alpha}_1$	1.5	1.4349	0.2692	0.2274
$\widehat{\alpha}_2$	0.5	0.5143	0.0684	0.0491
$\widehat{c}$	3	2.9441	0.2645	0.1365
$\widehat{\varphi}$	$\pi/4$	0.7906	0.1567	0.1214

Table 4.2: True parameter values (first column) and mean, RMSE and MAE of the estimates of the parameters of model (ii).

	TRUE	MEAN	RMSE	MAE
$\widehat{C}_1$	0.4	0.4072	0.0898	0.0690
$\widehat{C}_2$	0.8	0.8482	0.2187	0.1667
$\widehat{C}_3$	0.5	0.5003	0.1366	0.1085
$\widehat{\alpha}_1$	1.5	1.5144	0.0781	0.0594
$\widehat{\alpha}_2$	1.5	1.5043	0.1282	0.1054
$\widehat{\alpha}_3$	1	0.9694	0.1415	0.1082
$\widehat{\tau}_1$	1	1.0459	0.1250	0.0945
$\widehat{\tau}_2$	1	0.9916	0.0420	0.0320

Table 4.3: True parameter values (first column) and mean, RMSE and MAE of the estimates of the parameters of model (iii).

Summary

Summarising our results, we find that the GLSE estimates the model parameters very accurately. Bias and variance are largest for the parameter estimates of model (ii). There are two main reasons for this. Compared to model (i), for model (ii) we estimate two more parameters based on the same observation scheme. However, one is a direction, which to estimate is a non-trivial

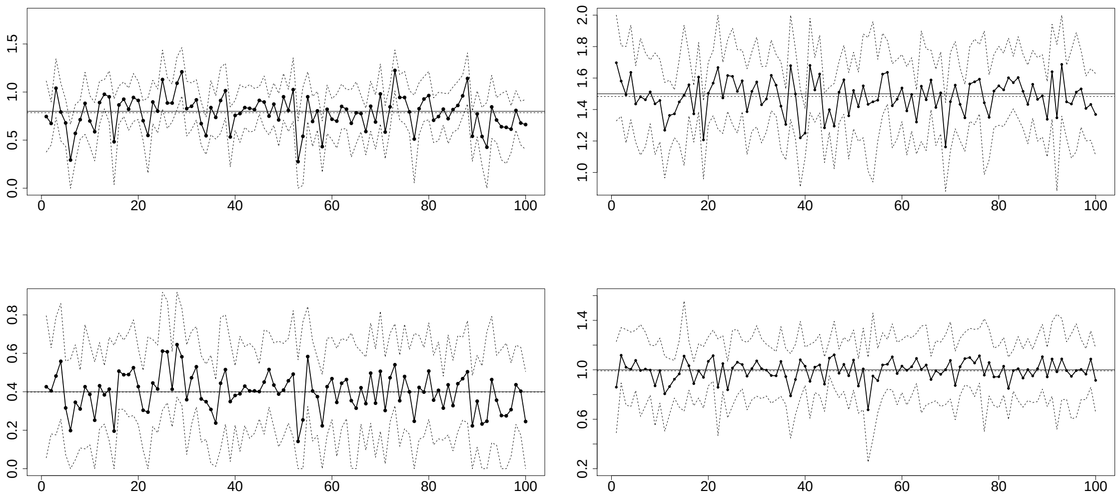

Figure 4.1: GLSEs of the parameters of model (i) for 100 simulated Brown-Resnick space-time processes together with pointwise 95%-subsampling confidence intervals (dotted). First row: C_1, α_1, second row: C_2, α_2. The middle solid line is the true parameter value and the middle dotted line represents the mean over all estimates.

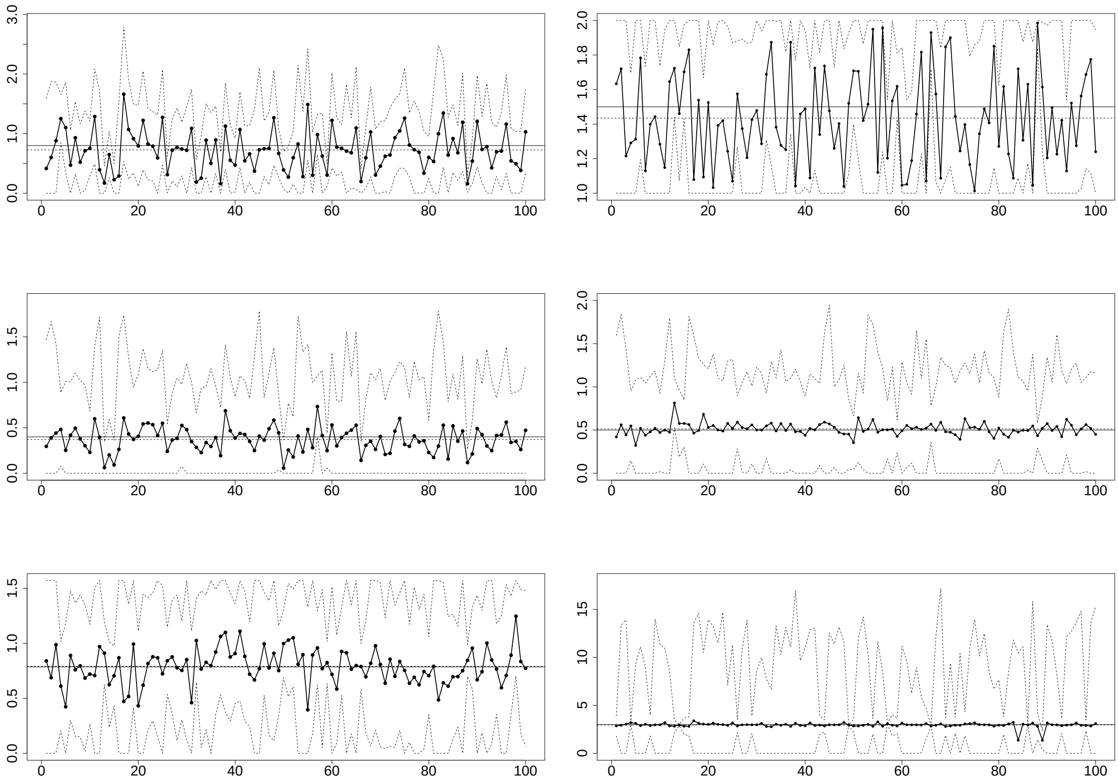

Figure 4.2: GLSEs of the parameters of model (ii) for 100 simulated Brown-Resnick space-time processes together with pointwise 95%-subsampling confidence intervals (dotted). First row: C_1, α_1, middle row: C_2, α_2, last row: φ and c. The middle solid line is the true value and the middle dotted line represents the mean over all estimates.

task and decreases the overall quality of the estimates. For the estimation of model (iii) the observation scheme is different; in particular, there is a relatively large number of both spatial and temporal observations available. In contrast, in the setting of models (i) and (ii) only the number of temporal observations is large.

It is obvious from Tables 4.1 and 4.2 that bias and variance of the spatial parameter estimates $\widehat{C}_1$ and $\widehat{\alpha}_1$ are considerably larger than bias and variance of the temporal parameter estimates $\widehat{C}_2$ and $\widehat{\alpha}_2$. Again this is due to the fact that only the number of temporal observations is large.

From Table 4.3 we read off that the variance of the estimates $\widehat{C}_1$ and $\widehat{\alpha}_1$, which correspond to the first spatial dimension, are considerably smaller than those of $\widehat{C}_2$ and $\widehat{\alpha}_2$. This is due to the lag vectors we included in the set $\mathcal{H}$, which show more variation with respect to the first dimension than with respect to the second.

Compared to likelihood-based methods computation time of semiparametric estimation is substantially lower. This is also found in Section 3.5, and Steinkohl [62], Chapter 6, however, for a much simpler model, where simple least squares estimation applied.

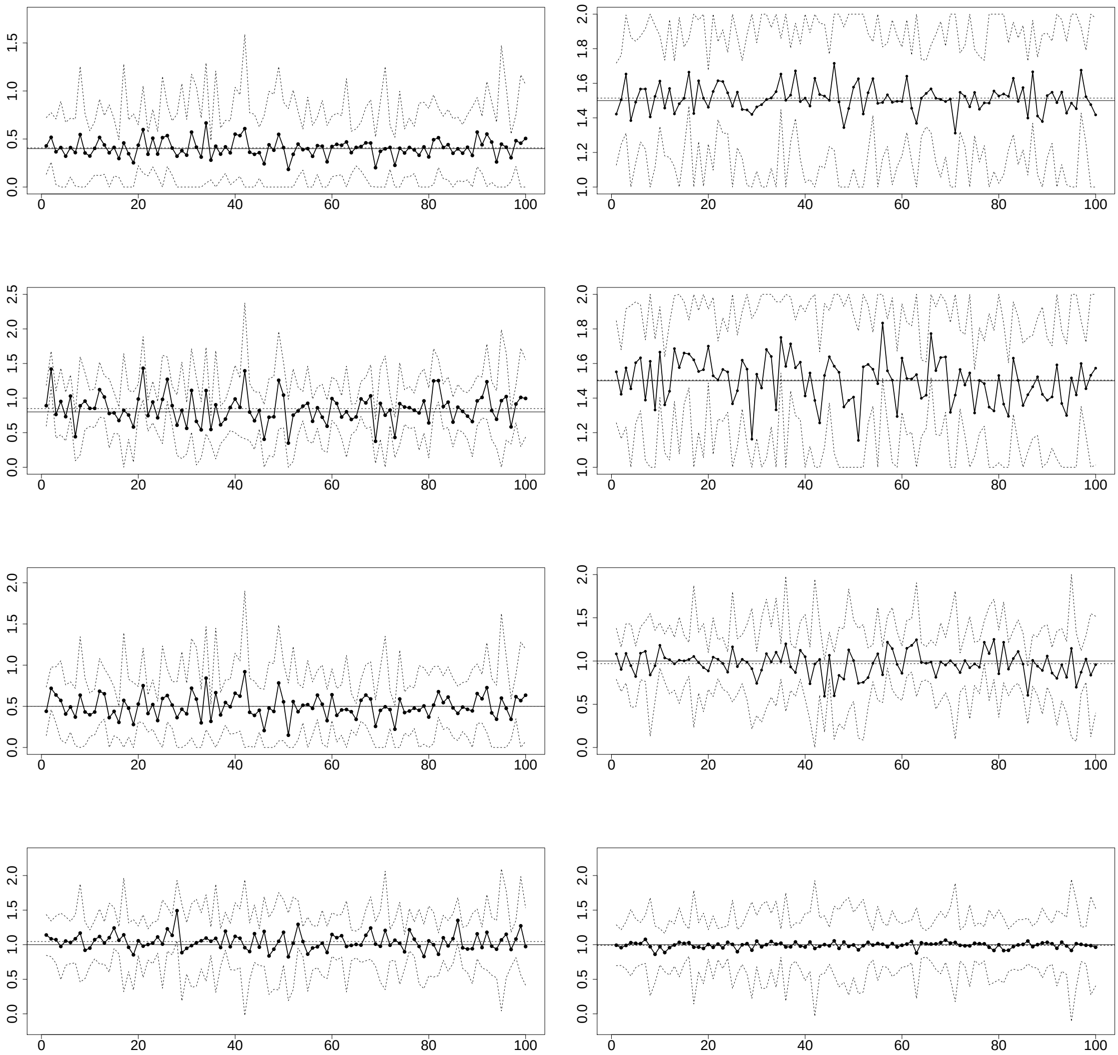

Figure 4.3: GLSEs of the parameters of model (iii) for 100 simulated Brown-Resnick space-time processes together with pointwise 95%-subsampling confidence intervals (dotted). First row: C_1, α_1, second row: C_2, α_2, third row: C_3, α_3, fourth row: τ_1, τ_2. The middle solid line is the true value and the middle dotted line represents the mean over all estimates.

Chapter 5

Anisotropic Brown-Resnick space-time processes: estimation and model assessment

Abstract

Spatially isotropic max-stable processes have been used to model extreme spatial or space-time observations. One prominent model is the Brown-Resnick process, which has been successfully fitted to time series, spatial data and space-time data. This chapter extends the process to possibly anisotropic spatial structures. For regular grid observations we prove strong consistency and asymptotic normality of pairwise maximum likelihood estimates for fixed and increasing spatial domain, when the number of observations in time tends to infinity. We also present a statistical test for isotropy versus anisotropy. We apply our test to precipitation data in Florida, and present some diagnostic tools for model assessment. Finally, we present a method to predict conditional probability fields and apply it to the data.

AMS 2010 Subject Classifications: primary: 62G32, 62M40, 62P12; secondary: 62F05, 62F12

Keywords: anisotropic space-time process; Brown-Resnick space-time process; hypothesis test for spatial isotropy; max-stable process; max-stable model check; pairwise likelihood; pairwise maximum likelihood estimate

5.1 Introduction

Max-stable processes, such as the Brown-Resnick process, have been successfully fitted to time series, spatial and recently to space-time data. Methods for inference include pairwise likelihood based on the bivariate density of the models (cf. Padoan et al. [54]), censored likelihood (cf. Wadsworth and Tawn [67]) or threshold-based approaches (cf. Engelke et al. [34]). In Davis et al. [20] a spatially isotropic Brown-Resnick space-time process is suggested and applied to precipitation data. Pairwise maximum likelihood estimates are shown to be strongly consistent and asymptotically normal, provided the domain of observations increases jointly in space and time. Their approach is restricted to isotropic spatial dependence.

In the present chapter we generalise the Brown-Resnick model to allow anisotropy in space. The new model allows for different extremal behaviour along orthogonal spatial directions. Anisotropy is often observed on Earth, for example in Middle Europe with its westerly winds or near the equator where trade winds involve predominant easterlies. All dependence parameters are summarised in the semivariogram of an underlying Gaussian space-time process. This semivariogram then defines the dependence structure of the max-stable process and, as a consequence, the tail dependence coefficient between two process values evaluated at two location and two time points.

Furthermore, since in real world applications, observations are often recorded over a large number of time points, but only at a comparably small number of spatial locations, we consider both a fixed and increasing spatial domain in combination with an increasing temporal domain. For both settings, fixed and increasing spatial domain, we prove strong consistency and asymptotic normality of the pairwise maximum likelihood estimates in the anisotropic model based on regular grid observations. This requires in particular to prove space-time and temporal mixing conditions in both settings for the anisotropic model.

We also provide tests for isotropy versus anisotropy again in both settings, which are designed for the new model. The asymptotic normality of the parameter estimates determines in principle the rejection areas of the test. However, the covariance matrices of the normal limit laws are not available in closed form. We formulate a subsampling procedure in the terminology of the Brown-Resnick space-time process and prove its convergence for fixed and increasing spatial domain.

We conclude with an analysis of space-time block maxima of radar rainfall measurements in Florida. Firstly, we present a simple procedure to test whether they originate from a max-stable process. As this cannot be rejected, we fit the Brown-Resnick space-time model to the data, using pairwise maximum likelihood estimation. Subsequently we apply the new isotropy test. Both the estimation and the test are based on the setting of a fixed spatial domain and increasing time series. In particular, since the Brown-Resnick space-time process satisfies the strong mixing conditions for increasing spatial and time domain as well as for fixed spatial and increasing time domain, the estimation and test procedure are independent of the specific setting: it works in both settings in exactly the same way, taking the different asymptotic covariance matrices into account. Finally, we assess the goodness of fit of the estimated model by a simulation diagnostics based on a large number of i.i.d. simulated anisotropic Brown-Resnick space-time processes. As a result, there is no statistical significance that the anisotropic Brown-Resnick space-time process with the fitted parameters should be rejected.

This chapter is organised as follows. In Section 5.2 we present the Brown-Resnick space-time model, which allows for anisotropic effects in space, and various dependence measures, including the parameterised dependence function. In Section 5.3 we compute the pairwise maximum likelihood estimates for the new model and prove their strong consistency and asymptotic normality for both settings, fixed and increasing spatial domain. Section 5.4 presents hypothesis tests for spatial isotropy and derives rejection areas based on a subsampling procedure. A data analysis is performed in Section 5.5 with focus on model assessment. The isotropy test rejects spatial isotropy for these data in favour of our new anisotropic model. Based on two other test procedures, we conclude that the anisotropic Brown-Resnick space-time process with the given dependence parameters is an appropriate model for the block-maxima data. We conclude by predicting conditional probability fields, which give the probability of a high value (for example of the amount of precipitation) at some space-time location given a high value at some other location.

5.2 Spatially anisotropic Brown-Resnick processes

Throughout the chapter we consider a *stationary Brown-Resnick space-time process* with representation

$$\eta(\boldsymbol{s},t) = \bigvee_{j=1}^{\infty} \left\{ \xi_j \, e^{W_j(\boldsymbol{s},t)-\delta(\boldsymbol{s},t)} \right\}, \quad (\boldsymbol{s},t) \in \mathbb{R}^d \times [0,\infty), \tag{5.1}$$

where $\{\xi_j : j \in \mathbb{N}\}$ are points of a Poisson process on $[0,\infty)$ with intensity $\xi^{-2}d\xi$, the *dependence function* δ is nonnegative and conditionally negative definite and $\{W_j(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\}$ are independent replicates of a Gaussian process
$\{W(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\}$ with stationary increments, $W(\boldsymbol{0},0) = 0$, $\mathbb{E}[W(\boldsymbol{s},t)] = 0$ and covariance function

$$\mathbb{C}ov[W(\boldsymbol{s}^{(1)},t^{(1)}), W(\boldsymbol{s}^{(2)},t^{(2)})] = \delta(\boldsymbol{s}^{(1)},t^{(1)}) + \delta(\boldsymbol{s}^{(2)},t^{(2)}) - \delta(\boldsymbol{s}^{(1)}-\boldsymbol{s}^{(2)},t^{(1)}-t^{(2)}).$$

Representation (5.1) goes back to de Haan [24] and Giné et al. [39]. Brown-Resnick processes have been studied by Brown and Resnick [8] in a time series context, as a spatial model by Kabluchko et al. [47], and in a space-time setting by Davis et al. [19] and Huser and Davison [43]. The univariate margins of the process η follow standard Fréchet distributions.

There are various quantities to describe the dependence in (5.1):

- In geostatistics, the dependence function δ is termed the *semivariogram* of the process $\{W(\boldsymbol{s},t)\}$: For $(\boldsymbol{s}^{(1)},t^{(1)}),(\boldsymbol{s}^{(2)},t^{(2)}) \in \mathbb{R}^d \times [0,\infty)$, it holds that

$$\mathbb{V}ar[W(\boldsymbol{s}^{(1)},t^{(1)}) - W(\boldsymbol{s}^{(2)},t^{(2)})] = 2\delta(\boldsymbol{s}^{(1)}-\boldsymbol{s}^{(2)},t^{(1)}-t^{(2)}).$$

- For $\boldsymbol{h} \in \mathbb{R}^d$ and $u \in \mathbb{R}$, the *tail dependence coefficient* $\chi(\boldsymbol{h},u)$ is given by (cf. Kabluchko et al. [47], Remark 25 or Davis et al. [19], Section 3)

$$\chi(\boldsymbol{h},u) := \lim_{y\to\infty} \mathbb{P}\big(\eta(\boldsymbol{s}^{(1)},t^{(1)}) > y \mid \eta(\boldsymbol{s}^{(2)},t^{(2)}) > y\big) = 2\left(1 - \Phi\left(\sqrt{\frac{\delta(\boldsymbol{h},u)}{2}}\right)\right), \tag{5.2}$$

where $\boldsymbol{h} = \boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)}$, $u = t^{(1)} - t^{(2)}$, and Φ denotes the standard normal distribution function.

- For $D = \{(\boldsymbol{s}^{(1)}, t^{(1)}), \dots, (\boldsymbol{s}^{(|D|)}, t^{(|D|)})\}$ and $\boldsymbol{y} = (y_1, \dots, y_{|D|}) > \boldsymbol{0}$ the finite-dimensional margins are given by

$$\mathbb{P}(\eta(\boldsymbol{s}^{(1)}, t^{(1)}) \le y_1, \eta(\boldsymbol{s}^{(2)}, t^{(2)}) \le y_2, \dots, \eta(\boldsymbol{s}^{(|D|)}, t^{(|D|)}) \le y_{|D|}) = e^{-V_D(\boldsymbol{y})}. \tag{5.3}$$

Here V_D denotes the *exponent measure*, which is homogeneous of order -1.

- The *extremal coefficient* ξ_D for any finite set $D \subset \mathbb{R}^d \times [0, \infty)$ is defined through

$$\mathbb{P}(\eta(\boldsymbol{s}^{(1)}, t^{(1)}) \le y, \eta(\boldsymbol{s}^{(2)}, t^{(2)}) \le y, \dots, \eta(\boldsymbol{s}^{(|D|)}, t^{(|D|)}) \le y) = e^{-\xi_D/y}, \quad y > 0;$$

i.e., $\xi_D = V_D(1, \dots, 1)$. If $|D| = 2$, then (cf. Beirlant et al. [3], Section 9.5.1)

$$\chi(\boldsymbol{s}^{(1)} - \boldsymbol{s}^{(2)}, t^{(1)} - t^{(2)}) = 2 - \xi_D.$$

In this chapter we assume the dependence function δ to be given for spatial lag $\boldsymbol{h}$ and time lag u by

$$\delta(\boldsymbol{h}, u) = \sum_{j=1}^{d} C_j |h_j|^{\alpha_j} + C_{d+1}|u|^{\alpha_{d+1}}, \quad (\boldsymbol{h}, u) = (h_1, \dots, h_d, u) \in \mathbb{R}^{d+1}, \tag{5.4}$$

with parameters $C_j > 0$ and $\alpha_j \in (0, 2]$ for $j = 1, \dots, d+1$.
Model (5.4) allows for different rates of decay of extreme dependence in different directions. This particularly holds along the axes of a d-dimensional spatial grid, but also for other directions. For example in the case $d = 2$, the decreases of dependence along the directions $(1, 2)$ and $(2, 1)$ differ. Model (5.4) can be generalised by a simple rotation to a setting, where not necessarily the axes, but other principal orthogonal directions play the major role. The rotation angle then needs to be estimated together with the other model parameters. A similar approach has been applied to introduce geometric or zonal anisotropy into a spatial isotropic model (see e.g. Blanchet and Davison [4], Section 4.2, or Engelke et al. [34], Section 5.2). For a justification of model (5.4) see Buhl [9], Sections 4.1 and 4.2. There it is shown that Brown-Resnick processes with this dependence function arise as limits of appropriately rescaled maxima of Gaussian processes with a large variety of correlation functions.

5.3 Pairwise maximum likelihood estimation

We extend the pairwise maximum likelihood procedure described in Davis et al. [20] for spatially isotropic space-time Brown-Resnick processes to the anisotropic case. We focus on the difference introduced by the spatial anisotropy and refer to the corresponding formulas in Davis et al. [20], where also a short introduction to composite likelihood estimation and further references can be found.

The pairwise likelihood function uses the bivariate distribution function of $(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u)) \stackrel{d}{=} (\eta(\mathbf{0},0),\eta(\boldsymbol{h},u))$ (equal in distribution by stationarity) for $\boldsymbol{h} \in \mathbb{R}^d$ and $u \in \mathbb{R}$, which is given as

$$G(y_1,y_2) = \exp\{-V(y_1,y_2)\}, \quad y_1,y_2>0, \tag{5.5}$$

where the exponent measure $V = V_D$ for $D = \{(\boldsymbol{s}^{(1)},t^{(1)}),(\boldsymbol{s}^{(2)},t^{(2)})\}$ has the representation

$$V(y_1,y_2) = \frac{1}{y_1}\Phi\left(\frac{\log(y_2/y_1)}{\sqrt{2\delta(\boldsymbol{h},u)}} + \sqrt{\frac{\delta(\boldsymbol{h},u)}{2}}\right) + \frac{1}{y_2}\Phi\left(\frac{\log(y_1/y_2)}{\sqrt{2\delta(\boldsymbol{h},u)}} + \sqrt{\frac{\delta(\boldsymbol{h},u)}{2}}\right), \tag{5.6}$$

which is a particular form of Eq. (2.7) in Hüsler and Reiss [44]. The dependence function δ is given by (5.4). For a derivation of (5.6) see for instance Oesting [53], Satz und Definition 2.4.

From this we can calculate the pairwise density $g(y_1,y_2) = g_{\boldsymbol{\theta}}(y_1,y_2)$ of G by differentiation. The parameter vector $\boldsymbol{\theta} = (C_1,\ldots,C_{d+1},\alpha_1,\ldots,\alpha_{d+1})$ lies in the parameter space

$$\Theta := \{(C_1,\ldots,C_{d+1},\alpha_1,\ldots,\alpha_{d+1}) : C_j \in (0,\infty), \alpha_j \in (0,2], j=1,\ldots,d+1\}.$$

We focus on data on a regular spatial grid and at equidistant time points. More precisely, we assume that the spatial observations lie on a regular d-dimensional lattice,

$$\mathcal{S}_M = \{\boldsymbol{s} = (s_1,\ldots,s_d) \in \{1,\ldots,M\}^d\}$$

for $M \in \mathbb{N}$, and that the time points are given by the set $\mathcal{T}_T = \{1,\ldots,T\}$ for $T \in \mathbb{N}$.

For the computation of the pairwise likelihood it is common not to include observations on all available space-time pairs, but only on those that lie within some prespecified spatio-temporal distance. This is motivated by the fact that pairs which lie sufficiently far apart in a space-time sense have little influence on the dependence parameters, see Nott and Rydén [52], Section 2.1. To express this notationally, we take inspiration from that paper and use a design mask adapted to the anisotropic setting; that is, for $\boldsymbol{r} = (r_1,\ldots,r_d) \in \mathbb{N}_0^d$, fix a set $\mathcal{H}_{\boldsymbol{r}}$ as one with maximum cardinality among all sets $\mathcal{H}'_{\boldsymbol{r}}$ that satisfy

$$\mathcal{H}'_{\boldsymbol{r}} \subseteq \left\{\boldsymbol{h} = (h_1,\ldots,h_d) \in \mathbb{Z}^d : |\boldsymbol{h}| \le \boldsymbol{r}\right\} \text{ such that } [\mathbf{0} \neq \boldsymbol{h} \in \mathcal{H}'_{\boldsymbol{r}} \Rightarrow -\boldsymbol{h} \notin \mathcal{H}'_{\boldsymbol{r}}], \tag{5.7}$$

where the operations $|\cdot|$ and $\le$ are taken componentwise. Note that this definition is not unique. We are now ready to define the pairwise log-likelihood function and the resulting estimate.

Definition 5.1 (Pairwise likelihood estimate)**.** *The pairwise log-likelihood function based on space-time pairs, whose maximum spatial lag is $\boldsymbol{r} \in \mathbb{N}_0^d$ and maximum time lag is $p \in \mathbb{N}_0$, such that $(\boldsymbol{r},p) \neq (\mathbf{0},0)$, is defined as*

$$PL^{(M,T)}(\boldsymbol{\theta}) := \sum_{\boldsymbol{s}\in\mathcal{S}_M} \sum_{t=1}^{T} \sum_{\substack{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}} \\ \boldsymbol{s}+\boldsymbol{h}\in\mathcal{S}_M}} \sum_{\substack{u=0 \\ t+u\le T}}^{p} \mathbb{1}_{\{(\boldsymbol{h},u)\neq(\mathbf{0},0)\}} \log\left\{g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u)\right)\right\}$$

$$= \sum_{\boldsymbol{s}\in\mathcal{S}_M}\sum_{t=1}^{T} q_{\boldsymbol{\theta}}(\boldsymbol{s},t;\boldsymbol{r},p) - \mathcal{R}^{(M,T)}(\boldsymbol{\theta}), \quad \boldsymbol{\theta}\in\Theta, \tag{5.8}$$

where

$$q_{\boldsymbol{\theta}}(\boldsymbol{s},t;\boldsymbol{r},p) := \sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}}\sum_{u=0}^{p} \mathbb{1}_{\{(\boldsymbol{h},u)\neq(\boldsymbol{0},0)\}} \log\left\{g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u)\right)\right\} \tag{5.9}$$

and

$$\begin{aligned}\mathcal{R}^{(M,T)}(\boldsymbol{\theta}) &:= \sum_{\boldsymbol{s}\in\mathcal{S}_M}\sum_{t=1}^{T}\sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}}\sum_{u=0}^{p} \mathbb{1}_{\{\boldsymbol{s}+\boldsymbol{h}\notin\mathcal{S}_M \text{ or } t+u>T\}} \log\left\{g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u)\right)\right\}\\ &= \sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}}\sum_{u=0}^{p}\sum_{(\boldsymbol{s},t)\in\mathcal{G}_{M,T}(\boldsymbol{h},u)} \log\left\{g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u)\right)\right\},\end{aligned} \tag{5.10}$$

with

$$\mathcal{G}_{M,T}(\boldsymbol{h},u) := \left\{(\boldsymbol{s},t)\in\mathcal{S}_M\times\mathcal{T}_T : \boldsymbol{s}+\boldsymbol{h}\notin\mathcal{S}_M \text{ or } t+u>T\right\}. \tag{5.11}$$

for $(\boldsymbol{h},u)\in\mathbb{N}^{d+1}$. *The pairwise maximum likelihood estimate (PMLE) is given by*

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}}\, PL^{(M,T)}(\boldsymbol{\theta}). \tag{5.12}$$

We derive the asymptotic properties of the PMLE for two scenarios. The first one is based on regularly spaced observations with an increasing spatio-temporal domain. For this scenario we follow the proofs in Davis et al. [20] and show that the properties of strong consistency and asymptotic normality also hold if the dependence structure δ allows for spatially anisotropic effects as in (5.4). In the second scenario, the observations are taken from a fixed spatial domain and an increasing temporal domain.

5.3.1 Increasing spatio-temporal domain

Lemma 5.2. *For* $(\boldsymbol{h},u)\in\mathcal{H}_{\boldsymbol{r}}\times\{0,\ldots,p\}$, *it holds that*

$$|\mathcal{G}_{M,T}(\boldsymbol{h},u)| \leq K_2(M^{d-1}T + M^d),$$

where K_2 *is a constant independent of* M *and* T.

Proof. The number of space-time points within the space-time observation area, from which some grid point outside the observation area is within a lag $(\boldsymbol{h},u)\in\mathcal{H}_{\boldsymbol{r}}\times\{0,\ldots,p\}$, is bounded by $2M^{d-1}T\sum_{j=1}^{d} r_j + M^d p$. Thus we obtain

$$|\mathcal{G}_{M,T}(\boldsymbol{h},u)| \leq 2M^{d-1}T\sum_{j=1}^{d} r_j + M^d p \leq K_2(M^{d-1}T + M^d),$$

where $K_2 := \max\left\{2\sum_{j=1}^d r_j, p\right\}$ is a constant independent of M and T. $\square$

Theorem 5.3 (Strong consistency for large M and T)**.** *Let $\left\{\eta(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\right\}$ be a Brown-Resnick process as in* (5.1) *with dependence structure*

$$\delta(\boldsymbol{h},u) = \sum_{j=1}^{d} C_j |h_j|^{\alpha_j} + C_{d+1}|u|^{\alpha_{d+1}}, \quad (\boldsymbol{h},u) \in \mathbb{R}^{d+1},$$

where $0 < \alpha_j \le 2$ and $C_j > 0$ for $j = 1,\dots,d+1$. Denote the parameter vector by

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

Assume that the true parameter vector $\boldsymbol{\theta}^\star$ lies in a compact set

$$\Theta^\star \subset \{(C_1,\dots,C_{d+1},\alpha_1,\dots,\alpha_{d+1}) : C_j \in (0,\infty), \alpha_j \in (0,2],\ j = 1,\dots,d+1\}. \tag{5.13}$$

Suppose that the following identifiability condition holds for all $(\boldsymbol{s},t) \in \mathcal{S}_M \times \mathcal{T}_T$:

$$\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}} \quad \Leftrightarrow \tag{5.14}$$
$$g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s},t), \eta(\boldsymbol{s}+\boldsymbol{h},t+u)\right) = g_{\tilde{\boldsymbol{\theta}}}\left(\eta(\boldsymbol{s},t), \eta(\boldsymbol{s}+\boldsymbol{h},t+u)\right), \quad \boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}}, \quad 0 \le u \le p.$$

Then, the PMLE

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

is strongly consistent:

$$\widehat{\boldsymbol{\theta}}^{(M,T)} \overset{\text{a.s.}}{\to} \boldsymbol{\theta}^\star \ \text{as } M,T \to \infty.$$

Proof. The proof uses the method of Wald [68]. One aim is to show that for some chosen maximum space-time lag $(\boldsymbol{r},p) \in \mathbb{N}_0^{d+1} \setminus \{\boldsymbol{0}\}$ and $\boldsymbol{\theta} \in \Theta^\star$,

$$\begin{aligned}&\frac{1}{M^d T} PL^{(M,T)}(\boldsymbol{\theta}) \\ &= \frac{1}{M^d T}\Big(\sum_{\boldsymbol{s}\in\mathcal{S}_M}\sum_{t=1}^{T} q_{\boldsymbol{\theta}}(\boldsymbol{s},t;\boldsymbol{r},p) - \mathcal{R}^{(M,T)}(\boldsymbol{\theta})\Big) \overset{\text{a.s.}}{\to} \mathrm{PL}(\boldsymbol{\theta}) := \mathbb{E}[q_{\boldsymbol{\theta}}(\boldsymbol{1},1;\boldsymbol{r},p)]\end{aligned}$$

as $M,T \to \infty$. This is done by verifying the following two limit results: Uniformly on $\Theta^\star$,

(A) $\dfrac{1}{M^d T}\sum_{\boldsymbol{s}\in\mathcal{S}_M}\sum_{t=1}^{T} q_{\boldsymbol{\theta}}(\boldsymbol{s},t;\boldsymbol{r},p) \overset{\text{a.s.}}{\to} \mathrm{PL}(\boldsymbol{\theta})$ as $M,T \to \infty$,

(B) $\dfrac{1}{M^d T}\mathcal{R}^{(M,T)}(\boldsymbol{\theta}) \overset{\text{a.s.}}{\to} 0$ as $M,T \to \infty$.

Furthermore, we need to show:

(C) The limit function $\mathrm{PL}(\boldsymbol{\theta})$ is uniquely maximised at the true parameter vector $\boldsymbol{\theta}^\star \in \Theta^\star$.

We show (A). The almost sure convergence holds because $q_{\boldsymbol{\theta}}(\cdot)$ is a measurable function of lagged versions of $\eta(\boldsymbol{s},t)$ for $\boldsymbol{s} \in \mathcal{S}_M$, $t \in \mathcal{T}_T$. Proposition 3 of Davis et al. [20] implies a strong law of large numbers. What remains to show is that the convergence is uniform on the compact parameter space $\Theta^\star$. This can be done by carefully following the lines of the proof of Theorem 1 of Davis et al. [20], adapting it to the spatially anisotropic setting. For details we refer to Buhl [9], Theorem 4.4. We find that there is a positive finite constant K_1, independent of $\boldsymbol{\theta}, M$ and T, such that

$$\mathbb{E}\big[\big|\log g_{\boldsymbol{\theta}}\big(\eta(\boldsymbol{s}^{(1)}, t^{(1)}), \eta(\boldsymbol{s}^{(2)}, t^{(2)})\big)\big|\big] < K_1, \quad (\boldsymbol{s}^{(1)}, t^{(1)}), (\boldsymbol{s}^{(2)}, t^{(2)}) \in \mathbb{N}^{d+1}, \tag{5.15}$$

and that $\mathbb{E}\big[\sup_{\boldsymbol{\theta}\in\Theta^\star} |q_{\boldsymbol{\theta}}(\mathbf{1}, 1; \boldsymbol{r}, p)|\big] < \infty$. Theorem 2.7 of Straumann [63] implies that the convergence is uniform.

Next we show (B). Using Proposition 3 of Davis et al. [20] and (5.15) we have that, uniformly on $\Theta^\star$,

$$\begin{aligned}
&\sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}} \sum_{u=0}^{p} \frac{1}{|\mathcal{G}_{M,T}(\boldsymbol{h}, u)|} \sum_{(\boldsymbol{s},t)\in\mathcal{G}_{M,T}(\boldsymbol{h},u)} \log\left\{g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u)\right)\right\} \\
&\overset{\text{a.s.}}{\to} \mathbb{E}\Big[\sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}} \sum_{u=0}^{p} \log\left\{g_{\boldsymbol{\theta}}\left(\eta(\mathbf{1}, 1), \eta(\mathbf{1}+\boldsymbol{h}, 1+u)\right)\right\}\Big] \text{ as } M, T \to \infty.
\end{aligned}$$

By Lemma 5.2 and (5.15) it follows that, uniformly on $\Theta^\star$,

$$\begin{aligned}
&\frac{1}{M^d T}|\mathcal{R}^{(M,T)}(\boldsymbol{\theta})| \\
&\leq K_2\Big(\frac{1}{M} + \frac{1}{T}\Big)\Big|\sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}} \sum_{u=0}^{p} \frac{1}{|\mathcal{G}_{M,T}(\boldsymbol{h}, u)|} \sum_{(\boldsymbol{s},t)\in\mathcal{G}_{M,T}(\boldsymbol{h},u)} \log\left\{g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u)\right)\right\}\Big| \\
&\overset{\text{a.s.}}{\to} 0 \quad \text{as } M, T \to \infty,
\end{aligned}$$

Finally, we prove (C). Let $\boldsymbol{\theta} \neq \boldsymbol{\theta}^\star$. For $\boldsymbol{s} \in \mathcal{S}_M$ and $t \in \mathcal{T}_T$, Jensen's inequality yields

$$\begin{aligned}
\mathbb{E}\left[\log\left\{\frac{g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u)\right)}{g_{\boldsymbol{\theta}^\star}\left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u)\right)}\right\}\right] &\leq \log\left\{\mathbb{E}\left[\frac{g_{\boldsymbol{\theta}}\left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u)\right)}{g_{\boldsymbol{\theta}^\star}\left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u)\right)}\right]\right\} \\
&= \log\Big\{\int_{(0,\infty)^2} \frac{g_{\boldsymbol{\theta}}(y_1, y_2)}{g_{\boldsymbol{\theta}^\star}(y_1, y_2)} g_{\boldsymbol{\theta}^\star}(y_1, y_2)\,\mathrm{d}(y_1, y_2)\Big\} \\
&= \log\Big\{\int_{(0,\infty)^2} g_{\boldsymbol{\theta}}(y_1, y_2)\,\mathrm{d}(y_1, y_2)\Big\} = 0,
\end{aligned}$$

and it directly follows from (5.9) that $\mathrm{PL}(\boldsymbol{\theta}) \leq \mathrm{PL}(\boldsymbol{\theta}^\star)$. As $\boldsymbol{\theta} \neq \boldsymbol{\theta}^\star$, the identifiability condition (5.14) yields (C). □

Remark 5.4. There are combinations of maximum space-time lags that lead to non-identifiable parameters, see Table 5.1. However, Theorem 5.3 still applies to all identifiable parameters (cf. Davis et al. [20], Remark 2).

Next we prove asymptotic normality of the PMLE defined in (5.12). As in the proof of The-

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

Table 5.1: Identifiable parameters for model (5.4) with $d = 2$ for some examples of maximum space-time lags (r_1, r_2, p).

orem 5.3 we follow the lines of proof of Davis et al. [20], Section 5, adapting the arguments to the anisotropic setting. We start with some basic results needed throughout the remainder of the section.

Lemma 5.5. *Assume that all conditions of Theorem 5.3 are satisfied. Then for* $\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)} \in \mathbb{R}^d$ *and* $t^{(1)}, t^{(2)} \in [0, \infty)$*, the following assertions hold componentwise:*

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

$$\mathbb{E}\left[\left|\nabla_{\boldsymbol{\theta}} \log g_{\boldsymbol{\theta}}(\eta(\boldsymbol{s}^{(1)}, t^{(1)}), \eta(\boldsymbol{s}^{(2)}, t^{(2)}))\right|^3\right] < \infty, \quad \boldsymbol{\theta} \in \Theta^\star.$$

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

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta} \in \Theta^\star} \left|\nabla^2_{\boldsymbol{\theta}} \log g_{\boldsymbol{\theta}}(\eta(\boldsymbol{s}^{(1)}, t^{(1)}), \eta(\boldsymbol{s}^{(2)}, t^{(2)}))\right|\right] < \infty.$$

Proof. Assume identifiability of all parameters C_j, α_j for $j = 1, \dots, d+1$. For $y_1, y_2 \in (0, \infty)$ and for $(\boldsymbol{h}, u) \in \mathbb{R}^{d+1} \setminus \{\boldsymbol{0}\}$ lengthy but simple calculations of derivatives of (5.5) yield

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

$$\frac{\partial \delta(\boldsymbol{h}, u)}{\partial C_j} = |h_j|^{\alpha_j}, \quad \frac{\partial \delta(\boldsymbol{h}, u)}{\partial \alpha_j} = C_j |h_j|^{\alpha_j} \log |h_j|, \quad j = 1, \dots d,$$

and

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

By compactness of the parameter space, as required in (5.13), we can bound those first partial derivatives as well as the second order partial derivatives from above and below. So it remains to show that for $\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)} \in S$ and $t^{(1)}, t^{(2)} \in T$,

$$\mathbb{E}_{\boldsymbol{\theta}^\star}\left[\left|\frac{\partial \log\{g_{\boldsymbol{\theta}}(\eta(\boldsymbol{s}^{(1)}, t^{(1)}), \eta(\boldsymbol{s}^{(2)}, t^{(2)}))\}}{\partial \delta(\boldsymbol{h}, u)}\right|^3\right] < \infty$$

and

$$\mathbb{E}_{\boldsymbol{\theta}^\star}\left[\sup_{\boldsymbol{\theta} \in \Theta^\star} \left|\frac{\partial^2 \log\{g_{\boldsymbol{\theta}}(\eta(\boldsymbol{s}^{(1)}, t^{(1)}), \eta(\boldsymbol{s}^{(2)}, t^{(2)}))\}}{\partial^2 \delta(\boldsymbol{h}, u)}\right|\right] < \infty,$$

where the function $\delta(\boldsymbol{h}, u)$ can be treated as a constant since it is bounded away from 0 by (5.13). Hence, for the rest of the proof we refer to that of Davis et al. [20], Lemma 1. □

For a central limit theorem we need certain mixing properties for a space-time setting (cf. Davis et al. [20], Section 5.1 and Huser and Davison [43], Section 3.2).

Definition 5.6 (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. Let d be some metric induced by a norm on $\mathbb{R}^{d+1}$. For $\Lambda_1, \Lambda_2 \subset \mathbb{Z}^d \times \mathbb{N}$ let*

$$d(\Lambda_1, \Lambda_2) := \inf\{d((\boldsymbol{s}^{(1)}, t^{(1)}), (\boldsymbol{s}^{(2)}, t^{(2)})) : (\boldsymbol{s}^{(1)}, t^{(1)}) \in \Lambda_1, (\boldsymbol{s}^{(2)}, t^{(2)}) \in \Lambda_2\}.$$

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

$$\begin{aligned} \alpha_{k,\ell}(n) := \sup\{&|\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| : \\ &A_1 \in \mathcal{F}_{\Lambda_1}, A_2 \in \mathcal{F}_{\Lambda_2}, |\Lambda_1| \leq k, |\Lambda_2| \leq \ell, d(\Lambda_1, \Lambda_2) \geq n\}, \end{aligned} \tag{5.16}$$

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

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

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

Recall from Eq. (5.2) that for $(\boldsymbol{h}, u) \in \mathbb{R}^{d+1}$ with δ as in (5.4) the tail dependence coefficient of the Brown-Resnick process is given by

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

Corollary 2.2 of Dombry and Eyi-Minko [27] links the α-mixing coefficients with the tail dependence coefficients, and we will use this for the next result.

Proposition 5.7. *Let $\{\eta(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\}$ be the Brown-Resnick process* (5.1) *with dependence function δ given by* (5.4)*. Then the process $\{\eta(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{Z}^d, t \in \mathbb{N}\}$ is α-mixing, where the mixing coefficients in (5.16) satisfy for $\mathcal{H}_{\boldsymbol{r}}$ as in* (5.7)

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

(2) $\alpha_{(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1),\infty}(n) = o(n^{-(d+1)})$ as $n \to \infty$,

(3) $\sum_{n=1}^{\infty} n^d \alpha_{(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1),(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1)}(n)^{\frac{1}{3}} < \infty$.

Proof. Note that for $(\boldsymbol{h}, u) \in \mathbb{R}^{d+1}$, by the equivalence of norms, for some positive constant L,

$$d((\boldsymbol{h}, u), (\boldsymbol{0}, 0)) \leq \frac{1}{L} \max\{|h_1|, \ldots, |h_d|, |u|\}$$

Therefore, for $n \in \mathbb{N}$, presuming $d((\boldsymbol{h},u),(\boldsymbol{0},0)) \geq n$ results in $\max\{|h_1|,\ldots,|h_d|,|u|\} \geq Ln$, so that by Corollary 2.2 and Eq. (3) of Dombry and Eyi-Minko [27] we get

$$\alpha_{k,\ell}(n) \leq 2k\ell \sup_{d((\boldsymbol{h},u),(\boldsymbol{0},0))\geq n} \chi(\boldsymbol{h},u) \leq 2k\ell \sup_{\max\{|h_1|,\ldots,|h_d|,|u|\}\geq Ln} \chi(\boldsymbol{h},u), \tag{5.17}$$

$$\alpha_{k,\infty}(n) \leq 2k \sum_{d((\boldsymbol{h},u),(\boldsymbol{0},0))\geq n} \chi(\boldsymbol{h},u) \leq 2k \sum_{\max\{|h_1|,\ldots,|h_d|,|u|\}\geq Ln} \chi(\boldsymbol{h},u). \tag{5.18}$$

In the following we use the notation $\|(\boldsymbol{h},u)\|_\infty := \max\{|h_1|,\ldots,|h_d|,|u|\}$ for $(\boldsymbol{h},u) \in \mathbb{Z}^d \times \mathbb{N}$. Using $1-\Phi(x) \leq \exp\{-\frac{1}{2}x^2\}$ for $x>0$ and Eq. (5.2) and (5.17), we find for all $k,\ell \geq 0$,

$$\begin{aligned}
\alpha_{k,\ell}(n) &\leq 4k\ell \sup_{\|(\boldsymbol{h},u)\|_\infty \geq Ln} \left(1-\Phi(\sqrt{\frac{\delta(\boldsymbol{h},u)}{2}}\right) \\
&\leq 4k\ell \sup_{\|(\boldsymbol{h},u)\|_\infty \geq Ln} \exp\left\{-\frac{\delta(\boldsymbol{h},u)}{4}\right\} \\
&= 4k\ell \sup_{\|(\boldsymbol{h},u)\|_\infty \geq Ln} \exp\left\{-\frac{1}{4}\left[C_1|h_1|^{\alpha_1}+\ldots+C_d|h_d|^{\alpha_d}+C_{d+1}|u|^{\alpha_{d+1}}\right]\right\} \\
&\leq 4k\ell \sup_{\|(\boldsymbol{h},u)\|_\infty \geq Ln} \exp\left\{-\frac{1}{4}\min\{|C_1|,\ldots,|C_{d+1}|\}\|(\boldsymbol{h},u)\|_\infty^{\min\{\alpha_1,\ldots,\alpha_{d+1}\}}\right\} \\
&\leq 4k\ell \exp\left\{-\frac{1}{4}\min\{|C_1|,\ldots,|C_{d+1}|\}(Ln)^{\min\{\alpha_1,\ldots,\alpha_{d+1}\}}\right\} \qquad (5.19) \\
&\to 0 \text{ as } n\to\infty.
\end{aligned}$$

This implies α-mixing.

By similar arguments we obtain by (5.18) for all $k \geq 0$,

$$\alpha_{k,\infty}(n) \leq 4k \sum_{\|(\boldsymbol{h},u)\|_\infty \geq Ln} \exp\left\{-\frac{1}{4}\min\{|C_1|,\ldots,|C_{d+1}|\}\|(\boldsymbol{h},u)\|_\infty^{\min\{\alpha_1,\ldots,\alpha_{d+1}\}}\right\}. \tag{5.20}$$

We use the above bounds to prove assertions (1)-(3).

(1) For $k+\ell \leq 4(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1)$ we have by (5.19),

$$\sum_{n=1}^\infty n^d \alpha_{k,\ell}(n) \leq 4k\ell \sum_{n=1}^\infty n^d \exp\left\{-\frac{1}{4}\min\{|C_1|,\ldots,|C_{d+1}|\}(Ln)^{\min\{\alpha_1,\ldots,\alpha_{d+1}\}}\right\} < \infty.$$

(2) First note that the number of grid points $(\boldsymbol{h},u) \in \mathbb{R}^{d+1}$ with $\|(\boldsymbol{h},u)\|_\infty = i$ for $i \in \mathbb{N}$ equals $(i+1)^{d+1} - i^{d+1}$, and is therefore of order $\mathcal{O}(i^d)$. We use (5.20) and a more precise estimate than in part (1) to obtain for sufficiently large n

$$\begin{aligned}
&n^{d+1}\alpha_{(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1),\infty}(n) \\
&\leq 4n^{d+1}(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1)
\end{aligned}$$

$$\sum_{\|(\boldsymbol{h},u)\|_\infty \geq Ln} \exp\left\{-\frac{1}{4}\min\{|C_1|,\ldots,|C_{d+1}|\}\|(\boldsymbol{h},u)\|_\infty^{\min\{\alpha_1,\ldots,\alpha_{d+1}\}}\right\}$$
$$\leq K_3 n^{d+1}(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1)\sum_{i=\lfloor Ln\rfloor}^{\infty} i^d \exp\Big\{-\frac{1}{4}\min\{C_1,\ldots,C_{d+1}\}i^{\min\{\alpha_1,\ldots,\alpha_{d+1}\}}\Big\}$$
$$\to 0 \text{ as } n\to\infty,$$

where K_3 is a positive constant. Convergence to 0 follows using the integral test for power series convergence and Lemma D.1, Eq. (D.1).

(3) We find, using again (5.19),

$$\begin{aligned}&\sum_{n=1}^{\infty} n^d \alpha_{(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1),(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1)}(n)^{\frac{1}{3}}\\ \leq&\big(4\big[(|\mathcal{H}_{\boldsymbol{r}}|+1)(p+1)\big]^2\big)^{\frac{1}{3}}\\ &\cdot\sum_{n=1}^{\infty} n^d \exp\Big\{-\frac{1}{12}\min\{C_1,\ldots,C_{d+1}\}(Ln)^{\min\{\alpha_1,\ldots,\alpha_{d+1}\}}\Big\}\\ <&\infty\end{aligned}$$

as in (1). □

Because of Lemma 5.5 and Proposition 5.7 the following central limit theorem of Bolthausen [6] holds.

Corollary 5.8. *Consider the process* $\{\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(\boldsymbol{s},t;\boldsymbol{r},p) : \boldsymbol{s}\in\mathbb{Z}^d, t\in\mathbb{N}\}$. *Then*

$$\frac{1}{M^{\frac{d}{2}}\sqrt{T}}\sum_{\boldsymbol{s}\in\mathcal{S}_M}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(\boldsymbol{s},t;\boldsymbol{r},p) \xrightarrow{d} \mathcal{N}(\boldsymbol{0},\Sigma_1) \text{ as } M,T\to\infty,$$

where

$$\Sigma_1 := \sum_{s_1=-\infty}^{\infty}\cdots\sum_{s_d=-\infty}^{\infty}\sum_{t=1}^{\infty}\operatorname{Cov}\left[\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(\boldsymbol{1},1;\boldsymbol{r},p), \nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(s_1,\ldots,s_d,t;\boldsymbol{r},p)\right]. \tag{5.21}$$

Now we formulate the main result of this section.

Theorem 5.9 (Asymptotic normality for large M and T)**.** *Assume the same conditions as in Theorem 5.3. Then*

$$\sqrt{M^dT}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^\star) \xrightarrow{d} \mathcal{N}(\boldsymbol{0},\tilde{\Sigma}_1) \text{ as } M,T\to\infty, \tag{5.22}$$

where $\tilde{\Sigma}_1 := F_1^{-1}\Sigma_1(F_1^{-1})^\top$ *with* Σ_1 *given in* (5.21) *and*

$$F_1 := \mathbb{E}\left[-\nabla^2_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(\boldsymbol{1},1;\boldsymbol{r},p)\right].$$

Proof. A Taylor expansion of the score function $\nabla_{\boldsymbol{\theta}} PL^{(M,T)}(\boldsymbol{\theta})$ around the true parameter vector

$\boldsymbol{\theta}^\star$ yields for some $\tilde{\boldsymbol{\theta}} \in [\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}^\star]$:

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

Therefore,

$$\begin{aligned}
M^{\frac{d}{2}}\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) &= -\Big(\frac{1}{M^d T}\nabla^2_{\boldsymbol{\theta}} PL^{(M,T)}(\tilde{\boldsymbol{\theta}})\Big)^{-1}\Big(\frac{1}{M^{\frac{d}{2}}\sqrt{T}}\nabla_{\boldsymbol{\theta}} PL^{(M,T)}(\boldsymbol{\theta}^\star)\Big) \\
&= -\Big(\frac{1}{M^d T}\sum_{\boldsymbol{s}\in\mathcal{S}_M}\sum_{t=1}^{T}\nabla^2_{\boldsymbol{\theta}} q_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{s},t;\boldsymbol{r},p) - \frac{1}{M^d T}\nabla^2_{\boldsymbol{\theta}}\mathcal{R}^{(M,T)}(\tilde{\boldsymbol{\theta}})\Big)^{-1} \\
&\quad \Big(\frac{1}{M^{\frac{d}{2}}\sqrt{T}}\sum_{\boldsymbol{s}\in\mathcal{S}_M}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(\boldsymbol{s},t;\boldsymbol{r},p) - \frac{1}{M^{\frac{d}{2}}\sqrt{T}}\nabla_{\boldsymbol{\theta}}\mathcal{R}^{(M,T)}(\boldsymbol{\theta}^\star)\Big) \\
&=: -(I_1 - I_2)^{-1}(J_1 - J_2).
\end{aligned}$$

Note the following:

- Corollary 5.8 implies that $J_1 \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma_1)$ as $M, T \to \infty$.

- Using representation (5.10) of the boundary term $\mathcal{R}^{(M,T)}(\cdot)$ and Lemma 5.2, we find

$$\begin{aligned}
\|J_2\| &= \frac{1}{M^{\frac{d}{2}}\sqrt{T}}\Big\|\sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}}\sum_{u=0}^{p}\sum_{(\boldsymbol{s},t)\in\mathcal{G}_{M,T}(\boldsymbol{h},u)}\nabla_{\boldsymbol{\theta}}\log\{g_{\boldsymbol{\theta}^\star}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))\}\Big\| \\
&\le \sqrt{K_2}\frac{\sqrt{M^{d-1}T + M^d}}{M^{\frac{d}{2}}\sqrt{T}} \\
&\quad \Big\|\sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}}\sum_{u=0}^{p}\frac{1}{\sqrt{|\mathcal{G}_{M,T}(\boldsymbol{h},u)|}}\sum_{(\boldsymbol{s},t)\in\mathcal{G}_{M,T}(\boldsymbol{h},u)}\nabla_{\boldsymbol{\theta}}\log\{g_{\boldsymbol{\theta}^\star}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))\}\Big\| \\
&\le \sqrt{K_2}(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{T}}) \\
&\quad \Big\|\sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}}\sum_{u=0}^{p}\frac{1}{\sqrt{|\mathcal{G}_{M,T}(\boldsymbol{h},u)|}}\sum_{(\boldsymbol{s},t)\in\mathcal{G}_{M,T}(\boldsymbol{h},u)}\nabla_{\boldsymbol{\theta}}\log\{g_{\boldsymbol{\theta}^\star}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))\}\Big\|
\end{aligned}$$

  In the same way as done in Corollary 5.8 for the process $\{\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}(\boldsymbol{s},t;\boldsymbol{r},p) : \boldsymbol{s}\in\mathbb{Z}^d, t\in\mathbb{N}\}$, we can apply Bolthausen's central limit theorem to the processes $\{\nabla_{\boldsymbol{\theta}}\log\{g_{\boldsymbol{\theta}^\star}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))\} : \boldsymbol{s}\in\mathbb{Z}^d, t\in\mathbb{N}\}$ for $\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}$, $u\in\{0,\ldots,p\}$. We conclude that

$$\sum_{\boldsymbol{h}\in\mathcal{H}_{\boldsymbol{r}}}\sum_{u=0}^{p}\frac{1}{\sqrt{|\mathcal{G}_{M,T}(\boldsymbol{h},u)|}}\sum_{(\boldsymbol{s},t)\in\mathcal{G}_{M,T}(\boldsymbol{h},u)}\nabla_{\boldsymbol{\theta}}\log\{g_{\boldsymbol{\theta}^\star}(\eta(\boldsymbol{s},t),\eta(\boldsymbol{s}+\boldsymbol{h},t+u))\}$$

  converges weakly to a normal distribution as $M, T \to \infty$, and it follows that $J_2 \xrightarrow{P} \mathbf{0}$ as $M, T \to \infty$.

- As $\{\eta(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{Z}^d, t \in \mathbb{N}\}$ is α-mixing, the process

$$\{\nabla^2_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}}(\boldsymbol{s},t;\boldsymbol{r},p) : \boldsymbol{s} \in \mathbb{Z}^d, t \in \mathbb{N}\}$$

  is α-mixing as a set of measurable functions of mixing lagged processes. Furthermore, as $\tilde{\boldsymbol{\theta}} \in [\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}^\star]$ and $\widehat{\boldsymbol{\theta}}$ is strongly consistent, we have that $I_1 \overset{\text{a.s.}}{\to} -F_1$ as $M, T \to \infty$. The convergence is uniform on $\Theta^\star$ by Lemma 5.5 which implies that

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

- Concerning I_2, the law of large numbers applied to

$$\left\{\nabla^2_{\boldsymbol{\theta}} \log\{g_{\boldsymbol{\theta}}(\eta(\boldsymbol{s},t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u))\} : \boldsymbol{s} \in \mathbb{Z}^d, t \in \mathbb{N}\right\}$$

  results in the fact that, in the same way as in part (B) of the proof of Theorem 5.3, $I_2 \overset{\text{a.s.}}{\to} \mathbf{0}$ as $M, T \to \infty$.

Finally, summarising these results, Slutzky's Lemma yields (5.22). □

5.3.2 Fixed spatial domain and increasing temporal domain

As before we compute the PMLE based on observations on the area $\mathcal{S}_M \times \mathcal{T}_T$, but now we consider M fixed, whereas T tends to infinity.

We define the temporal α-mixing coefficients (cf. Ibragimov and Linnik [45], Definition 17.2.1 or Bradley [7], Definition 1.6).

Definition 5.10 (Temporal mixing coefficients and temporal α-mixing)**.** *Let $\{\eta(\boldsymbol{s},t) : \boldsymbol{s} \in \mathcal{S}_M, t \in \mathbb{N}\}$ be a space-time process. Consider the metric $d(\cdot)$ of Definition 5.6.*

(1) Let $\mathcal{T}^{(1)}, \mathcal{T}^{(2)} \subset \mathbb{N}$. For $n \geq 0$ the temporal α-mixing coefficients are defined as

$$\begin{aligned}\alpha(n) := \sup\{&|P(A_1 \cap A_2) - P(A_1)P(A_2)| :\\ &A_1 \in \mathcal{F}_{\mathcal{S}_M\times\mathcal{T}^{(1)}}, A_2 \in \mathcal{F}_{\mathcal{S}_M\times\mathcal{T}^{(2)}}, d(\mathcal{S}_M \times \mathcal{T}^{(1)}, \mathcal{S}_M \times \mathcal{T}^{(2)}) \geq n\},\end{aligned} \tag{5.23}$$

where $\mathcal{F}_{\mathcal{S}_M\times\mathcal{T}^{(i)}} = \sigma(\eta(\boldsymbol{s},t) : (\boldsymbol{s},t) \in \mathcal{S}_M \times \mathcal{T}^{(i)})$ for $i = 1, 2$.

(2) $\{\eta(\boldsymbol{s},t) : \boldsymbol{s} \in \mathcal{S}_M, t \in \mathbb{N}\}$ is called temporally α-mixing, if

$$\alpha(n) \to 0, \quad n \to \infty. \tag{5.24}$$

Proposition 5.11. *Let $\{\eta(\boldsymbol{s},t) : \boldsymbol{s} \in \mathbb{R}^d, t \in [0,\infty)\}$ be the Brown-Resnick process* (5.1) *with dependence function δ given by* (5.4)*. Then the process $\{\eta(\boldsymbol{s},t) : \boldsymbol{s} \in \mathcal{S}_M, t \in \mathbb{N}\}$ is temporally α-mixing, where the mixing coefficients* (5.23) *satisfy*

$$\sum_{n=1}^{\infty} |\alpha(n)|^{\frac{1}{3}} < \infty. \tag{5.25}$$

Proof. We use Eq. (3) and Corollary 2.2 of Dombry and Eyi-Minko [27] and (5.2) to obtain for $n \in \mathbb{N}$

$$\begin{aligned}
&\alpha(n) \\
&\leq 2 \sup_{d(\mathcal{S}_M\times\mathcal{T}^{(1)},\mathcal{S}_M\times\mathcal{T}^{(2)})\geq n} \sum_{\substack{(\boldsymbol{s}^{(1)},t^{(1)})\\ \in\mathcal{S}_M\times\mathcal{T}^{(1)}}} \sum_{\substack{(\boldsymbol{s}^{(2)},t^{(2)})\\ \in\mathcal{S}_M\times\mathcal{T}^{(2)}}} \chi(\boldsymbol{s}^{(1)}-\boldsymbol{s}^{(2)},t^{(1)}-t^{(2)}) \\
&= 4 \sup_{d(\mathcal{S}_M\times\mathcal{T}^{(1)},\mathcal{S}_M\times\mathcal{T}^{(2)})\geq n} \sum_{\substack{(\boldsymbol{s}^{(1)},t^{(1)})\\ \in\mathcal{S}_M\times\mathcal{T}^{(1)}}} \sum_{\substack{(\boldsymbol{s}^{(2)},t^{(2)})\\ \in\mathcal{S}_M\times\mathcal{T}^{(2)}}} \\
&\quad \left(1-\Phi\left(\sqrt{\frac{1}{2}\big[C_1|s_1^{(1)}-s_1^{(2)}|^{\alpha_1}+\cdots+C_d|s_d^{(1)}-s_d^{(2)}|^{\alpha_d}+C_{d+1}|t^{(1)}-t^{(2)}|^{\alpha_{d+1}}\big]}\right)\right) \\
&\leq 4M^{2d} \sup_{d(\mathcal{S}_M\times\mathcal{T}^{(1)},\mathcal{S}_M\times\mathcal{T}^{(2)})\geq n} \sum_{\substack{(t^{(1)},t^{(2)})\\ \in\mathcal{T}^{(1)}\times\mathcal{T}^{(2)}}} \left(1-\Phi\left(\sqrt{\frac{1}{2}\big[C_{d+1}|t^{(1)}-t^{(2)}|^{\alpha_{d+1}}\big]}\right)\right) \\
&\leq 4M^{2d} \sup_{d(\mathcal{S}_M\times\mathcal{T}^{(1)},\mathcal{S}_M\times\mathcal{T}^{(2)})\geq n} \sum_{\substack{(t^{(1)},t^{(2)})\\ \in\mathcal{T}^{(1)}\times\mathcal{T}^{(2)}}} \exp\Big\{-\frac{1}{4}C_{d+1}|t^{(1)}-t^{(2)}|^{\alpha_{d+1}}\Big\},
\end{aligned}$$

where the last inequality follows from $1-\Phi(x) \leq \exp\{-\frac{1}{2}x^2\}$ for $x > 0$. We bound $\alpha(n)$ for large n further by

$$\alpha(n) \leq 4M^{2d} \sum_{t^{(1)}\in\{-\infty,\ldots,0\}} \sum_{t^{(2)}\in\{n,\ldots,\infty\}} \exp\Big\{-\frac{1}{4}C_{d+1}|t^{(1)}-t^{(2)}|^{\alpha_{d+1}}\Big\}.$$

In the double sum a temporal lag $u = |t^{(1)}-t^{(2)}| \geq n$ appears exactly $u-(n-1)$ times. This yields

$$\begin{aligned}
\alpha(n) &\leq 4M^{2d}\sum_{u=n}^{\infty}(u-(n-1))\exp\Big\{-\frac{1}{4}C_{d+1}u^{\alpha_{d+1}}\Big\} \\
&\leq 4M^{2d}\sum_{u=n}^{\infty}u\exp\Big\{-\frac{1}{4}C_{d+1}u^{\alpha_{d+1}}\Big\}.
\end{aligned}$$

Convergence of the series (5.25) now follows by the integral test and Lemma D.1. □

In the following we show that strong consistency of the PMLE also holds, if the spatial domain remains fixed.

Theorem 5.12 (Strong consistency for fixed M and large T)**.** *Assume the same conditions as in Theorem 5.3 restricted to the fixed space $\mathcal{S}_M$. Then the PMLE*

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

is strongly consistent, that is,

$$\widehat{\boldsymbol{\theta}}^{(M,T)} \overset{\text{a.s.}}{\to} \boldsymbol{\theta}^\star \text{ as } T\to\infty.$$

Proof. For $\boldsymbol{\theta} \in \Theta^\star$ and $t \in \mathbb{N}$, set

$$q_{\boldsymbol{\theta}}^M(t; \boldsymbol{r}, p) := \sum_{\boldsymbol{s} \in \mathcal{S}_M} \sum_{\substack{\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{r}} \\ \boldsymbol{s}+\boldsymbol{h} \in \mathcal{S}_M}} \sum_{\substack{u=0 \\ t+u \leq T}}^{p} \mathbb{1}_{\{(\boldsymbol{h},u) \neq (\boldsymbol{0},0)\}} \log \left\{ g_{\boldsymbol{\theta}} \left(\eta(\boldsymbol{s}, t), \eta(\boldsymbol{s}+\boldsymbol{h}, t+u) \right) \right\}.$$

Then

$$PL^{(M,T)}(\boldsymbol{\theta}) = \sum_{t=1}^{T} q_{\boldsymbol{\theta}}^M(t; \boldsymbol{r}, p).$$

Following carefully the lines of the proof of Theorem 5.3, the following conditions hold for fixed spatial domain:

(A) $\frac{1}{T} \sum_{t=1}^{T} q_{\boldsymbol{\theta}}^M(t; \boldsymbol{r}, p) \overset{\text{a.s.}}{\to} PL^M(\boldsymbol{\theta}) := \mathbb{E}[(q_{\boldsymbol{\theta}}^M(1; \boldsymbol{r}, p)]$ as $T \to \infty$ uniformly on the compact parameter space $\Theta^\star$. The main argument is that $q_{\boldsymbol{\theta}}^M(\cdot)$ is a function of temporally mixing lagged processes, then we apply again Theorem 2.7 of Straumann [63].

(B) The limit function $PL^M(\boldsymbol{\theta})$ is uniquely maximised at the true parameter vector $\boldsymbol{\theta}^\star \in \Theta^\star$.

□

Now we formulate the main result of this section.

Theorem 5.13 (Asymptotic normality for fixed M and large T)**.** *Assume the same conditions as in Theorem 5.3 restricted to the fixed space $\mathcal{S}_M$. Then*

$$\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) \overset{d}{\to} \mathcal{N}(\boldsymbol{0}, \tilde{\Sigma}_2) \text{ as } T \to \infty, \tag{5.26}$$

where $\tilde{\Sigma}_2 := F_2^{-1} \Sigma_2 (F_2^{-1})^\top$ *with*

$$F_2 := \mathbb{E}[-\nabla_{\boldsymbol{\theta}}^2 q_{\boldsymbol{\theta}^\star}^M(1; \boldsymbol{r}, p)]$$

and

$$\Sigma_2 := \mathbb{V}\text{ar}[\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}^M(1; \boldsymbol{r}, p)] + 2 \sum_{t=2}^{\infty} \mathbb{C}\text{ov}[\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}^M(1; \boldsymbol{r}, p), \nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}^M(t; \boldsymbol{r}, p)].$$

Proof. By its definition as a function of lagged temporally mixing processes, $(\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}^M(t; \boldsymbol{r}, p))_{t \in \mathbb{N}}$ is also temporally α-mixing with coefficients $\alpha'(n) = \alpha(n-p)$. Furthermore,

$$\mathbb{E}\left[\nabla_{\boldsymbol{\theta}} \log \left\{ g_{\boldsymbol{\theta}^\star} \left(\eta(\boldsymbol{0}, 0), \eta(\boldsymbol{h}, u) \right) \right\}\right] = 0, \quad (\boldsymbol{h}, u) \in \mathbb{N}_0^{d+1},$$

because Lemma 5.5 implies regularity conditions of the pairwise log-likelihood (5.8) allowing to interchange differentiation and integration. Now note that Lemma 5.5 and Proposition 5.11 imply that

- $\mathbb{E}[|\nabla_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}^\star}^M(t; \boldsymbol{r}, p)|^3] < \infty$ for $t \in \mathbb{N}$ and every maximum spatial lag $\boldsymbol{r}$ and time lag p, and that

- $\sum_{n=1}^{\infty} |\alpha'(n)|^{\frac{1}{3}} < \infty$.

Therefore, the conditions of Theorem 18.5.3 of Ibragimov and Linnik [45] (see also Bradley [7], Theorem 10.7) are satisfied and we conclude that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} q^M_{\boldsymbol{\theta}^\star}(t;\boldsymbol{r},p) \xrightarrow{d} \mathcal{N}(\mathbf{0},\Sigma_2) \text{ as } T \to \infty. \tag{5.27}$$

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

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

Therefore,

$$\begin{aligned}\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) &= -\Big(\frac{1}{T}\nabla^2_{\boldsymbol{\theta}} PL^{(M,T)}(\tilde{\boldsymbol{\theta}})\Big)^{-1}\Big(\frac{1}{\sqrt{T}}\nabla_{\boldsymbol{\theta}} PL^{(M,T)}(\boldsymbol{\theta}^\star)\Big) \\ &= -\Big(\frac{1}{T}\sum_{t=1}^{T} \nabla^2_{\boldsymbol{\theta}} q^M_{\tilde{\boldsymbol{\theta}}}(t;\boldsymbol{r},p)\Big)^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} q^M_{\boldsymbol{\theta}^\star}(t;\boldsymbol{r},p)\Big) \;=:\; -I^{-1}J.\end{aligned}$$

Note the following:

- (5.27) implies that $J \xrightarrow{d} \mathcal{N}(\mathbf{0},\Sigma_2)$ as $T \to \infty$.
- Uniform convergence holds because of Lemma 5.5 which implies that componentwise

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

  By temporal α-mixing, since $\tilde{\boldsymbol{\theta}} \in [\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}^\star]$, and $\widehat{\boldsymbol{\theta}}$ is strongly consistent, we have $I \overset{\text{a.s.}}{\to} -F_2$ as $T \to \infty$.

Finally, summarising those results, Slutzky's Lemma yields (5.26). □

Throughout this section we have proved asymptotic properties of the parameter estimates of model (5.4) by classical results for ML estimators in combination with a spatio-temporal central limit theorem. Such results can also be applied to other models like geometrically anisotropic models, provided the required rates for α-mixing hold.

5.4 Test for spatial isotropy

We use the results of Section 5.3 to formulate statistical tests for spatial isotropy versus anisotropy based on the model (5.4),

$$\delta(\boldsymbol{h},u) = \sum_{j=1}^{d} C_j |h_j|^{\alpha_j} + C_{d+1}|u|^{\alpha_{d+1}},$$

for spatial lags $(\boldsymbol{h}, u) = (h_1, \ldots, h_d, u) \in \mathbb{R}^{d+1}$. We derive the necessary results for $d = 2$. Generalisations to higher dimensions are possible, but notationally much more involved. Again we consider the two cases of an increasing and fixed spatial domain.

Due to the structure of model (5.4) a test for isotropy versus anisotropy is a test of

$$H_0 : \{C_1 = C_2 \text{ and } \alpha_1 = \alpha_2\} \quad \text{versus} \quad H_1 : \{C_1 \neq C_2 \text{ or } \alpha_1 \neq \alpha_2\}. \tag{5.28}$$

5.4.1 Increasing spatial domain

From Theorem 5.9 we know that, under suitable regularity conditions, the PMLE

$$\widehat{\boldsymbol{\theta}} = (\widehat{C}_1, \widehat{C}_2, \widehat{C}_3, \widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3)$$

is asymptotically normal; more precisely, for M^2 spatial observations on a regular grid and for T equidistant time points we have

$$M\sqrt{T} \begin{pmatrix} \widehat{C}_1 - C_1 \\ \widehat{C}_2 - C_2 \\ \widehat{C}_3 - C_3 \\ \widehat{\alpha}_1 - \alpha_1 \\ \widehat{\alpha}_2 - \alpha_2 \\ \widehat{\alpha}_3 - \alpha_3 \end{pmatrix} \quad \stackrel{d}{\to} \quad \mathcal{N}(\mathbf{0}, \tilde{\Sigma}_1) \text{ as } M, T \to \infty, \tag{5.29}$$

where $\tilde{\Sigma}_1 \in \mathbb{R}^{6\times 6}$ is the asymptotic covariance matrix given in Theorem 5.9.

Our test is based on the spatial parameters only. Moreover, we test the two equalities in H_0 separately and use Bonferroni's inequality to solve the multiple test problem.

Lemma 5.14. *Assume the conditions of Theorem 5.9. Setting* $A_1 := (-1, 1, 0, 0, 0, 0)$ *and* $A_2 := (0, 0, 0, -1, 1, 0)$, *we have that, as* $M, T \to \infty$,

$$M\sqrt{T}((\widehat{C}_2 - \widehat{C}_1) - (C_2 - C_1)) \stackrel{d}{\to} \mathcal{N}(0, A_1\tilde{\Sigma}_1 A_1^\top), \tag{5.30}$$

$$M\sqrt{T}((\widehat{\alpha}_2 - \widehat{\alpha}_1) - (\alpha_2 - \alpha_1)) \stackrel{d}{\to} \mathcal{N}(0, A_2\tilde{\Sigma}_1 A_2^\top). \tag{5.31}$$

Proof. We obtain the left hand side of (5.30) and (5.31) by multiplying A_1 and A_2 to (5.29), respectively. This yields the limits on the right hand side by the continuous mapping theorem. □

We define

$$\theta_C := (C_2 - C_1), \widehat{\theta}_C := (\widehat{C}_2 - \widehat{C}_1), \quad \theta_\alpha := (\alpha_2 - \alpha_1), \widehat{\theta}_\alpha := (\widehat{\alpha}_2 - \widehat{\alpha}_1).$$

Then the multiple test problem (5.28) becomes

$$H_{0,1} : \{\theta_C = 0\} \quad \text{versus} \quad H_{1,1} : \{\theta_C \neq 0\} \tag{5.32}$$

$$H_{0,2} : \{\theta_\alpha = 0\} \quad \text{versus} \quad H_{1,2} : \{\theta_\alpha \neq 0\}. \tag{5.33}$$

Since the variances in (5.30) and (5.31) are not known explicitly, we find the rejection areas of the two tests by subsampling as suggested in Politis et al. [56], Chapter 5. Their main Assumption 5.3.1, the existence of a weak limit law of the estimates, is satisfied by Lemma 5.14.

We formulate the subsampling procedure in the terminology of the space-time process $\{\eta(\boldsymbol{s},t) : \boldsymbol{s} \in \mathcal{S}_M, t \in \mathcal{T}_T\}$. We choose space-time block lengths $\boldsymbol{b} = (b_1, b_2, b_3) \geq (1,1,1)$ and the degree of overlap $\boldsymbol{e} = (e_1, e_2, e_3) \leq (M, M, T)$. The blocks are indexed by $\boldsymbol{i} = (i_1, i_2, i_3) \in \mathbb{N}^3$ with $i_j \leq q_j$ for $q_j := \lfloor \frac{M-b_j}{e_j} \rfloor + 1$, $j = 1,2$ and $q_3 := \lfloor \frac{T-b_j}{e_j} \rfloor + 1$. This results in a total number of $q = q_1 q_2 q_3$ blocks, which we summarise in the set

$$\begin{aligned} E_{\boldsymbol{i},\boldsymbol{b},\boldsymbol{e}} \quad = \quad & \big\{(s_1, s_2, t) \in \mathcal{S}_M \times \mathcal{T}_T : (i_j - 1)e_j + 1 \leq s_j \leq (i_j - 1)e_j + b_j, j = 1, 2, \\ & (i_3 - 1)e_3 + 1 \leq t \leq (i_3 - 1)e_3 + b_3\big\}. \end{aligned}$$

Now we estimate θ_C and θ_α based on all observations in a block, hence getting q different estimates, which we denote by $\widehat{\theta}_{C,\boldsymbol{b},\boldsymbol{i}}$ and $\widehat{\theta}_{\alpha,\boldsymbol{b},\boldsymbol{i}}$.

In order to find rejection areas for the isotropy test, we will use Lemma 5.14, and take care of the unknown variance in the normal limit by a subsampling result.

Theorem 5.15. *Denote by $\tau_{M,T} := M\sqrt{T}$ and $\tau_{\boldsymbol{b}} = \sqrt{b_1 b_2 b_3}$ the square roots of the number of observations in total and in each block, respectively. Assume that the conditions of Theorem 5.9 hold and, as $M, T \to \infty$,*

(i) $b_i \to \infty$ for $i = 1,2,3$, such that $b_i = o(M)$ for $i = 1,2$, and $b_3 = o(T)$ (hence, $\tau_{\boldsymbol{b}}/\tau_{M,T} \to 0$),

(ii) $\boldsymbol{e}$ does not depend on M or T.

In the following $\widehat{\theta}$ stands for either $\widehat{\theta}_C$ or $\widehat{\theta}_\alpha$. Define the empirical distribution function

$$L_{\boldsymbol{b},\widehat{\theta}}(x) := \frac{1}{q} \sum_{i_1=1}^{q_1} \sum_{i_2=1}^{q_2} \sum_{i_3=1}^{q_3} \mathbf{1}_{\{\tau_{\boldsymbol{b}} |\widehat{\theta}_{\boldsymbol{b},\boldsymbol{i}} - \widehat{\theta}| \leq x\}}, \quad x \in \mathbb{R}, \tag{5.34}$$

and the empirical quantile function

$$c_{\boldsymbol{b},\widehat{\theta}}(1-\beta) := \inf\left\{x \in \mathbb{R} : L_{\boldsymbol{b},\widehat{\theta}}(x) \geq 1 - \beta\right\}, \quad \beta \in (0,1). \tag{5.35}$$

Then the following statements hold for $M, T \to \infty$:

(1) Denote by $\Phi_\sigma(\cdot)$ the distribution function of a mean 0 normal random variable Z with variance

$$\sigma^2 = \begin{cases} A_1 \tilde{\Sigma}_1 A_1^\top, & \text{in case of } \widehat{\theta}_C, \\ A_2 \tilde{\Sigma}_1 A_2^\top, & \text{in case of } \widehat{\theta}_\alpha, \end{cases}$$

and recall that $2\Phi_\sigma(\cdot) - 1$ is the distribution function of $|Z|$. Then

$$L_{\boldsymbol{b},\widehat{\theta}}(x) \quad \xrightarrow{P} \quad 2\Phi_\sigma(x) - 1, \quad x \in \mathbb{R}.$$

(2) Set $J_{\widehat{\theta}}(x) := \mathbb{P}(\tau_{M,T}|\widehat{\theta} - \theta| \leq x)$ *for* $x \in \mathbb{R}$*, then*

$$\sup_{x \in \mathbb{R}} \left| L_{\boldsymbol{b},\widehat{\theta}}(x) - J_{\widehat{\theta}}(x) \right| \quad \xrightarrow{P} \quad 0.$$

(3) For $\beta \in (0,1)$*,*

$$\mathbb{P}\left(\tau_{M,T}|\widehat{\theta} - \theta| \leq c_{\boldsymbol{b},\widehat{\theta}}(1-\beta)\right) \to 1 - \beta. \tag{5.36}$$

Proof. We apply Corollary 5.3.1 of Politis et al. [56]. Their main Assumption 5.3.1; i.e., the existence of a continuous limit distribution, is satisfied by Lemma 5.14. Assumptions (i)-(ii) are also presumed by Politis et al. [56]. The required condition on the α-mixing coefficients is satisfied similarly as in the proof of Proposition 5.7 by Lemma D.1 and the result holds. □

From (5.36), we find rejection areas for the test statistics $\tau_{M,T}\widehat{\theta}$ at confidence level $\beta \in (0,1)$ as (recall that $\widehat{\theta}$ stands for either $\widehat{\theta}_C$ or $\widehat{\theta}_\alpha$)

$$\mathrm{Rej}_{\widehat{\theta}}^{(M,T)} := (-\infty, -c_{\boldsymbol{b},\widehat{\theta}}(1-\beta)) \cup (c_{\boldsymbol{b},\widehat{\theta}}(1-\beta), \infty) = [-c_{\boldsymbol{b},\widehat{\theta}}(1-\beta), c_{\boldsymbol{b},\widehat{\theta}}(1-\beta)]^c.$$

Bonferroni's inequality

$$\mathbb{P}(\text{reject } H_{0,1} \text{ or } H_{0,2}) \leq \mathbb{P}(\text{reject } H_{0,1}) + \mathbb{P}(\text{reject } H_{0,2}) \leq 2\beta,$$

applies and solves the multiple test problem.

5.4.2 Fixed spatial domain

First note that an analogue of Lemma 5.14 holds with rate $\sqrt{T}$ instead of $M\sqrt{T}$ and with the asymptotic covariance matrix $\tilde{\Sigma}_2$ as given in Theorem 5.13.

The subsampling statement corresponding to Theorem 5.15 then reads as follows.

Theorem 5.16. *Denote by* $\tau_T := \sqrt{T}$ *and* $\tau_{b_3} = \sqrt{b_3}$ *the square roots of the number of time points of observations in total and in each block, respectively. Assume that the conditions of Theorem 5.13 are satisfied and that Lemma 5.14 holds for* $T \to \infty$ *with rate* $\sqrt{T}$ *instead of* $M\sqrt{T}$ *and with the asymptotic covariance matrix* $\tilde{\Sigma}_2$ *as given in Theorem 5.13. Assume further that as* $T \to \infty$*,*

(i) $b_3 \to \infty$ *such that* $b_3 = o(T)$ *(hence,* $\tau_{b_3}/\tau_T \to 0$*),*

(ii) $\boldsymbol{e}$ *does not depend on* T*,*

(iii) $b_1, b_2 \to M$*.*

Let $\boldsymbol{b} = (b_1, b_2, b_3)$*,* $\tau_{\boldsymbol{b}} = \sqrt{b_1 b_2 b_3}$ *and* $\tau_{M,T} = M\sqrt{T}$*. With* $\tilde{\Sigma}_1$ *as in Theorem 5.15 replaced by* $M^2\tilde{\Sigma}_2$*, conclusions (a), (b), and (c) of Theorem 5.15 remain true as* T *tends to infinity.*

Proof. We apply Corollary 5.3.2 of Politis et al. [56]. The required temporal mixing condition is satisfied similarly as in the proof of Proposition 5.11 by Lemma D.1. □

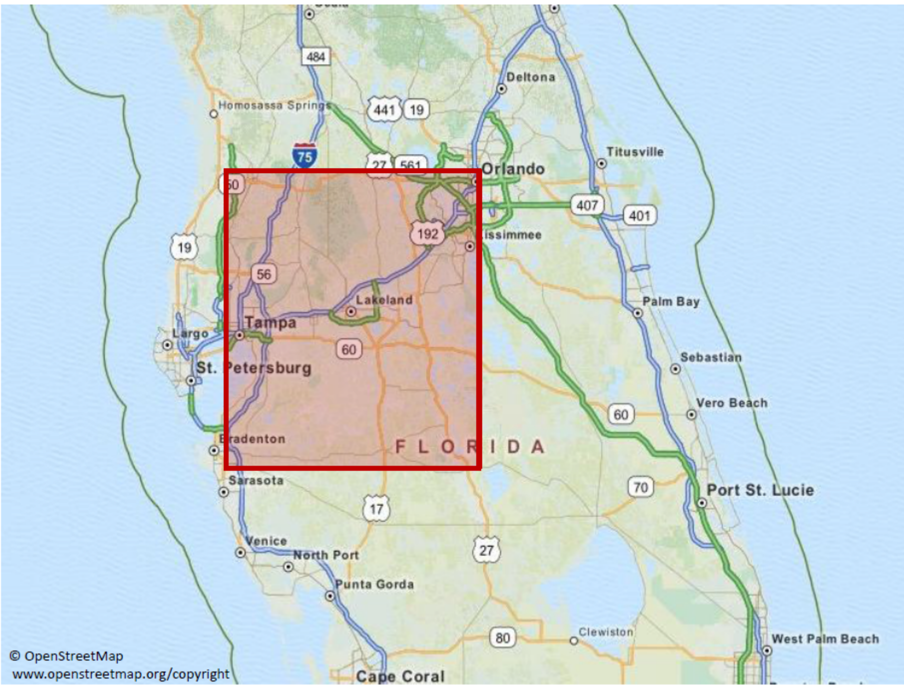

Figure 5.1: Rainfall observation area in Florida

Remark 5.17. We can in practice apply the same procedure of subsampling as in Section 5.4.1. This is justified by the fact that $\tau_{b_3}/\tau_T \to 0$ implies that $\tau_{\boldsymbol{b}}/\tau_{M,T} \to 0$ as $T \to \infty$ under conditions (i)-(iii) of Theorem 5.16. In particular, the rejection area for $\tau_T\widehat{\theta}$ (where again $\widehat{\theta}$ stands for either $\widehat{\theta}_C$ or $\widehat{\theta}_\alpha$) is found as

$$\mathrm{Rej}_{\widehat{\theta}}^{(T)} := \frac{1}{M}\mathrm{Rej}_{\widehat{\theta}}^{(M,T)}.$$

5.5 Data analysis

We fit the Brown-Resnick space-time process (5.1) with dependence structure given by the model (5.4) to radar rainfall data, which were provided by the Southwest Florida Water Management District (SWFWMD). The data used for the analysis are rainfall measurements on a square of 120km×120km in Florida (see Figure 5.1) over the years 1999-2004. The raw data consist of measurements in inches on a regular grid in space every two kilometres and every 15 minutes. Since there exist wet seasons and dry seasons with almost no rain we consider only the wet season June-September. Moreover, the area is basically flat with predominant easterly winds due to its closeness to the equator and, therefore, existing trade winds. Hence, (5.4) with parameters that possibly differ along both spatial axes fits well without introducing a rotation matrix.

5.5.1 Data transformation and marginal modelling

We carry out a block-maxima method in space and time as follows: We calculate cumulated hourly rainfall by adding up four consecutive measurements. Then we take block-maxima over 24 consecutive hours and over 10km×10km areas; i.e., the daily maxima over 25 locations, resulting in a 12×12 grid in space for all 6×122 days of the wet seasons giving a time series of dimension 12×12 and of length 732. Taking smaller areas than 10km×10km squares or a higher temporal resolution (e.g. 12-hour-maxima) results in observations that are not max-stable and the max-stability test described in Section 5.5.2 would reject.

By removing possible seasonal effects, we transform the data to stationarity. We obtain the observations

$$\{\tilde{\eta}((s_1, s_2), t) : s_1, s_2 = 1, \ldots, 12, t = 1, \ldots, 732\}. \tag{5.37}$$

Taking daily maxima removes for every location most of the dependence in the time series. This implies that marginal parameter estimates found by maximum likelihood estimation are consistent and asymptotically normal.

To give some details: for each fixed location (s_1, s_2), we fit a univariate *generalised extreme value distribution* (cf. Embrechts et al. [32], Definition 3.4.1) to the associated time series. The estimated shape parameters are all sufficiently close to 0 to motivate a Gumbel distribution as appropriate model. We therefore fit a Gumbel distribution $\Lambda_{\mu,\sigma}(x) = \exp\{-e^{-\frac{x-\mu}{\sigma}}\}$ with parameters $\mu = \mu(s_1, s_2) \in \mathbb{R}$ and $\sigma = \sigma(s_1, s_2) > 0$ and obtain estimates $\widehat{\mu} = \widehat{\mu}(s_1, s_2)$ and $\widehat{\sigma} = \widehat{\sigma}(s_1, s_2)$.

Depending on different statistical questions and methods, we transform (5.37) either to standard Gumbel or standard Fréchet margins. In the first case we set

$$\eta_1((s_1, s_2), t) := \frac{\tilde{\eta}((s_1, s_2), t) - \widehat{\mu}}{\widehat{\sigma}}, \quad t = 1, \ldots, 732, \tag{5.38}$$

and in the latter case, with $\Lambda_{\widehat{\mu},\widehat{\sigma}}$ denoting the Gumbel distribution with estimated parameters,

$$\eta_2((s_1, s_2), t) := -\frac{1}{\log\left\{\Lambda_{\widehat{\mu},\widehat{\sigma}}(\tilde{\eta}((s_1, s_2), t))\right\}}, \quad t = 1, \ldots, 732. \tag{5.39}$$

We assess the goodness of the marginal fits by qq-plots of the observations (5.38) versus the standard Gumbel quantiles for every spatial location. Figure 5.2 depicts the qq-plots at four exemplary spatial locations $(1, 1)$, $(6, 8)$, $(9, 4)$ and $(11, 10)$. * Confidence bounds are based on the Kolmogorov-Smirnov statistic (cf. Doksum and Sievers [26], Theorem 1 and Remark 1). All graphs show a reasonably good fit.

In the following data analysis we regard (5.39) as realisations of the space-time Brown-Resnick process (5.1) with dependence structure δ as in (5.4):

$$\delta(h_1, h_2, u) = C_1|h_1|^{\alpha_1} + C_2|h_2|^{\alpha_2} + C_3|u|^{\alpha_3}, \tag{5.40}$$

with $h_1 = s_1^{(1)} - s_1^{(2)}$, $h_2 = s_2^{(1)} - s_2^{(2)}$, $u = t^{(1)} - t^{(2)}$, for two spatial locations $\boldsymbol{s}^{(1)} = (s_1^{(1)}, s_2^{(1)})$ and $\boldsymbol{s}^{(2)} = (s_1^{(2)}, s_2^{(2)})$ and two time points $t^{(1)}$ and $t^{(2)}$.

*We use the R-package `extRemes` (Gilleland and Katz [38]).

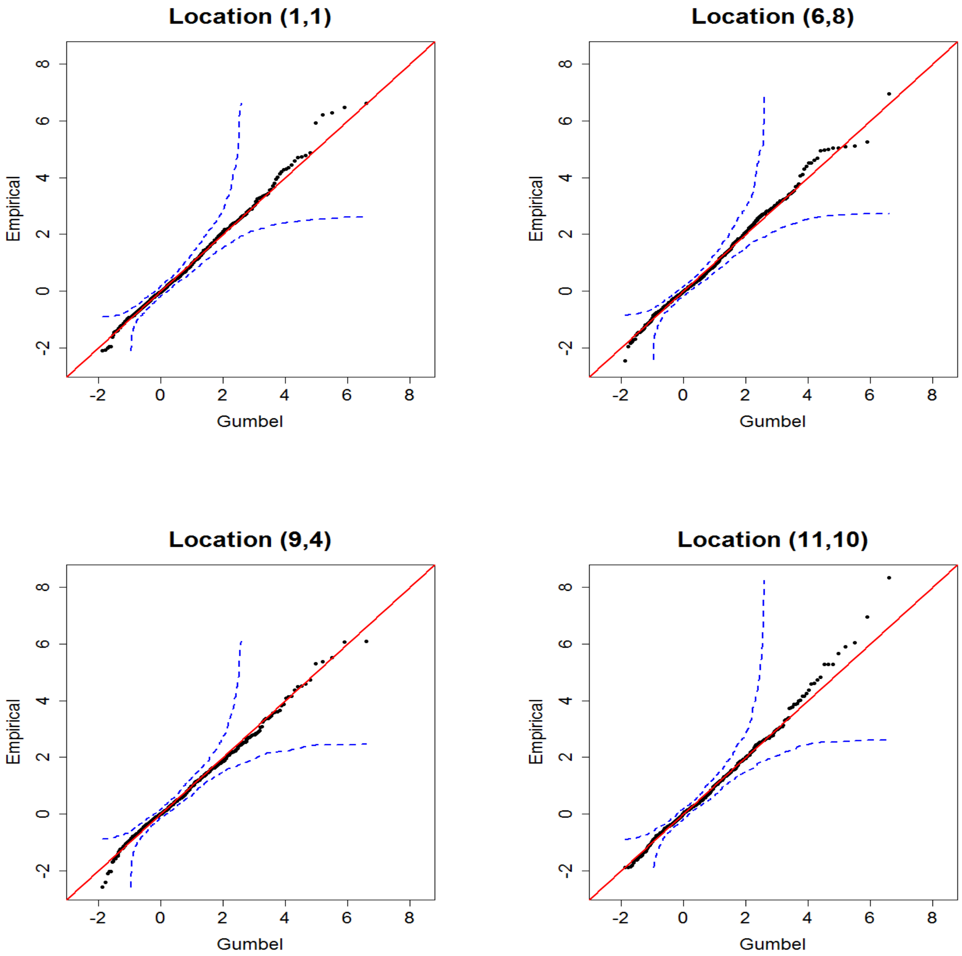

Figure 5.2: qq-plots of the Gumbel transformed time series values versus the standard Gumbel distribution for four locations: (1,1) (top left), (6,8) (top right), (9,4) (bottom left) and (11,10) (bottom right). Dashed blue lines mark 95% confidence bounds. Solid red lines correspond to no deviation.

5.5.2 Testing for max-stability in the data

We first want to check if the block-maxima data originate from a max-stable process. A diagnostic tool is based on a multivariate Gumbel model (cf. Wadsworth [66]), and we explain first the method in general. We assume a space-time model of a general spatial dimension $d \in \mathbb{N}$. As before, we denote the regular grid of space-time observations by

$$\mathcal{S}_M \times \mathcal{T}_T = \{1, \dots, M\}^d \times \{1, \dots, T\}.$$

We define a hypothesis test based on the standard Gumbel transformed space-time observations (5.38) by

$$H_0 : \{\eta_1(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in \mathbb{R}^d \times [0, \infty)\} \text{ is max-stable.} \tag{5.41}$$

Under H_0 all finite-dimensional margins are max-stable; particularly, for every $D \subseteq \mathcal{S}_M \times \mathcal{T}_T$, the multivariate distribution function of $\{\eta_1(\boldsymbol{s},t) : (\boldsymbol{s},t) \in D\}$ is given by

$$G_D(y_1, \ldots, y_{|D|}) = \exp\{-V_D(\mathrm{e}^{y_1}, \ldots, \mathrm{e}^{y_{|D|}})\}, \quad (y_1, \ldots, y_{|D|}) \in \mathbb{R}^{|D|},$$

where V_D is the exponent measure from (5.3). Since V_D is homogeneous of order -1, the random variable

$$\eta_D := \max\{\eta_1(\boldsymbol{s},t) : (\boldsymbol{s},t) \in D\}$$

has univariate Gumbel distribution function

$$\mathbb{P}(\eta_D \leq y) = G_D(y, \ldots, y) = \exp\{-\mathrm{e}^{-y} V_D(1, \ldots, 1)\} = \mathrm{e}^{-\mathrm{e}^{-(y-\mu_D)}}, \quad y \in \mathbb{R}; \tag{5.42}$$

i.e., $\mu_D := \log V_D(1, \ldots, 1)$ is the location parameter and, since $1 \leq V_D(1, \ldots, 1) \leq |D|$, we have $0 \leq \mu_D \leq \log |D|$. These considerations can be used to construct a graphical test for max-stability: First, choose different subsets D with the same fixed cardinality. Then extract several independent realisations of the random variables η_D from the data and test by means of a qq-plot, if they follow a Gumbel distribution.

We apply this test to the standardised Gumbel transformed data (5.38). As indicated above, taking daily maxima removes for every location most of the dependence in the time series. For this test we want to take every precaution to make sure that we work indeed with independent data. Preliminary tests show that spatial observations, which are a small number of B_2 days apart (to be specified below), show only very little time-dependence.

Consequently, we define time blocks of size B_1 of spatial observations, which are in turn separated by time blocks of size B_2 as

$$\mathcal{S}_M \times \mathcal{T}^{(i)} = \{1, \ldots, M\}^2 \times \{(i-1)(B_1 + B_2) + t : \quad t = 1, \ldots, B_1\}, \tag{5.43}$$

for $i = 1, \ldots, R = \lfloor \frac{T}{B_1 + B_2} \rfloor$. The numbers B_1 and B_2 need to be chosen in such a way that the blocks can be considered as independent. This results in R independent time blocks of length B_1 of spatial data and thus in R independent realisations of η_D for every $D \subseteq \mathcal{S}_M \times \{1, \ldots, B_1\}$. The procedure is illustrated in Figure 5.3.

We use these i.i.d. realisations to estimate μ_D for every D by maximum likelihood estimation restricted to $[0, \log |D|]$. Since the MLE of the location parameter of a Gumbel distribution is not unbiased (cf. Johnson et al. [46], Section 9.6), we perform a bias correction.

For the diagnostic we take $K \in \mathbb{N}$ and consider subsets D with cardinality $|D| = K$. As the total number $\binom{B_1 M^2}{K}$ of those subsets is in most cases intractably large, we randomly choose $m := \min\{R, \binom{B_1 M^2}{K}\}$ subsets and obtain in total $N = m \cdot R$ subsets, which we denote by $D_j^{(i)}$ for $j = 1, \ldots, m$ and $i = 1, \ldots, R$. For every $j = 1, \ldots, m$ we estimate μ_{D_j} by MLE based on the i.i.d. random variables $\eta_{D_j}^{(i)} := \eta_{D_j^{(i)}}$, $i = 1, \ldots, R$. Then we perform qq-plots of

$$\eta_{D_1}^{(1)} - \mu_{D_1}, \ldots, \eta_{D_m}^{(m)} - \mu_{D_m}$$

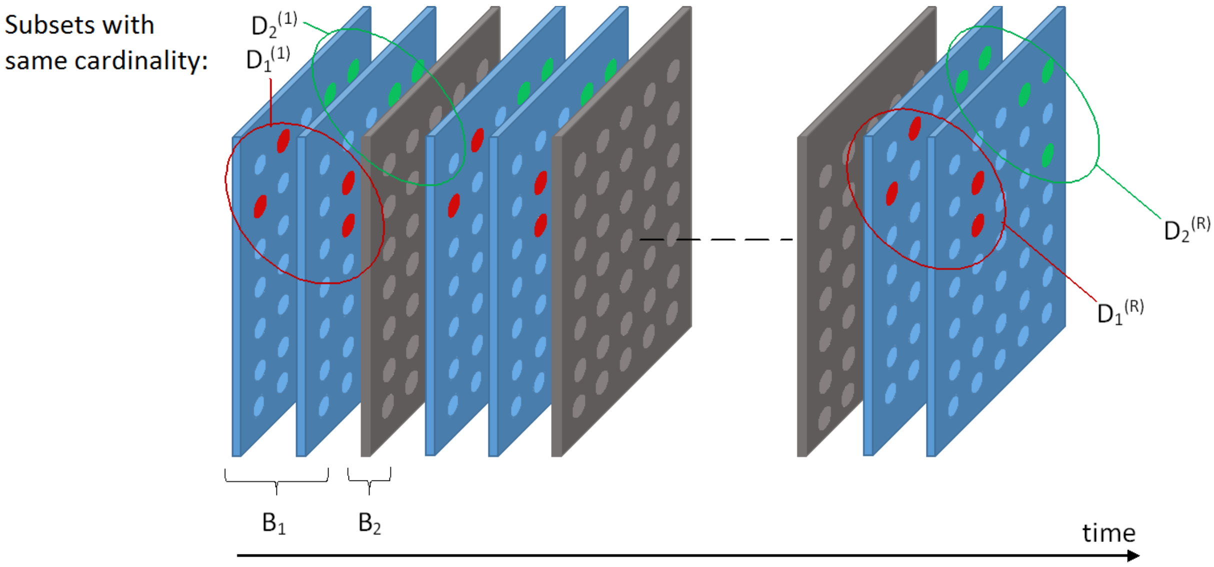

Figure 5.3: R independent realisations of η_D for different subsets D of the space-time observation area.

versus the standard Gumbel distribution. As a measure of variability of the estimates, non-parametric block bootstrap methods (cf. Politis and Romano [55], Section 3.2) are applied to obtain 95% pointwise confidence bounds. Using bootstrap methods, we preserve the dependence between different subsets D in the confidence intervals. Under H_0, the bisecting line should lie within these confidence bounds.

The Florida daily rainfall maxima show only little temporal dependence beyond one day. Hence we choose $B_1 = 2$ and $B_2 = 1$, which yields $R = \lfloor \frac{732}{3} \rfloor = 244$ mutually independent time blocks of spatial data. We perform the described procedure for $K = 2, 3, 4, 5$, which entails $m = R = 244$. Thus we obtain a total number of $N = 244^2 = 59\,536$ subsets. The power of this diagnostic test increases with K (cf. Wadsworth [66]) as it gets less likely to include sets of space-time points that are K-wise independent. Figure 5.4 shows the results for the different choices of K. The solid red bisecting lines lie inside the confidence bounds. Hence, there is no statistically significant evidence of the space-time process generating the data not to be max-stable.

5.5.3 Pairwise maximum likelihood estimation

We apply the pairwise maximum likelihood estimation to the standard Fréchet transformed data (5.39). The parameters to estimate are those of the function δ in (5.40); i.e., $C_1, C_2, C_3 \in (0, \infty)$ and $\alpha_1, \alpha_2, \alpha_3 \in (0, 2]$.

In the definition of the pairwise log-likelihood function (5.8), the maximum spatial and temporal lags are specified by the numbers r_1, r_2 and p, respectively. Immediately by model (5.40) for δ, the parameters of the three different dimensions (space and time) are separated in the extremal setting. This has also been noticed in Davis et al. [20], where a simulation study in Section 7 for the isotropic model shows that estimating the spatial and temporal parameter pairs individually leads to very good results in terms of root-mean-square error and mean absolute error. Hence, for example for parameter estimates for C_1 and α_1, we can set the maximum lags corresponding to the remaining parameters equal to 0 (i.e., we set $r_2 = p = 0$). This means that we basically

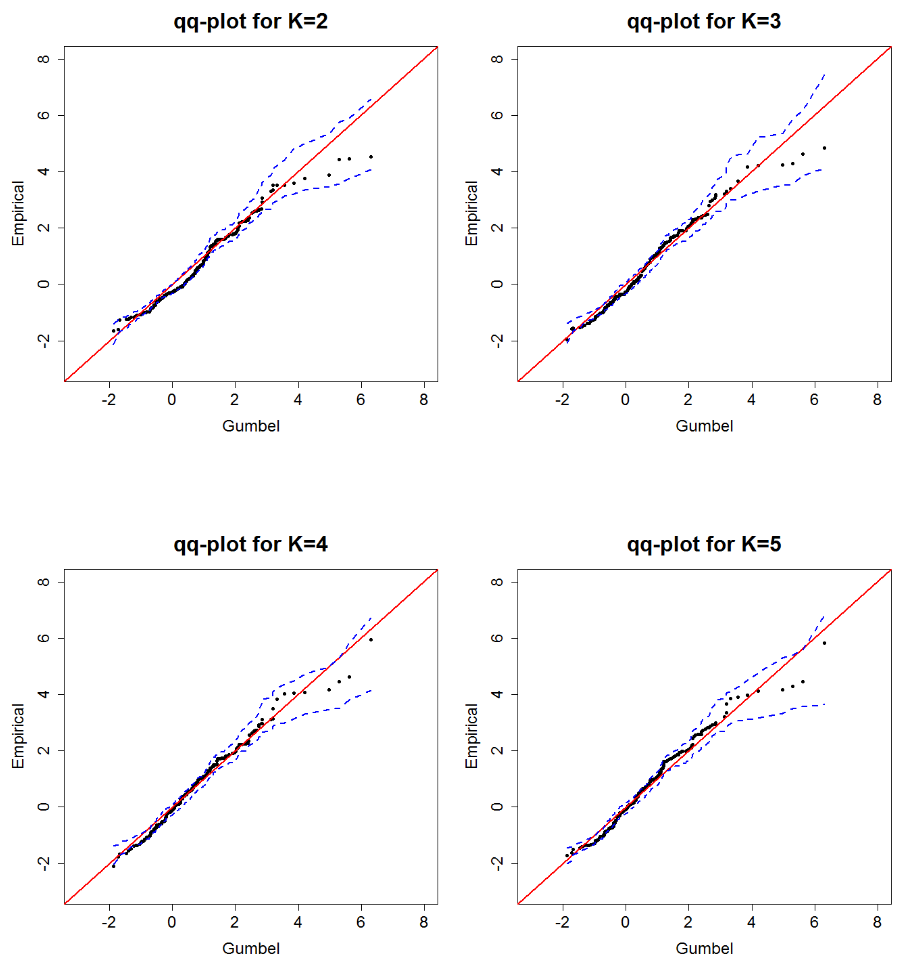

Figure 5.4: qq-plots of theoretical standard Gumbel quantiles versus the empirical quantiles (black dots). The latter correspond to the empirical distribution of maxima taken over groups of cardinality K. Dashed blue lines mark 95% pointwise confidence bounds obtained by block bootstrap. Solid red lines correspond to no deviation.

fit univariate models to the respective spatial and temporal parts of the dependence function (5.40). Hence, this separation simplifies the statistical estimation. However, proving asymptotic properties of the pairwise likelihood estimator in the special case of a univariate model would for instance still involve showing the required mixing conditions and thus not remove much of the complexity.

Furthermore, we know that we should not include too many lags in space or time into the likelihood, since independence effects can introduce a bias in the estimates, see for example Nott and Rydén [52], Section 2.1, or Huser and Davison [43], Section 4. On the other hand, an empirical analysis showed that extremal spatial dependence of the Florida daily rainfall maxima ranges up to lag 4 and extremal temporal dependence does not last more than one or two days, cf. Figure 7.2.6 in Steinkohl [62]. Hence, we perform the PMLE for maximum spatial and temporal lags up to 4 and 2, respectively, thus also assuring identifiability of all parameters according to Table 5.1. The results are summarised in Table 5.2. Setting r_1, r_2 or p equal to 1 results

max. lags	$\widehat{C}_i$	$\widehat{\alpha}_i$
(2,0,0)	0.6287 $[0.5928, 0.6646]$	0.9437 $[0.9065, 0.9808]$
(3,0,0)	0.6358 $[0.5989, 0.6728]$	0.8599 $[0.8189, 0.9009]$
(4,0,0)	0.6438 $[0.6051, 0.6825]$	0.8107 $[0.7690, 0.8525]$
(0,2,0)	0.7271 $[0.6492, 0.8050]$	0.9517 $[0.8715, 1.0320]$
(0,3,0)	0.7370 $[0.6586, 0.8154]$	0.8521 $[0.7737, 0.9305]$
(0,4,0)	0.7476 $[0.6677, 0.8275]$	0.7931 $[0.7039, 0.8822]$
(0,0,2)	4.8378 $[4.4282, 5.2474]$	0.1981 $[0.0177, 0.3784]$

Table 5.2: Estimates of the parameter pairs (C_1, α_1), (C_2, α_2) and (C_3, α_3) for different maximum spatial and temporal lags. Intervals below the point estimates are asymptotic 95%-confidence bounds based on subsampling.

in non-identifiability of the corresponding parameters α_1, α_2 or α_3, respectively; cf. Table 5.1. Therefore, they are not shown in Table 5.2.

The combination of a rather large estimate for $\widehat{C}_3$ and a rather small estimate for $\widehat{\alpha}_3$ indicates that there is only little extremal temporal dependence, see Steinkohl [62], Section 7.2. Asymptotic 95%-confidence intervals are based on asymptotic normality of the parameter estimates and estimated using subsampling methods (cf. Section 5.4).

5.5.4 Isotropic versus anisotropic model

Using the results of Section 5.4, we want to apply the test (5.28) for spatial isotropy to the hypothesis

$$H_0 : \{C_1 = C_2 \text{ and } \alpha_1 = \alpha_2\} \quad \text{versus} \quad H_1 : \{C_1 \neq C_2 \text{ or } \alpha_1 \neq \alpha_2\}.$$

For the block maxima of the precipitation data we have $d = 2$, $M = 12$ and $T = 732$. This corresponds to the situation of a fixed spatial domain with $\tau_T = \sqrt{732}$.

We use the spatial PMLEs based on maximum lags 2-4, which can be read off from Table 5.2. We obtain the rejection areas from Theorem 5.16. We choose $b_1 = b_2 = 5$, thus ensuring that the full range of spatial dependence is contained in the subsamples and simultaneously achieving that their number is large. Concerning the number of time points in each subsample, we take $b_3 = 600$. Here we choose a large number to ensure that Theorem 5.16, where $T \to \infty$, is applicable. This results in $\tau_{b_3} = \sqrt{b_3} = \sqrt{600}$. In order to obtain a large number of subsamples, we further choose $e_1 = e_2 = e_3 = 1$ as the degree of overlap.

max. lag	τ_T	$\widehat{C}_2-\widehat{C}_1$	$\tau_T(\widehat{C}_2-\widehat{C}_1)$	$\mathrm{Rej}^{(T)}_{\widehat{\theta}_C}$	**97.5%-CI for** C_2-C_1	**Reject** $C_1=C_2$
2	27.055	0.098	2.651	$[-2.400, 2.400]^c$	$[0.010, 0.187]$	yes
3	27.055	0.101	2.738	$[-2.392, 2.392]^c$	$[0.013, 0.190]$	yes
4	27.055	0.104	2.808	$[-2.393, 2.393]^c$	$[0.015, 0.192]$	yes

Table 5.3: Test results for parameters C_1 and C_2. All values are rounded to three positions after decimal point.

max. lag	τ_T	$\widehat{\alpha}_2-\widehat{\alpha}_1$	$\tau_T(\widehat{\alpha}_2-\widehat{\alpha}_1)$	$\mathrm{Rej}^{(T)}_{\widehat{\theta}_\alpha}$	**97.5%-CI for** $(\alpha_2-\alpha_1)$	**Reject** $\alpha_1=\alpha_2$
2	27.055	0.008	0.216	$[-2.162, 2.162]^c$	$[-0.072, 0.088]$	no
3	27.055	-0.008	-0.216	$[-2.130, 2.130]^c$	$[-0.087, 0.071]$	no
4	27.055	-0.018	-0.477	$[-2.342, 2.342]^c$	$[-0.104, 0.069]$	no

Table 5.4: Test results for parameters α_1 and α_2. All values are rounded to three positions after decimal point.

Tables 5.3 and 5.4 present the results of the two tests at individual confidence levels $\beta = 2.5\%$ giving a test for (5.28) at a confidence level $2\beta = 5\%$ by Bonferroni's inequality. The differences $(\widehat{C}_2 - \widehat{C}_1)$ and $(\widehat{\alpha}_2 - \widehat{\alpha}_1)$ can be obtained from Table 5.2.

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 of (5.28) at a confidence level of 5% and conclude that our data originate from a spatially anisotropic max-stable Brown-Resnick process. Further note the interesting fact that, although the asymptotic confidence interval for the difference $C_2 - C_1$ does not include 0, the individual intervals for C_1 and C_2 overlap, see Table 5.2. This is due to the fact that the individual confidence bounds are estimated independently of each other, whereas the estimated bounds for the difference reflect how far the parameter estimates lie apart in one fixed particular (sub)sample.

5.5.5 Model check

Finally, having fitted the Brown-Resnick space-time model (5.1) to the precipitation data, we want to assess the quality of the fit. We take inspiration from Section 7 of Davison et al. [22] and compare maxima taken over subsets of the space-time precipitation data with simulated counterparts.

Similarly as in Section 5.5.2, we consider subsets of the observations on a regular grid for L spatial locations and for time points $1, \ldots, B_1$,

$$D = \{(s_1^{(\ell)}, s_2^{(\ell)}, 1), \ldots, (s_1^{(\ell)}, s_2^{(\ell)}, B_1) : \ell = 1, \ldots, L\}.$$

We follow the procedure as in (5.43) to extract R independent realisations of $\{\eta_1(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in D\}$ from the standard Gumbel transformed space-time observations (5.38). This yields in turn R independent realisations of $\eta_D = \max\{\eta_1(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in D\}$, which we summarise in the ordered vector $\eta_{\text{data}} := (\eta_D^{(1)}, \dots, \eta_D^{(R)})$. Now we simulate a corresponding vector, denoted by $\widehat{\eta}_{\text{sim}} := (\widehat{\eta}_D^{(1)}, \dots, \widehat{\eta}_D^{(R)})$. To this end we need reliable Monte Carlo values as elements of $\widehat{\eta}_{\text{sim}}$. We obtain them by simulating empirical order statistics as follows. We simulate $m \cdot R$ independent copies of the Brown-Resnick space-time process on D with dependence structure δ as in (5.4) with the PMLEs from Table 5.2, where we take the estimates based on maximum lag 4 (for the spatial parameters) and 2 (for the temporal parameters), which are the maximum lags, where dependence is still present. We transform the univariate margins to standard Gumbel. This results in corresponding $m \cdot R$ independent simulations of η_D and we consider them as m blocks of size R. We order the R values in each block and define $\widehat{\eta}_D^{(i)}$ as the mean of all simulated ith order statistics for $i = 1, \dots, R$, which gives $\widehat{\eta}_{\text{sim}} := (\widehat{\eta}_D^{(1)}, \dots, \widehat{\eta}_D^{(R)})$.

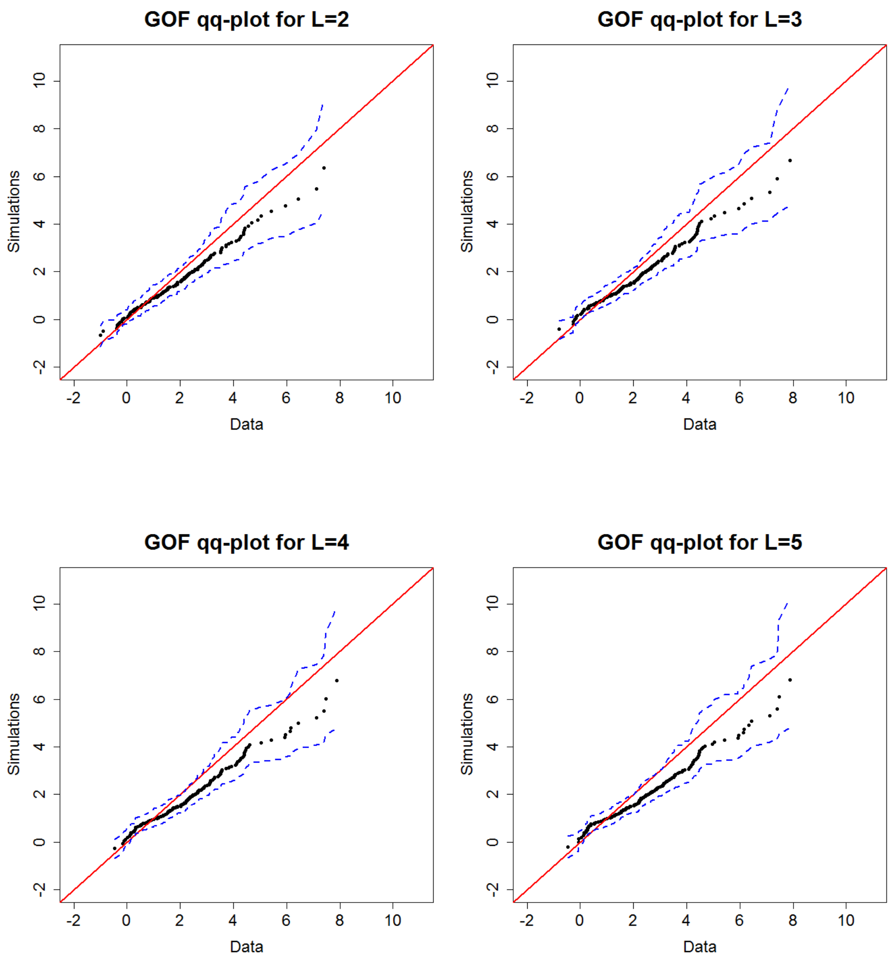

Figure 5.5: Goodness of fit qq-plots for different spatial locations and different L. Top left: $L = 2$: (1,1) and (1,2). Top right: $L = 3$: (1,1), (1,2) and (3,1). Bottom left: $L = 4$: (1,1), (1,2), (3,1) and (3,2). Bottom right: $L = 5$: (1,1), (1,2), (3,1), (3,2) and (2,1). PMLEs underlying the simulations are based on maximum spatial and temporal lags 4 and 2, respectively. Dashed blue lines mark 95% pointwise confidence bounds. Solid red lines correspond to no deviation.

The vectors η_{data} and $\widehat{\eta}_{\text{sim}}$ are compared by qq-plots. If the fit is good, the points in the plots lie approximately on the bisecting line. Pointwise 95%-confidence bands are determined by the

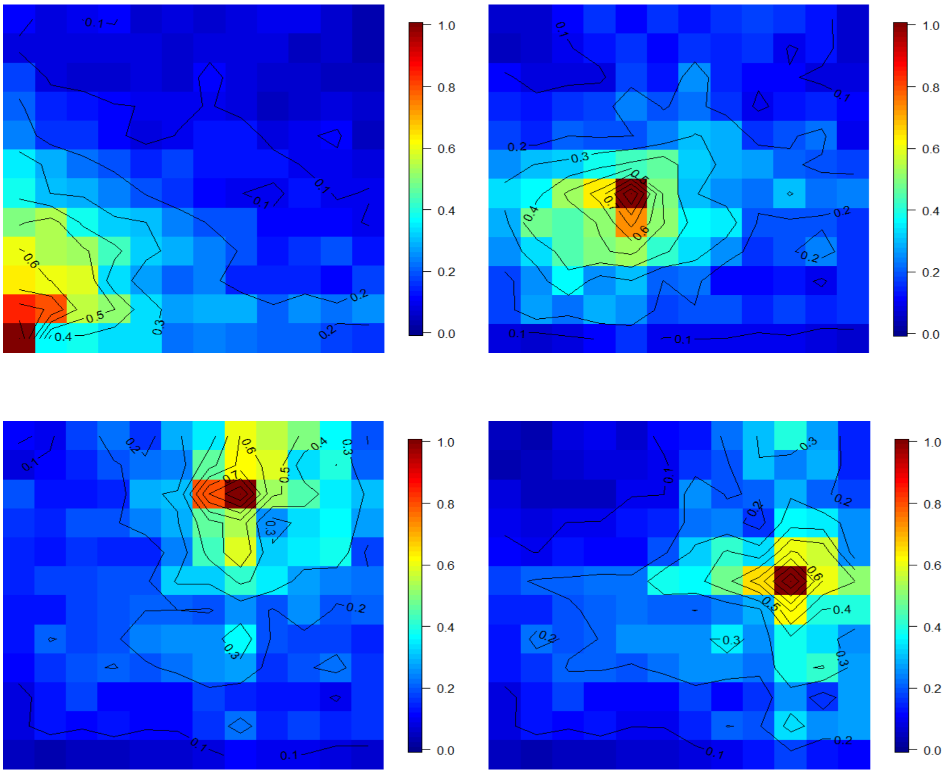

Figure 5.6: Predicted conditional probability fields based on daily maxima for reference space-time points (1,1,1), (5,6,1), (8,10,1) and (10,7,1) and rainfall levels $z = z^\star = 2.5$ (clockwise from the top left to the bottom right).

2.5% and the 97.5% quantiles of the simulated order statistics. As in Section 5.5.3, we choose $B_1 = 2$. The number of simulations is $N = m \cdot R = 100 \cdot 244 = 24400$. Figure 5.5 presents the results for four exemplary groups of locations. The plots reveal a good model fit.

We carried out the simulations using the exact method recently suggested in Dombry et al. [28], Sections 3.3 and 5.2. For an overview and comparison of different simulation methods for Brown-Resnick processes we refer to Leber [50].

5.5.6 Application: conditional probability fields

Based on the fitted model, we want to answer questions like: Given there is extreme rain at some space-time reference point $(s_1^\star, s_2^\star, t^\star) \in \{1, \ldots, 12\}^2 \times \{1, \ldots, 732\}$, what is the estimated probability of extreme rain at some prediction space-time point (s_1^p, s_2^p, t^p)? In other words, we want to estimate the probabilities

$$\mathbb{P}\left(\tilde{\eta}((s_1^p, s_2^p), t^p) > z \mid \tilde{\eta}((s_1^\star, s_2^\star), t^\star) > z^\star\right), \tag{5.44}$$

where $\{\tilde{\eta}((s_1, s_2), t) : s_1, s_2 = 1, \ldots, 12, t = 1, \ldots, 732\}$ are the stationary observations (5.37) and z and $z^\star$ are prediction and reference rainfall levels, respectively. Denote by $\Lambda_{\mu,\sigma}$ the Gumbel distribution with location and scale parameters μ and σ (cf. Section 5.5.1) and set $\widehat{\mu}^p := \widehat{\mu}(s_1^p, s_2^p)$,

$\widehat{\sigma}^p := \widehat{\sigma}(s_1^p, s_2^p)$, $\widehat{\mu}^\star := \widehat{\mu}(s_1^\star, s_2^\star)$ and $\widehat{\sigma}^\star := \widehat{\sigma}(s_1^\star, s_2^\star)$, which are the marginal Gumbel parameter estimates. Simple computations show that (5.44) can be estimated by

$$\frac{1}{1-\Lambda_{\widehat{\mu}^\star,\widehat{\sigma}^\star}(z^\star)}\Big(1-\Lambda_{\widehat{\mu}^\star,\widehat{\sigma}^\star}(z^\star)-\Lambda_{\widehat{\mu}^p,\widehat{\sigma}^p}(z)$$
$$+\exp\Big\{-\widehat{V}_D\Big(-\frac{1}{\log\big\{\Lambda_{\widehat{\mu}^p,\widehat{\sigma}^p}(z)\big\}},-\frac{1}{\log\big\{\Lambda_{\widehat{\mu}^\star,\widehat{\sigma}^\star}(z^\star)\big\}}\Big)\Big\}\Big),$$

where $\widehat{V}_D$ is the estimate of the exponent measure (5.6) obtained by plugging in the PMLEs of the parameters of the dependence function δ. Figure 5.6 shows four predicted conditional probability fields for the reference points $(1,1,1)$, $(5,6,1)$, $(8,10,1)$ and $(10,7,1)$ and for high empirical rainfall levels $z = z^\star = 2.5$. Because of the little temporal dependence in the daily maxima, we only consider equal time points for spatial predictions.

Appendix to Chapter 2

A.1 Taylor expansion for the pre-asymptotic extremogram

Lemma A.1. *Let the assumptions of Theorem 2.8 hold.*
(a) For $\boldsymbol{h} \in \mathbb{R}^d$ the true extremogram is given by

$$\rho_{AB}(\boldsymbol{h}) = \frac{a_1 a_2}{a_2 - a_1}\Big(- V_2(a_2, b_2) + V_2(a_2, b_1) + V_2(a_1, b_2) - V_2(a_1, b_1)\Big), \tag{A.1}$$

where $V_2(\cdot,\cdot) = V_2(\boldsymbol{h};\cdot,\cdot)$ is the bivariate exponent measure (cf. Beirlant et al. [3], Section 8.2.2) defined by

$$\mathbb{P}(X(\mathbf{0}) \le x_1, X(\boldsymbol{h}) \le x_2) = \exp\{-V_2(x_1, x_2)\}, \quad x_1, x_2 > 0.$$

For $A = (a, \infty)$ and $B = (b, \infty)$ we obtain

$$\rho_{AB}(\boldsymbol{h}) = a\Big(\frac{1}{a} + \frac{1}{b} - V_2(a, b)\Big). \tag{A.2}$$

(b) For fixed $\boldsymbol{h} \in \mathbb{R}^d$ and the sequence m_n satisfying the conditions of Theorem 2.6, the pre-asymptotic extremogram satisfies as $n \to \infty$,

$$\rho_{AB,m_n}(\boldsymbol{h}) = (1 + o(1))\Big[\rho_{AB}(\boldsymbol{h}) + \frac{1}{2m_n^d}\frac{a_1 a_2}{a_2 - a_1}\Big(V_2^2(a_2, b_2) + V_2^2(a_2, b_1) + V_2^2(a_1, b_2) + V_2^2(a_1, b_1)\Big)\Big]. \tag{A.3}$$

For $A = (a, \infty)$ and $B = (b, \infty)$ this reduces to

$$\rho_{AB,m_n}(\boldsymbol{h}) = (1 + o(1))\Big[\rho_{AB}(\boldsymbol{h}) + \frac{1}{2m_n^d a}(\rho_{AB}(\boldsymbol{h}) - 2\frac{a}{b})(\rho_{AB}(\boldsymbol{h}) - 1)\Big]. \tag{A.4}$$

Proof. Throughout the proof all asymptotic results hold as $n \to \infty$. Since $\{X(\boldsymbol{s} : \boldsymbol{s} \in \mathbb{R}^d\}$ has standard unit Fréchet margins, we can and do choose $a_n = n^d$ in (2.1) such that $\mathbb{P}(X(\mathbf{0}) > n^d) = 1 - \exp\{-n^{-d}\} \sim n^{-d}$.
(a) We first show (A.1). With this choice of a_n, equation (2.2) is equivalent to

$$\rho_{AB}(\boldsymbol{h}) = \lim_{n\to\infty} \frac{n^d \mathbb{P}(X(\mathbf{0}) \in n^d A, X(\boldsymbol{h}) \in n^d B)}{n^d \mathbb{P}(X(\mathbf{0}) \in n^d A)}.$$

We set $A = (a_1, a_2)$ and $B = (b_1, b_2)$. For the denominator we obtain by a first order Taylor expansion

$$\begin{aligned} n^d\mathbb{P}(X(\mathbf{0}) \in n^d(a_1,a_2)) =& n^d[\mathbb{P}(X(\mathbf{0}) \le n^d a_2) - \mathbb{P}(X(\mathbf{0}) \le n^d a_1)] \\ =& n^d\big[\exp\{-\frac{1}{n^d a_2}\} - \exp\{-\frac{1}{n^d a_1}\}\big] \\ =& \frac{1}{a_1} - \frac{1}{a_2} + \mathcal{O}(n^{-d}) \to \frac{1}{a_1} - \frac{1}{a_2} = \frac{a_2 - a_1}{a_1 a_2} > 0. \end{aligned} \tag{A.5}$$

Since by homogeneity $V_2(kx_1, kx_2)) = k^{-1}V_2(x_1, x_2)$ for $k > 0$, we find for the numerator

$$\begin{aligned} & n^d\mathbb{P}((X(\mathbf{0}), X(\boldsymbol{h})) \in n^d(a_1,a_2) \times n^d(b_1,b_2)) \\ =& n^d\Bigg[\exp\Big\{-\frac{1}{n^d}V_2(a_2,b_2)\Big\} - \exp\Big\{-\frac{1}{n^d}V_2(a_2,b_1)\Big\} \\ & -\exp\Big\{-\frac{1}{n^d}V_2(a_1,b_2)\Big\} + \exp\Big\{-\frac{1}{n^d}V_2(a_1,b_1)\Big\}\Bigg] \\ =& -V_2(a_2,b_2) + V_2(a_2,b_1) + V_2(a_1,b_2) - V_2(a_1,b_1) + \mathcal{O}(n^{-d}). \end{aligned} \tag{A.6}$$

This yields (A.1).
Furthermore, $V_2(a,\infty) = 1/a$, $V_2(\infty, b) = 1/b$ and $V_2(\infty,\infty) = 0$, see for instance Resnick [58], p. 268. Together with the fact that the denominator converges to $1/a$, this gives (A.2).

(b) For an estimate of the pre-asymptotic extremogram we need to improve the first order asymptotics of part (a). For an interval (a, b) we abbreviate $\Phi_n(a,b) := \exp\{-\frac{1}{m_n^d}V_2(a,b)\}$. From equation (2.9) together with (A.5) and (A.6) we obtain

$$\begin{aligned} \rho_{AB,m_n}(\boldsymbol{h}) &= \frac{\mathbb{P}(X(\mathbf{0}) \in m_n^d A, X(\boldsymbol{h}) \in m_n^d B)}{\mathbb{P}(X(\mathbf{0}) \in m_n^d A)} \\ &= \frac{\Phi_n(a_2,b_2) - \Phi_n(a_2,b_1) - \Phi_n(a_1,b_2) + \Phi_n(a_1,b_1)}{\exp\{-\frac{1}{a_2\, m_n^d}\} - \exp\{-\frac{1}{a_1\, m_n^d}\}} \\ &= \rho_{AB}(\boldsymbol{h}) + \frac{1}{\exp\{-\frac{1}{a_2\, m_n^d}\} - \exp\{-\frac{1}{a_1\, m_n^d}\}} \\ &\quad \Big[\Phi_n(a_2,b_2) - \Phi_n(a_2,b_1) - \Phi_n(a_1,b_2) + \Phi_n(a_1,b_1) \\ &\quad -\Big(\exp\Big\{-\frac{1}{a_2\, m_n^d}\Big\} - \exp\Big\{-\frac{1}{a_1\, m_n^d}\Big\}\Big)\rho_{AB}(\boldsymbol{h})\Big] \\ &= \rho_{AB}(\boldsymbol{h}) + \frac{a_1 a_2}{a_2 - a_1} m_n^d(1 + o(1)) \\ &\quad \Big[\Phi_n(a_2,b_2) - \Phi_n(a_2,b_1) - \Phi_n(a_1,b_2) + \Phi_n(a_1,b_1) \\ &\quad -\Big(\exp\Big\{-\frac{1}{a_2\, m_n^d}\Big\} - \exp\Big\{-\frac{1}{a_1\, m_n^d}\Big\}\Big)\rho_{AB}(\boldsymbol{h})\Big] \end{aligned} \tag{A.7}$$

By a second order Taylor expansion of Φ_n it follows that, using (A.1) and (A.5),

$$\rho_{AB,m_n}(\boldsymbol{h}) = (1 + o(1))\Big[\rho_{AB}(\boldsymbol{h}) +$$

$$\frac{1}{2m_n^d}\frac{a_1a_2}{a_2-a_1}\Big(V_2^2(a_2,b_2)+V_2^2(a_2,b_1)+V_2^2(a_1,b_2)+V_2^2(a_1,b_1)\Big)\Big].$$

This shows (A.3).

Now let $A=(a,\infty)$ and $B=(b,\infty)$. Then $a_1a_2/(a_2-a_1)=a_1+o(1)$ as $a_2\to\infty$ and the expression in the rectangular bracket in (A.7) becomes

$$\begin{aligned}\Big[\cdots\Big] &= 1-\exp\Big\{-\frac{1}{b\,m_n^d}\Big\}-\exp\Big\{-\frac{1}{a\,m_n^d}\Big\}+\exp\Big\{-\frac{1}{m_n^d}V_2(a,b)\Big\}\\ &\quad -\Big(1-\exp\Big\{-\frac{1}{a\,m_n^d}\Big\}\Big)\rho_{AB}(\boldsymbol{h}). \end{aligned}\tag{A.8}$$

Abbreviating $V_2:=V_2(a,b)$, a second order Taylor expansion gives with (A.2) for the right-hand side of (A.8),

$$\begin{aligned}&\Big(\frac{1}{a\,m_n^d}-\frac{1}{2a^2\,m_n^{2d}}\Big)+\Big(\frac{1}{b\,m_n^d}-\frac{1}{2b^2\,m_n^{2d}}\Big)-\Big(\frac{1}{m_n^d}V_2-\frac{1}{2m_n^{2d}}V_2^2\Big)\\ &-\Big(\frac{1}{m_n^d}-\frac{1}{2a\,m_n^{2d}}\Big)\Big(\frac{1}{a}+\frac{1}{b}-V_2\Big)+o(m_n^{-2d})\\ =\;&\frac{1}{2\,m_n^{2d}}\Big\{\Big(V_2^2-\frac{1}{a^2}-\frac{1}{b^2}\Big)+\frac{1}{a}\Big(\frac{1}{a}+\frac{1}{b}-V_2\Big)\Big\}+o(m_n^{-2d}).\end{aligned}\tag{A.9}$$

Solving (A.2) for V_2 gives $V_2=V_2(a,b)=\frac{1}{a}(1-\rho_{AB}(\boldsymbol{h}))+\frac{1}{b}$ such that we obtain for the expression in the curly brackets of (A.9),

$$\Big(\frac{1}{a}(1-\rho_{AB}(\boldsymbol{h}))+\frac{1}{b}\Big)^2-\frac{1}{a^2}-\frac{1}{b^2}+\frac{1}{a^2}\rho_{AB}(\boldsymbol{h}) \;=\; \frac{1}{a^2}(\rho_{AB}(\boldsymbol{h})-2\frac{a}{b})(\rho_{AB}(\boldsymbol{h})-1).$$

Going backwards with this expression proves (A.4). □

Appendix to Chapter 3

B.1 α-mixing of the Brown-Resnick space-time process

In the following we define α-mixing for spatial processes; see e.g. Doukhan [30] or Bolthausen [6].

Definition B.1. *For $d \in \mathbb{N}$, consider a strictly stationary random field $\left\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\right\}$ and let $d(\cdot,\cdot)$ be some metric induced by a norm on $\mathbb{R}^d$. For $\Lambda_1, \Lambda_2 \subset \mathbb{Z}^d$ set*

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

Further, for $i = 1, 2$ denote by $\mathcal{F}_{\Lambda_i} = \sigma\left\{X(\boldsymbol{s}), \boldsymbol{s} \in \Lambda_i\right\}$ the σ-algebra generated by $\{X(\boldsymbol{s}) : \ \boldsymbol{s} \in \Lambda_i\}$.

*(i) The α-*mixing coefficients *are defined for $k, l \in \mathbb{N} \cup \{\infty\}$ and $r \geq 0$ by*

$$\alpha_{k,l}(r) = \sup\left\{|\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| : \ A_i \in \mathcal{F}_{\Lambda_i}, |\Lambda_1| \leq k, |\Lambda_2| \leq l, d(\Lambda_1, \Lambda_2) \geq r\right\}, \tag{B.1}$$

where $|\Lambda_i|$ is the cardinality of the set Λ_i for $i = 1, 2$.

*(ii) The random field is called α-*mixing, *if $\alpha_{k,l}(r) \to 0$ as $r \to \infty$ for all $k, l \in \mathbb{N}$.*

For a strictly stationary max-stable processes Corollary 2.2 of Dombry and Eyi-Minko [27] shows that the α-mixing coefficients can be related to the extremogram of the max-stable process. Equations (B.2) and (B.3) follow as in the proofs of Proposition 5.7 and 5.11.

Proposition B.2. *For all fixed time points $t \in \mathbb{N}$ the random field $\left\{\eta(\boldsymbol{s}, t), \boldsymbol{s} \in \mathbb{Z}^2\right\}$ (3.1) is α-mixing with mixing coefficients satisfying*

$$\alpha_{k,l}(r) \leq 2kl \sup_{s \geq r} \chi(s, 0) \leq 4kle^{-\theta_1 r^{\alpha_1}/2}, \quad k, l \in \mathbb{N}, \ r \geq 0. \tag{B.2}$$

For all fixed locations $\boldsymbol{s} \in \mathbb{R}^2$ the time series $\{\eta(\boldsymbol{s}, t) : t \in [0, \infty)\}$ in (3.1) is α-mixing with mixing coefficients satisfying for some constant $c > 0$

$$\alpha(r) := \alpha_{\infty,\infty}(r) \leq c \sum_{u=r}^{\infty} ue^{-\theta_2 u^{\alpha_2}/2}, \quad r \geq 0. \tag{B.3}$$

We will make frequent use of the following simple result.

Lemma B.3. *Let $z \in \mathbb{N}$. For $(\theta, \alpha) \in \{(\theta_1, \alpha_1), (\theta_2, \alpha_2)\}$ and sufficiently large r such that the sequence $u^z e^{-\theta u^\alpha/2}$ is monotonously decreasing for $u \geq r$, we have*

$$g_z(r) = \sum_{u=r}^{\infty} u^z e^{-\theta u^\alpha/2} \leq c e^{-\theta r^\alpha/2} r^{z+1}, \quad r \in \mathbb{N}.$$

for some constant $c = c(z) > 0$.

Proof. An integral bound together with a change of variables yields

$$\begin{aligned}
g_z(r) &= r^z e^{-\theta r^\alpha/2} + \sum_{u=r+1}^{\infty} u^z e^{-\theta u^\alpha/2} \leq r^z e^{-\theta r^\alpha/2} + \int_r^\infty u^z e^{-\theta u^\alpha/2} du \\
&= r^z e^{-\theta r^\alpha/2} + \left(\frac{2}{\theta}\right)^{(z+1)/\alpha} \frac{1}{\alpha} \int_{\theta r^\alpha/2}^{\infty} t^{(z+1)/\alpha - 1} e^{-t} dt \\
&\leq r^z e^{-\theta r^\alpha/2} + c_1 \Gamma\left(\lceil (z+1)/\alpha \rceil, \theta r^\alpha/2\right) \\
&= r^z e^{-\theta r^\alpha/2} + c_1 \left(\lceil (z+1)/\alpha \rceil - 1\right)!\; e^{-\theta r^\alpha/2} \sum_{k=0}^{\lceil (z+1)/\alpha \rceil - 1} \frac{\theta^k r^{\alpha k}}{2^k k!} \\
&\leq r^z e^{-\theta r^\alpha/2} + c_2 e^{-\theta r^\alpha/2} r^{\alpha(\lceil (z+1)/\alpha \rceil - 1)} \\
&\leq c e^{-\theta r^\alpha/2} r^{z+1},
\end{aligned}$$

where $\Gamma(s, r) = \int_r^\infty t^{s-1} e^{-t} dt = (s-1)! e^{-r} \sum_{k=0}^{s-1} r^k/k!$, $s \in \mathbb{N}$, is the incomplete gamma function and $c_1, c > 0$ are constants depending on z. □

Appendix to Chapter 4

C.1 α-mixing with respect to the increasing dimensions

We need the concept of α-mixing for the process $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ with respect to $\mathbb{R}^w$. In a space-time setting with fixed spatial setting and increasing time series this is called *temporal* α-mixing.

Definition C.1 (α-mixing and α-mixing coefficients)**.** *Consider a strictly stationary process* $\left\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\right\}$ *and let* $d(\cdot,\cdot)$ *be some metric induced by a norm* $\|\cdot\|$ *on* $\mathbb{R}^d$*. For* $\Lambda_1, \Lambda_2 \subset \mathbb{Z}^w$ *define*

$$d(\Lambda_1, \Lambda_2) := \inf\left\{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| : \ \boldsymbol{s}_1 \in \mathcal{F} \times \Lambda_1, \boldsymbol{s}_2 \in \mathcal{F} \times \Lambda_2\right\}.$$

Further, for $i = 1, 2$ *denote by* $\sigma_{\mathcal{F}\times\Lambda_i} = \sigma\left\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{F} \times \Lambda_i\right\}$ *the* σ*-algebra generated by* $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{F} \times \Lambda_i\}$*.*

(i) We define the α-mixing coefficients with respect to $\mathbb{R}^w$ *for* $k_1, k_2 \in \mathbb{N}$ *and* $z \geq 0$ *as*

$$\alpha_{k_1,k_2}(z) := \sup\left\{|\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| : \ A_i \in \sigma_{\mathcal{F}\times\Lambda_i}, |\Lambda_i| \leq k_i, d(\Lambda_1, \Lambda_2) \geq z\right\}. \tag{C.1}$$

(ii) We call $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ α-mixing with respect to $\mathbb{R}^w$*, if* $\alpha_{k_1,k_2}(z) \to 0$ *as* $z \to \infty$ *for all* $k_1, k_2 \in \mathbb{N}$*.*

We have to control the dependence between vector processes $\{\boldsymbol{Y}(\boldsymbol{s}) = X_{B(\boldsymbol{s},\gamma)} : \boldsymbol{s} \in \Lambda_1'\}$ and $\{\boldsymbol{Y}(\boldsymbol{s}) = X_{B(\boldsymbol{s},\gamma)} : \boldsymbol{s} \in \Lambda_2'\}$ for subsets $\Lambda_i' \subset \mathbb{Z}^w$ with cardinalities $|\Lambda_1'| \leq k_1$ and $|\Lambda_2'| \leq k_2$.. This entails dealing with unions of balls $\Lambda_i = \cup_{\boldsymbol{s}\in\mathcal{F}\times\Lambda_i'} B(\boldsymbol{s}, \gamma)$. Since $\gamma > 0$ is some predetermined finite constant independent of n, we keep notation simple by redefining the α-mixing coefficients corresponding to the vector processes for $k_1, k_2 \in \mathbb{N}$ and $z \geq 0$ as

$$\begin{aligned}\alpha_{k_1,k_2}(z) :=& \sup\{|\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| : \\ & A_i \in \sigma_{\Lambda_i}, \ \Lambda_i = \cup_{\boldsymbol{s}\in\mathcal{F}\times\Lambda_i'} B(\boldsymbol{s}, \gamma), \ |\Lambda_i'| \leq k_i, d(\Lambda_1', \Lambda_2') \geq z\}.\end{aligned} \tag{C.2}$$

C.2 Proof of Theorem 4.8

The proof of Theorem 4.8 is divided into two parts. In the first part we prove a LLN and a CLT in Lemmas C.2 and C.3 for the estimators $\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}$ in (4.11). In the second part of the proof we derive the CLT for the empirical extremogram $\widehat{\rho}_{AB,m_n}$ in (4.9), and compute the asymptotic covariance matrix Π. The proof generalises corresponding proofs in Chapter 2 (where the observation area increases in all dimensions) in a non-trivial way. We recall the separation of every point and every lag in its components corresponding to the fixed domain, indicated by the sub index $\mathcal{F}$, and the remaining components, indicated by $\mathcal{I}$, from Assumption 4.4. In particular, we decompose $\boldsymbol{h}^{(i)} = (\boldsymbol{h}_{\mathcal{F}}^{(i)}, \boldsymbol{h}_{\mathcal{I}}^{(i)}) \in \mathcal{H}$.

The separation of the observation space with its fixed domain has to be introduced into the proofs given in Chapter 2, which is even in the regular grid situation highly non-trivial. We will give detailed references to those proofs, whenever possible, to support the understanding. On the other hand, if arguments just follow a previous proof line by line we avoid the details.

Part I: LLN and CLT for $\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}$

As in Section 2.5, we make use of a large/small block argument. For simplicity we assume that n^w/m_n^d is an integer and subdivide $\mathcal{D}_n$ into n^w/m_n^d non-overlapping d-dimensional large blocks $\mathcal{F} \times \mathcal{B}_i$ for $i = 1, \dots, n^w/m_n^d$, where the $\mathcal{B}_i$ are w-dimensional cubes with side lengths $m_n^{d/w}$. From those large blocks we then cut off smaller blocks, which consist of the first r_n elements in each of the w increasing dimensions. The large blocks are then separated (by these small blocks) with at least the distance r_n in all w increasing dimensions and shown to be asymptotically independent.

We divide the spatial lags in L_n into different sets according to the large and small blocks. Recall the notation of (4.12) and (4.16) and around. Observe that a spatial lag $(\boldsymbol{\ell}_{\mathcal{F}}, \boldsymbol{\ell}_{\mathcal{I}})$ with $\boldsymbol{\ell}_{\mathcal{I}} = (\ell_{\mathcal{I}}^{(1)}, \dots, \ell_{\mathcal{I}}^{(w)})$ appears in $L_{\mathcal{F}}^{(i)} \times L_n$ exactly $\mathrm{N}_{\mathcal{F}}^{(i)}(\boldsymbol{\ell}_{\mathcal{F}}) \prod_{j=1}^{w}(n - |\ell_{\mathcal{I}}^{(j)}|)$ times, where $\mathrm{N}_{\mathcal{F}}^{(i)}(\boldsymbol{\ell}_{\mathcal{F}}) = \mathrm{N}_{\mathcal{F}}^{(i,i)}(\boldsymbol{\ell}_{\mathcal{F}})$ is defined in (4.17). This term will replace $\prod_{j=1}^{d}(n - |h_j|)$ in the proofs of Chapter 2.

Lemma C.2. *Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ be a strictly stationary regularly varying process observed on $\mathcal{D}_n = \mathcal{F} \times \mathcal{I}_n$ as in (4.4). For $i \in \{1, \dots, p\}$, let $\boldsymbol{h}^{(i)} = (\boldsymbol{h}_{\mathcal{F}}^{(i)}, \boldsymbol{h}_{\mathcal{I}}^{(i)}) \in \mathcal{H} \subseteq B(\mathbf{0}, \gamma)$ for some $\gamma > 0$ be a fixed lag vector and use as before the convention that $(\boldsymbol{h}_{\mathcal{F}}^{(p+1)}, \boldsymbol{h}_{\mathcal{I}}^{(p+1)}) = \mathbf{0}$. Suppose that the following mixing conditions are satisfied.*

(1) $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ is α-mixing with respect to $\mathbb{R}^w$ with mixing coefficients $\alpha_{k_1,k_2}(\cdot)$ defined in (C.1).

(2) There exist sequences $m := m_n, r := r_n \to \infty$ with $m_n^d/n^w \to 0$ and $r_n^w/m_n^d \to 0$ as $n \to \infty$ such that (M3) and (M4i) hold.

Then for every fixed $i = 1, \dots, p+1$, as $n \to \infty$,

$$\mathbb{E}\big[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i)\big] \to \mu_{B(\mathbf{0},\gamma)}(D_i), \tag{C.3}$$

$$\mathbb{V}ar\big[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i)\big] \sim \frac{m_n^d}{n^w} \sigma^2_{B(\mathbf{0},\gamma)}(D_i), \tag{C.4}$$

with $\sigma^2_{B(\mathbf{0},\gamma)}(D_i)$ *specified in* (4.14). *If* $\mu_{B(\mathbf{0},\gamma)}(D_i) = 0$, *then* (C.4) *is interpreted as* $\mathbb{V}ar\big[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i)\big] = o(m_n^d/n^w)$. *In particular,*

$$\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i) \xrightarrow{P} \mu_{B(\mathbf{0},\gamma)}(D_i), \quad n \to \infty. \tag{C.5}$$

Proof of Lemma C.2. We suppress the superscript (i) of $\boldsymbol{h}^{(i)}$ (respectively $\boldsymbol{h}^{(i)}_{\mathcal{F}}$) for notational ease. Strict stationarity and relation (4.5) imply that

$$\mathbb{E}\Big[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i)\Big] = \frac{m_n^d}{n^w}\sum_{\boldsymbol{i}\in\mathcal{I}_n}\frac{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}})|}{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}})|}\mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_m}\in D_i\Big) = m_n^d\mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_m}\in D_i\Big) \to \mu_{B(\mathbf{0},\gamma)}(D_i).$$

As to the asymptotic variance, we start from (4.14), where it has been calculated that

$$\begin{aligned}\mathbb{V}ar\big[\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i)\big] &= \frac{m_n^{2d}}{n^{2w}|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}})|^2}\Big(|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}})|n^w\mathbb{V}ar\big[\mathbb{1}_{\{\frac{\boldsymbol{Y}(\mathbf{0})}{a_m}\in D_i\}}\big] \\ &\quad + \sum_{\boldsymbol{f},\boldsymbol{f}'\in\mathcal{F}(\boldsymbol{h}_{\mathcal{F}})}\ \sum_{\substack{\boldsymbol{i},\boldsymbol{i}'\in\mathcal{I}_n\\(\boldsymbol{f},\boldsymbol{i})\neq(\boldsymbol{f}',\boldsymbol{i}')}}\mathbb{C}ov\big[\mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})}{a_m}\in D_i\}},\mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{f}',\boldsymbol{i}')}{a_m}\in D_i\}}\big]\Big) \\ &=: A_1 + A_2. \end{aligned}\tag{C.6}$$

By (4.5) and since $\mathbb{P}(\boldsymbol{Y}(\mathbf{0})/a_m \in D_i) \to 0$,

$$A_1 = \frac{m_n^{2d}}{n^w|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}})|}\mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_m}\in D_i\Big)\Big(1-\mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_m}\in D_i\Big)\Big) \sim \frac{m_n^d}{n^w|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}})|}\mu_{B(\mathbf{0},\gamma)}(D_i)\to 0, \quad n\to\infty.$$

Counting the spatial lags as explained above this proof, for fixed $k \in \mathbb{N}$ we have by stationarity the analogy of Eq. (2.32).

$$\begin{aligned}\frac{n^w}{m_n^d}A_2 &= \frac{m_n^d}{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}})|^2}\Big(\sum_{\substack{\boldsymbol{\ell}_{\mathcal{I}}\in L_n\\0\le\|\boldsymbol{\ell}_{\mathcal{I}}\|\le k}} + \sum_{\substack{\boldsymbol{\ell}_{\mathcal{I}}\in L_n\\k<\|\boldsymbol{\ell}_{\mathcal{I}}\|\le r_n}} + \sum_{\substack{\boldsymbol{\ell}_{\mathcal{I}}\in L_n\\\|\boldsymbol{\ell}_{\mathcal{I}}\|> r_n}}\Big) \\ &\quad \sum_{\substack{\boldsymbol{\ell}_{\mathcal{F}}\in L^{(i)}_{\mathcal{F}}\\(\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}})\neq\mathbf{0}}}\mathrm{N}^{(i)}_{\mathcal{F}}(\boldsymbol{\ell}_{\mathcal{F}})\prod_{j=1}^{w}\big(1-\frac{|\ell^{(j)}_{\mathcal{I}}|}{n}\big)\mathbb{C}ov[\mathbb{1}_{\{\frac{\boldsymbol{Y}(\mathbf{0})}{a_m}\in D_i\}},\mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}})}{a_m}\in D_i\}}] \\ &=: A_{21}+A_{22}+A_{23}.\end{aligned}\tag{C.7}$$

Concerning A_{21} we have,

$$\begin{aligned}A_{21} =& \frac{m_n^d}{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}})|^2}\sum_{\substack{\boldsymbol{\ell}_{\mathcal{I}}\in L_n\\0\le\|\boldsymbol{\ell}_{\mathcal{I}}\|\le k}}\ \sum_{\substack{\boldsymbol{\ell}_{\mathcal{F}}\in L^{(i)}_{\mathcal{F}}\\(\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}})\neq\mathbf{0}}}\mathrm{N}^{(i)}_{\mathcal{F}}(\boldsymbol{\ell}_{\mathcal{F}})\prod_{j=1}^{w}\big(1-\frac{|\ell^{(j)}_{\mathcal{I}}|}{n}\big) \\ &\Big[\mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_m}\in D_i, \frac{\boldsymbol{Y}(\boldsymbol{\ell}_{\mathcal{F}},\boldsymbol{\ell}_{\mathcal{I}})}{a_m}\in D_i\Big) - \mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_m}\in D_i\Big)^2\Big].\end{aligned}$$

With (4.5) and (4.6) we obtain by dominated convergence,

$$\lim_{k\to\infty}\limsup_{n\to\infty} A_{21} = \frac{1}{|\mathcal{F}(\boldsymbol{h}_\mathcal{F})|^2}\sum_{\boldsymbol{\ell}_\mathcal{I}\in\mathbb{Z}^w}\sum_{\substack{\boldsymbol{\ell}_\mathcal{F}\in L_\mathcal{F}^{(i)}\\(\boldsymbol{\ell}_\mathcal{F},\boldsymbol{\ell}_\mathcal{I})\neq\boldsymbol{0}}} \mathrm{N}_\mathcal{F}^{(i)}(\boldsymbol{\ell}_\mathcal{F})\tau_{B(\boldsymbol{0},\gamma)\times B((\boldsymbol{\ell}_\mathcal{F},\boldsymbol{\ell}_\mathcal{I}),\gamma)}(D_i\times D_i). \tag{C.8}$$

As to A_{22}, observe that for all $n\geq 0$ we have $\prod_{j=1}^{w}(1-\frac{|\ell_\mathcal{I}^{(j)}|}{n})\leq 1$ for $\boldsymbol{\ell}_\mathcal{I}\in L_n$. Furthermore, since D_i is bounded away from $\boldsymbol{0}$, there exists $\epsilon>0$ such that $D_i\subset\{\boldsymbol{x}\in\overline{\mathbb{R}}^{|B(\boldsymbol{0},\gamma)|}:\|\boldsymbol{x}\|>\epsilon\}$. Hence, we obtain

$$\begin{aligned}|A_{22}| \leq & \frac{1}{|\mathcal{F}(\boldsymbol{h}_\mathcal{F})|^2}\sum_{\boldsymbol{\ell}_\mathcal{F}\in L_\mathcal{F}^{(i)}}\mathrm{N}_\mathcal{F}^{(i)}(\boldsymbol{\ell}_\mathcal{F})\sum_{\substack{\boldsymbol{\ell}_\mathcal{I}\in\mathbb{Z}^w\\k<\|\boldsymbol{\ell}_\mathcal{I}\|\leq r_n}}\Big\{m_n^d\mathbb{P}\Big(\|\boldsymbol{Y}(\boldsymbol{0})\|>\epsilon a_m,\|\boldsymbol{Y}(\boldsymbol{\ell}_\mathcal{F},\boldsymbol{\ell}_\mathcal{I})\|>\epsilon a_m\Big)\\ &+\frac{1}{m_n^d}\Big(m_n^d\mathbb{P}\Big(\frac{\boldsymbol{Y}(\boldsymbol{0})}{a_m}\in D_i\Big)\Big)^2\Big\}.\end{aligned}$$

which differs from the corresponding expression in Chapter 2 only by finite factors. Thus by an obvious modification of the arguments in that paper it follows that, using $r_n^w/m_n^d\to 0$ and condition (M3),

$$\lim_{k\to\infty}\limsup_{n\to\infty} A_{22}=0.$$

Using the definition (C.2) of α-mixing for $A_1=\{\boldsymbol{Y}(\boldsymbol{0})/a_m\in D_i\}$ and $A_2=\{\boldsymbol{Y}(\boldsymbol{\ell}_\mathcal{F},\boldsymbol{\ell}_\mathcal{I})/a_m\in D_i\}$, we obtain by (M4i),

$$|A_{23}|\leq\frac{1}{|\mathcal{F}(\boldsymbol{h}_\mathcal{F})|^2}\sum_{\boldsymbol{\ell}_\mathcal{F}\in L_\mathcal{F}^{(i)}}\mathrm{N}_\mathcal{F}^{(i)}(\boldsymbol{\ell}_\mathcal{F})m_n^d\sum_{\boldsymbol{\ell}_\mathcal{I}\in\mathbb{Z}^w:\|\boldsymbol{\ell}_\mathcal{I}\|>r_n}\alpha_{1,1}(\|\boldsymbol{\ell}_\mathcal{I}\|)\to 0,\quad n\to\infty. \tag{C.9}$$

Summarising these computations, we conclude from (C.7) and (C.8) that for $n\to\infty$,

$$A_2\sim\frac{m_n^d}{n^w}\sum_{\boldsymbol{\ell}_\mathcal{I}\in\mathbb{Z}^w}\frac{1}{|\mathcal{F}(\boldsymbol{h}_\mathcal{F})|^2}\sum_{\substack{\boldsymbol{\ell}_\mathcal{F}\in L_\mathcal{F}^{(i)}\\(\boldsymbol{\ell}_\mathcal{F},\boldsymbol{\ell}_\mathcal{I})\neq\boldsymbol{0}}}\mathrm{N}_\mathcal{F}^{(i)}(\boldsymbol{\ell}_\mathcal{F})\tau_{B(\boldsymbol{0},\gamma)\times B((\boldsymbol{\ell}_\mathcal{F},\boldsymbol{\ell}_\mathcal{I}),\gamma)}(D_i\times D_i),$$

and, therefore, (C.6) implies (C.4). Since $m_n^d/n^w\to 0$ as $n\to\infty$, equations (C.3) and (C.4) imply (C.5). □

Lemma C.3. *Let $\{X(\boldsymbol{s}):\boldsymbol{s}\in\mathbb{R}^d\}$ be a strictly stationary regularly varying process observed on $\mathcal{D}_n=\mathcal{F}\times\mathcal{I}_n$. For $i\in\{1,\ldots,p\}$, let $\boldsymbol{h}^{(i)}=(\boldsymbol{h}_\mathcal{F}^{(i)},\boldsymbol{h}_\mathcal{I}^{(i)})\in\mathcal{H}\subseteq B(\boldsymbol{0},\gamma)$ for some $\gamma>0$ be a fixed lag vector and take as before the convention that $(\boldsymbol{h}_\mathcal{F}^{(p+1)},\boldsymbol{h}_\mathcal{I}^{(p+1)})=\boldsymbol{0}$. Let the assumptions of Theorem 4.8 hold. Then for every fixed $i=1,\ldots,p+1$,*

$$\begin{aligned}\widehat{S}_{B(\boldsymbol{0},\gamma),m_n} &:= \sqrt{\frac{m_n^d}{n^w}}\sum_{\boldsymbol{i}\in\mathcal{I}_n}\Big[\frac{1}{|\mathcal{F}(\boldsymbol{h}_\mathcal{F})|}\Big(\sum_{\boldsymbol{f}\in\mathcal{F}(\boldsymbol{h}_\mathcal{F})}\mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})}{a_m}\in D_i\}}\Big)-\mathbb{P}\Big(\frac{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})}{a_m}\in D_i\Big)\Big]\\ &=\sqrt{\frac{n^w}{m_n^d}}\big[\widehat{\mu}_{B(\boldsymbol{0},\gamma),m_n}(D_i)-\mu_{B(\boldsymbol{0},\gamma),m_n}(D_i)\big]\xrightarrow{d}\mathcal{N}(0,\sigma^2_{B(\boldsymbol{0},\gamma)}(D_i)),\quad n\to\infty, \end{aligned}\tag{C.10}$$

with $\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i)$ *as in* (4.11), $\mu_{B(\mathbf{0},\gamma),m_n}(D_i)) := m_n^d \mathbb{P}(\boldsymbol{Y}(\mathbf{0})/a_m \in D_i)$ *and* $\sigma^2_{B(\mathbf{0},\gamma)}(D_i)$ *given in* (4.19).

Proof. Again we suppress the superscript (i) of $\boldsymbol{h}^{(i)}$ and $\boldsymbol{h}_{\mathcal{F}}^{(i)}$. As for the proof of consistency above, we generalise the proof of the CLT in Chapter 2 (based on Bolthausen [6]) to the new setting. We consider the process

$$\Big\{ \frac{\sqrt{m_n^d}}{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}})|} \Big(\sum_{\boldsymbol{f} \in \mathcal{F}(\boldsymbol{h}_{\mathcal{F}})} \mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})}{a_m} \in D_i\}} \Big) : \boldsymbol{i} \in \mathbb{Z}^w \Big\},$$

observed on the w-dimensional regular grid $\mathcal{I}_n$. In analogy to Eq. (2.37) define

$$I(\boldsymbol{i}) := \frac{1}{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}})|} \Big(\sum_{\boldsymbol{f} \in \mathcal{F}(\boldsymbol{h}_{\mathcal{F}})} \mathbb{1}_{\{\frac{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})}{a_m} \in D_i\}} \Big) - \mathbb{P}\Big(\frac{\boldsymbol{Y}(\mathbf{0})}{a_m} \in D_i \Big), \quad \boldsymbol{i} \in \mathcal{I}_n, \tag{C.11}$$

and note that by stationarity,

$$\widehat{S}_{B(\mathbf{0},\gamma),m_n} = \sqrt{\frac{m_n^d}{n^w}} \sum_{\boldsymbol{i} \in \mathcal{I}_n} I(\boldsymbol{i}). \tag{C.12}$$

The boundary condition required in Eq. (1) in Bolthausen [6] is satisfied for the regular grid $\mathcal{I}_n$. By the same arguments as in Chapter 2,

$$0 < \sigma^2_{B(\mathbf{0},\gamma)}(D_i) \sim \mathbb{V}ar[\widehat{S}_{B(\mathbf{0},\gamma),m_n}] \le \frac{m_n^d}{n^w} \sum_{\boldsymbol{i},\boldsymbol{i}' \in \mathbb{Z}^w} |\mathbb{E}[I(\boldsymbol{i})I(\boldsymbol{i}')]| < \infty, \tag{C.13}$$

such that $\sum_{\boldsymbol{i},\boldsymbol{i}' \in \mathbb{Z}^w} \mathbb{C}ov[I(\boldsymbol{i}), I(\boldsymbol{i}')] > 0$. Replacing $\mathcal{S}_n$ in Chapter 2 by $\mathcal{I}_n$ and n^d by n^w, we define

$$v_n := \frac{m_n^d}{n^w} \sum_{\substack{\boldsymbol{i},\boldsymbol{i}' \in \mathcal{I}_n \\ \|\boldsymbol{i}-\boldsymbol{i}'\| \le r_n}} \mathbb{E}\Big[I(\boldsymbol{i}) I(\boldsymbol{i}') \Big]. \tag{C.14}$$

and obtain by the same arguments that

$$\frac{v_n}{\mathbb{V}ar[\widehat{S}_{B(\mathbf{0},\gamma),m_n}]} = 1 - \frac{m_n^d}{n^w} \frac{1}{\sigma^2_{B(\mathbf{0},\gamma)}(D_i)} \sum_{\substack{\boldsymbol{i},\boldsymbol{i}' \in \mathcal{I}_n \\ \|\boldsymbol{i}-\boldsymbol{i}'\| > r_n}} \mathbb{E}[I(\boldsymbol{i})I(\boldsymbol{i}')](1 + o(1)).$$

Now note that

$$\frac{m_n^d}{n^w} \sum_{\substack{\boldsymbol{i},\boldsymbol{i}' \in \mathcal{I}_n \\ \|\boldsymbol{i}-\boldsymbol{i}'\| > r_n}} \mathbb{E}[I(\boldsymbol{i})I(\boldsymbol{i}')] \le \frac{1}{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}})|^2} \sum_{\boldsymbol{\ell}_{\mathcal{F}} \in L_{\mathcal{F}}^{(i)}} \mathrm{N}_{\mathcal{F}}^{(i)}(\boldsymbol{\ell}_{\mathcal{F}}) m_n^d \sum_{\boldsymbol{\ell}_{\mathcal{I}} \in \mathbb{Z}^q : \|\boldsymbol{\ell}_{\mathcal{I}}\| > r_n} \alpha_{1,1}(\|\boldsymbol{\ell}_{\mathcal{I}}\|) \to 0, \quad n \to \infty,$$

as in (C.9), with mixing coefficients defined in (C.2). Therefore,

$$v_n \sim \mathbb{V}ar[\widehat{S}_{B(\mathbf{0},\gamma),m_n}] \to \sigma^2_{B(\mathbf{0},\gamma)}(D_i), \quad n \to \infty. \tag{C.15}$$

The standardized quantities are again as in Chapter 2, with $\mathcal{S}_n$ replaced by $\mathcal{I}_n$ and n^d by n^w, by

$$\overline{S}_n := v_n^{-1/2}\widehat{S}_{B(\mathbf{0},\gamma),m_n} = v_n^{-1/2}\sqrt{\frac{m_n^d}{n^w}}\sum_{\boldsymbol{i}\in\mathcal{I}_n} I(\boldsymbol{i}) \quad \text{and} \quad \overline{S}_{\boldsymbol{i},n} := v_n^{-1/2}\sqrt{\frac{m_n^d}{n^w}}\sum_{\substack{\boldsymbol{i}'\in\mathcal{I}_n\\ \|\boldsymbol{i}-\boldsymbol{i}'\|\le r_n}} I(\boldsymbol{i}').$$

The proof continues as in Chapter 2, with n^d replaced by n^w, by estimating the quantities B_1, B_2 and B_3. The estimation of B_1 follows the same lines of the proof, resulting in

$$E[|B_1|^2] = \lambda^2 v_n^{-2}\Big(\frac{m_n^d}{n^w}\Big)^2 \sum_{\|\boldsymbol{i}-\boldsymbol{i}'\|\le r_n}\ \sum_{\|\boldsymbol{j}-\boldsymbol{j}'\|\le r_n} \mathbb{C}ov\big[I(\boldsymbol{i})I(\boldsymbol{i}'), I(\boldsymbol{j})I(\boldsymbol{j}')\big].$$

We use definition (C.2) of the α-mixing coefficients for

$$\Lambda_1' = \{\boldsymbol{i},\boldsymbol{i}'\} \quad \text{and} \quad \Lambda_2' = \{\boldsymbol{j},\boldsymbol{j}'\},$$

then $|\Lambda_1'|, |\Lambda_2'| \le 2$ and for $d(\Lambda_1',\Lambda_2')$ we consider the following two cases:

(1) $\|\boldsymbol{i}-\boldsymbol{j}\| \ge 3r_n$. Then $2r_n \le (2/3)\|\boldsymbol{i}-\boldsymbol{j}\|$ and $d(\Lambda_1',\Lambda_2') \ge \|\boldsymbol{i}-\boldsymbol{j}\| - 2r_n$. Since indicator variables are bounded and $\alpha_{2,2}$ is a decreasing function,

$$|\mathbb{C}ov\big[I(\boldsymbol{i})I(\boldsymbol{i}'), I(\boldsymbol{j})I(\boldsymbol{j}')\big]| \le 4\alpha_{2,2}\Big(\|\boldsymbol{i}-\boldsymbol{j}\| - 2r_n\Big) \le 4\alpha_{2,2}\Big(\frac{1}{3}\|\boldsymbol{i}-\boldsymbol{j}\|\Big).$$

(2) $\|\boldsymbol{i}-\boldsymbol{j}\| < 3r_n$. Set $z := \min\{\|\boldsymbol{i}-\boldsymbol{j}\|, \|\boldsymbol{i}-\boldsymbol{j}'\|, \|\boldsymbol{i}'-\boldsymbol{j}\|, \|\boldsymbol{i}'-\boldsymbol{j}'\|\}$, then $d(\Lambda_1',\Lambda_2') \ge z$ and, hence,

$$\mathbb{C}ov\big[I(\boldsymbol{i})I(\boldsymbol{i}'), I(\boldsymbol{j})I(\boldsymbol{j}')\big] \le 4\alpha_{k_1,k_2}(z), \quad 2 \le k_1 + k_2 \le 4.$$

Therefore,

$$\begin{aligned} E[|B_1|^2] \le& \frac{4\lambda^2}{v_n^2}\Big(\frac{m_n^d}{n^w}\Big)^2\Big[\sum_{\|\boldsymbol{i}-\boldsymbol{j}\|\ge 3r_n}\ \sum_{\substack{\|\boldsymbol{i}-\boldsymbol{i}'\|\le r_n\\ \|\boldsymbol{j}-\boldsymbol{j}'\|\le r_n}} \alpha_{2,2}\Big(\frac{1}{3}\|\boldsymbol{i}-\boldsymbol{j}\|\Big) + \sum_{\|\boldsymbol{i}-\boldsymbol{j}\|< 3r_n}\ \sum_{\substack{\|\boldsymbol{i}-\boldsymbol{i}'\|\le r_n\\ \|\boldsymbol{j}-\boldsymbol{j}'\|\le r_n}} \alpha_{k_1,k_2}(z)\Big] \\ \le& \frac{4\lambda^2}{v_n^2}\Big(\frac{m_n^d}{n^w}\Big)^2 n^w r_n^{2w}\Big[\sum_{\boldsymbol{\ell}_{\mathcal{I}}\in\mathbb{Z}^w:\|\boldsymbol{\ell}_{\mathcal{I}}\|\ge 3r_n} \alpha_{2,2}\Big(\frac{1}{3}\|\boldsymbol{\ell}_{\mathcal{I}}\|\Big) + \sum_{\boldsymbol{\ell}_{\mathcal{I}}\in\mathbb{Z}^w:\|\boldsymbol{\ell}_{\mathcal{I}}\|< 3r_n} \alpha_{k_1,k_2}(\|\boldsymbol{\ell}_{\mathcal{I}}\|)\Big]. \end{aligned}$$

The analogous argument as in Chapter 2 yields

$$E[|B_1|^2] = \mathcal{O}\Big(\frac{m_n^{2d} r_n^{2w}}{n^w}\Big) \to 0.$$

Next, $\mathbb{E}[|B_2|] \to 0$ as $n \to \infty$ by the same arguments as in Chapter 2, replacing $\mathcal{S}_n$ by $\mathcal{I}_n$ and n^d by n^w. Then we find for B_3 with the same replacements

$$\mathbb{E}[B_3] = v_n^{-\frac{1}{2}} m_n^{d/2} n^{w/2}\mathbb{E}\Big[I(\mathbf{0})\exp\Big\{i\lambda v_n^{-\frac{1}{2}}\sqrt{\frac{m_n^d}{n^w}}\sum_{\|\boldsymbol{i}\|>r_n} I(\boldsymbol{i})\Big\}\Big].$$

We use definition (C.2) of the α-mixing coefficients for

$$\Lambda_1' = \{\mathbf{0}\} \quad \text{and} \quad \Lambda_2' = \{\boldsymbol{i} \in \mathcal{I}_n : \|\boldsymbol{i}\| > r_n\},$$

such that $|\Lambda_1'| = 1$, $|\Lambda_2'| \le n^w$ and $d(\Lambda_1', \Lambda_2') > r_n$. Abbreviate

$$\eta(r_n) := \exp\Big\{ i\lambda v_n^{-\frac{1}{2}} \sqrt{\frac{m_n^d}{n^w}} \sum_{\|\boldsymbol{i}\| > r_n} I(\boldsymbol{i}) \Big\},$$

then $I(\mathbf{0})$ and $\eta(r_n)$ are measurable with respect to σ_{Λ_1} and σ_{Λ_2}, respectively, where $\Lambda_i = \cup_{\boldsymbol{s} \in \mathcal{F} \times \Lambda_i'} B(\boldsymbol{s}, \gamma)$. Now we apply Theorem 17.2.1 of Ibragimov and Linnik to obtain

$$|\mathbb{E}[B_3]| \le 4 v_n^{-1/2} m_n^{d/2} n^{w/2} \alpha_{1,n^w}(r_n) \to 0,$$

where convergence to 0 is guaranteed by condition (M4iii). $\square$

Part II: CLT for $\widehat{\rho}_{AB,m_n}$ and limit covariance matrix

Recall the definition of $\mathcal{H} = \{\boldsymbol{h}^{(1)}, \dots, \boldsymbol{h}^{(p)}\}$. For $i \in \{1, \dots, p\}$, write $\boldsymbol{h}^{(i)} = (\boldsymbol{h}_{\mathcal{F}}^{(i)}, \boldsymbol{h}_{\mathcal{I}}^{(i)})$ with respect to the fixed and increasing domains $\mathcal{F}$ and $\mathcal{I}_n$. Write further $\boldsymbol{h}_{\mathcal{F}}^{(i)} = (h_{\mathcal{F}}^{(i,1)}, \dots, h_{\mathcal{F}}^{(i,q)})$ and $\boldsymbol{h}_{\mathcal{I}}^{(i)} = (h_{\mathcal{I}}^{(i,1)}, \dots, h_{\mathcal{I}}^{(i,w)})$. Now we define the ratio

$$R_n(D_i, D_{p+1}) := \frac{\mathbb{P}(\boldsymbol{Y}(\mathbf{0})/a_m \in D_i)}{\mathbb{P}(\boldsymbol{Y}(\mathbf{0})/a_m \in D_{p+1})} = \frac{\mu_{B(\mathbf{0},\gamma),m_n}(D_i)}{\mu_{B(\mathbf{0},\gamma),m_n}(D_{p+1})}$$

and the corresponding empirical estimator

$$\begin{aligned}
\widehat{R}_n(D_i, D_{p+1}) &:= \frac{|\mathcal{F}| \sum_{\boldsymbol{i} \in \mathcal{I}_n} \sum_{\boldsymbol{f} \in \mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})} \mathbb{1}_{\{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})/a_m \in D_i\}}}{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})| \sum_{\boldsymbol{i} \in \mathcal{I}_n} \sum_{\boldsymbol{f} \in \mathcal{F}} \mathbb{1}_{\{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})/a_m \in D_{p+1}\}}} \\
&= \frac{\frac{m_n^d}{n^w} \sum_{\boldsymbol{i} \in \mathcal{I}_n} \frac{1}{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})|} \sum_{\boldsymbol{f} \in \mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})} \mathbb{1}_{\{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})/a_m \in D_i\}}}{\frac{m_n^d}{n^w} \sum_{\boldsymbol{i} \in \mathcal{I}_n} \frac{1}{|\mathcal{F}(\mathbf{0})|} \sum_{\boldsymbol{f} \in \mathcal{F}(\mathbf{0})} \mathbb{1}_{\{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})/a_m \in D_{p+1}\}}} = \frac{\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_i)}{\widehat{\mu}_{B(\mathbf{0},\gamma),m_n}(D_{p+1})},
\end{aligned}$$

using that $\mathcal{F}(\mathbf{0}) = \mathcal{F}$. Observe that

$$|\mathcal{D}_n(\boldsymbol{h}^{(i)})| = |\mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})| \prod_{j=1}^{w} (n - |h_{\mathcal{I}}^{(i,j)}|) \sim |\mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})| n^w, \quad n \to \infty.$$

Then the empirical extremogram as defined in (4.9) for μ-continuous Borel sets A, B in $\overline{\mathbb{R}} \backslash \{0\}$ satisfies as $n \to \infty$,

$$\widehat{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)}) = \frac{\frac{1}{|\mathcal{D}_n(\boldsymbol{h}^{(i)})|} \sum_{\boldsymbol{s} \in \mathcal{D}_n(\boldsymbol{h}^{(i)})} \mathbb{1}_{\{X(\boldsymbol{s})/a_m \in A, X(\boldsymbol{s}+\boldsymbol{h}^{(i)})/a_m \in B\}}}{\frac{1}{|\mathcal{D}_n|} \sum_{\boldsymbol{s} \in \mathcal{D}_n} \mathbb{1}_{\{X(\boldsymbol{s})/a_m \in A\}}}$$

$$\sim \frac{\frac{1}{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})|n^w} \sum_{\boldsymbol{i}\in\mathcal{I}_n(\boldsymbol{h}_{\mathcal{I}}^{(i)})} \sum_{\boldsymbol{f}\in\mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})} \mathbb{1}_{\{X(\boldsymbol{f},\boldsymbol{i})/a_m\in A, X(\boldsymbol{f}+\boldsymbol{h}_{\mathcal{F}}^{(i)},\boldsymbol{i}+\boldsymbol{h}_{\mathcal{I}}^{(i)})/a_m\in B\}}}{\frac{1}{|\mathcal{F}|n^w} \sum_{\boldsymbol{i}\in\mathcal{I}_n} \sum_{\boldsymbol{f}\in\mathcal{F}} \mathbb{1}_{\{X(\boldsymbol{f},\boldsymbol{i})/a_m\in D_{p+1}\}}}$$
$$\sim \frac{|\mathcal{F}| \sum_{\boldsymbol{i}\in\mathcal{I}_n} \sum_{\boldsymbol{f}\in\mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})} \mathbb{1}_{\{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})/a_m\in D_i\}}}{|\mathcal{F}(\boldsymbol{h}_{\mathcal{F}}^{(i)})| \sum_{\boldsymbol{i}\in\mathcal{I}_n} \sum_{\boldsymbol{f}\in\mathcal{F}} \mathbb{1}_{\{\boldsymbol{Y}(\boldsymbol{f},\boldsymbol{i})/a_m\in D_{p+1}\}}} = \widehat{R}_n(D_i, D_{p+1}),$$

by definition (4.7) of the sets D_i for $i = 1, \ldots, p$. The remaining proof follows exactly as that of Theorem 2.6, where in the last part the decomposition into a fixed and increasing grid has to be taken into account. □

C.3 Proof of Theorem 4.12

Throughout this proof, we suppress the sub index m_n of $\widehat{\rho}_{AB,m_n}$ and $\widehat{\rho}_{AB,m_n}$ for notational ease. The case, where $n^w/m_n^{3d} \to 0$ as $n \to \infty$, is covered by Theorem 4.11, so we assume that $n^w/m_n^{3d} \not\to 0$. Hence, by definition (4.25) we have to consider

$$\widetilde{\rho}_{AB}(\boldsymbol{h}) = \widehat{\rho}_{AB}(\boldsymbol{h}) - \frac{1}{2m_n^d a}\Big[(\widehat{\rho}_{AB}(\boldsymbol{h}) - 2\frac{a}{b})(\widehat{\rho}_{AB}(\boldsymbol{h}) - 1)\Big].$$

Observe that for $\boldsymbol{h} \in \mathcal{H} = \{\boldsymbol{h}^{(1)}, \ldots, \boldsymbol{h}^{(p)}\}$, as $n \to \infty$,

$$\begin{aligned}
&\widetilde{\rho}_{AB}(\boldsymbol{h}) - \rho_{AB}(\boldsymbol{h}) \\
&\quad= \widehat{\rho}_{AB}(\boldsymbol{h}) - \rho_{AB,m_n}(\boldsymbol{h}) + \rho_{AB,m_n}(\boldsymbol{h}) - \frac{1}{2m_n^d a}\Big[(\widehat{\rho}_{AB}(\boldsymbol{h}) - 2\frac{a}{b})(\widehat{\rho}_{AB}(\boldsymbol{h}) - 1)\Big] - \rho_{AB}(\boldsymbol{h}) \\
&\quad= (1+o(1))\Big\{\widehat{\rho}_{AB}(\boldsymbol{h}) - \rho_{AB,m_n}(\boldsymbol{h}) + \rho_{AB}(\boldsymbol{h}) + \frac{1}{2m_n^d a}\Big[(\rho_{AB}(\boldsymbol{h}) - 2\frac{a}{b})(\rho_{AB}(\boldsymbol{h}) - 1)\Big] \\
&\qquad - \frac{1}{2m_n^d a}\Big[(\widehat{\rho}_{AB}(\boldsymbol{h}) - 2\frac{a}{b})(\widehat{\rho}_{AB}(\boldsymbol{h}) - 1)\Big] - \rho_{AB}(\boldsymbol{h})\Big\}
\end{aligned}$$

Since the conditions of Theorem 4.8 are satisfied we have that

$$\sqrt{\frac{n^w}{m_n^d}}\Big[\widehat{\rho}_{AB}(\boldsymbol{h}^{(i)}) - \rho_{AB,m_n}(\boldsymbol{h}^{(i)})\Big]_{i=1,\ldots,p} \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \Pi)$$

and thus, by the continuous mapping theorem, it remains to show that for $\boldsymbol{h} \in \mathcal{H}$,

$$\sqrt{\frac{n^w}{4m_n^{3d}}}\frac{1}{a}\Big[(\widehat{\rho}_{AB}(\boldsymbol{h}) - 2\frac{a}{b})(\widehat{\rho}_{AB}(\boldsymbol{h}) - 1) - (\rho_{AB}(\boldsymbol{h}) - 2\frac{a}{b})(\rho_{AB}(\boldsymbol{h}) - 1)\Big] \xrightarrow{P} 0.$$

We rewrite the latter as

$$\begin{aligned}
&\sqrt{\frac{n^w}{4m_n^{3d}}}\frac{1}{a}\Big[(\widehat{\rho}_{AB}(\boldsymbol{h}) - 2\frac{a}{b})(\widehat{\rho}_{AB}(\boldsymbol{h}) - 1) - (\rho_{AB,m_n}(\boldsymbol{h}) - 2\frac{a}{b})(\rho_{AB,m_n}(\boldsymbol{h}) - 1) \\
&\qquad\qquad + (\rho_{AB,m_n}(\boldsymbol{h}) - 2\frac{a}{b})(\rho_{AB,m_n}(\boldsymbol{h}) - 1) - (\rho_{AB}(\boldsymbol{h}) - 2\frac{a}{b})(\rho_{AB}(\boldsymbol{h}) - 1)\Big] \\
&\quad =: A_1 + A_2.
\end{aligned}$$

As to A_1, we calculate

$$\sqrt{\frac{n^w}{4m_n^d}}\frac{1}{2\rho_{AB}(\boldsymbol{h})-(2\frac{a}{b}+1)}\Big[(\widehat{\rho}_{AB}(\boldsymbol{h})-2\frac{a}{b})(\widehat{\rho}_{AB}(\boldsymbol{h})-1)-(\rho_{AB,m_n}(\boldsymbol{h})-2\frac{a}{b})(\rho_{AB,m_n}(\boldsymbol{h})-1)\Big]$$
$$=\sqrt{\frac{n^w}{4m_n^d}}\frac{1}{2\rho_{AB}(\boldsymbol{h})-(2\frac{a}{b}+1)}\Big[\widehat{\rho}_{AB}(\boldsymbol{h})^2-(2\frac{a}{b}+1)\widehat{\rho}_{AB}(\boldsymbol{h})-\Big(\rho^2_{AB,m_n}(\boldsymbol{h})-(2\frac{a}{b}+1)\rho_{AB,m_n}(\boldsymbol{h})\Big)\Big]$$
$$=\sqrt{\frac{n^w}{4m_n^d}}\frac{1}{2\rho_{AB}(\boldsymbol{h})-(2\frac{a}{b}+1)}\Big[(\widehat{\rho}_{AB}(\boldsymbol{h})-\rho_{AB,m_n}(\boldsymbol{h}))(\widehat{\rho}_{AB}(\boldsymbol{h})+\rho_{AB,m_n}(\boldsymbol{h}))$$
$$-(2\frac{a}{b}+1)(\widehat{\rho}_{AB}(\boldsymbol{h})-\rho_{AB,m_n}(\boldsymbol{h}))\Big]$$
$$=\sqrt{\frac{n^w}{4m_n^d}}(\widehat{\rho}_{AB}(\boldsymbol{h})-\rho_{AB,m_n}(\boldsymbol{h}))\frac{\widehat{\rho}_{AB}(\boldsymbol{h})+\rho_{AB,m_n}(\boldsymbol{h})-(2\frac{a}{b}+1)}{2\rho_{AB}(\boldsymbol{h})-(2\frac{a}{b}+1)}.$$

By Theorem 4.8, the first term converges weakly to a normal distribution. Since $\widehat{\rho}_{AB}(\boldsymbol{h})\stackrel{P}{\to}\rho_{AB}(\boldsymbol{h})$ and $\rho_{AB,m_n}(\boldsymbol{h})\to\rho_{AB}(\boldsymbol{h})$ as $n\to\infty$, the second term converges to 1 in probability. Slutzky's theorem hence yields that $A_1\stackrel{P}{\to}0$. As to A_2, observe that

$$-\sqrt{\frac{4m_n^{3d}}{n^w}}aA_2=\rho^2_{AB}(\boldsymbol{h})-\rho^2_{AB,m_n}(\boldsymbol{h}))+(2\frac{a}{b}+1)(\rho_{AB,m_n}(\boldsymbol{h})-\rho_{AB}(\boldsymbol{h}))$$
$$=(1+o(1))\Big\{\rho^2_{AB}(\boldsymbol{h})-\Big[\rho_{AB}(\boldsymbol{h})+\frac{1}{2m_n^da}\Big[(\rho_{AB}(\boldsymbol{h})-2\frac{a}{b})(\rho_{AB}(\boldsymbol{h})-1)\Big]\Big]^2$$
$$+(2\frac{a}{b}+1)\Big[\rho_{AB}(\boldsymbol{h})+\frac{1}{2m_n^da}\Big[(\rho_{AB}(\boldsymbol{h})-2\frac{a}{b})(\rho_{AB}(\boldsymbol{h})-1)\Big]-\rho_{AB}(\boldsymbol{h})\Big]\Big\}$$
$$=(1+o(1))\Big\{\rho^2_{AB}(\boldsymbol{h})-\rho^2_{AB}(\boldsymbol{h})-\frac{\rho_{AB}(\boldsymbol{h})}{m_n^da}\Big[(\rho_{AB}(\boldsymbol{h})-2\frac{a}{b})(\rho_{AB}(\boldsymbol{h})-1)\Big]$$
$$-\frac{1}{4m_n^{2d}a^2}\Big[(\rho_{AB}(\boldsymbol{h})-2\frac{a}{b})(\rho_{AB}(\boldsymbol{h})-1)\Big]^2$$
$$+(2\frac{a}{b}+1)\Big[\rho_{AB}(\boldsymbol{h})+\frac{1}{2m_n^da}\Big[(\rho_{AB}(\boldsymbol{h})-2\frac{a}{b})(\rho_{AB}(\boldsymbol{h})-1)\Big]-\rho_{AB}(\boldsymbol{h})\Big]\Big\}$$
$$=(1+o(1))\Big\{\frac{1}{m_n^da}\Big[(\frac{a}{b}+\frac{1}{2}-\rho_{AB}(\boldsymbol{h}))\Big[(\rho_{AB}(\boldsymbol{h})-2\frac{a}{b})(\rho_{AB}(\boldsymbol{h})-1)\Big]$$
$$-\frac{1}{4m_n^da}\Big[(\rho_{AB}(\boldsymbol{h})-2\frac{a}{b})(\rho_{AB}(\boldsymbol{h})-1)\Big]^2\Big]\Big\}.$$

Therefore A_2 converges to 0 if and only if $\sqrt{n^w/m_n^{3d}}m_n^{-d}=\sqrt{n^w/m_n^{5d}}$ converges to 0. □

C.4 Proof of Theorem 4.16

We start with the proof of consistency and use a subsequence argument. Let $n'=n'(n)$ be some arbitrary subsequence of n. We show that there exists a further subsequence $n''=n''(n')$ such that $\widehat{\boldsymbol{\theta}}_{n'',V}\stackrel{\text{a.s.}}{\to}\boldsymbol{\theta}^\star$ as $n\to\infty$, which in turn implies (4.34).
By (G1) we have for $i=1,\dots,p$ that $\widehat{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)})\stackrel{P}{\to}\rho_{AB,\boldsymbol{\theta}^\star}(\boldsymbol{h}^{(i)})$ as $n\to\infty$. Hence, there

exists a subsequence n'' of n' such that

$$\big[\widehat{\rho}_{AB,m_{n''}}(\boldsymbol{h}^{(i)})\big]_{i=1,\dots,p} \overset{\text{a.s.}}{\to} \big[\rho_{AB,\boldsymbol{\theta}^\star}(\boldsymbol{h}^{(i)})\big]_{i=1,\dots,p}, \tag{C.16}$$

as $n \to \infty$. For $\boldsymbol{\theta} \in \Theta$, we define the column vector and the quadratic forms

$$\begin{aligned} g(\boldsymbol{\theta}) &:= \big[\rho_{AB,\boldsymbol{\theta}^\star}(\boldsymbol{h}^{(i)}) - \rho_{AB,\boldsymbol{\theta}}(\boldsymbol{h}^{(i)}) : i = 1,\dots,p\big]^\intercal_{i=1,\dots,p}, \\ Q(\boldsymbol{\theta}) &:= g(\boldsymbol{\theta})^T V(\boldsymbol{\theta}) g(\boldsymbol{\theta}) \quad \text{and} \quad \widehat{Q}_n(\boldsymbol{\theta}) := \widehat{\boldsymbol{g}}_n(\boldsymbol{\theta})^\intercal V(\boldsymbol{\theta}) \widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}), \end{aligned}$$

where we recall from (4.31) that $\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}) = \big[\widehat{\rho}_{AB,m_n}(\boldsymbol{h}^{(i)}) - \rho_{AB,\boldsymbol{\theta}}(\boldsymbol{h}^{(i)})\big]^\intercal_{i=1,\dots,p}$. Assumptions (G1) and (G3) imply that $Q(\boldsymbol{\theta}) > 0$ for $\boldsymbol{\theta}^\star \neq \boldsymbol{\theta} \in \Theta$ and that $Q(\boldsymbol{\theta}^\star) = 0$, so $\boldsymbol{\theta}^\star$ is the unique minimizer of Q. Smoothness and continuity of the functions $\rho_{AB,\boldsymbol{\theta}}(\boldsymbol{h}^{(i)})$ and $V(\boldsymbol{\theta})$ (Assumptions (G4) and (G5) with $z_1 = z_2 = 0$) and (C.16) yield

$$\widehat{\Delta}_{n''} := \sup_{\boldsymbol{\theta} \in \Theta} \{|\widehat{Q}_{n''}(\boldsymbol{\theta}) - Q(\boldsymbol{\theta})|\} \overset{\text{a.s.}}{\to} 0, \quad n \to \infty. \tag{C.17}$$

Now assume that there exists some $\omega \in \Omega$ such that (C.17) holds, but $\widehat{\boldsymbol{\theta}}_{n'',V}(\omega) \not\to \boldsymbol{\theta}^\star$. Then there exist $\epsilon > 0$ and a subsequence $n''' = n'''(n'')$ such that for all $n \geq 1$,

$$\|\widehat{\boldsymbol{\theta}}_{n''',V}(\omega) - \boldsymbol{\theta}^\star\| > \epsilon.$$

Thus,

$$\begin{aligned} &\widehat{Q}_{n'''}(\widehat{\boldsymbol{\theta}}_{n''',V}(\omega)) - \widehat{Q}_{n'''}(\boldsymbol{\theta}^\star) \\ &= -(Q(\widehat{\boldsymbol{\theta}}_{n''',V}(\omega)) - \widehat{Q}_{n'''}(\widehat{\boldsymbol{\theta}}_{n''',V}(\omega))) + Q(\widehat{\boldsymbol{\theta}}_{n''',V}(\omega)) - (\widehat{Q}_{n'''}(\boldsymbol{\theta}^\star) - Q(\boldsymbol{\theta}^\star)) - Q(\boldsymbol{\theta}^\star) \\ &\geq Q(\widehat{\boldsymbol{\theta}}_{n''',V}(\omega)) - Q(\boldsymbol{\theta}^\star) - 2\widehat{\Delta}_{n'''} = Q(\widehat{\boldsymbol{\theta}}_{n''',V}(\omega)) - 2\widehat{\Delta}_{n'''} \\ &\geq \inf\{Q(\boldsymbol{\theta}) : \|\boldsymbol{\theta} - \boldsymbol{\theta}^\star\| > \epsilon\} - 2\widehat{\Delta}_{n'''} > 0 \end{aligned}$$

for all $n \geq n_0$ for some $n_0 \geq 1$. But this contradicts the definition of $\widehat{\boldsymbol{\theta}}_{n''',V}$ as the minimizer of $\widehat{Q}_{n'''}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. Hence $\widehat{\boldsymbol{\theta}}_{n'',V} \overset{\text{a.s.}}{\to} \boldsymbol{\theta}^\star$ as $n \to \infty$ and this shows (4.34).

To prove the CLT (4.35), we introduce the following notation:

- We denote by $\boldsymbol{e}_\ell \in \mathbb{R}^k$ the ℓth unit vector.
- For $1 \leq i, j \leq p$, let $v_{ij}(\boldsymbol{\theta}) := (V(\boldsymbol{\theta}))_{ij}$ be the entry in the ith row and jth column of $V(\boldsymbol{\theta})$.
- Set $v_{ij}^{(\ell)}(\boldsymbol{\theta}) := \frac{\partial}{\partial \theta_\ell} v_{ij}(\boldsymbol{\theta})$ and $V^{(\ell)}(\boldsymbol{\theta}) := (v_{ij}^{(\ell)}(\boldsymbol{\theta}))_{1 \leq i,j \leq p}, \quad 1 \leq \ell \leq k.$

As $\widehat{\boldsymbol{\theta}}_{n,V}$ minimizes $\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta})^\intercal V(\boldsymbol{\theta}) \widehat{\boldsymbol{g}}_n(\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$, we obtain for $1 \leq \ell \leq k$,

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta_\ell} (\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta})^\intercal V(\boldsymbol{\theta}) \widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}))\Big|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_{n,V}} \\ &= \widehat{\boldsymbol{g}}_n(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal V^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V}) \widehat{\boldsymbol{g}}_n(\widehat{\boldsymbol{\theta}}_{n,V}) - \boldsymbol{\rho}_{AB}^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal [V(\widehat{\boldsymbol{\theta}}_{n,V}) + V(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal] \widehat{\boldsymbol{g}}_n(\widehat{\boldsymbol{\theta}}_{n,V}). \end{aligned} \tag{C.18}$$

Now define the $p \times k$-matrix $\widehat{\mathrm{P}}_{AB,n} := \int_0^1 \mathrm{P}_{AB}(u\boldsymbol{\theta}^\star + (1-u)\widehat{\boldsymbol{\theta}}_{n,V})\, \mathrm{d}u$, where the integral is taken componentwise. Assumptions (G4) and (G5) with $z_1 = z_2 = 1$ allow for a multivariate Taylor expansion of order 0 with integral remainder term of $\widehat{\boldsymbol{g}}_n(\widehat{\boldsymbol{\theta}}_{n,V})$ around the true parameter vector $\boldsymbol{\theta}^\star$, which yields

$$\widehat{\boldsymbol{g}}_n(\widehat{\boldsymbol{\theta}}_{n,V}) = \widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star) + \widehat{\mathrm{P}}_{AB,n} \cdot (\widehat{\boldsymbol{\theta}}_{n,V} - \boldsymbol{\theta}^\star).$$

Plugging this into (C.18) and rearranging terms, we find

$$\begin{aligned}
&\Big(- \boldsymbol{\rho}_{AB}^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal [V(\widehat{\boldsymbol{\theta}}_{n,V}) + V(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal]\widehat{\mathrm{P}}_{AB,n} + (\widehat{\boldsymbol{\theta}}_{n,V} - \boldsymbol{\theta}^\star)^\intercal \widehat{\mathrm{P}}_{AB,n}{}^\intercal V^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V})\widehat{\mathrm{P}}_{AB,n}\Big)(\widehat{\boldsymbol{\theta}}_{n,V} - \boldsymbol{\theta}^\star)\\
=&\boldsymbol{\rho}_{AB}^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal [V(\widehat{\boldsymbol{\theta}}_{n,V}) + V(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal]\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star) - \widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star)^\intercal V^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V})\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star)\\
&- \widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star)^\intercal [V^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V}) + V^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal]\widehat{\mathrm{P}}_{AB,n}(\widehat{\boldsymbol{\theta}}_{n,V} - \boldsymbol{\theta}^\star)
\end{aligned} \tag{C.19}$$

for $1 \leq \ell \leq k$. Defining $\widehat{R}_{n,V}$ as the $k \times k$-matrix whose ℓth row is given by

$$(\widehat{\boldsymbol{\theta}}_{n,V} - \boldsymbol{\theta}^\star)^\intercal \widehat{\mathrm{P}}_{AB,n}{}^\intercal V^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V})\widehat{\mathrm{P}}_{AB,n}, \quad 1 \leq \ell \leq k,$$

the system of equations (C.19) can be written in compact matrix form as

$$\begin{aligned}
&(\mathrm{P}_{AB}(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal [V(\widehat{\boldsymbol{\theta}}_{n,V}) + V(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal]\widehat{\mathrm{P}}_{AB,n} + \widehat{R}_{n,V})(\widehat{\boldsymbol{\theta}}_{n,V} - \boldsymbol{\theta}^\star)\\
=&- \mathrm{P}_{AB}(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal [V(\widehat{\boldsymbol{\theta}}_{n,V}) + V(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal]\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star) - \sum_{\ell=1}^{k} \widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star)^\intercal V^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V})\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star)\boldsymbol{e}_\ell\\
&- \sum_{\ell=1}^{k} \widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star)^\intercal [V^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V}) + V^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal]\widehat{\mathrm{P}}_{AB,n}(\widehat{\boldsymbol{\theta}}_{n,V} - \boldsymbol{\theta}^\star)\boldsymbol{e}_\ell.
\end{aligned} \tag{C.20}$$

Hence, multiplying (C.20) by $\sqrt{n^w/m_n^d}$ and rearranging terms, we have,

$$\begin{aligned}
&\sqrt{\frac{n^w}{m_n^d}}(\widehat{\boldsymbol{\theta}}_{n,V} - \boldsymbol{\theta}^\star)\\
=&- \{\mathrm{P}_{AB}(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal [V(\widehat{\boldsymbol{\theta}}_{n,V}) + V(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal]\widehat{\mathrm{P}}_{AB,n} + \widehat{R}_{n,V}\}^{-1}\\
&\quad \times \mathrm{P}_{AB}(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal [V(\widehat{\boldsymbol{\theta}}_{n,V}) + V(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal]\sqrt{\frac{n^w}{m_n^d}}\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star)\\
&- \{\mathrm{P}_{AB}(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal [V(\widehat{\boldsymbol{\theta}}_{n,V}) + V(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal]\widehat{\mathrm{P}}_{AB,n} + \widehat{R}_{n,V}\}^{-1} \sum_{\ell=1}^{k} \sqrt{\frac{n^w}{m_n^d}}\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star)^\intercal V^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V})\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star)\boldsymbol{e}_\ell\\
&- \{\mathrm{P}_{AB}(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal [V(\widehat{\boldsymbol{\theta}}_{n,V}) + V(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal]\widehat{\mathrm{P}}_{AB,n} + \widehat{R}_{n,V}\}^{-1}\\
&\quad \times \sum_{\ell=1}^{k} \sqrt{\frac{n^w}{m_n^d}}\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star)^\intercal [V^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V}) + V^{(\ell)}(\widehat{\boldsymbol{\theta}}_{n,V})^\intercal]\widehat{\mathrm{P}}_{AB,n}(\widehat{\boldsymbol{\theta}}_{n,V} - \boldsymbol{\theta}^\star)\boldsymbol{e}_\ell\\
&=: -A - B - C.
\end{aligned}$$

Observe that the smoothness conditions (G4) and (G5) and the rank condition (G6) ensure invertibility of the terms in curly brackets and boundedness of its inverse. For the remainder of the proof, we can hence use Slutsky's theorem; to this end note that, as $n \to \infty$:

- By conditions (G4i) and (G5i) with $z_1 = z_2 = 1$, the matrices $V(\boldsymbol{\theta})$ and $\mathrm{P}_{AB}(\boldsymbol{\theta})$ are continuous in $\boldsymbol{\theta}$, hence $V(\widehat{\boldsymbol{\theta}}_{n,V}) \xrightarrow{P} V(\boldsymbol{\theta}^\star)$ and $\mathrm{P}_{AB}(\widehat{\boldsymbol{\theta}}_{n,V}) \xrightarrow{P} \mathrm{P}_{AB}(\boldsymbol{\theta}^\star)$ by continuous mapping.
- Using (4.34), we find that $(\widehat{\boldsymbol{\theta}}_{n,V} - \boldsymbol{\theta}^\star) \xrightarrow{P} \mathbf{0}$, $\widehat{R}_{n,V} \xrightarrow{P} (\mathbf{0}, \ldots, \mathbf{0})$ and $\widehat{\mathrm{P}}_{AB,n} \xrightarrow{P} \mathrm{P}_{AB}(\boldsymbol{\theta}^\star)$.
- The previous bullet point directly implies that $C \xrightarrow{P} \mathbf{0}$.
- As to A, condition (G2) directly yields $\sqrt{\frac{n^w}{m_n^d}}\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Pi)$.
- Furthermore, $\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}^\star) \xrightarrow{P} \mathbf{0}$ by (G1) and therefore $B \xrightarrow{P} \mathbf{0}$.

Finally, summarising those results, with $B(\boldsymbol{\theta}^\star) = \left(\mathrm{P}_{AB}(\boldsymbol{\theta}^\star)^\intercal [V(\boldsymbol{\theta}^\star) + V(\boldsymbol{\theta}^\star)^\intercal]\mathrm{P}_{AB}(\boldsymbol{\theta}^\star)\right)^{-1}$, we obtain (4.35). $\square$

Appendix to Chapter 5

D.1 An auxiliary lemma

Lemma D.1. *The following two bounds hold true for $r \geq 1$, $\alpha \in (0,2]$ and $C > 0$:*

$$\int_y^\infty u^r \mathrm{e}^{-Cu^\alpha}\,\mathrm{d}u \sim \tfrac{1}{C\alpha} y^{r-\alpha+1} \mathrm{e}^{-Cy^\alpha}, \quad y \to \infty, \tag{D.1}$$

$$\int_1^\infty \Big(\int_y^\infty u^r \mathrm{e}^{-Cu^\alpha}\,\mathrm{d}u \Big)^{\frac{1}{3}} \mathrm{d}y < \infty. \tag{D.2}$$

Proof. First note that integrals of the form $\int_0^\infty u^r \mathrm{e}^{-Cu^\alpha}\,\mathrm{d}u$ are finite for every $r > -1$, $\alpha \in (0,2]$, and $C > 0$, since they are transformations of the gamma function $\Gamma(x) = \int_0^\infty t^{x-1}\mathrm{e}^{-t}\,\mathrm{d}t$, which exists for positive x. We prove (D.1) by an application of l'Hôpital's rule:

$$\lim_{y\to\infty} \frac{\int_y^\infty u^r \mathrm{e}^{-Cu^\alpha}\,\mathrm{d}u}{\frac{1}{C\alpha} y^{r-\alpha+1} \mathrm{e}^{-Cy^\alpha}} = \lim_{y\to\infty} \frac{-y^r \mathrm{e}^{-Cy^\alpha}}{\left(-y^r + \frac{r-\alpha+1}{C\alpha} y^{r-\alpha}\right) \mathrm{e}^{-Cy^\alpha}} = \lim_{y\to\infty} \frac{y^r}{y^r \left(1 - \frac{r-\alpha+1}{C\alpha} y^{-\alpha}\right)} = 1.$$

In order to prove (D.2) first note that it follows from (D.1) that for every $\epsilon > 0$ there exists $y_0 = y_0(\epsilon)$ such that for all $y \geq y_0$,

$$\Big(\int_y^\infty u^r \mathrm{e}^{-Cu^\alpha}\,\mathrm{d}u \Big)^{\frac{1}{3}} \leq (1+\epsilon) \Big(\frac{1}{C\alpha} \Big)^{\frac{1}{3}} y^{\frac{r-\alpha+1}{3}} \mathrm{e}^{-\frac{C}{3} y^\alpha}. \tag{D.3}$$

Now we split the double integral of (D.2) up into

$$\int_1^{y_0} \Big(\int_y^\infty u^r \mathrm{e}^{-Cu^\alpha}\,\mathrm{d}u \Big)^{\frac{1}{3}} \mathrm{d}y + \int_{y_0}^\infty \Big(\int_y^\infty u^r \mathrm{e}^{-Cu^\alpha}\,\mathrm{d}u \Big)^{\frac{1}{3}} \mathrm{d}y =: I_1 + I_2.$$

For I_1 we obtain

$$I_1 \leq \int_1^{y_0} \Big(\int_y^{y_0} u^r \mathrm{e}^{-Cu^\alpha}\,\mathrm{d}u \Big)^{\frac{1}{3}} \mathrm{d}y + \int_1^{y_0} \Big(\int_{y_0}^\infty u^r \mathrm{e}^{-Cu^\alpha}\,\mathrm{d}u \Big)^{\frac{1}{3}} \mathrm{d}y =: I_1^{(1)} + I_1^{(2)}.$$

$I_1^{(1)}$ is obviously finite, and to bound $I_1^{(2)}$ we use (D.3), which yields

$$I_1^{(2)} \leq (y_0 - 1)(1+\epsilon)\Big(\frac{1}{C\alpha}\Big)^{\frac{1}{3}} y_0^{\frac{r-\alpha+1}{3}} \mathrm{e}^{-\frac{C}{3} y_0^{\alpha}} < \infty.$$

Concerning I_2, note that

$$I_2 \leq (1+\epsilon)\Big(\frac{1}{C\alpha}\Big)^{\frac{1}{3}} \int\limits_{y_0}^{\infty} y^{\frac{r-\alpha+1}{3}} \mathrm{e}^{-\frac{C}{3} y^{\alpha}} \, \mathrm{d}y,$$

which is finite by finiteness of the gamma function. □

Bibliography

[1] D. W. K. Andrews. Estimation when a parameter is on a boundary. *Econometrica*, 67(6): 1341–1383, 1999.

[2] P. Asadi, A. C. Davison, and S. Engelke. Extremes on river networks. *Annals of Applied Statistics*, 9(4):2023–2050, 2015.

[3] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes, Theory and Applications.* Wiley Series in Probability and Statistics, John Wiley & Sons Ltd, Chichester, 2004.

[4] J. Blanchet and A. Davison. Spatial modeling of extreme snow depth. *Annals of Applied Statistics*, 5(3):1699–1724, 2011.

[5] E. Boergens, S. Buhl, D. Dettmering, C. Klüppelberg, and F. Seitz. Combination of multi-mission altimetry data along the mekong river with spatio-temporal kriging. *Journal of Geodesy*, 2016.

[6] E. Bolthausen. On the central limit theorem for stationary mixing random fields. *Annals of Probability*, 10(4):1047–1050, 1982.

[7] R. Bradley. *Introduction to Strong Mixing Conditions*, volume I. Kendrick Press, Heber City, Utah, 2007.

[8] B. Brown and S. Resnick. Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14(4):732–739, 1977.

[9] S. Buhl. *Modelling and Estimation of Extremes in Space and Time.* Master's thesis, Technische Universität München, 2013. Available under `https://mediatum.ub.tum.de/doc/1145694/1145694.pdf`.

[10] S. Buhl and A. Davison. Parametric change point detection in samples of random variables with adjustment for multiple testing under dependency. In preparation, 2017.

[11] S. Buhl and C. Klüppelberg. Anisotropic Brown-Resnick space-time processes: estimation and model assessment. *Extremes*, 19(4):627–660, 2016.

[12] S. Buhl and C. Klüppelberg. Limit theory for the empirical extremogram of random felds. Submitted. Preprint available under `https://arxiv.org/abs/1609.04961`, 2017.

[13] S. Buhl and C. Klüppelberg. Generalised least squares estimation of regularly varying space-time processes based on flexible observation schemes. Submitted, 2017.

[14] S. Buhl, R. Davis, C. Klüppelberg, and C. Steinkohl. Semiparametric estimation for isotropic max-stable space-time processes. Submitted. Preprint available under `https://arxiv.org/abs/1609.04967`, 2017.

[15] Y. Cho, R. Davis, and S. Ghosh. Asymptotic properties of the spatial empirical extremogram. *Scandinavian Journal of Statistics*, 43(3):757–773, 2016.

[16] S. Coles. *An Introduction to Statistical Modeling of Extreme Values.* Springer Series in Statistics, Springer, New York, 2001.

[17] R. Davis and T. Mikosch. The extremogram: A correlogram for extreme events. *Bernoulli*, 15(4):977–1009, 2009.

[18] R. Davis and S. Resnick. Basic properties and prediction of max-ARMA processes. *Advances in Applied Probability*, 21(4):781–803, 1989.

[19] R. Davis, C. Klüppelberg, and C. Steinkohl. Max-stable processes for extremes of processes observed in space and time. *Journal of the Korean Statistical Society*, 42(3):399–414, 2013.

[20] R. Davis, C. Klüppelberg, and C. Steinkohl. Statistical inference for max-stable processes in space and time. *Journal of the Royal Statistical Society B*, 75(5):791–819, 2013.

[21] R. Davis, T. Mikosch, and Y. Zhao. Measures of serial extremal dependence and their estimation. *Stochastic Processes and Their Applications*, 123(7):2575–2602, 2013.

[22] A. Davison, S. Padoan, and M. Ribatet. Statistical modelling of spatial extremes. *Statistical Science*, 27(2):161–186, 2012.

[23] R. de Fondeville and A. Davison. High-dimensional peaks-over-threshold inference for the Brown-Resnick process. Preprint available under `https://arxiv.org/abs/1605.08558`, 2016.

[24] L. de Haan. A spectral representation for max-stable processes. *Annals of Probability*, 12 (4):1194–1204, 1984.

[25] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction.* Springer Series in Operations Research and Financial Engineering, New York, 2006.

[26] K. Doksum and G. Sievers. Plotting with confidence: graphical comparisons of two populations. *Biometrika*, 63(3):421–434, 1976.

[27] C. Dombry and F. Eyi-Minko. Strong mixing properties of max-infinitely divisible random fields. *Stochastic Processes and their Applications*, 122(11):3790–3811, 2012.

[28] C. Dombry, S. Engelke, and M. Oesting. Exact simulation of max-stable processes. *Biometrika*, 103:303–317, 2016.

[29] C. Dombry, S. Engelke, and M. Oesting. Bayesian inference for multivariate extreme value distributions. Preprint available under `https://arxiv.org/abs/1611.05602`, 2016.

[30] P. Doukhan. *Mixing: Properties and Examples*. Springer, New York, 1994.

[31] J. Einmahl, A. Kiriliouk, and J. Segers. A continuous updating weighted least squares estimator of tail dependence in high dimension. Preprint available under `https://arxiv.org/abs/1601.04826`, 2016.

[32] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin, 1997.

[33] P. Embrechts, E. Koch, and C. Robert. Space-time max-stable models with spectral separability. *Advances in Applied Probability*, 48(A):77–97, 2016.

[34] S. Engelke, A. Malinowski, Z. Kabluchko, and M. Schlather. Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. *Journal of the Royal Statistical Society B*, 77 (1):239–265, 2015.

[35] V. Fasen, C. Klüppelberg, and M. Schlather. High-level dependence in time series models. *Extremes*, 13(1):1–33, 2010.

[36] R. Fisher and L. Tippett. On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24: 180–190, 1928.

[37] M. Genton, Y. Ma, and H. Sang. On the likelihood function of Gaussian max-stable processes. *Biometrika*, 98(2):481–488, 2011.

[38] E. Gilleland and R. W. Katz. New software to analyze how extremes change over time. *Eos*, 92(2):13–14, 2011.

[39] E. Giné, M. G. Hahn, and P. Vatan. Max-infinitely divisible and max-stable sample continuous processes. *Probability Theory and Related Fields*, 87:139–165, 1990.

[40] H. Hult and F. Lindskog. Extremal behavior of regularly varying stochastic processes. *Stochastic Processes and their Applications*, 115:249–274, 2005.

[41] H. Hult and F. Lindskog. Regular variation for measures on metric spaces. *Publications de l'Institut Mathématique (Beograd)*, 80:121–140, 2006.

[42] R. Huser and A. Davison. Composite likelihood estimation for the Brown-Resnick process. *Biometrika*, 100(2):511–518, 2013.

[43] R. Huser and A. Davison. Space-time modelling of extreme events. *Journal of the Royal Statistical Society B*, 76(2):439–461, 2014.

[44] J. Hüsler and R.-D. Reiss. Maxima of normal random vectors: between independence and complete dependence. *Statistics and Probability Letters*, 7(4):283–286, 1989.

[45] I. Ibragimov and Y. Linnik. *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff, 1971.

[46] N. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions - Volume 2*. Wiley, New York, 2nd edition, 1995.

[47] Z. Kabluchko, M. Schlather, and L. de Haan. Stationary max-stable fields associated to negative definite functions. *Annals of Probability*, 37(5):2042–2065, 2009.

[48] S. N. Lahiri, Y. Lee, and N. Cressie. On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. *Journal of Statistical Planning and Inference*, 103(1):65–85, 2002.

[49] M. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer Verlag, New York, 1983.

[50] D. Leber. *Comparison of Simulation Methods of Brown-Resnick processes*. Master's thesis, Technische Universität München, 2015. Available under `https://mediatum.ub.tum.de/doc/1286600/1286600.pdf`.

[51] B. Li, M. Genton, and M. Sherman. On the asymptotic joint distribution of sample space-time covariance estimators. *Bernoulli*, 14(1):208–248, 2008.

[52] D. J. Nott and T. Rydén. Pairwise likelihood methods for inference in image models. *Biometrika*, 86(3):661–676, 1999.

[53] M. Oesting. *Simulationsverfahren für Brown-Resnick-Prozesse*. Diploma thesis, Georg-August-Universität zu Göttingen, 2009. Available under `https://arxiv.org/abs/0911.4389v2`.

[54] S. Padoan, M. Ribatet, and S. Sisson. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277, 2010.

[55] D. N. Politis and J. P. Romano. Nonparametric resampling for homogeneous strong mixing random fields. *Journal of Multivariate Analysis*, 47:301–328, 1993.

[56] D. N. Politis, J. P. Romano, and M. Wolf. *Subsampling*. Springer, New York, 1999.

[57] S. Resnick. Point processes, regular variation and weak convergence. *Advances in Applied Probability*, 18(1):66–138, 1986.

[58] S. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer, New York, 1987.

[59] S. Resnick. *Heavy-Tail Phenomena, Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering, 2007.

[60] M. Schlather. RandomFields, contributed package on random field simulation for R. `http://cran.r-project.org/web/packages/RandomFields`.

[61] R. L. Smith. Max-stable processes and spatial extremes. Unpublished manuscript, University of North California, 1990.

[62] C. Steinkohl. *Statistical Modelling of Extremes in Space and Time using Max-Stable Processes.* Dissertation, Technische Universität München, München, 2013. Available under `https://mediatum.ub.tum.de/doc/1120541/1120541.pdf`.

[63] D. Straumann. *Estimation in Conditionally Heteroscedastic Time Series Models.* Lecture Notes in Statistics, Springer, Berlin, 2004.

[64] E. Thibaud and T. Opitz. Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102(4):855–870, 2015.

[65] E. Thibaud, J. Aalto, D. Cooley, A. Davison, and J. Heikkinen. Bayesian inference for the BrownResnick process, with an application to extreme low temperatures. *Annals of Applied Statistics*, 10(4):2303–2324, 2016.

[66] J. Wadsworth. On the occurrence times of componentwise maxima and bias in likelihood inference for multivariate max-stable distributions. *Biometrika*, 102(3):705–711, 2015.

[67] J. Wadsworth and J. Tawn. Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, 101(1):1–15, 2014.

[68] A. Wald. Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics*, 20(4):595–601, 1946.