

Brown-Resnick Processes: Analysis, Inference and Generalizations

Dissertation
zur Erlangung des mathematisch-naturwissenschaftlichen
Doktorgrades
„Doctor rerum naturalium“
der Georg-August-Universität Göttingen

im Promotionsprogramm Mathematik
der Georg-August University School of Sciences (GAUSS)

vorgelegt von
Sebastian Engelke
aus Göttingen

Göttingen 2012

Betreuungsausschuss

Prof. Dr. Martin Schlather, Institut für Mathematik, Universität Mannheim

Prof. Dr. Anja Sturm, Institut für Math. Stochastik, Göttingen

Mitglieder der Prüfungskommission

Referent:

Prof. Dr. Martin Schlather, Institut für Mathematik, Universität Mannheim

Koreferentin:

Prof. Dr. Anja Sturm, Institut für Math. Stochastik, Göttingen

Weitere Mitglieder der Prüfungskommission:

Jun.-Prof. Dr. Andrea Krajina, Institut für Math. Stochastik, Göttingen

Prof. Dr. Gert Lube, Institut für Num. und Ang. Mathematik, Göttingen

Prof. Dr. Ingo Witt, Mathematisches Institut, Göttingen

Jun.-Prof. Dr. Zakhar Kabluchko, Institut für Stochastik, Universität Ulm

Tag der mündlichen Prüfung: 14. Dezember 2012

Preface

First of all, I would like to thank my supervisor Prof. Martin Schlather for his constant support, his availability at any time of the day and the endless inspiring discussions that ultimately resulted in this thesis. He always left me free to follow my own ideas and, at the same time, he provided me with guidance whenever I needed it. Further, I am grateful to my coauthor Prof. Zakhar Kabluchko for his permanent support, his motivating ideas and his ability to explain even complicated facts in an intuitive way. I would also like to thank Prof. Anja Sturm for co-supervising my thesis and for introducing me to many new areas of stochastics in her excellent lectures and seminars.

I am deeply indebted to *Deutsche Telekom Stiftung* who made my research possible. Without both its financial and intangible support, many cooperations, conferences and experiences would have been unthinkable.

I am grateful to Prof. Anthony Davison and his working group for interesting discussions on extreme value statistics and pleasant lunch times during my stay at *École Polytechnique Fédérale de Lausanne*.

Furthermore, I warmly thank my coauthors for fruitful cooperations. In particular, working together with Alexander Malinowski and Marco Oesting on an extremely interesting research project based on an idea of Prof. Zakhar Kabluchko was always enjoyable and resulted in the publications [Engelke *et al.* \(2012c,d\)](#).

I also appreciated numerous discussions with Kirstin Strokorb and Thomas Rippl on several parts of this thesis.

I am particularly grateful to my fellow PhD students for countless lively conversations, refreshing breaks and great times inside and outside university.

Finally, I would like to thank my friends and my family for encouraging me to pursue my interests and for supporting me at all times. Their never-ending confidence and genuine friendship provide the basis for every decision in both personal and professional life. Above all, I am deeply grateful to Silvia, Gert, Felix and Jana, without whom this thesis would not have been possible. Thank you.

Contents

1	Introduction	1
1.1	Extreme value theory	1
1.2	Stochastic modeling and extreme value statistics	3
1.3	Overview	4
2	An equivalent representation of the Brown-Resnick process	7
2.1	Theorem and proof	7
2.2	Simulation method	12
2.3	Error bounds	12
3	Estimation of Hüsler-Reiss distributions and Brown-Resnick processes	19
3.1	Hüsler-Reiss distributions and Brown-Resnick processes	21
3.2	Estimation	24
3.3	Simulation study	31
3.4	Application: Wind speed data	36
3.5	Discussion	41
3.6	Proofs	42
4	A generalization of the Brown-Resnick process in terms of Lévy processes	45
4.1	The one-dimensional case	46
4.2	Convergence results	52
4.3	The d -dimensional case: Lévy-Mori random fields	59
4.4	Estimation	63
5	Maxima of independent, non-identically distributed Gaussian vectors	69
5.1	The bivariate case	70
5.2	The multivariate case	76
5.3	Application to Brown-Resnick processes	77
5.4	Proofs	80
	Bibliography	93

1 Introduction

Gaussian random fields are the basis for modeling spatial data in meteorology, geology and many other fields. Owing to the central limit theorem, they are well-suited for statements on everyday phenomena such as the average amount of rainfall in a certain area. In contrast, this framework can not be employed to derive sensible results linked to rare events, as for instance the total amount of precipitation in the drainage basin of a river during extreme rainfall. Extreme value theory has been developed to tackle this problem. The role of Gaussian fields for averages is taken over by *max-stable processes* as appropriate models for extreme spatial observations.

The aim of this thesis is to explore and develop further a remarkable max-stable process introduced by [Brown & Resnick \(1977\)](#), and, ultimately, to provide a better understanding of its probabilistic properties. In the rapid-growing field of research of spatial extreme value theory, the Brown-Resnick process and its generalizations in [Kabluchko *et al.* \(2009\)](#) are believed to take an outstanding role in modeling extreme events of natural phenomena such as large floods, heat waves or windstorms.

1.1 Extreme value theory

A classical challenge in probability theory is to analyze the asymptotic behavior of sums $\sum_{i=1}^n X_i$, as $n \rightarrow \infty$, where $(X_i)_{i \in \mathbb{N}}$ are independent copies of the random variable X . In this direction, the central limit theorem and its various generalizations give deep insight into theoretical properties of limit distributions and domains of attraction, and they provide the foundation for statistical inference.

The key goal of extreme value theory is to investigate *maxima* instead of *sums*, that is to consider the weak convergence of the distributions of

$$c_n^{-1} \left(\max_{i=1, \dots, n} X_i - d_n \right), \quad (1.1)$$

as $n \rightarrow \infty$, for sequences $c_n > 0$, $d_n \in \mathbb{R}$. In fact, [Fisher & Tippett \(1928\)](#) and [Gnedenko \(1943\)](#) showed that the only possible, non-degenerate limits of (1.1) (up to linear transformations) are given by the generalized extreme value (GEV) distribution (cf. [Embrechts *et al.*, 1997](#), Def. 3.4.1)

$$F_\xi(x) = \begin{cases} \exp \left\{ -(1 + \xi x)^{-1/\xi} \right\}, & \xi \neq 0 \\ \exp \left\{ -\exp(-x) \right\}, & \xi = 0 \end{cases}, \quad \text{where } 1 + \xi x > 0. \quad (1.2)$$

The cases $\xi < 0$, $\xi = 0$ and $\xi > 0$ correspond to the Weibull, the Gumbel and the Fréchet

distribution, respectively. If (1.1) converges to F_ξ for some $\xi \in \mathbb{R}$, X is said to be in the *max-domain of attraction* of F_ξ , and F_ξ is termed an *extreme value distribution*. The slightly more general notion of a *max-stable distribution*, which includes degenerate random variables, is used exchangeably in the sequel (Resnick, 2008).

In the last decades, univariate extreme value theory has been extended to the multivariate setting in order to allow for modeling dependencies between extreme values at different locations. Similarly to the univariate case, a d -variate random vector is called max-stable if it is the weak limit of (1.1) for suitable sequences $\mathbf{c}_n, \mathbf{d}_n \in \mathbb{R}^d, \mathbf{c}_n > 0$, and \mathbb{R}^d -valued random vectors $\mathbf{X}_i, i \in \mathbb{N}$. Here, maxima as well as algebraic operations are always understood componentwise. Clearly, a d -variate max-stable distribution has GEV margins. Multivariate extreme value theory typically focuses on modeling the extremal dependence structure and thus, without loss of generality, throughout this thesis, almost all results are stated for Gumbel margins only (cf. Resnick, 2008, Prop. 5.10). The reader is referred to de Haan & Ferreira (2006), Embrechts *et al.* (1997) and Resnick (2008) for details on univariate and multivariate extreme value theory.

More recently, this approach has been generalized to max-stable processes, where all finite-dimensional distributions are in the class of multivariate extreme value distributions. An early example of such a process has already figured in Brown & Resnick (1977) who analyzed pointwise maxima of a sequence $(W_i)_{i \in \mathbb{N}}$ of independent standard Brownian motions $\{W(t) : t \in \mathbb{R}\}$ on the real line. More precisely, they introduced the stochastic process

$$\Psi(t) = \max_{i \in \mathbb{N}} [X_i + W_i(t) - |t|/2], \quad t \in \mathbb{R}, \quad (1.3)$$

where $\sum_{i \in \mathbb{N}} \delta_{X_i}$ is an independent Poisson point process (PPP) on the real axis with intensity $e^{-x} dx, x \in \mathbb{R}$, which is called the Gumbel point process in the sequel. For details on PPPs see Daley & Vere-Jones (2003). Besides its max-stability, an astonishing fact regarding the *Brown-Resnick process* Ψ is that it turns out to be stationary, even though the drifted Brownian motions $W_i(t) - |t|/2$ tend to $-\infty$ for all $i \in \mathbb{N}$, almost surely. For $t \geq 0$, this can be seen by perceiving the Brownian motion as a homogeneous Markov process and noting that the σ -finite measure with density $e^{-x} dx$ is invariant (Brown, 1970). As it is less intuitive why “mirroring” the Brownian motion is the correct choice for the negative real axis, considering Lévy processes with suitable drifts instead of $W(t) - |t|/2$ in (1.3) can provide more insight (cf. Chapter 4). See Bertoin (1996) and Sato (1999) for an introduction to Lévy processes.

In fact, these “Lévy-Brown-Resnick” processes have already been defined and analyzed in Stoev (2008) on the positive real axis. It was shown therein that the corresponding max-stable process is stationary and mixing. Further, the question was raised whether these processes admit a mixed moving maxima (MMM) representation, which refers to the following construction principle for max-stable processes (cf. Smith (1990); Stoev (2008)): For a measurable stochastic process $\{F(t) : t \in \mathbb{R}^d\}$ suppose that $\mathbb{E} \int_{\mathbb{R}^d} \exp(F(t)) dt = 1$ and let $\sum_{i \in \mathbb{N}} \delta_{(S_i, V_i)}$ be a PPP on $\mathbb{R}^d \times \mathbb{R}$ with intensity $ds e^{-v} dv$. Then, the process

$$\zeta(t) = \max_{i \in \mathbb{N}} [V_i + F_i(t - S_i)], \quad t \in \mathbb{R}^d, \quad (1.4)$$

is called a MMM process, where F_i , $i \in \mathbb{N}$, are independent copies of the spectral function F . Besides the theoretical insight, the knowledge of such a representation also facilitates the simulation of ζ , which generally is a challenging problem for max-stable processes (Dombry *et al.*, 2012; Oesting *et al.*, 2012; Wang & Stoev, 2011). For general Lévy processes such a representation was still unknown, whereas in the special case (1.3) of a Brownian motion (and also for other Gaussian processes) Kabluchko *et al.* (2009) proved the existence. Surprisingly, even in this case, the distribution of the shape function F has never been computed.

1.2 Stochastic modeling and extreme value statistics

For extreme value statistics, it is crucial to have appropriate random field models on \mathbb{R}^d that take into account the dependencies between different locations in space. Even though there is profound literature on theoretical properties of max-stable random fields (de Haan, 1984; de Haan & Pickands, 1986; Giné *et al.*, 1990; Kabluchko, 2009b; Kabluchko & Schlather, 2010; Stoev, 2008; Stoev & Taqqu, 2005; Wang & Stoev, 2010), concrete examples which are both flexible and statistically tractable are scarce. An important construction principle in this direction are the MMM processes in (1.4). For instance, using a deterministic parabola for F yields the well-known Smith process (Smith (1990), see also de Haan & Pereira (2006) for other choices for F). The widely-used approach in Schlather (2002) uses stationary Gaussian random fields to obtain max-stable processes parameterized by positive definite functions. This class is both flexible and easily applicable to spatial data sets, but it has the limitation of always being non-ergodic (for recent applications see Blanchet & Davison (2011); Huser & Davison (2012)).

For practical applications, the original Brown-Resnick process Ψ is unfeasible as it is only defined on the real line and it has a fixed dependence structure. It attracted new attention from both probabilists (Das & Hashorva, 2012; Kabluchko, 2011a; Stoev, 2008) and extreme value statisticians (Bacro & Gaetan, 2012; Davison *et al.*, 2012; Turkman *et al.*, 2010) when Kabluchko *et al.* (2009) replaced the Brownian motion W in (1.3) by a general centered Gaussian random field $\{Y(t) : t \in \mathbb{R}^d\}$ on \mathbb{R}^d with stationary increments and variance $\sigma^2(t) = \mathbb{E}[Y(t)]^2$, giving rise to the process

$$\xi(t) = \max_{i \in \mathbb{N}} [X_i + Y_i(t) - \sigma^2(t)/2], \quad t \in \mathbb{R}^d, \quad (1.5)$$

where $(Y_i)_{i \in \mathbb{N}}$ are independent copies of Y and $\sum_{i \in \mathbb{N}} \delta_{X_i}$ is a Gumbel point process. Even in this general setting, the resulting max-stable process ξ keeps the stationarity property and its finite dimensional distributions follow the Hüsler-Reiss distribution (Hüsler & Reiss, 1989). Apart from that, the only possible limits of maxima of suitably normalized and rescaled Gaussian random fields lie in this extended class of Brown-Resnick processes (cf. Kabluchko *et al.* (2009, Thm. 20), Kabluchko (2011a)). Thus, since Gaussianity is the common assumption for frequent observations, the processes in (1.5) constitute promising and natural models for spatial extremes.

Statistical inference for max-stable processes from observations in its MDA is currently an active field of research (cf. Davison & Gholamrezaee (2012); de Haan & Pereira (2006); Padoan

et al. (2010), for instance). Estimation methods typically use aggregated data obtained as the block maximum over many observations. These methods discard all data below the maximum, although additional information might be contained in the non-aggregated, single “large” observations. For this reason, inference is still particularly problematic for small amounts of data, since in that case only very few extreme events can be extracted (see Chapter 3 for details).

For most max-stable processes a closed form of the multivariate likelihoods is unavailable and thus, estimation methods often rely on summary statistics. For instance, a widely used measure of dependence for extreme value distributions is the *extremal correlation* ρ . For a bivariate max-stable random vector (X_1, X_2) with Gumbel margins, $\rho \in [0,1]$ is determined by

$$\mathbb{P}(X_1 \leq u, X_2 \leq u) = \mathbb{P}(X_1 \leq u)^{2-\rho},$$

for some (and hence all) $u \in \mathbb{R}$. The quantity ρ measures the degree of tail dependence with limit cases $\rho = 0$ and $\rho = 1$ corresponding to complete independence and complete dependence, respectively. For a stationary, max-stable process ξ on \mathbb{R}^d , the *extremal correlation function* $\rho(h)$ is defined as the extremal correlation of $(\xi(0), \xi(h))$, for $h \in \mathbb{R}^d$ (Schlather & Tawn (2003)). For instance, for the Brown-Resnick process in (1.5), this function is given by

$$\rho(h) = 2 \left[1 - \Phi \left(\sqrt{\gamma(h)}/2 \right) \right], \quad h \in \mathbb{R}^d, \quad (1.6)$$

where γ is the variogram of Y and Φ denotes the standard normal distribution function. Since there are model-independent estimators for extremal correlation functions, e.g., the madogram in Cooley *et al.* (2006), it is a useful tool for model checking.

1.3 Overview

The first part of this thesis contributes to the understanding and applicability of the Brown-Resnick process in (1.3) and its generalizations (1.5). In the second part, the underlying construction principle is extended in two different directions to more general frameworks.

1.3.1 Analysis and Inference

In Chapter 2 we explicitly compute the distribution of the shape functions F in the MMM representation of the original Brown-Resnick process Ψ . It equals a well-known diffusion, namely a standard Brownian motion with drift conditional on taking negative values only. As an application, an efficient simulation method for Ψ is presented together with error bounds that specify the rate of convergence of this algorithm.

Following the idea of using not only the aggregated data, Chapter 3 introduces a new approach of inference based on a multivariate peaks-over-threshold method. It is shown that, for any process in the MDA of the extended Brown-Resnick process in (1.5), suitably defined conditional increments asymptotically follow a multivariate Gaussian distribution. This leads to computationally efficient estimates of Hüsler-Reiss distributions which enable

parametric inference for Brown-Resnick processes, even if only a small data set is available. As an application, a non-isotropic Brown-Resnick process is fitted to the extremes of 12-year data of daily wind speed measurements.

1.3.2 Generalizations

Based on the results in [Stoev \(2008\)](#), Chapter 4 is concerned with a more complete definition of Brown-Resnick processes based on Lévy processes, particularly the non-trivial extension to the whole real line. Amongst others, it is shown that these “Lévy-Brown-Resnick” processes naturally arise as limits of maxima of stationary α -stable Ornstein-Uhlenbeck processes. In particular, this extends the Gaussian ($\alpha = 2$) result in [Brown & Resnick \(1977\)](#) to the α -stable case with $\alpha \in (1, 2)$. Furthermore, the question in [Stoev \(2008\)](#) on the existence of a MMM representation is answered. Using the generalization of Lévy processes to random fields introduced in [Mori \(1992\)](#), also a spatial version of the “Lévy-Brown-Resnick” process is constructed.

Chapter 5 is devoted to the study of a triangular array $X_{i,n}, n \in \mathbb{N}, 1 \leq i \leq n$, of independent \mathbb{R}^d -valued Gaussian random vectors with covariance matrices $\Sigma_{i,n}$. Necessary conditions are given under which the row-wise maxima converge to some max-stable distribution which generalizes the class of Hüsler-Reiss distributions. In the bivariate case, the conditions will also be sufficient. Using these results, new models for bivariate extremes are derived explicitly. Moreover, a new class of stationary, max-stable processes is introduced, which can be seen as max-mixtures of the processes in (1.5), and whose finite dimensional margins coincide with the above limit distributions. As an application, it is shown that these processes realize a large set of extremal correlation functions (1.6). This set includes all functions $\psi(\sqrt{\gamma(h)})$, $h \in \mathbb{R}^d$, where ψ is a completely monotone function and γ is an arbitrary variogram.

1.3.3 Submissions and Publications

The main theorem of Chapter 2 has been published in [Engelke et al. \(2011\)](#).

The results of Chapter 3 have been submitted to the *Journal of the Royal Statistical Society: Series B* ([Engelke et al., 2012c](#)). This is joint work with Alexander Malinowski, a fellow PhD student, in which I have made the major contribution. Another outcome of this collaboration with reversed roles is the paper [Engelke et al. \(2012d\)](#), which is not part of this thesis. It studies statistical inference for max-stable processes that admit a MMM or a certain incremental representation and has been submitted to the *Applied Probability Trust*.

Parts of Chapter 4 on “Lévy-Brown-Resnick” have already been presented on the *7th Conference on Extreme Value Analysis* in Lyon in 2011 and are based on the manuscript [Engelke et al. \(2012a\)](#). Section 4.4 is joint work with Prof. Anthony Davison from *École Polytechnique Fédérale de Lausanne*.

Chapter 5 is based on the paper [Engelke et al. \(2012b\)](#) which has been submitted to *Bernoulli* and is now being revised.

There are two further publications which were prepared during the time as PhD student. [Engelke & Woerner \(2013\)](#) analyzes fractional Lévy processes and is published in *Stochastics*

and Dynamics. [Engelke & Schlather \(2011\)](#) reviews the book [Qian \(2010\)](#) and has been published in the *Biometrical Journal*.

2 An equivalent representation of the Brown-Resnick process

The original construction of the Brown-Resnick process Ψ in (1.3) turns out to be impractical for applications such as simulation methods. When sampling the Gumbel point process on the real axis top down and attaching a drifted Brownian motion to each point, the convergence to a stationary process is extremely slow. The reason for this is that the Brownian motions converge rapidly to $-\infty$ and only because the Poisson points become very dense at $-\infty$, there are eventually processes that contribute to the maximum far away from the origin. It is thus crucial to obtain exact and efficient simulation methods. Alternative techniques have already been considered in Oesting *et al.* (2012) which are also applicable to the generalization of Ψ in (1.5). See also Dombry *et al.* (2012) for conditional simulation of Brown-Resnick processes.

In the situation of the process in (1.3) the Markovian structure of Brownian motion can beneficially be used to establish a MMM representation (1.4) of Ψ . In this chapter we show that the distribution of the corresponding shape function F equals a well known diffusion process. Namely, it is a standard Brownian motion with drift $-|t|/2$ conditioned on not taking positive values.

This chapter is organized as follows. First, we state and proof the main theorem on the MMM representation of Ψ in Section 2.1. In Section 2.2 we then discuss how this result can be used for efficient simulation methods. Finally, in Section 2.3 we provide error bounds on the accuracy of this algorithm.

2.1 Theorem and proof

Theorem 1. *Let B be a Brownian motion on the real line with drift $-|t|/2$ and $B(0) = 0$. Then the following two processes have the same distribution:*

1.

$$\Psi_1(t) = \max_{i \in \mathbb{N}} [X_i + B_i(t)], \quad t \in \mathbb{R}, \quad (2.1)$$

where B_i , $i \in \mathbb{N}$, are independent copies of B and $\sum_{i \in \mathbb{N}} \delta_{X_i}$ is a Gumbel point process independent of the B_i , $i \in \mathbb{N}$.

2.

$$\Psi_2(t) = \max_{i \in \mathbb{N}} [V_i + R_i(t - S_i)], \quad t \in \mathbb{R}, \quad (2.2)$$

where $\sum_{i \in \mathbb{N}} \delta_{(S_i, V_i)}$ is a Poisson point process on $\mathbb{R} \times \mathbb{R}$ with intensity $\frac{1}{2} dt e^{-x} dx$, $t, x \in \mathbb{R}$, independent of the R_i , $i \in \mathbb{N}$. The R_i 's are independent copies of the process

$$R(t) = 1_{t \leq 0} R_-(-t) + 1_{t \geq 0} R_+(t),$$

where R_- and R_+ are independent processes which follow for $t \geq 0$ the law of the non-positive diffusion determined by the stochastic differential equation

$$d\tilde{R}(t) = dW(t) + \frac{1}{2} \coth(\tilde{R}(t)/2) dt, \quad (2.3)$$

starting at $R_+(0) = R_-(0) = 0$, where W is a standard Brownian motion.

Remark 1. For $y \leq 0$, the diffusion defined by (2.3) and started at $\tilde{R}(0) = y$ has the same distribution as a Brownian motion $\{B(t)\}_{t \geq 0}$ with drift $-t/2$ and $B(0) = y$ conditioned on not taking positive values for all $t \geq 0$. For details see Lemma 55.1 in [Rogers & Williams \(2000\)](#).

In fact, for $y = 0$, it follows from Theorem 3 in [Rogers & Pitman \(1981\)](#) that the opposite of the diffusion, namely $-\tilde{R}$, has the same distribution as a three dimensional Bessel process of drifting Brownian motion. More precisely, it can be obtained by taking the radial part of a Brownian motion in \mathbb{R}^3 with drift $1/2$ in the direction of the first coordinate axis. This is particularly important with regard to efficient simulation methods.

Remark 2. In fact, we will prove more than the equality in law of Ψ_1 and Ψ_2 . We will show that the random families of functions $\{X_i + B_i(\cdot)\}_{i \in \mathbb{N}}$ and $\{V_i + R_i(\cdot - S_i)\}_{i \in \mathbb{N}}$ considered as point processes on $C(\mathbb{R})$, the space of continuous functions on \mathbb{R} , have the same law.

Proof of Theorem 1. Let us first set up some notation. For the Brownian motion B with drift $-|t|/2$ and $B(0) = 0$ let $Y = \sup_{s \in \mathbb{R}} B(s)$ and $T = \arg \sup_{s \in \mathbb{R}} B(s)$. Moreover, denote by $Z(t) = B(t + T) - Y$ the process B whose supremum is shifted into the origin.

Consider the Poisson point process $\sum_{i \in \mathbb{N}} \delta_{(X_i, B_i)}$ on $\mathbb{R} \times C(\mathbb{R})$ with intensity $e^{-x} dx \mathbb{P}_B(df)$, where $C(\mathbb{R})$ denotes the space of continuous functions on \mathbb{R} and \mathbb{P}_B is the law of B . The mapping

$$\eta : \mathbb{R} \times C(\mathbb{R}) \rightarrow C(\mathbb{R}) \times \mathbb{R} \times \mathbb{R}, (u, f) \mapsto (f((\cdot) + \arg \sup f) - \sup f, \arg \sup f, \sup f + u)$$

is measurable (see proof of Lemma 2.13 in [Oesting et al. \(2012\)](#) for details), where the $\arg \sup$ is the infimum of all points where the supremum is attained. The key idea of the proof is to compute the intensity measure Φ of the Poisson point process on $C(\mathbb{R}) \times \mathbb{R} \times \mathbb{R}$ induced by η , namely

$$\sum_{i \in \mathbb{N}} \delta_{(B_i((\cdot) + \arg \sup B_i) - \sup B_i, \arg \sup B_i, \sup B_i + X_i)}. \quad (2.4)$$

The main difficulty here is to find the distribution of the random element (Z, T, Y) which is essential in the later calculations. More precisely, for $t_1 \leq \dots \leq t_n \in \mathbb{R}$, $y_1, \dots, y_n < 0$,

$n \in \mathbb{N}$, $t \in \mathbb{R}$ and $y \in \mathbb{R}_+$ we will later establish the equation

$$\begin{aligned} d\mathbb{P}(Z(t_1) = y_1, \dots, Z(t_n) = y_n, T = t, Y = y) \\ = d\mathbb{P}(R(t_1) = y_1, \dots, R(t_n) = y_n, R(-t) = -y) \frac{e^{-y}}{2}, \end{aligned} \quad (2.5)$$

where R is the two-sided diffusion process which satisfies (2.3) and starts in 0. Thus, the distribution of the shifted Brownian motion Z essentially equals the one of the re-weighted diffusion R . In order to compute the measure Φ let $M \times I \times D \in \mathcal{B}(C(\mathbb{R})) \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ be a measurable set with $M = C_{t_1, \dots, t_n}(E) = \{g \in C(\mathbb{R}) : (g(t_1), \dots, g(t_n)) \in E\}$, where $E \in \mathcal{B}(\mathbb{R}^n)$. Furthermore, denote by $(\Omega, \mathcal{A}, \mathbb{P})$ the underlying probability space.

$$\begin{aligned} \Phi(M \times I \times D) &= \int_{\mathbb{R}} e^{-x} \mathbb{P}((Z, \arg \sup B, \sup B) \in M \times I \times D - (0, 0, x)) dx \\ &= \int_{\mathbb{R}} e^{-x} \mathbb{P}((Z(t_1), \dots, Z(t_n)) \in E, \arg \sup B \in I, \sup B \in D - x) dx \\ &= \int_{\mathbb{R}} e^{-x} \int_I \int_D \int_E \mathbb{P}(Z(t_1) = y_1, \dots, Z(t_n) = y_n, \arg \sup B = t, \sup B = y - x) \\ &\quad d(y_1, \dots, y_n) dy dt dx \\ &= \int_{\mathbb{R}} e^{-x} \int_I \int_D \int_E \mathbb{P}(R(t_1) = y_1, \dots, R(t_n) = y_n, R(-t) = -(y - x)) \frac{1}{2} e^{-(y-x)} \\ &\quad d(y_1, \dots, y_n) dy dt dx \\ &= \int_D \frac{1}{2} e^{-y} \int_I \int_E \int_{\mathbb{R}} \mathbb{P}(R(t_1) = y_1, \dots, R(t_n) = y_n, R(-t) = -(y - x)) \\ &\quad dx d(y_1, \dots, y_n) dt dy \\ &= \int_D \frac{1}{2} e^{-y} \int_I \int_E \mathbb{P}(R(t_1) = y_1, \dots, R(t_n) = y_n) d(y_1, \dots, y_n) dt dy \\ &= \mathbb{P}_R(C_{t_1, \dots, t_n}(E)) \int_I \frac{1}{2} dt \int_D e^{-y} dy, \end{aligned}$$

where the fourth equality follows from (2.5) and \mathbb{P}_R is the law of R on $C(\mathbb{R})$.

Therefore, the intensity Φ of the Poisson point process (2.4) is given by $\mathbb{P}_R(df) \frac{1}{2} dt e^{-x} dx$. On the other hand, if we consider a Poisson point process $\sum_{i \in \mathbb{N}} \delta_{(S_i, V_i)}$ with intensity $\frac{1}{2} dt e^{-x} dx$ and independent marks $R_i \sim \mathbb{P}_R$, $i \in \mathbb{N}$, then we also obtain a Poisson point process $\sum_{i \in \mathbb{N}} \delta_{(R_i, S_i, V_i)}$ with intensity $\mathbb{P}_R(df) \frac{1}{2} dt e^{-x} dx$. This implies the equality in distribution of (2.1) and (2.2).

Let us now proceed with the proof of equation (2.5). To this end, we will first relate the distribution of (T, Y) to the one-dimensional distributions of the diffusion R . In a second step, we compute the distribution of the shifted Brownian motion Z conditional on (T, Y) . By formulae 2.1.1.4 and 2.1.13.4 in Borodin & Salminen (1996) it follows that the bivariate random variable (T, Y) has the density

$$d\mathbb{P}(T = t, Y = y) = f(|t|, y)(1 - e^{-y}) dt dy, \quad t \in \mathbb{R}, y \in \mathbb{R}_+, \quad (2.6)$$

where f is given by

$$f(t, y) = \frac{y}{\sqrt{2\pi t^{3/2}}} \exp\left(-\frac{(y+t/2)^2}{2t}\right) = -e^{-y}f(t, -y).$$

Note that $(1 - e^{-y})$ is the probability that $\{B(s) : s \geq 0\}$ never hits $y \in \mathbb{R}_+$. On the other hand, we can explicitly compute the entrance law and the transition densities of the diffusion $\{R(t)\}_{t \geq 0}$ satisfying (2.3) and started at $R(0) = 0$. For $x, y < 0$ let

$$p_t(x, y) dy = d\mathbb{P}\left(\sup_{0 \leq s \leq t} B(s) \leq 0, B(t) = y | B(0) = x\right), \quad t > 0.$$

By Lemma 55.1 in Rogers & Williams (2000) we conclude that

$$p_t(x, y) = (2\pi t)^{-1/2} e^{-(y-x)/2-t/8} \left[e^{-(y-x)^2/(2t)} - e^{-(y+x)^2/(2t)} \right]$$

and

$$p_t(x, y) = e^{-(y-x)} p_t(y, x). \quad (2.7)$$

Thus, formula 2.1.1.4(1) in Borodin & Salminen (1996) shows that the transition density of R has the form

$$q_t(x, y) dy = d\mathbb{P}(B(t) = y | B(0) = x, \sup_{0 \leq s \leq \infty} B(s) \leq 0) = (1 - e^x)^{-1} p_t(x, y) (1 - e^y) dy,$$

for $t > 0, x, y < 0$. Consequently, the entrance law $q_t(0, \cdot)$ of R , i.e., $q_t(0, y) dy = d\mathbb{P}(R(t) = y)$ for $t > 0$ and $y < 0$ is given by

$$\begin{aligned} q_t(0, y) &= \lim_{x \rightarrow 0} q_t(x, y) = \frac{1}{\sqrt{2\pi t}} e^{-y/2-t/8} (1 - e^y) \lim_{x \rightarrow 0} \frac{e^{-(y-x)^2/(2t)} - e^{-(y+x)^2/(2t)}}{1 - e^x} \\ &= \frac{-2y}{\sqrt{2\pi t^{3/2}}} e^{-(y+t/2)^2/(2t)} (1 - e^y) \\ &= -2f(t, y) (1 - e^y). \end{aligned} \quad (2.8)$$

Note that this together with (2.6) implies for $t \in \mathbb{R}$ and $y \in \mathbb{R}_+$

$$\begin{aligned} d\mathbb{P}(T = t, Y = y) &= f(|t|, y) (1 - e^{-y}) dt dy \\ &= -f(|t|, -y) (1 - e^{-y}) e^{-y} dt dy = d\mathbb{P}(R(-t) = -y) \frac{e^{-y}}{2}, \end{aligned} \quad (2.9)$$

We now look more closely at the distribution of Z conditional on the maximum (T, Y) . To this end, without loss of generality, let $D_{t,y} = \{T = t, Y = y\}$ with $t, y \geq 0$. By Williams' path decomposition of Brownian motion (Theorem 55.9 in Rogers & Williams (2000)), $\{Z(s) : -t \leq s \leq 0 | D_{t,y}\}$ is independent of $\{Z(s) : s > 0 | D_{t,y}\}$ and the latter has the same distribution as $\{R(s) : s > 0 | R(0) = 0\}$, where R satisfies (2.3). Furthermore, by the construction of Z we have that the process $\{Z(s) : s \leq -t | D_{t,y}\}$ is independent of $\{Z(s) : s > -t | D_{t,y}\}$.

For $p, l, m \in \mathbb{N}$ and $r_1 \leq \dots \leq r_p \leq -t \leq s_1 \leq \dots \leq s_l \leq 0 \leq u_1 \leq \dots \leq u_m$ define random events

$$\begin{aligned} A_1 &= \{Z(r_1) = w_1, \dots, Z(r_p) = w_p\}, \\ A_2 &= \{Z(s_1) = x_1, \dots, Z(s_l) = x_l\}, \\ A_3 &= \{Z(u_1) = z_1, \dots, Z(u_m) = z_m\}, \end{aligned}$$

where $w_1, \dots, w_p, x_1, \dots, x_l, z_1, \dots, z_m \leq 0$. The above considerations yield

$$d\mathbb{P}(A_1 \cap A_2 \cap A_3 | D_{t,y}) = d\mathbb{P}(A_1 | D_{t,y}) d\mathbb{P}(A_2 | D_{t,y}) d\mathbb{P}(A_3 | D_{t,y}). \quad (2.10)$$

As already mentioned above, we have

$$d\mathbb{P}(A_3 | D_{t,y}) = d\mathbb{P}(R(u_1) = z_1, \dots, R(u_m) = z_m | R(0) = 0).$$

Moreover,

$$\begin{aligned} d\mathbb{P}(A_1 | D_{t,y}) &= d\mathbb{P}(B(r_1 + t) = w_1 + y, \dots, B(r_p + t) = w_p + y | B(0) = 0, B(s) \leq y \forall s \leq 0) \\ &= d\mathbb{P}(B(r_1 + t) = w_1, \dots, B(r_p + t) = w_p | B(0) = -y, B(s) \leq 0 \forall s \leq 0) \\ &= d\mathbb{P}(R(r_1 + t) = w_1, \dots, R(r_p + t) = w_p | R(0) = -y) \\ &= d\mathbb{P}(R(r_1) = w_1, \dots, R(r_p) = w_p | R(-t) = -y), \end{aligned}$$

where we used Remark 1 and the fact that R is Markov.

For the second factor in (2.10) we first compute explicitly $d\mathbb{P}(A_2 \cap D_{t,y})$. To this end, put $s_0 = -t$, $x_0 = -y$, $\Delta s_i = s_i - s_{i-1}$ and $\xi = (x_1, \dots, x_l, t, y)$.

$$\begin{aligned} d\mathbb{P}(A_2 \cap D_{t,y}) &= d\mathbb{P}(Z(s_1) = x_1, \dots, Z(s_l) = x_l, T = t, Y = y) \\ &= f(|s_l|, -x_l)(1 - e^{x_0}) \prod_{i=1}^l p_{\Delta s_i}(x_{i-1}, x_i) d\xi \\ &= f(|s_l|, -x_l)(1 - e^{x_0}) e^{-(x_l - x_0)} \prod_{i=1}^l p_{\Delta s_i}(x_i, x_{i-1}) d\xi \\ &= -f(|s_l|, x_l)(1 - e^{x_0}) e^{x_0} \prod_{i=1}^l p_{\Delta s_i}(x_i, x_{i-1}) d\xi \\ &= -f(|s_l|, x_l)(1 - e^{x_l}) e^{x_0} \prod_{i=1}^l q_{\Delta s_i}(x_i, x_{i-1}) d\xi \\ &= \frac{e^{-y}}{2} q_{|s_l|}(0, x_l) \prod_{i=1}^l q_{\Delta s_i}(x_i, x_{i-1}) d\xi \\ &= d\mathbb{P}(R(s_1) = x_1, \dots, R(s_l) = x_l, R(-t) = -y) \frac{e^{-y}}{2}, \end{aligned}$$

where we used (2.7) and (2.8). The second equation follows by an elementary computation

of the transition probabilities for the shifted Brownian motion Z . Starting from the point (s_0, x_0) it moves to (s_l, x_l) and then attains the one-sided maximum at the origin (with probability $f(|s_l|, -x_l)$, see (2.6)). At the same time, the process B stays below level y on the negative half-axis, which has probability $(1 - e^{x_0})$. Thus, (2.9) yields

$$d\mathbb{P}(A_2|D_{t,y}) = d\mathbb{P}(R(s_1) = x_1, \dots, R(s_l) = x_l | R(-t) = -y).$$

Putting the parts together, formula (2.5) follows from (2.9) and (2.10). □

2.2 Simulation method

The representation of the Brown-Resnick process in (2.2) turns out to be a promising basis for simulations. Schlather (2002) gives an algorithm for the simulation of MMM processes where the spectral functions have finite support. Here, this approach can be adopted, the only difference being that in our case the spectral function, i.e., the conditional negative Brownian motion, does not have finite support. Diffusions from anywhere outside the simulation window can therefore influence the value of the Brown-Resnick process inside the simulation window. However, since the diffusions quickly tend to negative infinity, the probability of this event is exponentially decreasing with the size of the interval which is added on both sides of the simulation window. This error probability can therefore be made arbitrarily small. In fact, the calculations below show that, independent of the size of the simulation window, additional intervals of length 45 on both sides ensure that the probability of not drawing a path from the desired distribution is less than 0.5%. Furthermore, paths starting in these additional intervals do not need to be simulated on the whole enlarged area since knowing the diffusion entrance law enables us to directly jump into the simulation window. Owing to these facts, the algorithm based on the MMM representation is about 60 times faster than the more general algorithms considered in Oesting *et al.* (2012).

Another advantage of this simulation method is its flexibility. In particular, the technique is not restricted to grids but can easily be performed on any set of points. Moreover, also large simulation windows are feasible since the computational costs increase first quadratically, then linearly, for increasing simulation intervals.

2.3 Error bounds

We seek to simulate a Brown-Resnick process Ψ according to the above MMM representation in (2.2). To this end, we fix $a, b \in \mathbb{R}$ with $a < b$ and $n \in \mathbb{N}$. We are interested in the Brown-Resnick process on the interval $[a, b]$. Let $\Theta = \sum_{i \in \mathbb{N}} \delta_{(S_i, V_i)}$ be the PPP in the second part of Theorem 1. We simulate the Poisson points of Θ on a larger interval $[a - n, b + n]$ in the first component and give an estimate for the probability that points from $\mathbb{R} \setminus [a - n, b + n]$ have an impact on the Brown-Resnick process on $[a, b]$.

The process Θ can be split up into two independent Poisson point processes, $\sum_{i \in \mathbb{N}} \delta_{(T_i, Y_i, \tilde{R}_i)}$

with intensity $\mathbf{1}_{[a-n, b+n]}(t) \frac{1}{2} dt e^{-x} dx \mathbb{P}_R(df)$ and $\sum_{j \in \mathbb{N}} \delta_{(U_j, Z_j, R_j)}$ with intensity

$$\Lambda(dt \times dx \times df) = \mathbf{1}_{\mathbb{R} \setminus [a-n, b+n]}(t) \frac{1}{2} dt e^{-x} dx \mathbb{P}_R(df).$$

Let Ψ^n denote the approximation of Ψ generated only by the former Poisson point process, i.e.

$$\Psi^n(t) = \max_{i \in \mathbb{N}} Y_i + \tilde{R}_i(t - T_i), \quad t \in \mathbb{R},$$

and note that the true Brown-Resnick process generated by both Poisson point processes equals

$$\Psi(t) = \max \left(\max_{i \in \mathbb{N}} Y_i + \tilde{R}_i(t - T_i), \max_{j \in \mathbb{N}} Z_j + R_j(t - U_j) \right), \quad t \in \mathbb{R}.$$

The following theorem gives an error bound for the probability that a point outside the interval $[a - n, b + n]$ influences the process Ψ on $[a, b]$.

Theorem 2. *For some $n_0 \in \mathbb{N}$ suppose that we have already simulated Ψ^{n_0} and put $C = \min_{t \in [a, b]} \Psi^{n_0}(t)$. The probability that the approximation Ψ^n , $n \geq n_0$, differs from the max-stable process Ψ on the interval $[a, b]$ decreases exponentially fast in n , conditionally on C . More precisely,*

$$\limsup_{n \rightarrow \infty} \frac{\sqrt{2\pi} \mathbb{P}(\Psi(t) \neq \Psi^n(t) \text{ for a } t \in [a, b] \mid C)}{640 \cdot e^{-C} n^{-3/2} e^{-n/8}} \leq 1,$$

as n tends to ∞ .

Proof. First, note that by the disintegration theorem (Kallenberg, 2002, Thm. 6.4)

$$\begin{aligned} \mathbb{P}(\Psi(t) \neq \Psi^n(t) \text{ for a } t \in [a, b] \mid C) &\leq \mathbb{P} \left(\exists j \in \mathbb{N} \text{ with } \max_{s \in [a, b]} Z_j + R_j(s - U_j) > C \mid C \right) \\ &= 1 - \exp(-\Lambda(A_{n,C})), \end{aligned}$$

where

$$A_{n,C} = \left\{ (t, x, f) \in \mathbb{R} \times \mathbb{R} \times C(\mathbb{R}) : t \notin [a - n, b + n], \max_{s \in [a, b]} x + f(s - t) > C \right\}.$$

Since for $\Lambda(A_{n,C})$ close to 0 we have $1 - \exp(-\Lambda(A_{n,C})) \approx \Lambda(A_{n,C})$, we need to find an error bound for $\Lambda(A_{n,C})$. Clearly, we have

$$\Lambda(A_{n,C}) = A_{n,C,a} \cup A_{n,C,a}^* \cup B_{n,C,b} \cup B_{n,C,b}^*, \quad (2.11)$$

where

$$A_{n,C,a} = \{(t,x,f) \in \mathbb{R} \times \mathbb{R} \times C(\mathbb{R}) : t < a - n, x + f(a - t) > C\}$$

$$A_{n,C,a}^* = \left\{ (t,x,f) \in \mathbb{R} \times \mathbb{R} \times C(\mathbb{R}) : t < a - n, x + f(a - t) \leq C, \max_{s \in [a,b]} x + f(s - t) > C \right\},$$

and the sets $B_{n,C,b}$ and $B_{n,C,b}^*$ being defined analogously for diffusions starting from $(b+n, \infty)$. Thus, we evaluate

$$\begin{aligned} \Lambda(A_{n,C,a}) &= \int_{-\infty}^{a-n} \int_C \frac{1}{2} e^{-x} \int_{C-x}^0 q_{0,|t-s|}(y) dy dx dt \\ &= \int_{-\infty}^{a-n} \int_{-\infty}^0 \frac{-y}{\sqrt{2\pi}|t-s|^{3/2}} \exp\left(-\frac{(y+|t-s|/2)^2}{2|t-s|}\right) (1-e^y) \int_{C-y}^{\infty} e^{-x} dx dy dt \\ &= \int_{-\infty}^{a-n} \int_{-\infty}^0 e^{-C} \frac{-y}{\sqrt{2\pi}|t-s|^{3/2}} \exp\left(-\frac{(y-|t-s|/2)^2}{2|t-s|}\right) (1-e^y) dy dt \\ &\leq \int_{-\infty}^{a-n} \frac{e^{-C}}{|t-s|} \int_{-\infty}^0 \frac{-y}{\sqrt{2\pi}|t-s|} \exp\left(-\frac{(y-|t-s|/2)^2}{2|t-s|}\right) dy dt \\ &= \int_{-\infty}^{a-n} \frac{e^{-C}}{|t-s|} \int_{-\infty}^{-|t-s|^{1/2}/2} \frac{-(z|t-s|^{1/2} + |t-s|/2)}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz dt. \end{aligned}$$

Splitting the expression at the + into two integrals, we obtain for first one

$$\begin{aligned} &\int_{\mathbb{R}} \mathbf{1}_{(-\infty, a-n)}(t) \int_{\mathbb{R}} \mathbf{1}_{(-\infty, -|t-a|^{1/2}/2)}(z) \frac{-z|t-a|^{1/2} e^{-C}}{\sqrt{2\pi}|t-a|} \exp\left(-\frac{z^2}{2}\right) dz dt \\ &= \frac{e^{-C}}{\sqrt{2\pi}} \int_{-\infty}^{-n^{1/2}/2} (-z) \exp\left(-\frac{z^2}{2}\right) \int_{a-4z^2}^{a-n} \frac{1}{|t-a|^{1/2}} dt dz \\ &= \frac{2e^{-C}}{\sqrt{2\pi}} \int_{-\infty}^{-n^{1/2}/2} (2z^2 + n^{1/2}z) \exp\left(-\frac{z^2}{2}\right) dz \\ &= \frac{2e^{-C}}{\sqrt{2\pi}} \left(2\sqrt{2\pi}\Phi(-n^{1/2}/2) + n^{1/2} \exp(-n/8) - n^{1/2} \exp(-n/8)\right) \\ &= 4e^{-C}\Phi(-n^{1/2}/2). \end{aligned}$$

On the other hand, we have:

$$\begin{aligned}
& \int_{\mathbb{R}} \mathbf{1}_{(-\infty, a-n)}(t) \int_{\mathbb{R}} \mathbf{1}_{(-\infty, -|t-a|^{1/2}/2)}(z) \frac{-|t-a|e^{-C}}{2\sqrt{2\pi}|t-a|} \exp\left(-\frac{z^2}{2}\right) dz dt \\
&= \frac{-e^{-C}}{2\sqrt{2\pi}} \int_{-\infty}^{-n^{1/2}/2} \exp\left(-\frac{z^2}{2}\right) \int_{a-4z^2}^{a-n} dt dz \\
&= \frac{-2e^{-C}}{\sqrt{2\pi}} \int_{-\infty}^{-n^{1/2}/2} z^2 \exp\left(-\frac{z^2}{2}\right) dz + \frac{ne^{-C}}{2} \Phi(-n^{1/2}/2) \\
&= \left(\frac{n}{2} - 2\right) e^{-C} \Phi(-n^{1/2}/2) - \frac{n^{1/2}e^{-C}}{\sqrt{2\pi}} \exp(-n/8).
\end{aligned}$$

Together, this yields

$$\Lambda(A_{n,C,a}) \leq \frac{e^{-C}}{\sqrt{2\pi}} \left(\left(2 + \frac{n}{2}\right) \sqrt{2\pi} \Phi(-n^{1/2}/2) - n^{1/2} \exp(-n/8) \right).$$

Recall the definition of R via the stochastic differential equation (2.3). By Proposition 2.18 in Karatzas & Shreve (1991) it follows for $y < 0$ that

$$\mathbb{P}\left(R(u) \leq W(u) - u/2, \forall 0 \leq u < \infty \mid R(0) = W(0) = y\right) = 1,$$

where W is a standard Brownian motion. Consequently, for any $y < x < 0$ and $s > 0$

$$\mathbb{P}\left(\max_{u \in [0,s]} R(u) > x \mid R(0) = y\right) \leq \mathbb{P}\left(\max_{u \in [0,\infty)} W(u) - u/2 > x \mid W(0) = y\right) = \exp(y-x),$$

where the last equality follows from formula 2.1.1.4(1) in Borodin & Salminen (1996). Thus,

$$\begin{aligned}
\Lambda(A_{n,C,a}^*) &= \int_{-\infty}^{a-n} \int_C^{\infty} \frac{1}{2} e^{-x} \int_{-\infty}^{C-x} q_{0,|t-a|}(y) \mathbb{P}\left(\max_{u \in [0,b-a]} R(u) > C-x \mid R(0) = y\right) dy dx dt \\
&\leq \int_{-\infty}^{a-n} \int_C^{\infty} \frac{1}{2} e^{-x} \int_{-\infty}^{C-x} q_{0,|t-a|}(y) e^{y-C+x} dy dx dt \\
&= \int_{-\infty}^{a-n} \int_{-\infty}^0 \frac{-y}{\sqrt{2\pi}|t-a|^{3/2}} \exp\left(-\frac{(y+|t-a|/2)^2}{2|t-a|}\right) (1-e^y) e^{y-C} \int_C^{C-y} dx dy dt \\
&\leq \int_{-\infty}^{a-n} \frac{e^{-C}}{|t-a|} \int_{-\infty}^0 \frac{y^2}{\sqrt{2\pi}|t-a|} \exp\left(-\frac{(y-|t-a|/2)^2}{2|t-a|}\right) dy dt \\
&= \int_{-\infty}^{a-n} \frac{e^{-C}}{|t-a|} \int_{-\infty}^{-|t-a|^{1/2}/2} \frac{(z^2|t-a| + z|t-a|^{3/2} + |t-a|^2/4)}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz dt
\end{aligned}$$

Hence, we need to compute the three integrals obtained by splitting the sum:

$$\begin{aligned}
& \int_{\mathbb{R}} \mathbf{1}_{(-\infty, a-n)}(t) \int_{\mathbb{R}} \mathbf{1}_{(-\infty, -|t-a|^{1/2}/2)}(z) \frac{z^2 |t-a| e^{-C}}{\sqrt{2\pi} |t-a|} \exp\left(-\frac{z^2}{2}\right) dz dt \\
&= \frac{e^{-C}}{\sqrt{2\pi}} \int_{-\infty}^{-n^{1/2}/2} z^2 \exp\left(-\frac{z^2}{2}\right) \int_{a-4z^2}^{a-n} dt dz \\
&= \frac{e^{-C}}{\sqrt{2\pi}} \int_{-\infty}^{-n^{1/2}/2} (4z^4 - nz^2) \exp\left(-\frac{z^2}{2}\right) dz \\
&= \frac{e^{-C}}{\sqrt{2\pi}} \left((12-n)\sqrt{2\pi}\Phi(-n^{1/2}/2) + 6n^{1/2} \exp(-n/8) \right),
\end{aligned}$$

$$\begin{aligned}
& \int_{\mathbb{R}} \mathbf{1}_{(-\infty, a-n)}(t) \int_{\mathbb{R}} \mathbf{1}_{(-\infty, -|t-a|^{1/2}/2)}(z) \frac{z|t-a|^{3/2} e^{-C}}{\sqrt{2\pi} |t-a|} \exp\left(-\frac{z^2}{2}\right) dz dt \\
&= \frac{e^{-C}}{\sqrt{2\pi}} \int_{-\infty}^{-n^{1/2}/2} z \exp\left(-\frac{z^2}{2}\right) \int_{a-4z^2}^{a-n} (a-t)^{1/2} dt dz \\
&= \frac{e^{-C}}{\sqrt{2\pi}} \int_{-\infty}^{-n^{1/2}/2} (-16/3z^4 - 2/3n^{3/2}z) \exp\left(-\frac{z^2}{2}\right) dz \\
&= \frac{e^{-C}}{\sqrt{2\pi}} \left(-16\sqrt{2\pi}\Phi(-n^{1/2}/2) - 8n^{1/2} \exp(-n/8) \right),
\end{aligned}$$

$$\begin{aligned}
& \int_{\mathbb{R}} \mathbf{1}_{(-\infty, a-n)}(t) \int_{\mathbb{R}} \mathbf{1}_{(-\infty, -|t-a|^{1/2}/2)}(z) \frac{|t-a|^2 e^{-C}}{4\sqrt{2\pi} |t-a|} \exp\left(-\frac{z^2}{2}\right) dz dt \\
&= \frac{e^{-C}}{4\sqrt{2\pi}} \int_{-\infty}^{-n^{1/2}/2} \exp\left(-\frac{z^2}{2}\right) \int_{a-4z^2}^{a-n} (a-t) dt dz \\
&= \frac{e^{-C}}{\sqrt{2\pi}} \int_{-\infty}^{-n^{1/2}/2} \left(2z^4 - \frac{1}{8}n^2\right) \exp\left(-\frac{z^2}{2}\right) dz \\
&= \frac{e^{-C}}{\sqrt{2\pi}} \left(\left(6 - \frac{1}{8}n^2\right)\sqrt{2\pi}\Phi(-n^{1/2}/2) + \left(3n^{1/2} + \frac{n^{3/2}}{4}\right) \exp(-n/8) \right),
\end{aligned}$$

where we used the equalities

$$\begin{aligned}
& \int_{-\infty}^{-n^{1/2}/2} z e^{-z^2/2} dz = -\exp(-n/8), \\
& \int_{-\infty}^{-n^{1/2}/2} z^2 e^{-z^2/2} dz = \sqrt{2\pi}\Phi(-n^{1/2}/2) + \frac{n^{1/2}}{2} \exp(-n/8), \\
& \int_{-\infty}^{-n^{1/2}/2} z^4 e^{-z^2/2} dz = 3\sqrt{2\pi}\Phi(-n^{1/2}/2) + \frac{3n^{1/2}}{2} \exp(-n/8) + \frac{n^{3/2}}{8} \exp(-n/8).
\end{aligned}$$

Putting the parts together yields the following estimate.

$$\begin{aligned} & \int_{-\infty}^{a-n} \int_C \frac{1}{2} e^{-x} \int_{-\infty}^{C-x} q_{0,|t-a|}(y) \mathbb{P} \left(\max_{u \in [0, b-a]} R(u) > C-x \mid R(0) = y \right) dy dx dt \\ & \leq \frac{e^{-C}}{\sqrt{2\pi}} \left((2-n - \frac{1}{8}n^2) \sqrt{2\pi} \Phi(-n^{1/2}/2) + (n^{1/2} + \frac{1}{4}n^{3/2}) \exp(-n/8) \right) \end{aligned}$$

A symmetry argument yields the same results for diffusions coming from the interval $(b+n, \infty)$, i.e. $\Lambda(A_{n,C,a}) = \Lambda(B_{n,C,b})$ and $\Lambda(A_{n,C,a}^*) = \Lambda(B_{n,C,b}^*)$. Thus, by (2.11) we obtain:

$$\begin{aligned} \Lambda(A_{n,C}) & \leq 2 \left[\int_{-\infty}^{a-n} \int_C \frac{e^{-x}}{2} \int_{C-x}^0 q_{0,|t-a|}(y) dy dx dt \right. \\ & \quad \left. + \int_{-\infty}^{a-n} \int_C \frac{e^{-x}}{2} \int_{-\infty}^{C-x} q_{0,|t-a|}(y) \mathbb{P} \left(\max_{u \in [0, b-a]} R(u) > C-x \mid R(0) = y \right) dy dx dt \right] \\ & \leq \frac{2e^{-C}}{\sqrt{2\pi}} \left((2 + \frac{1}{2}n) \sqrt{2\pi} \Phi(-n^{1/2}/2) - n^{1/2} \exp(-n/8) \right) \\ & \quad + (2-n - \frac{1}{8}n^2) \sqrt{2\pi} \Phi(-n^{1/2}/2) + (n^{1/2} + \frac{1}{4}n^{3/2}) \exp(-n/8) \\ & = \frac{e^{-C}}{\sqrt{2\pi}} \left((8-n - \frac{1}{4}n^2) \sqrt{2\pi} \Phi(-n^{1/2}/2) + \frac{1}{2}n^{3/2} \exp(-n/8) \right) \end{aligned} \quad (2.12)$$

The last expression can be simplified by using the asymptotic expansion of the normal distribution (cf. formula 8.254 in [Gradshteyn & Ryzhik \(2007\)](#)) function $\Phi(-n^{1/2}/2)$ for large n :

$$\Phi(-n^{1/2}/2) = \frac{\exp(-n/8)}{\sqrt{2\pi}} \left[2n^{-1/2} - 2^3 n^{-3/2} + 3 \cdot 2^5 n^{-5/2} - 15 \cdot 2^7 n^{-7/2} + R(n) \right], \quad (2.13)$$

where $R(n) = O(n^{-9/2} \exp(-n/8))$, as $n \rightarrow \infty$. Plugging (2.13) into (2.12) gives

$$\Lambda(A_{n,C}) \leq \frac{e^{-C} e^{-n/8}}{\sqrt{2\pi}} \left(10 \cdot 2^5 n^{-3/2} + O(n^{-5/2}) \right).$$

This concludes the proof. □

3 Estimation of Hüsler-Reiss distributions and Brown-Resnick processes

Univariate extreme value theory is concerned with the limits of linearly normalized maxima of i.i.d. observations as in (1.1), namely the max-stable distributions in (1.2) (cf. de Haan & Ferreira (2006)). Statistical inference of the parameters is well-developed and usually based on one of the following two approaches. Maximum likelihood estimation is applied to blockwise maxima of the original data, where a typical block size in environmental applications is one year. On the other hand, the *peaks-over-threshold* (POT) method fits a suitable Poisson point process to all data that exceed a certain high threshold and thus follow approximately a generalized Pareto distribution (cf. Davison & Smith (1990)). The advantage of the latter approach is that it avoids discarding extreme values within the blocks that are below the maximum but nevertheless contain information on the parameters.

When interested in the joint extreme behavior of multivariate quantities, there are different possibilities of ordering the data, though, the most common procedure is taking componentwise maxima. In multivariate extreme value theory, a random process $\{\xi(t) : t \in T\}$ with some index set T is called *max-stable*, if there exists a sequence $(\eta_i)_{i \in \mathbb{N}}$ of independent copies of a process $\{\eta(t) : t \in T\}$ and functions $c_n(t) > 0$, $d_n(t) \in \mathbb{R}$, $n \in \mathbb{N}$, such that the convergence

$$\xi(t) = \lim_{n \rightarrow \infty} c_n^{-1}(t) \left(\max_{i=1}^n \eta_i(t) - d_n(t) \right), \quad t \in T, \quad (3.1)$$

holds in the sense of finite-dimensional distributions. In this case, the process η is said to be in the max-domain of attraction (MDA) of ξ . Typically, T is a finite set or $T = \mathbb{R}^d$, $d \in \mathbb{N}$, for the multivariate or the spatial case, respectively. Both theory and inference are considerably more demanding than in the univariate framework due to the fact that no finite-dimensional parametric model captures every possible dependence structure of a multivariate max-stable distribution (cf. Resnick (2008)). Similarly to the univariate case, a standard approach for parameter estimation of the max-stable process ξ from data in its MDA is via componentwise block maxima, which ignores much of the information contained in the original data. Moreover, even if the exact max-stable process is available, maximum likelihood (ML) estimation is problematic since typically only the bivariate densities of max-stable distributions are known in closed form. Composite likelihood (CL) approaches are common tools to avoid this difficulty (cf. Padoan *et al.* (2010), Davison & Gholamrezaee (2012)).

Only recently, multivariate POT methods have attracted increased attention. In contrast to the univariate case, the definition of exceedances over a certain threshold is ambiguous. For instance, Rootzén & Tajvidi (2006) define a multivariate generalized Pareto distribution (MGPD) as the limit distribution of some multivariate random vector in the MDA of a

max-stable distribution, conditional on the event that at least one of the components is large. A simulation study in [Bacro & Gaetan \(2012\)](#) shows, that these MGPD perform well in many situations, yet, again only bivariate densities in a CL framework are used since multivariate densities are unknown. Alternatively, exceedances can be defined as the event that the norm of the random vector is large, giving rise to the spectral measure (cf. [Coles & Tawn \(1991\)](#)). [Engelke et al. \(2012d\)](#) have recently proposed to condition a fixed component on exceeding a high threshold, which enables new methods of inference for processes that admit a certain incremental or a mixed moving maxima representation.

With regard to practical application such as modeling extreme wind speed or precipitation data, max-stable models need to find a compromise between flexibility and tractability. There are several parametric families of multivariate extreme-value distributions (see [Kotz & Nadarajah \(2000\)](#)) and only few max-stable models in the spatial domain (cf. [de Haan & Pereira \(2006\)](#); [Schlather \(2002\)](#); [Smith \(1990\)](#)). For most of them, statistical inference is difficult and time-intensive. Furthermore, except for the max-stable process ξ in [\(3.1\)](#) itself, usually no further processes η in the MDA of attraction of ξ are known and thus, it lacks a theoretical connection between modeling the daily processes η and modeling the extremal process ξ .

In many applications such as geostatistics it is natural to assume that the data is normally distributed. Under this assumption, the only possible non-trivial limit for extreme observations is the d -variate Hüsler-Reiss distribution (cf. [Hüsler & Reiss \(1989\)](#); [Kablichko \(2011a\)](#)). In fact, [Hashorva \(2006\)](#) and [Hashorva et al. \(2012\)](#) show that also other distributions are attracted by the Hüsler-Reiss distribution. Hence, we can expect good fits of this model if the daily data is close to normality. Recently, it has been shown that the class of Brown-Resnick processes ([Brown & Resnick \(1977\)](#); [Kablichko et al. \(2009\)](#)) constitutes the spatial analog of the Hüsler-Reiss distributions since the latter occur as finite-dimensional marginals of the Brown-Resnick process. The research on both theoretical properties (cf. [Dombry et al. \(2011\)](#); [Oesting et al. \(2012\)](#) for simulation methods) and practical applications (e.g., [Davison et al. \(2012\)](#)) of these processes is actively ongoing at present. Statistical inference, however, was so far limited to the CL methods based on bivariate densities.

In this chapter, we propose new estimation methods based on a POT approach for data in the MDA of Hüsler-Reiss distributions and Brown-Resnick processes. Similarly to [Engelke et al. \(2012d\)](#), we consider *extremal increments*, i.e., increments of the data with respect to a fixed component, conditional on the event, that this component is large. The great advantage of this approach is the fact that the extremal increments turn out to be multivariate Gaussian distributed. This enables, for instance, ML estimation with the full multivariate density function as well as parameter estimation based on functionals of the Gaussian distribution. Moreover, the concept of extremal increments as well as estimators derived from spectral densities are shown to be suitable tools for fitting a Brown-Resnick process based on a parametric family of variograms.

The remainder of the chapter is organized as follows. Section [3.1](#) comprises the definitions and some general properties of Hüsler-Reiss distributions and Brown-Resnick processes. In Section [3.2](#), we provide a result on weak convergence of suitably transformed and conditioned variables in the MDA of the Hüsler-Reiss distribution, which is the basis for our estimation

methods. It is used to derive the specific asymptotic distribution for extremal increments (Section 3.2.1) and for conditioning in the spectral sense (Section 3.2.2). In both cases, non-parametric estimation as well as parametric fitting of Brown-Resnick processes are considered. A simulation study is presented in Section 3.3, which compares the performance of the different estimators from the preceding section. As an application, in Section 3.4 we analyze daily wind speed data from the Netherlands and use our new methods of inference to model spatial extreme events. Proofs of the theoretical results can be found at the end of this chapter.

3.1 Hüsler-Reiss distributions and Brown-Resnick processes

In this section we briefly review some details on Hüsler-Reiss distributions and Brown-Resnick processes and define *extremal coefficient functions* as a dependence measure for max-stable processes.

3.1.1 Hüsler-Reiss distributions

It is well known that the standard normal distribution Φ is in the max-domain of attraction of the Gumbel distribution, i.e.

$$\lim_{n \rightarrow \infty} \Phi(b_n + x/b_n)^n = \exp(-\exp(-x)), \quad \text{for all } x \in \mathbb{R},$$

where $b_n, n \in \mathbb{N}$, is a sequence of normalizing constants defined by $b_n = n\phi(b_n)$, where ϕ is the standard normal density. By Theorem 1.5.3 in Leadbetter *et al.* (1983) it is given as

$$b_n := \sqrt{2 \log n} - \frac{(1/2) \log \log n + \log(2\sqrt{\pi})}{\sqrt{2 \log n}} + o((\log n)^{-1/2}). \quad (3.2)$$

Sibuya (1960) showed that the maxima of i.i.d. bivariate normal distributions with correlation $\rho < 1$ asymptotically always become independent. However, for triangular arrays with i.i.d. entries where the correlation in the different rows approaches 1 with an appropriate speed, Hüsler & Reiss (1989) proved that the row-wise maxima converge to a new class of max-stable bivariate distributions, namely

$$F_\lambda(x, y) = \exp \left[-\Phi \left(\lambda + \frac{x-y}{2\lambda} \right) e^{-y} - \Phi \left(\lambda + \frac{y-x}{2\lambda} \right) e^{-x} \right], \quad x, y \in \mathbb{R}. \quad (3.3)$$

Here $\lambda \in [0, \infty]$ parameterizes the dependence in the limit, 0 and ∞ corresponding to complete dependence and asymptotic independence, respectively. In fact, Kabluchko *et al.* (2009) provide a simple argument that these are also the only possible limit points for such triangular arrays.

The higher-dimensional Hüsler-Reiss distributions were also introduced in Hüsler & Reiss (1989) as the limit of suitably normalized Gaussian random vectors. Suppose that the correlation matrix Σ_n in the n -th row of a triangular array of $(k+1)$ -variate, zero-mean,

unit-variance Gaussian distributions satisfies

$$A = \lim_{n \rightarrow \infty} b_n^2 (\mathbf{1} \cdot \mathbf{1}^\top - \Sigma_n) / 2 \in \mathcal{D}, \quad (3.4)$$

where $\mathbf{1} = (1, \dots, 1)^\top \in \mathbb{R}^{k+1}$ and $\mathcal{D} \subset [0, \infty)^{(k+1) \times (k+1)}$ denotes the space of symmetric, strictly conditionally negative definite matrices

$$\mathcal{D} = \left\{ (a_{i,j})_{0 \leq i, j \leq k} = A \in [0, \infty)^{(k+1) \times (k+1)} : \mathbf{x}^\top A \mathbf{x} < 0 \text{ for all } \mathbf{x} \in \mathbb{R}^{k+1} \setminus \{\mathbf{0}\} \text{ s.t.} \right. \\ \left. \sum_{i=0}^k x_i = 0, a_{i,j} = a_{j,i}, a_{i,i} = 0 \text{ for all } 0 \leq i, j \leq k \right\}.$$

Then the normalized row-wise maxima converge to the $(k+1)$ -variate Hüsler-Reiss distribution which is completely characterized by the matrix A . Note that $(\mathbf{1} \cdot \mathbf{1}^\top - \Sigma_n)$ automatically lies in \mathcal{D} if Σ_n is non-degenerate, $n \in \mathbb{N}$. For any matrix $A = (\lambda_{i,j}^2)_{0 \leq i, j \leq k} \in \mathcal{D}$, define a family of positive definite matrices by

$$\Psi_{l, \mathbf{m}}(A) = 2 \left(\lambda_{m_i, m_0}^2 + \lambda_{m_j, m_0}^2 - \lambda_{m_i, m_j}^2 \right)_{1 \leq i, j \leq l},$$

where l runs over $1, \dots, k$ and $\mathbf{m} = (m_0, \dots, m_l)$ with $0 \leq m_0 < \dots < m_l \leq k$. The distribution function of the $(k+1)$ -dimensional Hüsler-Reiss distribution with standard Gumbel margins is then given by

$$H_A(\mathbf{x}) = \exp \left\{ \sum_{l=0}^k (-1)^{l+1} \sum_{\mathbf{m}: 0 \leq m_0 < \dots < m_l \leq k} h_{l, \mathbf{m}, A}(x_{m_1}, \dots, x_{m_l}) \right\}, \quad \mathbf{x} \in \mathbb{R}^{k+1}, \quad (3.5)$$

where

$$h_{l, \mathbf{m}, A}(y_0, \dots, y_l) = \int_{y_0}^{\infty} S \left\{ \left(y_i - z + 2\lambda_{m_i, m_0}^2 \right)_{i=1, \dots, l} | \Psi_{l, \mathbf{m}}(A) \right\} e^{-z} dz,$$

for $1 \leq l \leq k$ and $h_{0, \mathbf{m}, A}(y) = \exp(-y)$ for $\mathbf{m} \in \{0, \dots, k\}$. Furthermore, for $q \in \mathbb{N}$ and $\Psi \in \mathbb{R}^{q \times q}$ positive definite, $S(\cdot | \Psi)$ denotes the so-called survivor function of a q -dimensional normal random vector with mean vector $\mathbf{0}$ and covariance matrix Ψ , i.e., if $\mathbf{Y} \sim N(\mathbf{0}, \Psi)$ and $\mathbf{x} \in \mathbb{R}^q$, then $S(\mathbf{x} | \Psi) = \mathbb{P}(Y_1 > x_1, \dots, Y_q > x_q)$. In the bivariate case, the distribution function (3.5) simplifies to F_λ , where $\lambda = \lambda_{0,1} \in [0, \infty]$. Note that the class of Hüsler-Reiss distributions is closed in the sense that the lower-dimensional margins of H_A are again Hüsler-Reiss distributed with parameter matrix consisting of the respective entries in A . Consequently, the distribution of the bivariate sub-vector of the i -th and j -th component only depends on the parameter $\lambda_{i,j}$. Thus, one can modify this parameter (subject to the restriction $A \in \mathcal{D}$) without affecting the other components. This flexibility was demanded in Cooley *et al.* (2010) as a desirable property of multivariate extreme value models that most models do not possess, unfortunately.

Remark 3. *The k -variate Hüsler-Reiss distribution is usually given by its distribution function H_Λ . The density for $k \geq 3$ is rather complicated and involves multivariate integration. Hence, for maximum likelihood estimation based on block maxima, only the bivariate or sometimes the trivariate (cf. [Genton et al. \(2011\)](#)) densities are used in the framework of a composite likelihood approach.*

3.1.2 Brown-Resnick processes

We review the definition of the extended class of Brown-Resnick processes in (1.5) in more detail. For $T = \mathbb{R}^d$, $d \geq 1$, let $\{Y(t) : t \in T\}$ be a centered Gaussian process with stationary increments. Further, let $\gamma(t) = \mathbb{E}(Y(t) - Y(0))^2$ and $\sigma^2(t) = \mathbb{E}(Y(t))^2$ be the variogram and the variance of Y , $t \in \mathbb{R}^d$, respectively. Then, for a Gumbel point process $\sum_{i \in \mathbb{N}} \delta_{U_i}$ on \mathbb{R} and i.i.d. copies $Y_i \sim Y$, $i \in \mathbb{N}$, the process

$$\xi(t) = \max_{i \in \mathbb{N}} \left[U_i + Y_i(t) - \sigma^2(t)/2 \right], \quad t \in \mathbb{R}^d, \quad (3.6)$$

is max-stable, stationary and its distribution only depends on the variogram γ . For the special case where Y is a Brownian motion, the process ξ was already introduced by [Brown & Resnick \(1977\)](#). Its generalization in (3.6) is called *Brown-Resnick process associated to the variogram γ* ([Kabluchko et al. \(2009\)](#)). Since any conditionally negative definite function can be used as variogram, Brown-Resnick processes constitute an extremely flexible class of max-stable random fields. Moreover, the subclass associated to the family of fractal variograms $\gamma_{\alpha,s}(\cdot) = \|\cdot/s\|^\alpha$, $\alpha \in (0,2]$, $s \in (0,\infty)$, arises as limits of pointwise maxima of suitably rescaled and normalized, independent, *stationary* and isotropic Gaussian random fields (cf. [Kabluchko et al. \(2009\)](#)). Here $\|\cdot\|$ denotes the Euclidean norm. The model by [Smith \(1990\)](#) is another frequently used special case of Brown-Resnick processes, which corresponds to the class of variograms $\gamma(h) = \|h\Sigma^{-1}h\|$, for $h \in \mathbb{R}^d$ and an arbitrary covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$.

We remark that the finite-dimensional marginal distribution at locations $t_0, \dots, t_k \in \mathbb{R}^d$ of a Brown-Resnick process is the Hüsler-Reiss distribution H_Λ with $\Lambda = (\gamma(t_i - t_j)/4)_{0 \leq i, j \leq k}$.

3.1.3 Extremal coefficient function

Recall from Chapter 1 the definition of extremal correlation ρ . In this chapter we use a similar quantity called *extremal coefficient* θ . For a bivariate max-stable random vector (X_1, X_2) with identically distributed margins, $\theta \in [1,2]$ is determined by

$$\mathbb{P}(X_1 \leq u, X_2 \leq u) = \mathbb{P}(X_1 \leq u)^\theta,$$

for some suitable $u \in \mathbb{R}$, i.e. $\theta = 2 - \rho$. Similarly, for a stationary, max-stable process ξ on \mathbb{R}^d , the *extremal coefficient function* $\theta(h)$ is defined as $\theta(h) = 2 - \rho(h)$, for $h \in \mathbb{R}^d$.

For the bivariate Hüsler-Reiss distribution (3.3) we have $F_\lambda(u, u) = \exp(-2\Phi(\lambda)e^{-u})$ and thus, the extremal coefficient equals $\theta = 2\Phi(\lambda)$. Hence, for Hüsler-Reiss distributions, there is a one-to-one correspondence between the parameter $\lambda \in [0, \infty]$ and the set of extremal coefficients. Similarly, the extremal coefficient function of the Brown-Resnick process in (3.6)

is given by

$$\theta(h) = 2\Phi\left(\sqrt{\gamma(h)}/2\right), \quad h \in \mathbb{R}^d. \quad (3.7)$$

3.2 Estimation

In this section, we propose new estimators for the parameter matrix Λ of the Hüsler-Reiss distribution and use them to fit Brown-Resnick processes based on a parametric family of variograms. We will consider both estimation based on extremal increments and estimation in the spectral domain.

Suppose that $\mathbf{X}_i = (X_i^{(0)}, \dots, X_i^{(k)})$, $i = 1, \dots, n$, are independent copies of a random vector $\mathbf{X} \in \mathbb{R}^{k+1}$ in the MDA of the Hüsler-Reiss distribution H_Λ with some parameter matrix $\Lambda = (\lambda_{i,j}^2)_{0 \leq i, j \leq k} \in \mathcal{D}$. Recall that H_Λ has standard Gumbel margins. Without loss of generality, we assume that \mathbf{X} has standard exponential margins. Otherwise we could consider $(U_0(X_i^{(0)}), \dots, U_k(X_i^{(k)}))$, where $U_i = -\log(1 - F_i)$, and F_i is the cumulative distribution function of the i -th marginal of \mathbf{X} (cf. Prop. 5.15 in Resnick (2008)). In the sequel, we denote by $\tilde{\mathbf{X}}_n = \mathbf{X} - \log n$ and $\tilde{\mathbf{X}}_{i,n} = \mathbf{X}_i - \log n$ the rescaled data such that the empirical point process $\Pi_n = \sum_{i=1}^n \delta_{\tilde{\mathbf{X}}_{i,n}}$ converges in distribution to a Poisson point process Π on $E = [-\infty, \infty)^{k+1} \setminus \{-\infty\}$ with intensity measure $\mu([-\infty, \mathbf{x}]^C) = -\log H_\Lambda(\mathbf{x})$ (Prop. 3.21 in Resnick (2008)), as $n \rightarrow \infty$. Based on this convergence of point processes, the following theorem provides the conditional distribution of those data which are extreme in some sense.

Theorem 3. *For $m \in \mathbb{N}$ and a metric space S , let $g : \mathbb{R}^{k+1} \rightarrow S$ be a measurable transformation of the data and assume that it satisfies the invariance property $g(x + a \cdot \mathbf{1}) = g(x)$ for any $a \in \mathbb{R}$ and $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^{k+1}$. Further, let $u(n) > 0$, $n \in \mathbb{N}$, be a sequence of real numbers such that $\lim_{n \rightarrow \infty} u(n)/n = 0$. Then, for all Borel sets $B \in \mathcal{B}(S)$ and $A \in \mathcal{B}(E)$ bounded away from $-\infty$,*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ g(\tilde{\mathbf{X}}_n) \in B \mid \tilde{\mathbf{X}}_n \in A - \log u(n) \right\} = Q_{g,A}(B), \quad (3.8)$$

for some probability measure $Q_{g,A}$ on S .

Remark 4. *Note that due to the invariance property of g , the transformed data is independent of the rescaling, i.e. $g(\tilde{\mathbf{X}}_{i,n}) = g(\mathbf{X}_i)$, for all $i = 1, \dots, n$, $n \in \mathbb{N}$.*

In the above theorem, $u(n)$ only has to satisfy $u(n)/n \rightarrow 0$, as n tends to ∞ . However, for practical applications it is advisable to choose $u(n)$ in such a manner that also $\lim_{n \rightarrow \infty} u(n) = \infty$, since this ensures that the cardinality of the index set of extremal observations

$$I_A = \{i \in \{1, \dots, n\} : \tilde{\mathbf{X}}_{i,n} \in A - \log u(n)\}, \quad (3.9)$$

tends to ∞ as $n \rightarrow \infty$, almost surely.

Theorem 3 implies that for all extreme events, the transformed data $\{g(\mathbf{X}_i) : i \in I_A\}$ approximately follow the distribution $Q_{g,A}$. Clearly, $Q_{g,A}$ depends on the choices for g and

A and in the subsequent sections we encounter different possibilities for which the limit (3.8) can be computed explicitly. Furthermore, if g and A are chosen suitably, the distribution $Q_{g,A}$ will still contain all information on the parameter matrix Λ . Our estimators will therefore be based on the set of transformed data $\{g(\mathbf{X}_i) : i \in I_A\}$ and the knowledge of their asymptotic distribution $Q_{g,A}$. For instance, a maximum likelihood approach can be applied using the fact that Π_n converges to Π . If, for a particular realization of the \mathbf{X}_i , $I_A = \{i_1, \dots, i_N\}$ for some $N \leq n$, $i_1, \dots, i_N \in \{1, \dots, n\}$, and $g(\mathbf{X}_i) = \mathbf{s}_i$, $i = 1, \dots, n$, a canonical approach is to maximize the likelihood

$$\begin{aligned} L_{g,A}(\Lambda; \mathbf{s}_1, \dots, \mathbf{s}_n) &= \mathbb{P} \left\{ |I_A| = N, g(\mathbf{X}_{i_j}) \in d\mathbf{s}_{i_j}, j = 1, \dots, N \right\} \\ &= \mathbb{P} (|I_A| = N) \prod_{j=1}^N \mathbb{P} \left\{ g(\mathbf{X}) \in d\mathbf{s}_{i_j} \mid \tilde{\mathbf{X}}_n \in A - \log u(n) \right\}. \end{aligned}$$

With the Poisson approximation $\sum_{i=1}^n \mathbf{1}\{\tilde{\mathbf{X}}_{i,n} \in A - \log u(n)\} \approx \text{Pois}\{\mu(A - \log u(n))\}$ and the convergence (3.8) we obtain

$$L_{g,A}(\Lambda; \mathbf{s}_1, \dots, \mathbf{s}_n) \approx \exp\{-\mu(A - \log u(n))\} \frac{\mu(A - \log u(n))^N}{N!} \prod_{j=1}^N Q_{g,A}(d\mathbf{s}_{i_j}). \quad (3.10)$$

If the ML approach is unfeasible, estimation of Λ can also be based on other suitably chosen functionals of the conditional distribution of $g(\mathbf{X})$, for instance on the variance of $g(\mathbf{X})$.

3.2.1 Inference based on extremal increments

In this subsection, we apply Theorem 3 with g mapping the data to its increments w.r.t. a fixed index, i.e., $g : \mathbb{R}^{k+1} \rightarrow \mathbb{R}^k$, $\mathbf{x} \mapsto \Delta \mathbf{x} = (x^{(1)} - x^{(0)}, \dots, x^{(k)} - x^{(0)})$. In particular, g satisfies the invariance property $g(x + a \cdot \mathbf{1}) = g(x)$ for any $a \in \mathbb{R}$. Consequently, our estimators are based on the incremental distribution of those data which are extreme in the sense specified by the set A . The following theorem provides the limiting distribution $Q_{g,A}$ for two particular choices of A , namely $A_1 = (0, \infty) \times \mathbb{R}^k$ and $A_2 = [-\infty, \mathbf{0}]^C$.

Theorem 4. *Let \mathbf{X} be in the MDA of H_Λ with some $\Lambda \in \mathcal{D}$, and suppose that the sequence $u(n)$ is chosen as in Theorem 3. Then, we have the following convergences in distribution.*

1. For $k \in \mathbb{N}$,

$$\left(X^{(1)} - X^{(0)}, \dots, X^{(k)} - X^{(0)} \mid \tilde{X}_n^{(0)} > -\log u(n) \right) \xrightarrow{d} \mathcal{N}(M, \Sigma), \quad n \rightarrow \infty,$$

where $\mathcal{N}(M, \Sigma)$ denotes the multivariate normal distribution with mean vector $M = -\text{diag}(\Psi_{k,(0,\dots,k)}(\Lambda))/2$ and covariance matrix $\Sigma = \Psi_{k,(0,\dots,k)}(\Lambda)$.

2. For the bivariate case, i.e., $k = 1$,

$$\left(X^{(1)} - X^{(0)} \mid \tilde{X}_n^{(0)} > -\log u(n) \text{ or } \tilde{X}_n^{(1)} > -\log u(n) \right) \xrightarrow{d} Z, \quad n \rightarrow \infty,$$

where Z is a real-valued random variable with density given by

$$g_\lambda(t) = \frac{1}{4\lambda\Phi(\lambda)} \phi\left(\lambda - \frac{|t|}{2\lambda}\right), \quad t \in \mathbb{R}, \lambda = \lambda_{0,1}.$$

Here, Φ and ϕ denote the standard normal distribution function and density, respectively.

Remark 5. The positive definite matrix $\Sigma = \Psi_{k,(0,\dots,k)}(\Lambda)$ contains all information on Λ . In fact, the transformation

$$\Lambda(\Sigma) = \frac{1}{4} \left(\begin{array}{c|c} 0 & \text{diag}(\Sigma)^\top \\ \hline \text{diag}(\Sigma) & \mathbf{1} \text{diag}(\Sigma)^\top + \text{diag}(\Sigma)\mathbf{1}^\top - 2\Sigma \end{array} \right)$$

recovers the matrix $\Lambda = (\lambda_{i,j}^2)_{0 \leq i,j \leq k}$.

Based on the convergence results in Theorem 4 we propose various estimation procedures for both multivariate Hüsler-Reiss distributions (non-parametric case) and Brown-Resnick processes with a parameterized family of variograms (parametric case).

Non-parametric multivariate case

For the likelihood based approach in (3.10) we first consider the extremal set $A_1 = (0, \infty) \times \mathbb{R}^k$ and put $N_1 = |I_{A_1}|$. By part one of Theorem 4 we have

$$\begin{aligned} -\log L(\Lambda; \mathbf{s}_1, \dots, \mathbf{s}_n) &\approx -\log \left\{ \exp(-u(n)) \frac{u(n)^{N_1}}{N_1!} \prod_{j=1}^{N_1} \phi_{M(\Lambda), \Sigma(\Lambda)}(\mathbf{s}_{i_j}) \right\} \\ &\propto \frac{N_1}{2} \log \det \Sigma(\Lambda) + \frac{1}{2} \sum_{j=1}^{N_1} \left\{ (\mathbf{s}_{i_j} - M(\Lambda))^\top \Sigma(\Lambda)^{-1} (\mathbf{s}_{i_j} - M(\Lambda)) \right\}, \end{aligned} \quad (3.11)$$

where \mathbf{s}_i is the realization of $\Delta \mathbf{X}_i$, $i = 1, \dots, n$ and $\phi_{M(\Lambda), \Sigma(\Lambda)}$ is the density of the normal distribution with mean vector $M(\Lambda) = -\text{diag}(\Psi_{k,(0,\dots,k)}(\Lambda))/2$ and covariance matrix $\Sigma(\Lambda) = \Psi_{k,(0,\dots,k)}(\Lambda)$. The corresponding maximum likelihood estimator is given by

$$\hat{\Lambda}_{\text{MLE}} = \arg \min_{\Lambda \in \mathcal{D}} \left[\frac{N_1}{2} \log \det \Sigma(\Lambda) + \frac{1}{2} \sum_{j=1}^{N_1} \left\{ (\mathbf{s}_{i_j} - M(\Lambda))^\top \Sigma(\Lambda)^{-1} (\mathbf{s}_{i_j} - M(\Lambda)) \right\} \right]. \quad (3.12)$$

Notice that for this particular choice of A , the asymptotic value of $\mathbb{P}(|I_{A_1}| = N_1)$ does not depend on the parameter matrix Λ . Hence, this ML ansatz coincides with simply maximizing the likelihood of the increments without considering the number of points exceeding the

threshold. In the bivariate case, i.e., $k = 1$ and $A_1 = (0, \infty) \times \mathbb{R}$, (3.11) simplifies to

$$-\log L(\lambda; s_1, \dots, s_n) \propto \frac{N_1 \lambda^2}{2} + N_1 \log \lambda + \frac{1}{8\lambda^2} \sum_{j=1}^{N_1} s_{i_j}^2, \quad (3.13)$$

and the minimizer of (3.13) can be given in explicit form:

$$\hat{\lambda}_{\text{MLE}}^2 = \frac{1}{2} \left\{ -1 + \sqrt{1 + \frac{1}{N_1} \sum_{j=1}^{N_1} (\Delta \mathbf{X}_{i_j})^2} \right\}. \quad (3.14)$$

Staying in the bivariate case, for the choice $A_2 = [-\infty, \mathbf{0}]^C$, we put $N_2 = |I_{A_2}|$ and by part two of Theorem 4,

$$\begin{aligned} -\log L(\lambda; s_1, \dots, s_n) &\approx -\log \left\{ \exp(-2\Phi(\lambda)u(n)) \frac{(2\Phi(\lambda)u(n))^{N_2}}{N_2!} \prod_{j=1}^{N_2} g_\lambda(s_{i_j}) \right\} \\ &\propto 2\Phi(\lambda)u(n) + \frac{N_2 \lambda^2}{2} + \frac{1}{8\lambda^2} \sum_{j=1}^{N_2} s_{i_j}^2. \end{aligned}$$

Numerical optimization can be applied to obtain the estimator

$$\hat{\lambda}_{\text{MLE2}}^2 = \arg \min_{\theta \geq 0} \left\{ 2\Phi(\sqrt{\theta})u(n) + \frac{N_2 \theta}{2} + \frac{1}{8\theta} \sum_{j=1}^{N_2} (\Delta \mathbf{X}_{i_j})^2 \right\}.$$

While the above likelihood-based estimators (except for (3.14)) require numerical optimization, the following approach is computationally much more efficient: A natural estimator for $\Sigma = \Psi_{k,(0,\dots,k)}(\Lambda) \in \mathbb{R}^{k \times k}$ based on the first part of Theorem 4 is given by the empirical covariance $\hat{\Sigma}$ of the extremal increments $\Delta \mathbf{X}_i = (X_i^{(1)} - X_i^{(0)}, \dots, X_i^{(k)} - X_i^{(0)})$ for $i \in I_{A_1}$, i.e.

$$\hat{\Sigma} = \frac{1}{N_1} \sum_{j=1}^{N_1} (\Delta \mathbf{X}_{i_j} - \hat{\mu})(\Delta \mathbf{X}_{i_j} - \hat{\mu})^\top, \quad \hat{\mu} = \frac{1}{N_1} \sum_{j=1}^{N_1} \Delta \mathbf{X}_{i_j}. \quad (3.15)$$

By Remark 5 this also gives an estimator $\hat{\Lambda}_{\text{Var}} = \Lambda(\hat{\Sigma})$ for the parameter matrix Λ , which we call the *variance-based estimator*. Apart from its simple form, another advantage of (3.15) is that $\hat{\Sigma}$ is automatically a positive definite matrix and hence, $\hat{\Lambda}_{\text{Var}}$ is conditionally negative definite and therefore a valid matrix for a $(k+1)$ -variate Hüsler-Reiss distribution. Note that (3.15) is not the maximum likelihood estimator (MLE) for Σ since the mean of the conditional distribution of $\Delta \mathbf{X}_i$ depends on the diagonal of Σ . The MLE of Σ is instead given by optimizing (3.12) w.r.t. Σ , which, to our knowledge, does not admit a closed analytical form.

Applying (3.15) with $k = 1$ yields the bivariate variance-based estimator

$$\hat{\lambda}_{\text{Var}}^2 = \frac{1}{4N_1} \sum_{j=1}^{N_1} (X_{i_j}^{(1)} - X_{i_j}^{(0)} - \hat{\mu})^2, \quad \hat{\mu} = \frac{1}{N_1} \sum_{j=1}^{N_1} (X_{i_j}^{(1)} - X_{i_j}^{(0)}). \quad (3.16)$$

Since the mean of the extremal increments is also directly related to the parameter λ , another sensible estimator might be

$$\hat{\lambda}_{\text{mean}}^2 = -\frac{1}{2N_1} \sum_{j=1}^{N_1} (X_{i_j}^{(1)} - X_{i_j}^{(0)}). \quad (3.17)$$

Parametric approach for Brown-Resnick processes

Statistical inference for Brown-Resnick processes as in (3.6) is usually based on fitting a parametric variogram model $\{\gamma_\vartheta : \vartheta \in \Theta\}$, $\Theta \subset \mathbb{R}^j$, $j \in \mathbb{N}$, to point estimates of the extremal coefficient function (3.7) based on the madogram. Alternatively, composite likelihood approaches are used in connection with block maxima of bivariate data (Davison & Gholamrezaee (2012)).

Since for $t_0, \dots, t_k \in \mathbb{R}^d$, the vector $(\xi(t_0), \dots, \xi(t_k))$ with ξ being a Brown-Resnick process associated to the variogram $\gamma : \mathbb{R}^d \rightarrow [0, \infty)$ is Hüsler-Reiss distributed with parameter matrix

$$\Lambda = (\gamma(t_i - t_j)/4)_{0 \leq i, j \leq k},$$

the above estimators enable parametric estimation of Brown-Resnick processes. In fact, replacing Λ in (3.12) by

$$\Lambda(\vartheta) = (\gamma_\vartheta(t_i - t_j)/4)_{0 \leq i, j \leq k} \quad (3.18)$$

leads to the ML estimator

$$\hat{\vartheta}_{\text{MLE}} = \arg \min_{\vartheta \in \Theta} \{-\log L(\Lambda(\vartheta); \mathbf{s}_1, \dots, \mathbf{s}_n)\}$$

with L as in (3.11). Note that, other than in classical extreme value statistics, here the use of higher-dimensional densities is feasible and promises a gain in accuracy.

Estimation of ϑ can also be based on any of the bivariate estimators $\hat{\lambda}_{\text{MLE}}^2$, $\hat{\lambda}_{\text{MLE2}}^2$, $\hat{\lambda}_{\text{Var}}^2$, $\hat{\lambda}_{\text{mean}}^2$, or on the multivariate estimator $\hat{\Lambda}_{\text{Var}}$ by “projecting” the latter matrix or the matrix consisting of all bivariate estimates onto the set of matrices $\{(\gamma_\vartheta(t_i - t_j)/4)_{0 \leq i, j \leq k} : \vartheta \in \Theta\}$, i.e.,

$$\hat{\vartheta}_{\text{PROJ}} = \arg \min_{\vartheta \in \Theta} \left\| \left(\hat{\lambda}_{i_j}^2 - \gamma_\vartheta(t_i - t_j)/4 \right)_{0 \leq i, j \leq k} \right\|, \quad (3.19)$$

where $\|\cdot\|$ can be any matrix norm.

Similar to Bacro & Gaetan (2012), the bivariate estimators can readily be used in a parametric

composite likelihood framework.

3.2.2 Inference based on spectral densities

As at the beginning of Section 3.2, let \mathbf{X}_i , $i = 1, \dots, n$, be a sequence of independent copies of \mathbf{X} , already standardized to exponential margins, in the MDA of the max-stable distribution H_Λ . Since we work in the spectral domain in this section, we will switch to standard Fréchet margins with distribution function $\exp(-1/y)$, $y \geq 0$. More precisely, we consider the vectors $\mathbf{Y} = \exp(\mathbf{X})$ and $\mathbf{Y}_i = \exp(\mathbf{X}_i)$, $i = 1, \dots, n$, which are in the MDA of the Hüsler-Reiss distribution $G_\Lambda(\mathbf{x}) = H_\Lambda(\log \mathbf{x})$, $\mathbf{x} \geq \mathbf{0}$, with standard Fréchet margins.

The most convenient tool to characterize the dependence structure of a multivariate extreme value distribution is via its spectral measure. To this end, let $\tilde{\mathbf{Y}}_n = \mathbf{Y}/n$ and $\tilde{\mathbf{Y}}_{i,n} = \mathbf{Y}_i/n$ denote the rescaled data such that the point process $P_n = \sum_{i=1}^n \delta_{\tilde{\mathbf{Y}}_{i,n}}$ converges, as $n \rightarrow \infty$, to a non-homogeneous Poisson point process P on $[0, \infty)^{k+1} \setminus \{\mathbf{0}\}$ with intensity measure $\nu([\mathbf{0}, \mathbf{x}]^C) = -\log G_\Lambda(\mathbf{x})$. Transforming a vector $\mathbf{x} = (x_0, \dots, x_k) \in [0, \infty)^{k+1} \setminus \{\mathbf{0}\}$ to its pseudo-polar coordinates

$$r = \|\mathbf{x}\|, \quad \boldsymbol{\omega} = r^{-1}\mathbf{x}, \quad (3.20)$$

for any norm $\|\cdot\|$ on \mathbb{R}^{k+1} , we can rewrite ν as a measure on $(0, \infty) \times S_k$, where S_k is the k -dimensional unit simplex $S_k = \{\mathbf{y} \geq \mathbf{0} : \|\mathbf{y}\| = 1\}$. Namely, we have $\nu(d\mathbf{x}) = r^{-2}dr \times M(d\boldsymbol{\omega})$, where the measure M is called the spectral measure of G_Λ and embodies the dependence structure of the extremes. For our purposes, it is most convenient to choose the L_1 -norm, i.e., $\|\mathbf{x}\|_1 = \sum_{i=0}^k |x_i|$. In this case, for the set $A_{r_0} = \{\mathbf{x} \in [0, \infty)^{k+1} \setminus \{\mathbf{0}\} : \|\mathbf{x}\|_1 > r_0\}$, $r_0 > 0$, we obtain

$$\nu(A_{r_0}) = r_0^{-1}M(S_k) = r_0^{-1} \cdot (k+1), \quad (3.21)$$

since the measure M satisfies $\int_{S_k} \omega_i M(d\boldsymbol{\omega}) = 1$ for $i = 0, \dots, k$. Hence, the ν -measure of A_{r_0} does not depend on the parameters of the specific model chosen for M . The distribution function can be written as

$$G_\Lambda(\mathbf{x}) = \exp \left\{ - \int_{S_k} \max \left(\frac{\omega_0}{x_0}, \dots, \frac{\omega_k}{x_k} \right) M(d\boldsymbol{\omega}) \right\}, \quad \mathbf{x} \geq \mathbf{0}.$$

As the space of all spectral measures is infinite-dimensional, there is a need of parametric models which are analytically tractable and at the same time flexible enough to approximate the dependence structure in real data sufficiently well. Parametric models are usually given in terms of their spectral density h of the measure M . The book by Kotz & Nadarajah (2000) gives an overview of parametric multivariate extreme value distributions, most of them, however, being only valid in the bivariate case. For the multivariate case only few models are known, e.g., the logistic distribution and its extensions (Joe, 1990; Tawn, 1990) and the Dirichlet distribution Coles & Tawn (1991). The recent interest in this topic resulted in new multivariate parametric models (Boldi & Davison (2007); Cooley *et al.* (2010)) as well as in general construction principles for multivariate spectral measures (Ballani & Schlather

(2011)). All these approaches have in common that they propose models for multivariate max-stable distributions in order to fit data obtained by exceedances over a certain threshold or by block maxima.

Given a parametric model for the spectral density $h(\cdot; \vartheta)$, we have the analog result as in Theorem 3 for the Fréchet case with $A = A_{r_0}$ and $g: \mathbb{R}^{k+1} \rightarrow S_k, \mathbf{x} \mapsto \mathbf{x}/\|\mathbf{x}\|_1$, which now satisfies the multiplicative invariance property $g(a \cdot \mathbf{x}) = g(\mathbf{x})$, for all $a \in \mathbb{R}$. The Fréchet version of (3.8) for this choice of g and A reads as

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \mathbf{Y}/\|\mathbf{Y}\|_1 \in B \mid \tilde{\mathbf{Y}}_n \in A_{r_0}/u(n) \right\} = \frac{M(B)}{M(S_k)} = \frac{1}{k+1} \int_B h(\boldsymbol{\omega}; \vartheta) d\boldsymbol{\omega}, \quad (3.22)$$

for all $B \in \mathcal{B}(S_k)$ and $u(n)$, $n \in \mathbb{N}$, as in Theorem 3. Based on this conditional distribution of those \mathbf{Y}_i for which the sum $\|\mathbf{Y}_i\|_1$ is large, similarly to (3.10) we obtain the likelihood

$$\begin{aligned} L_{A_{r_0}}(\vartheta; (r_1, \boldsymbol{\omega}_1), \dots, (r_n, \boldsymbol{\omega}_n)) \\ \approx \exp\{-\nu(A_{r_0}/u(n))\} \frac{\nu(A_{r_0}/u(n))^{|I_0|}}{|I_0|!} \prod_{i \in I_0} r_i^{-2} (k+1)^{-1} h(\boldsymbol{\omega}_i; \vartheta) \\ \propto \prod_{i \in I_0} h(\boldsymbol{\omega}_i; \vartheta), \end{aligned} \quad (3.23)$$

where $\{(r_i, \boldsymbol{\omega}_i) : 1 \leq i \leq n\}$ are the pseudo-polar coordinates of $\{\tilde{\mathbf{Y}}_{i,n} : 1 \leq i \leq n\}$ as in (3.20) and I_0 is the set of all indices $1 \leq i \leq n$ with $\tilde{\mathbf{Y}}_{i,n} \in A_{r_0}/u(n)$. Note that the proportional part in (3.23) only holds because the ν -measure of A_{r_0} is independent of the model parameter ϑ .

For the Hüsler-Reiss distribution it is possible to write down the spectral density $h(\cdot; \Lambda)$ explicitly.

Proposition 1. *For any matrix $\Lambda = (\lambda_{i,j}^2)_{0 \leq i,j \leq k} \in \mathcal{D}$ the Hüsler-Reiss distribution can be written as*

$$G(\mathbf{x}) = \exp \left\{ - \int_{S_k} \max \left(\frac{\omega_0}{x_0}, \dots, \frac{\omega_k}{x_k} \right) h(\boldsymbol{\omega}; \Lambda) d\boldsymbol{\omega} \right\},$$

with spectral density

$$h(\boldsymbol{\omega}, \Lambda) = \frac{1}{\omega_0^2 \cdots \omega_k (2\pi)^{k/2} |\det \Sigma|^{1/2}} \exp \left(-\frac{1}{2} \tilde{\boldsymbol{\omega}}^\top \Sigma^{-1} \tilde{\boldsymbol{\omega}} \right), \quad \boldsymbol{\omega} \in S_k, \quad (3.24)$$

where $\Sigma = \Psi_{k,(0,\dots,k)}(\Lambda)$ and $\tilde{\boldsymbol{\omega}} = (\log \frac{\omega_i}{\omega_0} + 2\lambda_{i,0}^2 : 1 \leq i \leq k)^\top$.

Non-parametric, multivariate case

Based on the explicit expression for the spectral density of the Hüsler-Reiss distribution in (3.24), we define the estimator $\hat{\Lambda}_{\text{SPEC}}$ of Λ as the matrix in \mathcal{D} that maximizes the likelihood

in (3.23), i.e.,

$$\hat{\Lambda}_{\text{SPEC}} = \arg \min_{\Lambda \in \mathcal{D}} \left(\frac{|I_0|}{2} \log \det \Psi_{k,(0,\dots,k)}(\Lambda) + \frac{1}{2} \sum_{i \in I_0} \tilde{\omega}_i^\top \Psi_{k,(0,\dots,k)}(\Lambda)^{-1} \tilde{\omega}_i \right). \quad (3.25)$$

In the bivariate case, the spectral density in (3.24) simplifies to

$$h(\boldsymbol{\omega}; \lambda) = \frac{1}{2\lambda\omega_0^2\omega_1(2\pi)^{1/2}} \exp \left(-\frac{(\log \frac{\omega_1}{\omega_0} + 2\lambda^2)^2}{8\lambda^2} \right)$$

and the corresponding estimator can be given in explicit form:

$$\hat{\lambda}_{\text{SPEC}}^2 = \frac{1}{2} \left[-1 + \sqrt{1 + \frac{1}{|I_0|} \sum_{i \in I_0} \left\{ \log \left(\tilde{\mathbf{Y}}_i^{(1)} / \tilde{\mathbf{Y}}_i^{(0)} \right) \right\}^2} \right]. \quad (3.26)$$

Note that the estimators (3.25) and (3.26) have exactly the same form as the maximum likelihood estimators (3.12) and (3.14), respectively, for the extremal increments. However, the specification of the set A differs and so does the choice of extreme data that is plugged in.

Parametric approach for Brown-Resnick processes

Analogously to Section 3.2.1, we obtain a parametric estimate of the dependence structure of a Brown-Resnick process based on a parametric family of variograms by replacing Λ on the right-hand side of (3.25) by $\Lambda(\vartheta)$ defined in (3.18). This yields

$$\hat{\vartheta}_{\text{SPEC}} = \arg \min_{\vartheta \in \Theta} \left\{ -\log L_{A_{r_0}}(\Lambda(\vartheta); (r_1, \boldsymbol{\omega}_1), \dots, (r_n, \boldsymbol{\omega}_n)) \right\}. \quad (3.27)$$

3.3 Simulation study

We compare the performance of the different parametric and non-parametric estimation procedures of Brown-Resnick processes and Hüsler-Reiss distributions proposed in the previous section via a simulation study.

In the first instance, we consider bivariate data that is in the MDA of the Hüsler-Reiss distribution with known dependence parameter $\lambda = \lambda_{0,1}$. For simplicity, we simulate data from the Hüsler-Reiss distribution itself, which does not mean that the thresholding procedure via the set A becomes obsolete. All estimators rely on considering only extremal events and hence, there is no obvious advantage over using any other data being in the MDA of H_λ . We compare the estimators $\hat{\lambda}_{\text{MLE}}^2$, $\hat{\lambda}_{\text{MLE2}}^2$, $\hat{\lambda}_{\text{Var}}^2$, $\hat{\lambda}_{\text{mean}}^2$ and $\hat{\lambda}_{\text{SPEC}}^2$ from Section 3.2 for different sample sizes $n \in \{500, 8000, 100000\}$. The sequence of thresholds $u(n)$ is chosen in such a way that the number of exceedances $k(n)$ increases to ∞ , but at the same time, the corresponding quantile $q(n) = 1 - k(n)/n$ approaches 1, as $n \rightarrow \infty$. In addition to the new threshold based estimators, we include the classical estimators, which use block maxima,

namely the madogram estimator $\hat{\lambda}_{\text{mado}} = \Phi^{-1}(\hat{\theta}_{\text{mado}}/2)$ (Cooley *et al.*, 2006) and the ML estimator $\hat{\lambda}_{\text{HRMLE}}^2$ of the bivariate Hüsler-Reiss distribution. To model a year of (dependent) data, we choose a block size of 150 which is of order of but less than 365.

The pseudo-code of the exact simulation setup is the following:

1. for $\lambda^2 \in \{k \cdot 0.025 : k = 1, \dots, 30\}$
2. for $n \in \{500, 8000, 100000\}$
3. simulate n bivariate Hüsler-Reiss distributions
with parameter λ
4. for $\hat{\lambda}^2 \in \{\hat{\lambda}_{\text{MLE}}^2, \hat{\lambda}_{\text{MLE2}}^2, \hat{\lambda}_{\text{Var}}^2, \hat{\lambda}_{\text{mean}}^2, \hat{\lambda}_{\text{SPEC}}^2, \hat{\lambda}_{\text{mado}}^2, \hat{\lambda}_{\text{HRMLE}}^2\}$
5. estimate λ^2 through $\hat{\lambda}^2$
6. obtain an estimate of the corresponding extremal
coefficient θ through $\hat{\theta} = \theta(\hat{\lambda}) = 2\Phi(\hat{\lambda})$
7. repeat (1)-(6) 500 times

Since the finite-dimensional margins of a Brown-Resnick process are Hüsler-Reiss distributed, we can easily implement step (1) by simulating a one-dimensional Brown-Resnick process with variogram $\gamma(h) = |h|$ on the interval $[0, 3]$. Since we consider bivariate Hüsler-Reiss distributions for different values of λ^2 lying on a fine grid, we visualize the estimates $\hat{\theta}$ as functions of the true λ^2 (Figure 3.1). However, it is important to remark that estimation in this first part of the study is exclusively based on the bivariate distributions. For each value of λ^2 , we repeat simulation and estimation 500 times. Figure 3.1 shows the pointwise mean value of the extremal coefficient and the corresponding empirical 95% confidence intervals. As expected, in finite samples, all estimators based on multivariate POT methods underestimate the true degree of extremal dependence since they are based on an asymptotic distribution with non-zero mean while the simulated data come from a stationary process. As the sample size n and the threshold $u(n)$ increase, all estimators approach the true value. Among the POT-based estimators, $\hat{\lambda}_{\text{SPEC}}^2$ seems to be at least as good as the other estimators, uniformly for all values of λ^2 under consideration. $\hat{\lambda}_{\text{Var}}^2$ performs well for small values of λ^2 but is more biased than other estimators for large values of λ^2 . The good performance of $\hat{\lambda}_{\text{mean}}^2$ for large values of λ^2 might be due to the fact that it only uses first moments of the extremal increments and is hence less sensible to aberration of the finite sample distribution from the asymptotic distribution. Compared to the two estimators based on block maxima, the POT-based estimators all perform well even for small data sets, which is a great advantage for many applications. Moreover, the variances of the POT-estimates are generally smaller than those based on block maxima, since more data can be used. Finally, note that the POT-based estimation does not exploit the fact, that the simulated data in the max-domain of attraction is in fact the max-stable distribution itself. The speed of convergence may though differ when using data from other models in the MDA. In contrast, $\hat{\lambda}_{\text{mado}}^2$ and $\hat{\lambda}_{\text{HRMLE}}^2$ do profit from simulating i.i.d. realizations of the max-stable distribution

itself since then, the blockwise maxima are exactly Hüsler-Reiss distributed and not only an approximation as in the case of real data.

In the second part of the simulation study we examine the performance of parametric estimates of Brown-Resnick processes using the same data as above. While the true variogram is $\gamma(h) = |h|$, we estimate the parameter vector (α, s) for the family of variograms $\gamma_{\alpha,s}(h) = \|h/s\|^\alpha$, $\alpha \in (0, 2]$, $s > 0$. We compare the following three estimators: the spectral estimator $(\alpha, s)_{\text{SPEC}}$, given by (3.27) and using the full multivariate density; the composite likelihood estimator $(\alpha, s)_{\text{SPEC, CL}}$, defined as the maximizer of the product of all bivariate spectral densities, implicitly assuming independence of all tuples of locations; and the least squares estimator $(\alpha, s)_{\text{PROJ, LS}}$, given by (3.19) for the Euclidean norm, where $\hat{\lambda}_{\text{MLE}}^2$ serves as non-parametric input. The estimated values of α and s are compared in the right column of Figure 3.2. The left panel shows the corresponding extremal coefficient functions for α and s representing the mean, the 5% sample quantile and the 95% sample quantile from the 500 repetitions, respectively.

The estimator $(\alpha, s)_{\text{SPEC}}$, which incorporates the full multivariate information, performs best both in the sense of minimal bias and minimal variance. Especially estimation of the shape parameter of the variogram gains stability when using higher-dimensional densities. The projection estimator seems to have the largest bias and the largest variance. The results remain very similar if we replace $\hat{\lambda}_{\text{MLE}}^2$ by one of the other non-parametric estimators. Let us finally remark that all three estimators can be modified by considering only small distances for inference. Then, since the approximation error of the asymptotic conditional distribution decays for smaller distances, this can substantially improve the accuracy in a simulation framework, but might distort the results in real data situations.

Figure 3.1: Estimated extremal coefficients compared to the true ones of bivariate Hüsler-Reiss distributions. 500 repetitions. Block size for block maxima is 150. Left: $\hat{\theta}$ vs. λ^2 . Right: relative difference of $\hat{\theta}$ to the true value of θ .

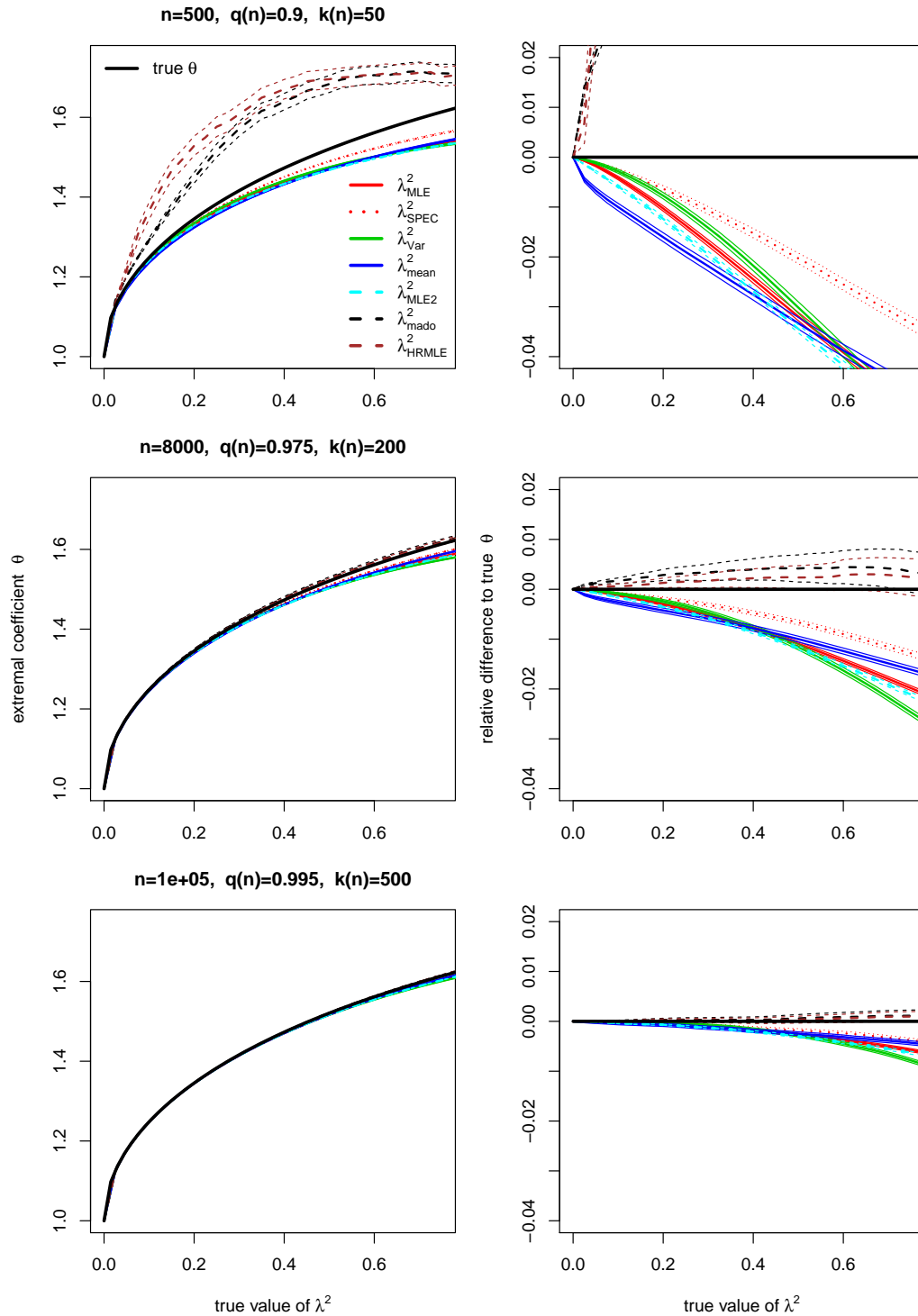
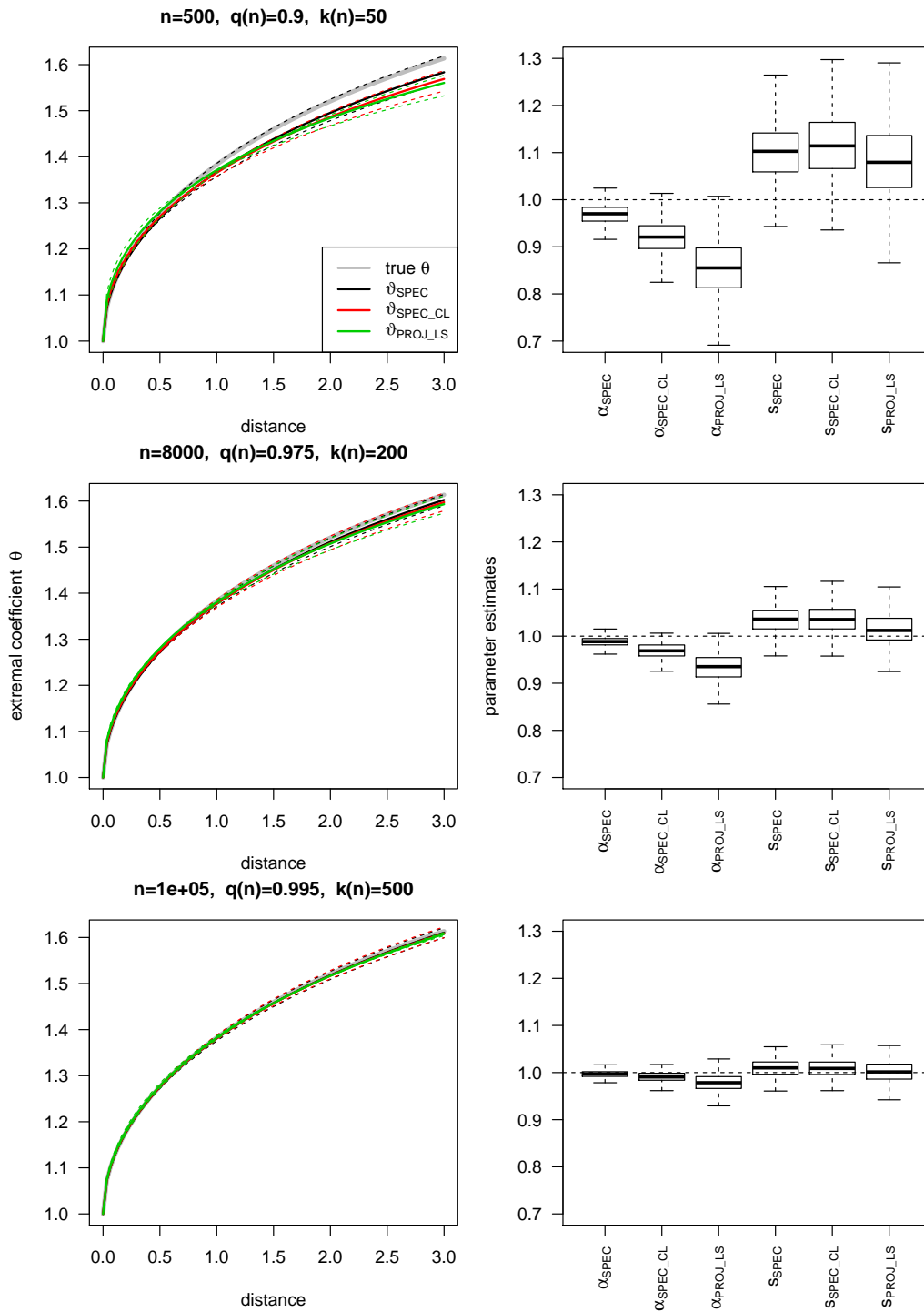


Figure 3.2: Parametric fit of Brown-Resnick process.



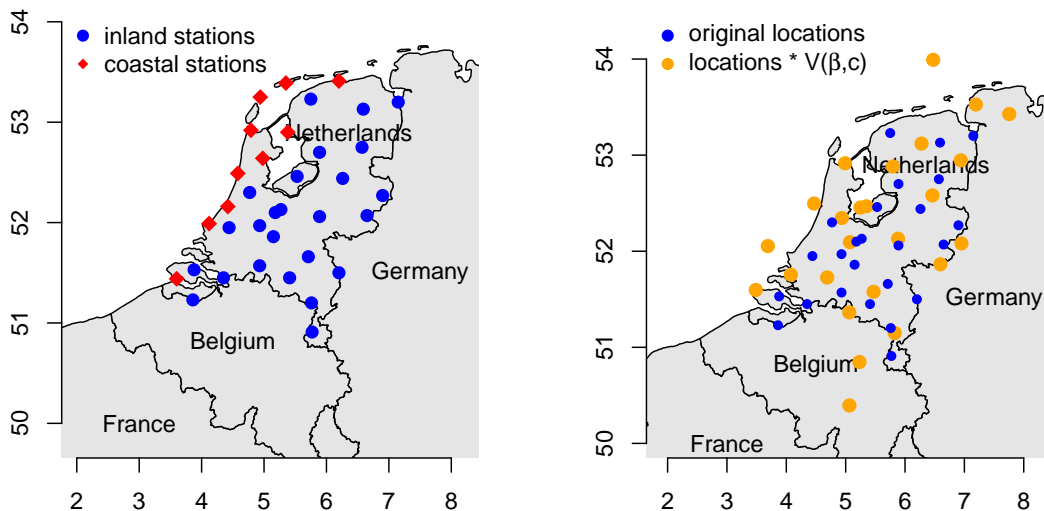
3.4 Application: Wind speed data

We apply the above theory of estimating Hüsler-Reiss distributions to wind speed data provided by the Royal Netherlands Meteorological Institute. We use the daily maxima of wind speeds measured at 35 meteorological stations $x_1, \dots, x_{35} \in \mathcal{X}$, where the set $\mathcal{X} \subset \mathbb{R}^2$ denotes the geographical coordinates of the Netherlands. The data cover a 23-year period of 8172 days from 1990/01/01 to 2012/05/12. Figure 3.3 provides an overview of the spatial locations of the stations.

3.4.1 Stationarity assumption with zonal anisotropy

In the sequel, we use the data to fit a stationary Brown-Resnick process based on the parametric family of variograms $\gamma_{\alpha,s}(h) = \|h/s\|^\alpha$, $\alpha \in (0, 2]$, $s > 0$. As mentioned in Section 3.1.2, this subclass of Brown-Resnick processes is a natural choice, since they arise as the max-limits of suitably rescaled, stationary Gaussian random fields. The stationarity assumption, however, turns out to be unrealistic, since stations close to the coast exhibit weaker extremal dependence to neighboring stations than inland stations. This is illustrated in the left panel of Figure 3.4, where the estimated bivariate extremal coefficients based on $\hat{\lambda}_{\text{MLE}}^2$ of all stations are compared to those without the coastal stations. Hence, we restrict our analysis to the 25 inland stations, say $T = \{x_1, \dots, x_{25}\}$, when fitting a stationary Brown-Resnick process. We therefore need to estimate the shape parameter α and the scale parameter s of the corresponding parameter matrix $\Lambda_{\alpha,s}$ of the Brown-Resnick process on T ,

Figure 3.3: Left: locations of the 35 meteorological stations. Right: locations of the 25 non-coast stations before and after multiplication with the anisotropy matrix $V(\hat{\beta}, \hat{c})$.



given by $A_{\alpha,s} = (\gamma_{\alpha,s}(x_i - x_j)/4)_{1 \leq i,j \leq 25}$.

While the above class of variograms assumes isotropy of the underlying process, meteorological data and particularly wind speed data can be expected to exhibit a main direction of spatial dependence. We capture this anisotropy by introducing a transformed space $\tilde{\mathcal{X}} = V\mathcal{X}$ (cf. right panel of Figure 3.3), where

$$V = V(\beta, c) = \begin{pmatrix} \cos \beta & -\sin \beta \\ c \sin \beta & c \cos \beta \end{pmatrix}, \quad \beta \in [0, 2\pi], c > 0,$$

is a rotation and dilution matrix; Blanchet & Davison (2011) recently applied this idea to the extremal Gaussian process of Schlather (2002). The new parametric variogram model becomes $A_{\vartheta} = (\gamma_{\alpha,s}(Vx_i - Vx_j)/4)_{1 \leq i,j \leq 25}$, where $\vartheta = (\alpha, s, \beta, c)$ is the vector of parameters. As in the above simulation study, we apply the three estimators

$$\hat{\vartheta}_{\text{PROJ, LS}} = \arg \min_{\vartheta \in \Theta} \|(\hat{\lambda}_{\text{MLE}, ij}^2)_{1 \leq i,j \leq 25} - A_{\vartheta}\|_2, \quad \hat{\vartheta}_{\text{SPEC}}, \quad \hat{\vartheta}_{\text{SPEC, CL}}. \quad (3.28)$$

For all estimators, the data is first normalized as described at the beginning of Section 3.2 and the threshold $u(n)$ is chosen in such a way that, out of the 8172 days, all data above the 97.5%-quantile are labeled as extremal. Note that these numbers coincide with the second set of parameters $(n, q(n))$ in the simulation study. Hence, the middle row of Figure 3.1 provides a rough estimate of the estimation error.

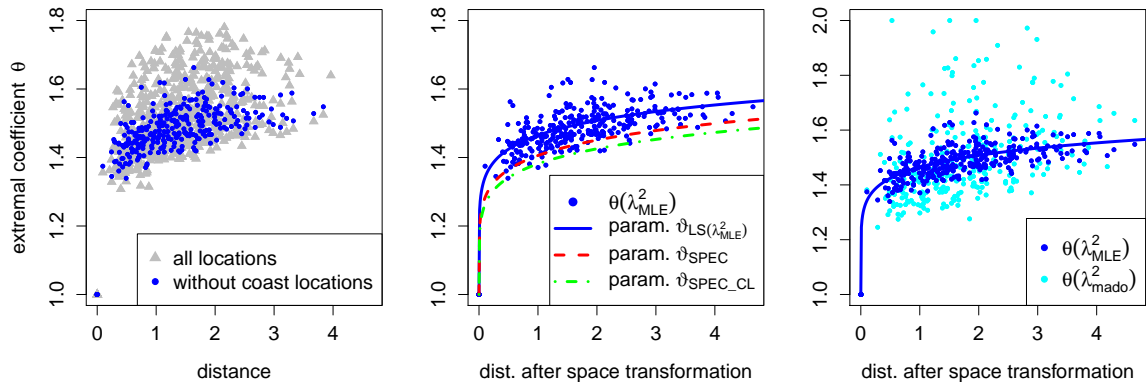
The estimation results and standard deviations for the parameters (α, s, β, c) are given in Table 3.1. The middle panel of Figure 3.4 illustrates the effect of transforming the space via the matrix V and displays the fitted extremal coefficient functions for the three estimators in (3.28). Moreover, the right panel shows the estimates of pairwise extremal coefficients based on $\hat{\lambda}_{\text{MLE}}^2$ and the model-independent madogram estimator, where the latter exhibits a considerably larger variation. In Figure 3.5, we illustrate the effect of transforming the space via the matrix $V(\beta, c)$ on the extremal coefficient function and on a typical realization of the corresponding Brown-Resnick process.

In order to validate the reliability of the estimated model parameters ϑ , we re-simulate data in the MDA of the three fitted Brown-Resnick models. Similarly to the simulation study, we use 8172 realizations of the Brown-Resnick process itself (which is clearly in its own MDA) for the daily data. As index set, we use the transformed locations $V(\hat{\beta}, \hat{c})T$ on which the Brown-Resnick process is isotropic. Based on this new data, we apply the estimation procedure exactly as for the real data to obtain new estimates for ϑ and thus for the extremal coefficient function. This is repeated 100 times and the results for the

Table 3.1: Estimation results. The values for the standard deviation are obtained from simulating and re-estimating the respective models 100 times.

estimator	α	s	β	c
$\hat{\vartheta}_{\text{PROJ, LS}}$	0.296 (0.0193)	0.234 (0.0744)	0.379 (0.532)	1.67 (0.1761)
$\hat{\vartheta}_{\text{SPEC}}$	0.338 (0.0166)	0.687 (0.1797)	0.456 (0.439)	2.21 (0.1596)
$\hat{\vartheta}_{\text{SPEC, CL}}$	0.346 (0.0234)	1.025 (0.4806)	0.144 (0.520)	1.61 (0.1846)

Figure 3.4: Estimated extremal coefficients based on $\hat{\lambda}_{MLE}^2$ against distance between the stations. Left panel: original locations with and without coast stations. Middle panel: transformed locations (only non-coast stations), extremal coefficient functions corresponding to the parameters in Table 3.1 are included. Right panel: comparison to madogram estimator.



three different estimators in (3.28) are shown in Figure 3.6. In agreement with the results of the simulation study, the multivariate estimator $\hat{\nu}_{SPEC}$ seems to be most reliable since the re-estimated extremal coefficient functions are close to the true value of the simulation. In contrast, the composite likelihood estimator $\hat{\nu}_{SPEC_CL}$ significantly underestimates the true degree of extremal dependence. This is probably a result of the false assumption of independence of bivariate densities which underlies the concept of composite likelihoods.

Figure 3.5: Level lines of the extremal coefficient function and realizations of the fitted Brown-Resnick process. Left: Without transformation. Right: After space transformation.

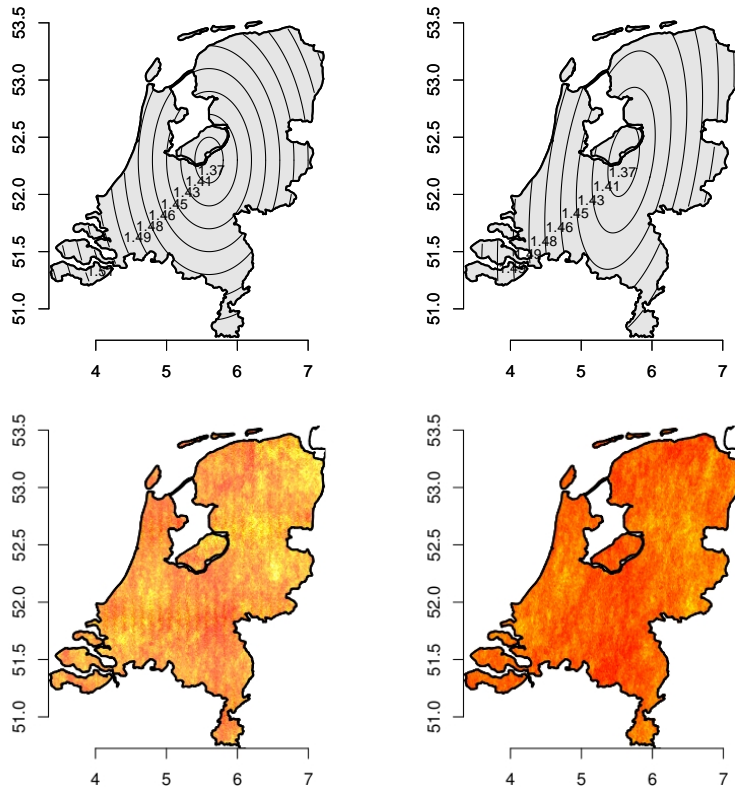
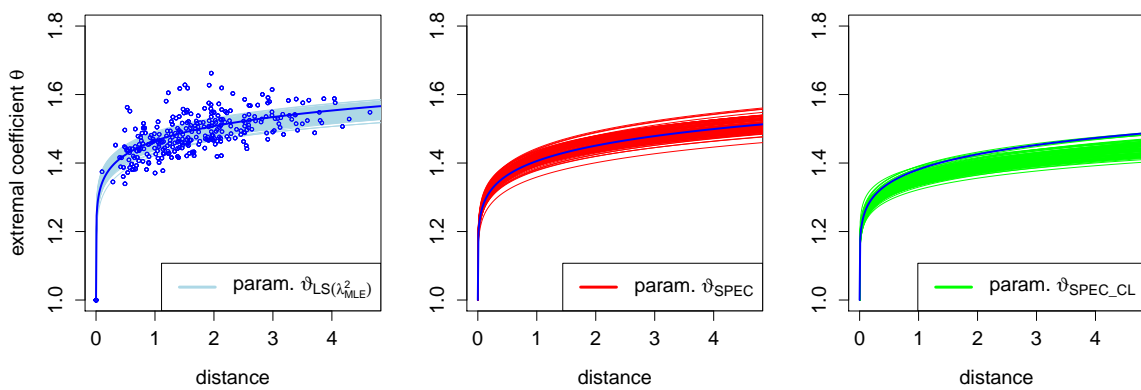


Figure 3.6: Validation of estimation: Fitted extremal coefficient functions for 100 simulations of 8172 Brown-Resnick processes on the transformed locations according to the estimated parameters.



3.4.2 Non-stationarity

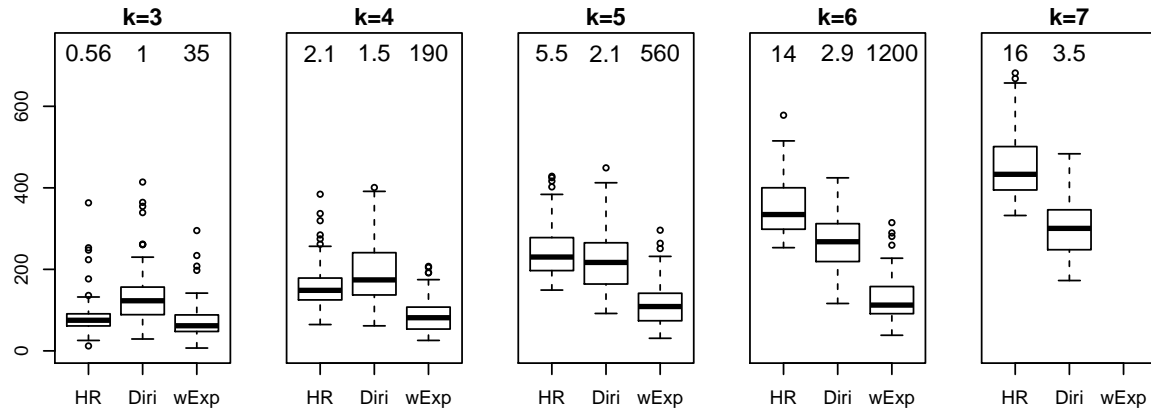
In the previous subsection we excluded the 10 coastal stations from fitting the stationary Brown-Resnick model since they exhibit a different extremal dependence structure than the 25 inland stations. Here we fit a multivariate Hüsler-Reiss distribution as extreme value model, which does not rely on any stationarity assumption. In particular, for T' being any subset of the 35 locations $x_1, \dots, x_{35} \in \mathcal{X}$, we estimate the $k(k-1)/2$ parameters of the dependence matrix $A \in \mathbb{R}^{k \times k}$, where $k = |T'|$. To this end, we can use any of the three newly proposed estimators \hat{A}_{Var} , \hat{A}_{MLE} and \hat{A}_{SPEC} . While \hat{A}_{Var} is given in explicit form and hence computationally very efficient and applicable to arbitrary dimensions, the latter two estimators require numerical optimization. Fortunately, the respective likelihood functions can be still evaluated much faster than most of the commonly used spectral density models. For the ML algorithm, \hat{A}_{Var} and, since the class of Hüsler-Reiss distributions is closed, also the lower-dimensional parameter estimates provide reasonable starting values.

In what follows, we use \hat{A}_{Var} as a starting value for the numerical optimization of \hat{A}_{SPEC} . We compare the likelihood values of the Hüsler-Reiss model fit to those of two other parametric models for spectral densities, namely the Dirichlet model (Coles & Tawn, 1991) and the weighted exponential model (Ballani & Schlather, 2011). The comparison is based on randomly drawing $k = 3, 4, 5, 6$ and 7 out of the 35 stations and fitting all three models. This is repeated 100 times. The weighted exponential model seems to fit worst for all $k \in \{3, 4, 5, 6\}$. Note that numerical optimization for this model involves a rather complicated likelihood and is extremely time-consuming. This is why the weighted exponential model is only included for $k \in \{3, 4, 5, 6\}$. The Hüsler-Reiss model seems to outperform the Dirichlet model for $k \geq 5$, which is not completely surprising since the Dirichlet model has only k parameters, while the Hüsler-Reiss model has $k(k-1)/2$ parameters encoding the extremal dependence. The results are summarized by Figure 3.7, which shows boxplots of the maximum likelihood values for each of the 100 choices of stations, and Table 3.2, which shows the percentage of cases in which the Hüsler-Reiss model outperforms the Dirichlet and the weighted exponential model.

Table 3.2: Fraction of cases in which the Hüsler-Reiss model outperforms the Dirichlet and the weighted exponential model.

number of stations	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$
$\mathbb{P}(L_{\text{HR}} > L_{\text{Diri}})$	0.10	0.25	0.73	1.00	0.99
$\mathbb{P}(L_{\text{HR}} > L_{\text{wExp}})$	0.79	0.96	0.93	1.00	–

Figure 3.7: Comparison of different spectral density models based on the maximized likelihood. The numbers above the boxes show the average computing time (in seconds) for the numerical maximization.



3.5 Discussion

This chapter presents several new estimators for the Hüsler-Reiss distribution and its spatial analog, the Brown-Resnick process. The methods are based on asymptotic conditional distributions of random variables in the MDA of the Hüsler-Reiss model. Within the framework of multivariate peaks-over-threshold, it is shown how conditioning on different extreme events leads to different estimators. In particular, the concept of extremal increments turns out to be fruitful, since for the Hüsler-Reiss model the increments conditioned on a fixed component being large are approximately multivariate Gaussian distributed. This enables very efficient inference even for high dimensions. The simulation study shows, that the proposed estimators perform well, both in terms of bias and variance. Especially for small data sets they outperform classical block methods. Moreover, the non-parametric, bivariate estimators are a suitable tool for exploratory data analysis (such as distinguishing between coastal and inland stations in Section 3.4.1), since they are computationally efficient and yield reliable results.

With regard to spatial extreme value statistics, one of the most promising models is the class of Brown-Resnick processes, due to their flexibility in connection with parametric families of variograms. The chapter provides several methods for parametric fitting of these models. Particularly the good performance of the multivariate spectral estimator suggests that using higher-dimensional densities better captures the shape of the underlying variogram than methods based on bivariate distributions only. Also for multivariate analysis of non-stationary data, the Hüsler-Reiss model is shown to be both well fitting and applicable in high-dimensions due to low computational costs of the estimators (Section 3.4.2).

While the simulation study in Section 3.3 already provides some empirical evidence for the consistency of the proposed estimators, a deeper analysis of the theoretical properties such as speed of convergence is left for future research. The main difficulty is to find appropriate

assumptions such that the conditional increments converge not only in distribution but also in L_1 or L_2 .

The idea of including all single extreme events into statistical inference, in connection with the concept of conditional increments, might also be applicable to other max-stable models such as mixed moving maxima processes.

Supplementary material

The raw data for the application can be downloaded from <http://www.knmi.nl>.

3.6 Proofs

Proof of Theorem 3. Note that

$$\begin{aligned} \mathbb{P}\left\{g(\tilde{\mathbf{X}}_n) \in B \mid \tilde{\mathbf{X}}_n \in A - \log u(n)\right\} &= \mathbb{P}\left\{g(\tilde{\mathbf{X}}_n + \log u(n)) \in B \mid \tilde{\mathbf{X}}_n \in A - \log u(n)\right\} \\ &= \frac{\mathbb{P}\left\{\tilde{\mathbf{X}}_n \in (g^{-1}(B) - \log u(n)) \cap (A - \log u(n))\right\}}{\mathbb{P}\left\{\tilde{\mathbf{X}}_n \in A - \log u(n)\right\}} \\ &= \frac{n/u(n)\mathbb{P}\left\{\mathbf{X} - \log(n/u(n)) \in g^{-1}(B) \cap A\right\}}{n/u(n)\mathbb{P}\left\{\mathbf{X} - \log(n/u(n)) \in A\right\}}. \end{aligned}$$

Thus, applying Prop. 5.17 in Resnick (2008), we obtain

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{g(\tilde{\mathbf{X}}_n) \in B \mid \tilde{\mathbf{X}}_n \in A - \log u(n)\right\} = \frac{\mu(g^{-1}(B) \cap A)}{\mu(A)},$$

and the measure $Q_{g,A}$ is given by

$$Q_{g,A}(B) = \frac{\mu(g^{-1}(B) \cap A)}{\mu(A)}, \quad B \in \mathcal{B}(S).$$

□

Proof of Theorem 4. 1. The density of the exponent measure μ of the Hüsler-Reiss distribution (3.5) is given by

$$\mu(dx_0, \dots, dx_k) = \frac{e^{-x_0}}{(2\pi)^{\frac{k}{2}} |\det \Sigma|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{y}^\top \Sigma^{-1} \mathbf{y}\right) dx_0 \dots dx_k, \quad x_0, \dots, x_k \in \mathbb{R},$$

where $\mathbf{y} = (x_1 - x_0 + 2\lambda_{1,0}^2, \dots, x_k - x_0 + 2\lambda_{k,0}^2)^\top$ and $\Sigma = \Psi_{k,(0,\dots,k)}(\lambda)$. For $\mathbf{s} = (s_1, \dots, s_k) \in \mathbb{R}^k$, $B_{\mathbf{s}} = \{\mathbf{y} \in \mathbb{R}^k : y_i \leq s_i, i = 1, \dots, k\}$, note that $g^{-1}(B_{\mathbf{s}}) = \{\mathbf{x} \in$

$\mathbb{R}^{k+1} : x_i - x_0 \leq s_i, i = 1, \dots, k$. Thus, for $A_1 = (0, \infty) \times \mathbb{R}^k$, we obtain

$$\begin{aligned} \mu(g^{-1}(B_{\mathbf{s}}) \cap A_1) &= \int_0^\infty e^{-x_0} \int_{-\infty}^{x_0+s_1} \dots \int_{-\infty}^{x_0+s_k} \frac{\exp\left(-\frac{1}{2}\mathbf{y}^\top \Sigma^{-1}\mathbf{y}\right)}{(2\pi)^{\frac{k}{2}} |\det \Sigma|^{1/2}} dx_k \dots dx_0 \\ &= \Phi_{M, \Sigma}(s_1, \dots, s_k), \end{aligned}$$

where $\Phi_{M, \Sigma}$ is the cumulative distribution function of a k -variate normal distribution with mean $M = (-2\lambda_{1,0}^2, \dots, -2\lambda_{k,0}^2)^\top$ and covariance matrix $\Sigma = \Psi_{k,(0,\dots,k)}(A)$. Since $\mu(A_1) = 1$ and the family of sets $\{B_{\mathbf{s}}, \mathbf{s} \in \mathbb{R}^k\}$ is a generator of the Borel σ -algebra $\mathcal{B}(\mathbb{R}^k)$ on \mathbb{R}^k , the first assertion follows from the proof of Theorem 3.

2. In the bivariate case the density of μ simplifies to

$$\mu(dx, dy) = \frac{e^{-x}}{2\lambda} \phi\left(\lambda + \frac{y-x}{2\lambda}\right) dx dy, \quad x, y \in \mathbb{R},$$

with $\lambda = \lambda_{0,1}$. We consider the set $A_2 = [-\infty, \mathbf{0}]^C \subset \mathbb{R}^2$ and note that $\mu(A_2) = 2\Phi(\lambda)$. It thus suffices to compute $\mu(g^{-1}(B_t) \cap A_2)$ for $t \in \mathbb{R}$. For $t < 0$ we have

$$\mu(g^{-1}(B_t) \cap A_2) = \int_0^\infty \int_{-\infty}^{x+t} \frac{e^{-x}}{2\lambda} \phi\left(\lambda + \frac{y-x}{2\lambda}\right) dy dx = \Phi\left(\lambda + \frac{t}{2\lambda}\right),$$

and similarly for $t > 0$,

$$\mu(g^{-1}(B_t) \cap A_2) = \mu(A_2) - \Phi\left(\lambda - \frac{t}{2\lambda}\right).$$

By the above considerations and the proof of Theorem 3 this yields

$$\begin{aligned} &\lim_{n \rightarrow \infty} \mathbb{P}\left\{X^{(1)} - X^{(0)} \leq t \mid \tilde{X}_n^{(0)} > \log u(n) \text{ or } \tilde{X}_n^{(1)} > \log u(n)\right\} \\ &= \begin{cases} \frac{1}{2\Phi(\lambda)} \Phi\left(\lambda + \frac{t}{2\lambda}\right) & \text{for } t < 0, \\ 1 - \frac{1}{2\Phi(\lambda)} \Phi\left(\lambda - \frac{t}{2\lambda}\right) & \text{for } t > 0. \end{cases} \end{aligned}$$

In other words, $X^{(1)} - X^{(0)}$ conditional on either $\tilde{X}_n^{(0)}$ or $\tilde{X}_n^{(1)}$ being large converges in distribution to some random variable Z with density

$$g_\lambda(t) = \frac{1}{4\lambda\Phi(\lambda)} \phi\left(\lambda - \frac{|t|}{2\lambda}\right), \quad t \in \mathbb{R}.$$

□

Proof of Proposition 1. By Theorem 1 in Coles & Tawn (1991) we can compute the spectral

density h as a derivative of the exponent measure $\nu(\mathbf{x}) = -\log G_\Lambda(\mathbf{x})$, namely

$$\frac{\partial \nu(\mathbf{x})}{\partial x_0 \dots \partial x_k} = - \left(\sum_{i=0}^k x_i \right)^{-k} h \left(\frac{x_0}{\sum_{i=0}^k x_i}, \dots, \frac{x_k}{\sum_{i=0}^k x_i}; \Lambda \right).$$

Since all but one summands of the exponent measure ν vanish, it suffices to evaluate

$$\frac{\partial}{\partial x_0 \dots \partial x_k} (-1)^k \int_{\log x_0} \int_{\log x_1 - z + 2\lambda_{1,0}^2} \dots \int_{\log x_k - z + 2\lambda_{k,0}^2} \phi(\mathbf{y}|\Sigma) dy_1 \dots dy_k e^{-z} dz, \quad (3.29)$$

where $\phi(\cdot|\Sigma)$ is the density function of the k -dimensional normal distribution with covariance matrix Σ . Carrying out this computation yields (3.24). \square

4 A generalization of the Brown-Resnick process in terms of Lévy processes

We recall definition (1.3) of the original process in [Brown & Resnick \(1977\)](#) as

$$\max_{i \in \mathbb{N}} [X_i + W_i(t) - |t|/2], \quad t \in \mathbb{R}, \quad (4.1)$$

where $W_i, i \in \mathbb{N}$, are independent copies of a standard Brownian motion on \mathbb{R} and $\sum_{i \in \mathbb{N}} \delta_{X_i}$ is an independent Gumbel point process. They showed that the max-stable process (4.1) is stationary and that it is the limit of pointwise maxima of independent Gaussian Ornstein-Uhlenbeck processes.

The Brownian motion is a Markov process with stationary and independent increments. A natural approach is to generalize (4.1) to the class of all such stochastic processes which have stationary and independent increments, namely Lévy processes (cf. [Bertoin \(1996\)](#); [Sato \(1999\)](#)). In fact, [Stoev \(2008\)](#) introduced the process

$$\eta^+(t) = \max_{i \in \mathbb{N}} [X_i + L_i^+(t) - t\psi^+(1)], \quad t \geq 0, \quad (4.2)$$

where $\sum_{i \in \mathbb{N}} \delta_{X_i}$ is as above and $L_i^+, i \in \mathbb{N}$, are independent copies of a Lévy process L^+ with finite exponential moments and $\psi^+(1) = \log \mathbb{E}e^{L^+(1)}$. They showed that η^+ is max-stable and stationary on $[0, \infty)$. However, somehow counter-intuitively, their results cannot be generalized by simply reflecting the Lévy processes at the y -axis (as it is seemingly done with the Brownian motion in (4.1)). One can easily show that this would not yield a stationary process on \mathbb{R} . Moreover, only few theoretical properties of the process η^+ are known. For instance, it was asked in [Stoev \(2008\)](#) whether the one-sided process in (4.2) has a representation as a mixed moving maxima (MMM) process.

In this chapter we give a more complete definition of Brown-Resnick processes based on Lévy processes not only on the whole real line but also in terms of random fields on \mathbb{R}^d . By analyzing their properties we aim at providing a deeper understanding of the two-sided version of (4.2), also illuminating the construction (4.1).

By a suitable change of measure, in Section 4.1 we introduce an exponentially transformed Lévy process L^- , which already figured in [Kablichko \(2011b\)](#) in the construction of a certain α -stable process. This process will give rise to an extension η^- of the process η^+ on the negative real axis. Several properties of the combined process η on the whole real line are derived and, specifically, the question on a MMM representation is answered. Section 4.2 is concerned with the convergence of maxima of i.i.d. realizations of certain stochastic processes to η . In particular, extending the result on the Gaussian Ornstein-Uhlenbeck process in [Brown & Resnick \(1977\)](#), we present a similar result for α -stable Ornstein-Uhlenbeck processes.

In Section 4.3 we use the one-dimensional results to extend the process η to d -dimensional random fields. To this end, the notion of linearly additive, infinitely divisible random fields introduced by Mori (1992) is crucial. Some ideas on statistical inference for the max-stable process η are given in Section 4.4.

4.1 The one-dimensional case

Throughout this section, $\{L^+(t) : t \geq 0\}$ will be a Lévy process, i.e. a stochastic process with stationary and independent increments such that $L^+(0) = 0$ holds a.s. and L^+ already is a version which has càdlàg paths (right-continuous with finite left limits). Denote by $(0, \sigma^2, \nu^+)$ the characteristic triplet of L^+ , where σ^2 is the variance of the Brownian component and ν^+ is the Lévy measure. The drift is irrelevant and thus assumed to equal 0. Further, let L^+ satisfy the moment condition

$$\Psi^+(1) = \log \mathbb{E}e^{L^+(1)} < \infty. \quad (4.3)$$

The natural space for the sample paths of a Lévy process is the Skorokhod space of càdlàg paths. For an interval $I \subset \mathbb{R}$ it is defined as

$$D_{\mathbb{R}}(I) = \{f : I \rightarrow \mathbb{R}, f \text{ has càdlàg sample paths}\},$$

which becomes a Polish space when equipped with the Skorokhod metric (cf. Ethier & Kurtz (1986)). By $\mathcal{B}(D_{\mathbb{R}}(I))$ we will denote the corresponding Borel σ -algebra on $D_{\mathbb{R}}(I)$. For instance, with this notation, L^+ is a random element in the Skorokhod space $D_{\mathbb{R}}([0, \infty))$.

4.1.1 Definition of the process L^-

With the Lévy process L^+ above, put $Z^+(t) = L^+(t) - t\Psi^+(1)$, such that $\mathbb{E}e^{Z^+(t)} = 1$ for $t \geq 0$. We can now define a transformed Lévy process $\{Z^-(t) : t \geq 0\}$ that arises from an exponential change of measure (cf. Kabluchko (2011b)). Since Z^+ is a Markov process, say on the filtered space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \geq 0})$, there is a corresponding semigroup of transition kernels $(\mu_t^+)_{t \geq 0}$ given by

$$\mathbb{P}\left(Z^+(t) \in B | \mathcal{F}_s\right) = \mu_{t-s}^+(Z^+(s), B) = \mu_{t-s}^+(0, B - Z^+(s)), \quad 0 \leq s < t,$$

for any $B \in \mathcal{B}(\mathbb{R})$, where the second equality holds since Z^+ is a Lévy process. For details on Markov processes see Chapter 8 in Kallenberg (2002). The relation

$$\mu_t^-(x, B) = \mathbb{E}^x \mathbf{1}_B(Z^+(t)) e^{Z^+(t)-x} = \int_{\mathbb{R}} \mathbf{1}_B(y) e^{y-x} \mu_t^+(x, dy), \quad x \in \mathbb{R}, B \in \mathcal{B}(\mathbb{R}), \quad (4.4)$$

defines another Markov process $(Z^-(t))_{t \geq 0}$ with transition kernels $(\mu_t^-)_{t \geq 0}$. Indeed, it is easy to verify that the Chapman-Kolmogorov relation $\mu_{t+s}^- = \mu_t^- \mu_s^-$ is satisfied and, moreover,

that this semigroup is space homogeneous, since

$$\mu_t^-(x, B) = \mathbb{E}^x \mathbf{1}_B(Z^+(t)) e^{Z^+(t)-x} = \mathbb{E}^0 \mathbf{1}_B(Z^+(t) + x) e^{Z^+(t)} = \mu_t^-(0, B - x). \quad (4.5)$$

Thus, $(Z^-(t))_{t \geq 0}$ is in fact a Lévy process.

One can check that (4.5) implies that $-Z^-(t) = L^-(t) - t\Psi^-(1)$, $t \geq 0$, where L^- is a Lévy process with characteristic triplet $(0, \sigma^2, \nu^-)$, with $\nu^-(dx) = e^{-x}\nu^+(-dx)$, and $\Psi^-(1) = \log \mathbb{E} e^{L^-(1)}$. Note that $\Psi^-(1)$ is well-defined since L^- satisfies the exponential moment condition (4.3). We define the combined two-sided process $\{Z(t) : t \in \mathbb{R}\}$ with sample paths in $D_{\mathbb{R}}(\mathbb{R})$ by

$$Z(t) = \begin{cases} L^+(t) - t\Psi^+(1) & \text{for } t \geq 0, \\ L^-((-t)_-) - (-t)\Psi^-(1) & \text{for } t < 0, \end{cases} \quad (4.6)$$

where L^+ and L^- are assumed to be independent and $(u)_-$ denotes the left limit at $u > 0$.

Remark 6. For instance, if L^+ is a standard Brownian motion $\{W^+(t) : t \geq 0\}$, i.e. $Z^+(t) = W^+(t) - t/2$, $t \geq 0$, then $Z^-(t) \stackrel{d}{=} W^+(t) + t/2$, $t \geq 0$ and thus

$$Z(t) = W(t) - |t|/2, \quad t \in \mathbb{R},$$

for a standard Brownian motion W on \mathbb{R} . This shows that in the construction (4.1), the Brownian motion is not “mirrored” but in fact transformed by an exponential change of measure to obtain the process W^- on the negative real axis.

Remark 7. Note that the process Z is not a Lévy process on the real axis. It still has independent increments, however, they are no longer stationary since $\nu^-(-dx) = e^x\nu^+(dx) \neq \nu^+(dx)$, unless $\nu \equiv 0$.

Remark 8. For any Lévy process L^+ satisfying the moment condition (4.3), the associated process Z has the property

$$\lim_{|t| \rightarrow \infty} Z(t) = -\infty, \quad \text{almost surely.}$$

This follows from $\mathbb{E} e^{Z(t)} = 1$, $t \in \mathbb{R}$, and Exercise VI.4 in Bertoin (1996)

Remark 9. Suppose L^+ is a spectrally negative Lévy process, i.e. the support of its Lévy measure is contained in $(-\infty, 0]$ (cf. Chapter VII in Bertoin (1996) for details). Clearly, L^+ satisfies the moment condition (4.3). In this case there is a concrete interpretation of the associated process Z^- obtained by the exponential change of measure (4.4) applied to Z^+ . In fact, Lemma VII.7 in Bertoin (1996) states that for any $t \geq 0$ and $A \in \mathcal{F}_t$

$$\mathbb{P}_{Z^-}(A) = \lim_{x \rightarrow \infty} \mathbb{P}_{Z^+}(A \mid \sup_{t \geq 0} Z^+(t) > x),$$

where \mathbb{P}_{Z^-} is the law of Z^- . In other words, Z^- is nothing else than the process Z^+ (which drifts to $-\infty$) conditioned to drift to ∞ .

4.1.2 The Brown-Resnick process η based on Z

We are now in a position to extend the construction in (4.2) to the whole real line and define a two-sided version of the Brown-Resnick process based on Lévy processes. Here, we assume that $\sigma^2 = 0$, since the Gaussian case has already been studied in Brown & Resnick (1977). In fact, the assertion of this theorem is a consequence of Prop. 2.8 in Kabluchko (2009a), but we give a more elementary proof to avoid the technical methodology.

Theorem 5. *Let $(Z(t))_{t \in \mathbb{R}}$ be defined as above based on L^+ satisfying the moment condition (4.3). Further, let $\sum_{i=1}^{\infty} \delta_{X_i}$ be a Poisson point process with intensity $e^{-x} dx$ and let Z_i be independent copies of Z , $i \in \mathbb{N}$. Then, the process*

$$\eta(t) = \max_{i \in \mathbb{N}} (X_i + Z_i(t)), \quad t \in \mathbb{R}, \quad (4.7)$$

is max-stable and stationary on \mathbb{R} .

Proof. The max-stability follows directly from the construction. First note that by Theorem 25.17 in Sato (1999), $\mathbb{E}e^{uZ(t)} = 1$ implies that $\mathbb{E}e^{uZ(t)} < \infty$, for all $u \in [0,1]$ and $t \in \mathbb{R}$. We apply the stationarity criterion in Prop. 6 in Kabluchko *et al.* (2009). To this end, let $n \in \mathbb{N}$, $t_1 \leq \dots \leq t_j \leq 0 \leq t_{j+1} \leq \dots \leq t_n$ and $u_1, \dots, u_n \in [0,1]$ such that $\sum_{i=1}^n u_i = 1$. Since $(Z(t_1), \dots, Z(t_j))$ and $(Z(t_{j+1}), \dots, Z(t_n))$ are independent by the definition of Z , we compute their Laplace exponents separately. Here, let $\mathbb{R}_0 = \mathbb{R} \setminus \{0\}$. Using the definition of Z^- in (4.5) we obtain

$$\begin{aligned} & \log \mathbb{E} \exp (u_1 Z(t_1) + \dots + u_j Z(t_j)) \\ &= \log \mathbb{E} \exp \left(-u_1 (Z^-((-t_1)_-) - Z^-((-t_2)_-)) - (u_1 + u_2) (Z^-((-t_2)_-) - Z^-((-t_3)_-)) - \dots - (u_1 + \dots + u_j) Z^-((-t_j)_-) \right) \\ &= (t_2 - t_1) \int_{\mathbb{R}_0} e^{(1-u_1)x} - 1 - (1-u_1)x \mathbf{1}_{|x| \leq 1} \nu^+(dx) + \\ & \quad \dots + (-t_j) \int_{\mathbb{R}_0} e^{(1-(u_1+\dots+u_j))x} - 1 - (1-(u_1+\dots+u_j))x \mathbf{1}_{|x| \leq 1} \nu^+(dx) \\ & \quad - \left(\sum_{i=1}^{j-1} (t_{i+1} - t_i) (1 - (u_1 + \dots + u_i)) - t_j (1 - (u_1 + \dots + u_j)) \right) \Psi^+(1) \\ &= (t_2 - t_1) \int_{\mathbb{R}_0} e^{(u_2+\dots+u_n)x} - 1 - (u_2 + \dots + u_n)x \mathbf{1}_{|x| \leq 1} \nu^+(dx) + \\ & \quad \dots + (-t_j) \int_{\mathbb{R}_0} e^{(u_{j+1}+\dots+u_n)x} - 1 - (u_{j+1} + \dots + u_n)x \mathbf{1}_{|x| \leq 1} \nu^+(dx) \\ & \quad + \left(\sum_{i=1}^n t_1 u_i - \sum_{i=1}^j t_i u_i \right) \Psi^+(1), \end{aligned}$$

where we used $\sum_{i=1}^n u_i = 1$. Furthermore,

$$\begin{aligned}
& \log \mathbb{E} \exp \left(u_{j+1}(L^+(t_{j+1}) - \Psi^+(t_{j+1})) + \dots + u_n(L^+(t_n) - \Psi^+(t_n)) \right) \\
&= \log \mathbb{E} \exp \left(u_n(L^+(t_n) - L^+(t_{n-1})) + (u_n + u_{n-1})(L^+(t_{n-1}) - L^+(t_{n-2})) + \right. \\
&\quad \left. \dots + (u_n + \dots + u_{j+1})L^+(t_{j+1}) - \sum_{i=j+1}^n t_i u_i \Psi^+(1) \right) \\
&= (t_n - t_{n-1}) \int_{\mathbb{R}_0} e^{u_n x} - 1 - u_n x \mathbf{1}_{|x| \leq 1} \nu(dx) + \\
&\quad \dots + t_{j+1} \int_{\mathbb{R}_0} e^{(u_{j+1} + \dots + u_n)x} - 1 - (u_{j+1} + \dots + u_n)x \mathbf{1}_{|x| \leq 1} \nu^+(dx) \\
&\quad - \sum_{i=j+1}^n t_i u_i \Psi^+(1).
\end{aligned}$$

Putting the parts together yields

$$\begin{aligned}
& \log \phi_{t_1, \dots, t_n}(u_1, \dots, u_n) = \log \mathbb{E} \exp (u_1 Z(t_1) + \dots + u_n Z(t_n)) \\
&= (t_n - t_{n-1}) \int_{\mathbb{R}_0} e^{u_n x} - 1 - u_n x \mathbf{1}_{|x| \leq 1} \nu^+(dx) + \\
&\quad \dots + (t_2 - t_1) \int_{\mathbb{R}_0} e^{(u_2 + \dots + u_n)x} - 1 - (u_2 + \dots + u_n)x \mathbf{1}_{|x| \leq 1} \nu^+(dx) \\
&\quad + \sum_{i=2}^n (t_1 - t_i) u_i \Psi^+(1).
\end{aligned}$$

This expression does no longer depend on the assumption that t_1, \dots, t_j are negative and t_{j+1}, \dots, t_n are positive but is true for any $t_1, \dots, t_n \in \mathbb{R}$. Since it only depends on differences of two locations, we clearly have for any $h \in \mathbb{R}$ that

$$\phi_{t_1+h, \dots, t_n+h}(u_1, \dots, u_n) = \phi_{t_1, \dots, t_n}(u_1, \dots, u_n).$$

This concludes the proof. \square

Similar to the Gaussian case, the next result shows that the paths of the Brown-Resnick process look essentially like those of the underlying Lévy processes.

Proposition 2. *For any compact interval $K \subset \mathbb{R}$ only finitely many paths contribute to the process $\{\eta(t) : t \in K\}$ from Theorem 5 almost surely. In particular, η is again a process with càdlàg sample paths.*

Proof. Without loss of generality put $K = [-T, T]$, for some $T > 0$. Let $\sum_{i=1}^{\infty} \delta_{(X_i, Z_i)}$ be a Poisson point process with intensity $e^{-x} dx \mathbb{P}_Z$ and suppose that the processes Z and Z_i , $i \in \mathbb{N}$, are as in Theorem 5. For $k \in \mathbb{Z}$ define the random number

$$I_k = \#\{i \in \mathbb{N} : \sup_{t \in [-T, T]} (X_i + Z_i(t)) > k\},$$

which is Poisson distributed with parameter $\lambda_k \in [0, \infty]$. Observe that

$$\begin{aligned} \lambda_k &\leq e^{-k} \int_{\mathbb{R}} e^z \left(\mathbb{P} \left[\sup_{t \in [-T, 0]} Z(t) > z \right] + \mathbb{P} \left[\sup_{t \in [0, T]} Z(t) > z \right] \right) dz \\ &\leq e^{-k} \left(\mathbb{E} \exp \left\{ \sup_{t \in [-T, 0]} Z(t) \vee 0 \right\} + \mathbb{E} \exp \left\{ \sup_{t \in [0, T]} Z(t) \vee 0 \right\} \right) < \infty, \end{aligned}$$

where for $a, b \in \mathbb{R}$ we put $a \vee b = \max(a, b)$. For the last inequality we used the assumption that the exponential moments of Z exist together with Theorem 25.18 of [Sato \(1999\)](#) applied to the sub-multiplicative function $x \mapsto \exp(x \vee 0)$. Thus, I_k is almost surely finite.

Moreover, η is almost surely càdlàg as the maximum of finitely many càdlàg functions. \square

Recall the definition of the extremal correlation function from Chapter 1 as a measure of extremal dependence. [Kabluchko & Schlather \(2010\)](#) showed that a process with extremal correlation function ρ is mixing if and only if

$$\lim_{t \rightarrow \infty} \rho(t) = 0.$$

Proposition 3. *Let the processes η and Z be as in Theorem 5. The extremal correlation function of η is given by*

$$\rho(t) = 2 - \mathbb{E} \left[1 \vee e^{Z(t)} \right], \quad t \in \mathbb{R},$$

and, moreover, η is mixing.

Proof. Recall from Remark 8 that $\lim_{|t| \rightarrow \infty} Z(t) = -\infty$, almost surely. The first assertion follows by a simple calculation. For the second claim we use $\mathbb{E} e^{Z(t)} = 1$ for all $t \geq 0$ and dominated convergence to obtain

$$\mathbb{E} \left[1 \vee e^{Z(t)} \right] = \mathbb{P}(Z(t) < 0) + 1 - \mathbb{E} \mathbf{1}_{Z(t) < 0} e^{Z(t)} \rightarrow 2,$$

for $t \rightarrow \infty$. \square

The mixing property was already shown in [Stoev \(2008\)](#) for the one-sided process (4.2). However, an open question therein was whether the stationary process in (4.2) has a representation as a MMM process as defined in (1.4). The following theorem answers this question.

Theorem 6. *Let the processes η and Z be as in Theorem 5. The Brown-Resnick process η admits a MMM representation.*

Proof. Note that the Borel σ -algebra on $D_{\mathbb{R}}(\mathbb{R})$ is induced by the projections $\pi_s : D \rightarrow \mathbb{R}$, $f \mapsto f(s)$, for $s \in \mathbb{Q}$ (cf. Theorem 14.5 in [Billingsley \(1968\)](#)). The subset D^{\downarrow} of functions in

$D_{\mathbb{R}}(\mathbb{R})$ that tend to negative infinity for $|t| \rightarrow \infty$ is measurable. Indeed, we note that

$$D^{\downarrow} = \bigcap_{n \in \mathbb{N}} \bigcup_{m \in \mathbb{N}} \bigcap_{\substack{s \in \mathbb{Q} \\ |s| \geq m}} \left\{ \pi_s^{-1}((-\infty, -n]) \right\},$$

by the right continuity of paths in $D_{\mathbb{R}}(\mathbb{R})$. Furthermore, Remark 8 implies that $\mathbb{P}_Z(D^{\downarrow}) = 1$, where P_Z denotes the law of Z on $D_{\mathbb{R}}(\mathbb{R})$. Let $N_Z^* = \sum_{i=1}^{\infty} \delta_{(X_i, Z_i)}$ be a Poisson point process on $\mathbb{R} \times D^{\downarrow}$ with intensity $e^{-x} dx \mathbb{P}_Z$. We will show below that the mapping

$$\begin{aligned} \varphi : \mathbb{R} \times D^{\downarrow} &\longrightarrow \mathbb{R} \times \mathbb{R} \times D^{\downarrow} \\ (x, f) &\mapsto \left(x + \sup_{s \in \mathbb{R}} f(s), \arg \sup f, f(\cdot + \arg \sup f) - \sup_{s \in \mathbb{R}} f(s) \right) \end{aligned} \quad (4.8)$$

is measurable. Applied to the points of N_Z^* it then induces a Poisson point process

$$\sum_{i=1}^{\infty} \delta_{(X_i + \sup_{s \in \mathbb{R}} Z_i(s), \arg \sup Z_i, Z_i(\cdot + \arg \sup Z_i) - \sup_{s \in \mathbb{R}} Z_i(s))},$$

on $\mathbb{R} \times \mathbb{R} \times D^{\downarrow}$ with a certain intensity measure Γ . Using the same estimate as in the proof of Prop. 2 we conclude that

$$\Gamma([0,1] \times [0,1] \times D^{\downarrow}) \leq \int_{\mathbb{R}} e^x \mathbb{P} \left(\sup_{t \in [0,1]} Z(t) > x \right) dx < \infty.$$

This fact and the arguments in the proof of Theorem 14 in [Kabluchko et al. \(2009\)](#) show that Γ factorizes and is given by $c e^{-x} dx dt d\mathbb{Q}$, where c is a finite constant (depending on Z) and \mathbb{Q} is a probability measure on D^{\downarrow} . This gives the required MMM representation in (1.4).

For the measurability of φ in (4.8) we first note that due to the right continuity and the fact that the supremum of a function $f \in D^{\downarrow}$ is finite we have

$$\sup_{s \in \mathbb{R}} f(s) = \sup_{s \in \mathbb{Q}} f(s) = \sup_{s \in \mathbb{Q}} \pi_s(f),$$

and thus, \sup is measurable on D^{\downarrow} as the supremum of countably many measurable functions.

Furthermore we define the $\arg \sup$ to be the supremum of those points where the supremum is attained and those where it is approached arbitrarily close. More precisely, let

$$\arg \sup : D^{\downarrow} \longrightarrow \mathbb{R}, \quad f \mapsto \lim_{n \rightarrow \infty} \sup_{s \in \mathbb{Q}} h_s^n(f),$$

where $h_s^n : D^{\downarrow} \rightarrow \mathbb{R}$ for $s \in \mathbb{Q}$ and $n \in \mathbb{N}$ is defined by $f \mapsto s \mathbf{1}_{\{\sup f - f(s) < 1/n\}}$. Since the h_s^n are measurable, so is $\arg \sup$.

For the measurability of the translation by the $\arg \sup$ we note that the mapping $\gamma : D^{\downarrow} \times \mathbb{R} \rightarrow D^{\downarrow}$, $(f, t) \mapsto f(\cdot + t)$ is measurable thus also the concatenation of measurable

functions

$$\begin{aligned} D^\downarrow &\longrightarrow \mathbb{R} \times D^\downarrow \xrightarrow{\gamma} D^\downarrow \\ f &\mapsto (\arg \sup f, f) \mapsto f(\cdot + \arg \sup f). \end{aligned}$$

This establishes the measurability of φ and concludes the proof. \square

Remark 10. *In general it is not possible to explicitly compute the distribution \mathbb{Q} of the process F in the MMM representation (1.4) for general processes η . However, in the special case where Z is a standard Brownian motion with drift, Engelke et al. (2011) showed that \mathbb{Q} is the law of a well-known diffusion process, namely a Brownian motion with drift conditioned to stay negative (cf. Chapter 2).*

4.2 Convergence results

In this section we show that the stationary processes η defined in (4.7) arises as the max-limit of certain suitably rescaled and normalized stochastic processes. For the sake of simplicity, we restrict our analysis to processes on $[-1,1]$, the case $(-\infty, \infty)$ being similar.

Denote by $M_p(E)$ the space of all Radon point measures on the complete, separable and second countable metric space $(E, \mathcal{B}(E))$. $M_p(E)$ is equipped with the σ -algebra $\mathcal{M}_p(E)$ induced by the family of all projections $\pi_A : M_p(E) \rightarrow [0, \infty]$, $\tau \mapsto \tau(A)$, for $A \in \mathcal{B}(E)$.

Further, for a stochastic process $\{R(t) : t \in T\}$ on an index set T and $\mathbf{t} = (t_1, \dots, t_m) \in T^m$, $m \in \mathbb{N}$, we use the shorthand notation $R(\mathbf{t}) = (R(t_1), \dots, R(t_m))$.

4.2.1 Extremes of independent Lévy processes

Motivated by the results in Kabluchko (2009a) for *sums* of independent geometric Lévy processes, in this section we consider the *extremes* in a similar situation. Suppose that L_i , $i \in \mathbb{N}$, are independent copies of a non-deterministic Lévy process $\{L(t) : t \geq 0\}$ satisfying the moment condition

$$\Psi(u) = \log \mathbb{E} e^{uL(1)} < \infty, \text{ for all } u \in \mathbb{R},$$

and suppose that $L(1)$ is non-lattice. The function Ψ is strictly convex and infinitely differentiable with $\Psi(0) = 0$. Let λ be given by $\lambda = I(\Psi'(1)) = \Psi'(1) - \Psi(1)$, where I is the Legendre-Fenchel transform of Ψ (for details see Kabluchko (2009a)), and suppose that s_n , $n \in \mathbb{N}$, is a sequence of non-negative real numbers such that

$$\lim_{n \rightarrow \infty} \frac{\log n}{s_n} = \lambda.$$

Further, put $Z_i(t) = L_i(t) - t\Psi(1)$, $t \in \mathbb{R}$, $i \in \mathbb{N}$, and define normalizing constants by

$$b_n = s_n I^{-1} \left(\frac{\log n - \log(\sqrt{2\pi\Psi''(1)s_n})}{s_n} \right), \quad n \in \mathbb{N}.$$

Theorem 7. *As $n \rightarrow \infty$, we have the following convergence*

$$\max_{i=1,\dots,n} Z_i(s_n + t) - b_n \xrightarrow{d} \eta(t), \quad t \in [-1,1],$$

in the sense of finite dimensional distributions, where η is the process in (4.7).

Proof. Let us first consider the one-sided case on $[0,1]$. Note that by the Lévy property $Z_i(s_n + \cdot) - Z_i(s_n)$ is independent of $Z_i(s_n)$ and it has the same distribution as $Z_i(\cdot)$, $i \in \mathbb{N}$. Thus, for $u \in \mathbb{R}$ and $A \in \mathcal{B}(D_{\mathbb{R}}[0,1])$ we have the convergence

$$\begin{aligned} n\mathbb{P}((Z_1(s_n) - b_n, Z_1(s_n + \cdot) - Z_1(s_n)) \in (u, \infty) \times A) \\ = n\mathbb{P}(Z_1(s_n) - b_n \in (u, \infty)) \cdot \mathbb{P}(Z_1(\cdot) \in A) \rightarrow e^{-u}\mathbb{P}(Z_1(\cdot) \in A), \end{aligned}$$

as $n \rightarrow \infty$, since $n\mathbb{P}(Z_1(s_n) - b_n \in (u, \infty)) \rightarrow e^{-u}$ by Lemma 10.3 in [Kabluchko \(2011b\)](#). Thus, since weak convergence on $D_{\mathbb{R}}[0,1]$ implies convergence of finite dimensional distributions, it follows by Prop. 3.21 in [Resnick \(2008\)](#) that for any $m \in \mathbb{N}$ and $\mathbf{t} = (t_1, \dots, t_m) \in [0,1]^m$, the point processes

$$M_n^* = \sum_{i=1}^n \delta_{(Z_i(s_n) - b_n, Z_i(s_n + \mathbf{t}) - Z_i(s_n))},$$

converges weakly on $\mathbb{R} \times \mathbb{R}^m$ to the PPP $N_{Z(\mathbf{t})}^* = \sum_{i=1}^{\infty} \delta_{(X_i, Z_i(\mathbf{t}))}$ with intensity $e^{-x} dx \mathbb{P}_{Z(\mathbf{t})}$. We observe that with $\mathcal{M} = \{\psi \in M_p(\mathbb{R} \times \mathbb{R}^m) : \psi(\{\mathbf{x} \in \mathbb{R}^{m+1} : \mathbf{x} \leq \mathbf{0}\}^C) < \infty\} \in \mathcal{M}_p(\mathbb{R} \times \mathbb{R}^m)$ we have $\mathbb{P}(N_{Z(\mathbf{t})}^* \in \mathcal{M}) = 1$, where A^C denotes the complement of the set A . Applying the continuous map

$$\kappa : \mathcal{M} \rightarrow \mathbb{R}^m, \quad \sum_{i \in \mathbb{N}} \delta_{(y_i, \mathbf{s}_i)} \mapsto \max_{i \in \mathbb{N}} y_i + \mathbf{s}_i \quad (4.9)$$

to M_n^* , together with the continuous mapping theorem, yields the one-sided version of the theorem. Here, the maxima in (4.9) are to be understood componentwise. The two-sided result can now be deduced from the one-sided version following the lines of Section 10.6 in [Kabluchko \(2009a\)](#). \square

4.2.2 Ornstein-Uhlenbeck processes driven by totally skewed α -stable Lévy processes

In this section we show that η also arises as the max-limit of independent Ornstein-Uhlenbeck processes which are driven by totally skewed α -stable Lévy processes, for $\alpha \in (1,2)$. [Brown & Resnick \(1977\)](#) proved a similar result for Gaussian ($\alpha = 2$) Ornstein-Uhlenbeck processes attracted by the original Brown-Resnick process.

Totally skewed α -stable random variables

Let us first recall some definitions on α -stable processes (cf. [Samorodnitsky & Taqqu \(1994\)](#)). A real-valued random variable X is said to have an α -stable distribution $S_{\alpha}(\sigma, \beta, \mu)$ with

parameters $\alpha \in (1,2)$, $\sigma \geq 0$, $\beta \in [-1,1]$ and $\mu \in \mathbb{R}$ if its characteristic function has the form

$$\mathbb{E} \exp(i\theta X) = \exp \{-\sigma^\alpha |\theta|^\alpha (1 - i\beta \text{sign}(\theta) \tan(\pi\alpha/2)) + i\mu\theta\}, \theta \in \mathbb{R}.$$

In general, α -stable distributions possess heavy polynomial tails and are thus not in the max-domain of attraction of the Gumbel distribution. An exception, which we will focus on, is the case of α -stable random variables that are totally skewed to the left, i.e. $\beta = -1$. In fact, if X is $S_\alpha(1, -1, 0)$ distributed with $\alpha \in (1,2)$, then it has a positive density on the whole real line $(-\infty, \infty)$ (unlike the case $\alpha < 1$, where the mass is concentrated on $(-\infty, 0]$). In this case, its upper tail $\bar{F} = 1 - F$ is light and satisfies as $x \rightarrow \infty$

$$\bar{F}(x) = \mathbb{P}(X > x) \sim (\alpha - 1)A_\alpha x^{-\alpha/(2(\alpha-1))} \exp\{-B_\alpha x^{\alpha/(\alpha-1)}\},$$

for some positive constants

$$\begin{aligned} A_\alpha &= [2\pi(\alpha - 1)^3]^{-1/2} [|\alpha| |\cos(\alpha\pi/2)|]^{1/(2\alpha-2)}, \\ B_\alpha &= (\alpha - 1)\alpha^{-\alpha/(\alpha-1)} |\cos(\alpha\pi/2)|^{1/(\alpha-1)}, \end{aligned}$$

(cf. Albin (1993), Samorodnitsky & Taqqu (1994, Eq. 1.2.11)). Similarly, for the density f we have

$$f(x) \sim \alpha A_\alpha B_\alpha x^{(2-\alpha)/(2(\alpha-1))} \exp\{-B_\alpha x^{\alpha/(\alpha-1)}\},$$

as $x \rightarrow \infty$. In particular, note that X satisfies the moment condition (4.3), since $\alpha/(\alpha-1) > 2$, for all $\alpha \in (1,2)$. Thus, we can rewrite the tail as

$$\bar{F}(x) = c \exp \left\{ - \int_z^x \frac{g(t)}{a(t)} dt \right\}, \quad z < x < \infty,$$

for some $c, z \in \mathbb{R}$, a measurable function g satisfying $\lim_{t \rightarrow \infty} g(t) = 1$ and a positive, absolutely continuous function a with density a' having $\lim_{t \rightarrow \infty} a'(t) = 0$. More precisely, by Theorem 3.3.26 in Embrechts *et al.* (1997) the function a can be chosen as

$$a(t) = \bar{F}(t)/f(t) = \frac{\alpha - 1}{\alpha B_\alpha} t^{-1/(\alpha-1)}, \quad t \in \mathbb{R}.$$

Suppose that $X_i, i \in \mathbb{N}$, are independent copies of X . Then, with $d_n = F^{\leftarrow}(1 - n^{-1})$ and $c_n = a(d_n)$ it holds that (cf. Theorem 3.3.26 in Embrechts *et al.* (1997))

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(c_n^{-1} \left(\max_{i=1, \dots, n} X_i - d_n \right) \leq x \right) = \exp(\exp(-x)), x \in \mathbb{R}, \quad (4.10)$$

where F^{\leftarrow} is the pseudo-inverse of F . Note in particular, that $d_n \rightarrow \infty$ and thus $c_n \rightarrow 0$ as n tends to ∞ .

Lévy Ornstein-Uhlenbeck processes

To any totally skewed α -stable random variable X with distribution $S_\alpha(1, -1, 0)$, $\alpha \in (1, 2)$, there is a Lévy process $\{L_\alpha(t) : t \geq 0\}$ such that $L_\alpha(1) \stackrel{d}{=} X$. It is easy to see that the process L_α is $1/\alpha$ -self-similar, i.e., the processes $\{L_\alpha(ct) : t \geq 0\}$ and $\{c^{1/\alpha}L_\alpha(t) : t \geq 0\}$ have the same finite dimensional distributions (cf. [Samorodnitsky & Taquu \(1994, Section 3.1\)](#)).

Based on this, we define the associated Ornstein-Uhlenbeck process $\{Y_\alpha(t) : -\infty < t < \infty\}$ by

$$Y_\alpha(t) = e^{-t/\alpha}L_\alpha(e^t).$$

By the self-similarity of L_α we note that the process Y_α is stationary with $S_\alpha(1, -1, 0)$ margins. In fact, for any $h \in \mathbb{R}$ we observe (cf. [Breiman \(1968\)](#))

$$\begin{aligned} \{Y_\alpha(t+h) : -\infty < t < \infty\} &= \{e^{-h/\alpha}e^{-t/\alpha}L_\alpha(e^he^t) : -\infty < t < \infty\} \\ &\stackrel{d}{=} \{e^{-t/\alpha}L_\alpha(e^t) : -\infty < t < \infty\} \\ &= \{Y_\alpha(t) : -\infty < t < \infty\}. \end{aligned}$$

Convergence to the max-stable process

In order to obtain convergence of the Ornstein-Uhlenbeck process to the process η in (4.7), we need to define a version of Y_α that is rescaled in time and normalized in space. More precisely, for c_n, d_n as above, let

$$W_i^n(t) = c_n^{-1}(Y_i(s_n t) - d_n), \quad t \in [-1, 1],$$

where Y_i , $i \in \mathbb{N}$, are independent copies of Y_α and s_n is given by

$$s_n = \Psi(1)\alpha c_n d_n^{-1} = \Psi(1)\frac{\alpha - 1}{B_\alpha}d_n^{-\alpha/(\alpha-1)} = c_n^\alpha. \quad (4.11)$$

Here, $\Psi(1) = \log \mathbb{E} \exp(L_\alpha(1)) = |\cos(\alpha\pi/2)|$ by Prop. 1.2.12 in [Samorodnitsky & Taquu \(1994\)](#). Note that s_n converges to 0 as $n \rightarrow \infty$. Since the process L_α satisfies the moment condition (4.3), we can define the corresponding two-sided process Z_α according to (4.6).

Theorem 8. *As n tends to ∞ we have convergence*

$$\max_{i=1, \dots, n} W_i^n(t) \xrightarrow{d} \eta(t), \quad t \in [-1, 1], \quad (4.12)$$

in the sense of finite dimensional distributions, where η is the process in (4.7) based on the process Z_α .

Proof. We start to show the result on the positive interval $[0, 1]$. First, note that by the Lévy property we have for $i \in \mathbb{N}$ the stochastic representation $L_i(e^{s_n t}) \stackrel{d}{=} L_i(1) + L_i^*(e^{s_n t} - 1)$,

where L_i^* are independent copies of the Lévy process L_α , also independent of the L_i . This yields the identity in distribution on $D_{\mathbb{R}}[0,1]$

$$\begin{aligned} W_i^n(t) &= c_n^{-1}(e^{-s_n t/\alpha} L_i(e^{s_n t}) - d_n) \\ &\stackrel{d}{=} c_n^{-1}(L_i(1) - d_n) + \left[c_n^{-1} e^{-s_n t/\alpha} L_i^*(e^{s_n t} - 1) - t\Psi(1) \right] \\ &\quad + \left[c_n^{-1} L_i(1)(e^{-s_n t/\alpha} - 1) + t\Psi(1) \right] \\ &= X_{i,n} + Z_{i,n}(t) + \Delta_{i,n}(t), \quad t \in [0,1], \end{aligned}$$

where $X_{i,n}$, $Z_{i,n}$ and $\Delta_{i,n}$ are defined as the three summands, respectively. The first step is to prove the following weak convergence on $\mathbb{R} \times D_{\mathbb{R}}[0,1]$ (equipped with the Borel σ -algebra)

$$n\mathbb{P}((X_{1,n}, Z_{1,n} + \Delta_{1,n}) \in (u, \infty) \times A) \rightarrow e^{-u} \mathbb{P}_{Z_\alpha}(A), \quad (4.13)$$

for $u \in \mathbb{R}$ and $A \in \mathcal{B}(D_{\mathbb{R}}[0,1])$ with $\mathbb{P}_{Z_\alpha}(\partial A) = 0$. To this end, let us establish the weak convergence

$$n\mathbb{P}((X_{1,n}, Z_{1,n}) \in (u, \infty) \times A) = n\mathbb{P}(X_{1,n} \in (u, \infty)) \cdot \mathbb{P}(Z_{1,n} \in A) \rightarrow e^{-u} \mathbb{P}_{Z_\alpha}(A), \quad (4.14)$$

where we used independence of $X_{1,n}$ and $Z_{1,n}$ for the first equation. From (4.10) it follows that $n\mathbb{P}(X_{1,n} \in (u, \infty)) \rightarrow \exp(-u)$, as $n \rightarrow \infty$ and we now seek to prove the weak convergence

$$\{Z_{1,n}(t) : t \in [0,1]\} \xrightarrow{d} \{L_1^*(t) - t\Psi(1) : t \in [0,1]\} = \{Z_\alpha(t) : t \in [0,1]\}, \quad (4.15)$$

on $D_{\mathbb{R}}[0,1]$, as $n \rightarrow \infty$. In fact, we can even show that this convergence holds almost surely. To this end, recall the definition of the Skorokhod metric in [Ethier & Kurtz \(1986\)](#). Define a sequence of strictly increasing, Lipschitz continuous functions for $n \in \mathbb{N}$ by

$$\lambda_n : [0,1] \rightarrow [0,1], \quad t \mapsto (e^{s_n t} - 1)/(e^{s_n} - 1).$$

It is easy to check that

$$\gamma(\lambda_n) = \sup_{0 \leq t < s \leq 1} \left| \log \frac{\lambda_n(s) - \lambda_n(t)}{s - t} \right| \rightarrow 0, \quad n \rightarrow \infty.$$

Furthermore, we observe that by the self-similarity of L_1^*

$$\{Z_{1,n}(t) : t \in [0,1]\} \stackrel{d}{=} \left\{ [c_n^{-\alpha}(e^{s_n} - 1)]^{1/\alpha} e^{-s_n t/\alpha} L_1^*(\lambda_n(t)) - t\Psi(1) : t \in [0,1] \right\} \quad (4.16)$$

and also

$$\sup_{t \in [0,1]} \left| [c_n^{-\alpha}(e^{s_n} - 1)]^{1/\alpha} e^{-s_n t/\alpha} L_1^*(\lambda_n(t)) - L_1^*(\lambda_n(t)) \right| \rightarrow 0, \quad n \rightarrow \infty,$$

almost surely, since $c_n^{-\alpha}(e^{s_n} - 1) = c_n^{-\alpha}(s_n + o(s_n)) = 1 + o(1)$ by (4.11). Thus, by the definition of the Skorokhod metric, $[c_n^{-\alpha}(e^{s_n} - 1)]^{1/\alpha} e^{-s_n t/\alpha} L_1^*(\lambda_n(\cdot))$ converges almost surely

to L_1^* in $D_{\mathbb{R}}[0,1]$, as $n \rightarrow \infty$. Since almost sure convergence implies weak convergence, this together with (4.16) yields (4.15), and thus (4.14).

In order to see that (4.14) implies (4.13) we first show that

$$\|\Delta_{1,n}\|_{\infty} = \sup_{t \in [0,1]} |c_n^{-1} L_1(1)(e^{-s_n t/\alpha} - 1) + t\Psi(1)| \rightarrow 0, \quad n \rightarrow \infty, \quad (4.17)$$

in probability, conditionally on $X_{1,n}$. Here $\|\cdot\|_{\infty}$ denotes the sup-norm on $D_{\mathbb{R}}[0,1]$. In fact, for arbitrary $\epsilon > 0$ and $a < b \in \mathbb{R}$

$$\begin{aligned} \mathbb{P}(\|\Delta_{1,n}\|_{\infty} > \epsilon \mid X_{1,n} \in [a,b]) &\leq \mathbb{P}\left(\|c_n^{-1} d_n(e^{-s_n t/\alpha} - 1) + t\Psi(1)\|_{\infty} > \epsilon/2 \mid X_{1,n} \in [a,b]\right) \\ &\quad + \mathbb{P}\left(\|X_{1,n}(e^{-s_n t/\alpha} - 1)\|_{\infty} > \epsilon/2 \mid X_{1,n} \in [a,b]\right) \end{aligned}$$

Clearly, the first summand tends to 0 as $n \rightarrow \infty$, since by the definition of s_n in (4.11) and the Taylor approximation of exp we have

$$\sup_{t \in [0,1]} |c_n^{-1} d_n(e^{-s_n t/\alpha} - 1) + t\Psi(1)| = \sup_{t \in [0,1]} |c_n^{-1} d_n(-c_n d_n^{-1} t\Psi(1) + o(c_n d_n^{-1})) + t\Psi(1)| \rightarrow 0.$$

For the second summand we observe

$$\begin{aligned} \mathbb{P}\left(\|X_{1,n}(e^{-s_n t/\alpha} - 1)\|_{\infty} > \epsilon/2 \mid X_{1,n} \in [a,b]\right) \\ \leq \mathbb{P}\left(\|\max\{|a|, |b|\}(e^{-s_n t/\alpha} - 1)\|_{\infty} > \epsilon/2\right) \rightarrow 0, \end{aligned}$$

as $n \rightarrow \infty$, since $s_n \rightarrow 0$. We are now in the position to prove that (4.17) together with (4.14) implies

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_{1,n} + \Delta_{1,n} \in B \mid X_{1,n} \in A) = \mathbb{P}_{Z_{\alpha}}(B), \quad (4.18)$$

for all sets $B \in \mathcal{B}(D_{\mathbb{R}}[0,1])$ with $\mathbb{P}_{Z_{\alpha}}(\partial B) = 0$ and $A \in \mathcal{B}(\mathbb{R})$. This assertion is similar to Lemma 7.3 of Das & Hashorva (2012), where only the space $C[0,1]$ of continuous functions equipped with the sup-norm is considered. Our proof below in the case of the Skorokhod space $D_{\mathbb{R}}[0,1]$ follows essentially the lines of their proof.

For $U \in \mathcal{B}(D_{\mathbb{R}}[0,1])$ and $\epsilon > 0$ define the sets

$$\begin{aligned} U_{\epsilon} &= \{f \in D_{\mathbb{R}}[0,1] : f = g + h \text{ for some } g \in U, h \in D_{\mathbb{R}}[0,1] \text{ and } \|h\|_{\infty} \leq \epsilon\}, \\ U_{-\epsilon} &= \{f \in U : f + h \in U \text{ for all } h \in D_{\mathbb{R}}[0,1] \text{ with } \|h\|_{\infty} \leq \epsilon\}. \end{aligned}$$

For any $\epsilon > 0$ with $\mathbb{P}_{Z_{\alpha}}(\partial B_{\epsilon}) = \mathbb{P}_{Z_{\alpha}}(\partial B_{-\epsilon}) = 0$ we observe

$$\begin{aligned} \mathbb{P}(Z_{1,n} + \Delta_{1,n} \in B \mid X_{1,n} \in A) &\geq \mathbb{P}(Z_{1,n} + \Delta_{1,n} \in B, \|\Delta_{1,n}\|_{\infty} < \epsilon \mid X_{1,n} \in A) \\ &\geq \mathbb{P}(Z_{1,n} \in B_{-\epsilon}, \|\Delta_{1,n}\|_{\infty} < \epsilon \mid X_{1,n} \in A) \\ &= \mathbb{P}(Z_{1,n} \in B_{-\epsilon}) \mathbb{P}(\|\Delta_{1,n}\|_{\infty} < \epsilon \mid X_{1,n} \in A) \\ &\rightarrow \mathbb{P}_{Z_{\alpha}}(B_{-\epsilon}), \quad n \rightarrow \infty, \end{aligned}$$

where we used (4.15) and the fact that $Z_{1,n}$ is independent of both $\Delta_{1,n}$ and $X_{1,n}$, and, as shown above, that $\lim_{n \rightarrow \infty} \mathbb{P}(\|\Delta_{1,n}\|_\infty < \epsilon | X_{1,n} \in A) = 1$. Similarly, we can show that $\limsup_{n \rightarrow \infty} \mathbb{P}(Z_{1,n} + \Delta_{1,n} \in B | X_{1,n} \in A) \leq \mathbb{P}_{Z_\alpha}(B_\epsilon)$. Consequently, we have

$$\begin{aligned} \mathbb{P}_{Z_\alpha} \left(\lim_{m \rightarrow \infty} B_{-\epsilon_m} \right) &\leq \liminf_{n \rightarrow \infty} \mathbb{P}(Z_{1,n} + \Delta_{1,n} \in B | X_{1,n} \in A) \\ &\leq \limsup_{n \rightarrow \infty} \mathbb{P}(Z_{1,n} + \Delta_{1,n} \in B | X_{1,n} \in A) \leq \mathbb{P}_{Z_\alpha} \left(\lim_{m \rightarrow \infty} B_{\epsilon_m} \right), \end{aligned}$$

for a sequence $\epsilon_m \searrow 0$ such that $\mathbb{P}_{Z_\alpha}(\partial B_{\epsilon_m}) = \mathbb{P}_{Z_\alpha}(\partial B_{-\epsilon_m}) = 0$, $m \in \mathbb{N}$. Since any open set in the Skorokhod topology is also open in the $\|\cdot\|_\infty$ -norm, it follows that the interior (in the Skorokhod metric) B^o of B is contained in the set $\lim_{m \rightarrow \infty} B_{-\epsilon_m}$. Similarly, as any accumulation point in the $\|\cdot\|_\infty$ -norm is also an accumulation point in the Skorokhod topology, the set $\lim_{m \rightarrow \infty} B_{\epsilon_m}$ is contained in $B \cup \partial B$. Since $\mathbb{P}_{Z_\alpha}(B^o) = \mathbb{P}_{Z_\alpha}(B \cup \partial B) = \mathbb{P}_{Z_\alpha}(B)$, this yields

$$\mathbb{P}_{Z_\alpha} \left(\lim_{m \rightarrow \infty} B_{\epsilon_m} \right) = \mathbb{P}_{Z_\alpha} \left(\lim_{m \rightarrow \infty} B_{-\epsilon_m} \right) = \mathbb{P}_{Z_\alpha}(B),$$

and thus, (4.18) and consequently (4.13) follow.

To complete the proof for the one-sided case, let $m \in \mathbb{N}$ and $\mathbf{t} = (t_1, \dots, t_m) \in [0,1]^m$ be arbitrary and consider the point processes

$$N_n^* = \sum_{i=1}^n \delta_{(X_{i,n}, Z_{i,n}(\mathbf{t}) + \Delta_{i,n}(\mathbf{t}))}.$$

Since weak convergence on $D_{\mathbb{R}}[0,1]$ implies convergence of finite dimensional distributions, it follows from Prop. 3.21 in Resnick (2008) and (4.13) that N_n^* converges weakly to the PPP $\sum_{i \in \mathbb{N}} \delta_{(X_i, Z_i(\mathbf{t}))}$ with intensity $e^{-x} dx \mathbb{P}_{Z_\alpha}(\mathbf{t})$. Applying the continuous mapping theorem with κ in (4.9) to this weak convergence of point processes gives the desired result on $[0,1]$.

For the two-sided convergence of (4.12) on $[-1,1]$ we first note that by the same arguments as above this convergence holds on $[0,2]$, or equivalently,

$$\max_{i=1, \dots, n} W_i^n(\cdot + 1) \xrightarrow{d} \eta(\cdot + 1), \text{ on } [-1,1].$$

By the stationarity of the W_i^n and η this yields

$$\max_{i=1, \dots, n} W_i^n(\cdot) \xrightarrow{d} \eta(\cdot), \text{ on } [-1,1].$$

□

Remark 11. Both, Theorem 7 and 8 are stated for convergence of finite dimensional distributions. We though believe that even weak convergence on the Skorokhod space $D_{\mathbb{R}}[-1,1]$ holds true. In the proofs of the two theorems we needed to specialize to finite dimensional margins since the statements in Resnick (2008), specifically Prop. 3.21, are valid only for point processes on locally compact, second countable spaces. The Skorokhod space $D_{\mathbb{R}}[-1,1]$

is however not locally compact and therefore another technique is necessary. To this end, the weak convergence in (4.13) can probably be used together with bounds on the lower order statistics, similar as in the proof of Theorem 1.4 in [Kabluchko et al. \(2009\)](#), to establish the desired result.

4.3 The d -dimensional case: Lévy-Mori random fields

We now extend the results to d -dimensional random fields having infinitely divisible distributions. Let us first set up the framework. We will follow the approach of [Mori \(1992\)](#) who proved a representation theorem for stochastically continuous, infinitely divisible, linearly additive random fields $\{\xi(t) : t \in \mathbb{R}^d\}$ without Gaussian components. By linearly additive we understand that for every pair $a, b \in \mathbb{R}^d$ the stochastic process $\{\xi(a + \lambda b) : \lambda \in \mathbb{R}\}$ has independent increments.

In order to describe these random fields explicitly we consider the space of all $(d - 1)$ -dimensional hyperplanes in \mathbb{R}^d , denoted by H_d . Every pair $(s, \theta) \in \mathbb{R} \times S^{d-1}$ represents a hyperplane $\{x \in \mathbb{R}^d : \langle x, \theta \rangle = s\}$, where (s, θ) and $(-s, -\theta)$ correspond to the same hyperplane. We can identify H_d with the half cylinder

$$C_{d+} = (\mathbb{R}_+ \times S^{d-1}) \cup (\{0\} \times S_+^{d-1}),$$

where $\mathbb{R}_+ = (0, \infty)$ and S_+^{d-1} is the set of all $x \in S^{d-1}$ whose first non-zero component is positive. Moreover, define the set

$$A||B = \{h \in H_d : h \text{ separates } A \text{ and } B\}.$$

A measure m on H_d is called bundleless if $m(\{h \in H_d : h \cap \{a\} \neq \emptyset\}) = 0$ for all $a \in \mathbb{R}^d$.

We now consider a measure Λ on the space $H_d \times \mathbb{R}_0$ which satisfies the following conditions:

(A1) For all $A \in \mathcal{B}_0(H_d)$, the space of all relatively compact Borel sets in H_d , we have

$$\begin{aligned} \int_{\mathbb{R}_0} |x|^2 \wedge 1 \Lambda(A \times dx) &< \infty, \\ \int_{|x|>1} e^x \Lambda(A \times dx) &< \infty. \end{aligned}$$

(A2) For every $a > 0$, the measure $\Lambda_a(\cdot) = \Lambda(\cdot \times \{|x| > a\})$ is locally finite and bundleless.

Every such measure Λ satisfying the two above conditions determines an infinitely divisible random measure $Y = Y_\Lambda$ on H_d , i.e., a family of infinitely divisible random variables $\{Y(A) : A \in \mathcal{B}_0(H_d)\}$ which satisfy the following conditions:

- (i) $Y(A_1)$ and $Y(A_2)$ are independent for disjoint $A_1, A_2 \in \mathcal{B}_0(H_d)$,
- (ii) $Y(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} Y(A_i)$ a.s. if $A_i \in \mathcal{B}_0(H_d)$ are disjoint and $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}_0(H_d)$,

(iii) for $A \in \mathcal{B}_0(H_d)$

$$\mathbb{E} \exp\{uY(A)\} = \exp \left\{ \int_{\mathbb{R}_0} \left(e^{ux} - 1 - ux \mathbf{1}_{|x| \leq 1} \right) \Lambda(A \times dx) \right\}, \quad u \in [0,1].$$

Proposition 4.1 in [Mori \(1992\)](#) states that for each such measure Λ satisfying (A1) and (A2) we can define a random field I_Λ by

$$I_\Lambda(t) = Y(0||t), \quad t \in \mathbb{R}^d, \quad (4.19)$$

where Y is the infinitely divisible random measure associated to Λ . Furthermore, I_Λ is infinitely divisible, stochastically continuous, linearly additive and it has no Gaussian component.

In order to define a measure Λ on $H_d \times \mathbb{R}_0$ we want to use the half cylinder representation of H_d . To this end, let τ be the natural bijection from H_d to C_{d+} given by $\tau h = (s, \theta)$ if $h = \{u \in \mathbb{R}^d : \langle u, \theta \rangle = s\}$. Moreover, define the mapping

$$\tilde{\tau} : H_d \times \mathbb{R}_0 \rightarrow C_{d+} \times \mathbb{R}_0$$

by $\tilde{\tau}(h, u) = (\tau h, u)$ for $h \in H_d$ and $u \in \mathbb{R}_0$.

Following Def. 4 in [Kablichko et al. \(2009\)](#), we call a process $\{\zeta(t) : t \in \mathbb{R}^d\}$ *Brown-Resnick stationary* if for a Poisson point process $\sum_{i=1}^{\infty} \delta_{X_i}$ on the real line with intensity $e^{-x} dx$ and independent copies ζ_i , $i \in \mathbb{N}$, the process

$$\max_{i \in \mathbb{N}} (X_i + \zeta_i(t)), \quad t \in \mathbb{R}^d,$$

is stationary. We are now in a position to state the main result of this section.

Theorem 9. *Let ν be a measure on $S^{d-1} \times \mathbb{R}_0$ such that*

$$\nu(d(-\theta) \times dx) = e^{-x} \nu(d\theta \times d(-x)). \quad (4.20)$$

Furthermore, suppose that ν satisfies

$$\begin{aligned} \int_{\mathbb{R}_0} |x|^2 \wedge 1 \nu(S^{d-1} \times dx) &< \infty, \\ \int_{|x|>1} e^x \nu(S^{d-1} \times dx) &< \infty. \end{aligned}$$

Define the measure Λ on $H_d \times \mathbb{R}_0$ by

$$\Lambda \tilde{\tau}^{-1}(A) = \begin{cases} (\lambda \times \nu)(A) & \text{if } A \in \mathcal{B}(\mathbb{R}_+ \times S^{d-1} \times \mathbb{R}_0) \\ 0 & \text{if } A \in \mathcal{B}(\{0\} \times S_+^{d-1} \times \mathbb{R}_0), \end{cases}$$

where λ is the Lebesgue measure on \mathbb{R}_+ . Then Λ satisfies (A1) and (A2) and the process

$$\zeta(t) = I_\Lambda(t) - \log \mathbb{E} e^{I_\Lambda(t)}, \quad t \in \mathbb{R}^d,$$

is Brown-Resnick stationary, where I_Λ is defined as in (4.19).

Proof. Fix $n \in \mathbb{N}$, $t_1, \dots, t_n \in \mathbb{R}^d$ and $u_1, \dots, u_n \in [0, 1]$ with $\sum_{i=1}^n u_i = 1$. We verify the stationarity criterion by [Kabluchko et al. \(2009\)](#) for the process ζ .

To this end, let \mathcal{N}_n be the set of all nonempty subsets of $\{1, \dots, n\}$. For $J \in \mathcal{N}_n$ define the set $A_J \subset H_d$ by

$$A_J = (\{0\} \cup \{t_j, j \notin J\}) \cup \{t_k, k \in J\}$$

and $u_J = \sum_{k \in J} u_k$. Further, let

$$\psi(u, x) = e^{ux} - 1 - ux \mathbf{1}_{|x| \leq 1}, \quad u \in [0, 1], x \in \mathbb{R}_0.$$

First, consider the easiest example for ν , namely for fixed $\alpha \in S^{d-1}$ and $z \in \mathbb{R}_0$

$$\nu^{\alpha, z}(d\theta \times dx) = \delta_\alpha(d\theta) \delta_z(dx) + e^{-x} \delta_{-\alpha}(d\theta) \delta_{-z}(dx),$$

which satisfies (4.20). Indeed,

$$\begin{aligned} \nu^{\alpha, z}(d(-\theta) \times dx) &= \delta_{-\alpha}(d\theta) \delta_z(dx) + e^{-x} \delta_\alpha(d\theta) \delta_{-z}(dx) \\ &= e^{-x} x \left(e^{-(-x)} \delta_{-\alpha}(d\theta) \delta_{-z}(d(-x)) + \delta_\alpha(d\theta) \delta_z(d(-x)) \right) \\ &= e^{-x} \nu^{\alpha, z}(d\theta \times d(-x)). \end{aligned}$$

This measure induces the measure $\Lambda^{\alpha, z}$ and the random field $\zeta^{\alpha, z}$ which essentially is a Poisson process in the direction of α which is constant on lines orthogonal to α . It follows from Corollary 9 in [Kabluchko et al. \(2009\)](#) and Theorem 5 that $\zeta^{\alpha, z}$ is Brown-Resnick stationary.

Let Y be the infinitely divisible random measure determined by Λ , then we conclude by geometrical considerations that

$$\sum_{i=1}^n u_i I_\Lambda(t_i) = \sum_{J \in \mathcal{N}_n} u_J Y(A_J) \quad \text{a.s.}$$

and that the sets A_J , $J \in \mathcal{N}_n$, are disjoint (cf. [Mori \(1992\)](#)). Thus,

$$\log \mathbb{E} \exp \left\{ \sum_{i=1}^n u_i I_\Lambda(t_i) \right\} = \sum_{J \in \mathcal{N}_n} \log \mathbb{E} \exp \{ u_J Y(A_J) \}$$

and we compute for $J \in \mathcal{N}_n$

$$\begin{aligned}
\log \mathbb{E} \exp \{u_J Y(A_J)\} &= \int_{\mathbb{R}_0} \psi(u_J, x) \Lambda(A_J \times dx) \\
&= \int_{A_J \times \mathbb{R}_0} \psi(u_J, x) \Lambda(dh \times dx) \\
&= \int_{\tilde{\tau}(A_J \times \mathbb{R}_0)} \psi(u_J, x) \Lambda \tilde{\tau}^{-1}(ds \times d\theta \times dx) \\
&= \int_{\tau(A_J) \times \mathbb{R}_0} \psi(u_J, x) ds \nu(d\theta \times dx) \\
&= \int_{S^{d-1} \times \mathbb{R}_0} \int_{\mathbb{R}_+} \mathbf{1}_{\tau(A_J)}(s, \theta) \psi(u_J, x) ds \nu(d\theta \times dx) \\
&= \int_{S_+^{d-1} \times \mathbb{R}_0} \int_{\mathbb{R}_+} \mathbf{1}_{\tau(A_J)}(s, \theta) \psi(u_J, x) + \mathbf{1}_{\tau(A_J)}(s, -\theta) e^x \psi(u_J, -x) ds \nu(d\theta \times dx),
\end{aligned}$$

where the last equality follows from (4.20). Applying this formula to the process $I_{\Lambda^{\alpha, z}}$ we get

$$\log \mathbb{E} \exp \left\{ \sum_{i=1}^n u_i I_{\Lambda^{\alpha, z}}(t_i) \right\} = \sum_{J \in \mathcal{N}_n} \int_{\mathbb{R}_+} \mathbf{1}_{\tau(A_J)}(s, \alpha) \psi(u_J, z) + \mathbf{1}_{\tau(A_J)}(s, -\alpha) e^z \psi(u_J, -z) ds.$$

Furthermore, for fixed $i \in \{1, \dots, n\}$, by putting $u_i = 1$ and $u_j = 0$, $j \neq i$, we obtain

$$\begin{aligned}
\log \mathbb{E} e^{I_{\Lambda}(t_i)} &= \int_{S_+^{d-1} \times \mathbb{R}_0} \sum_{J \in \mathcal{N}_n} \int_{\mathbb{R}_+} \mathbf{1}_{\tau(A_J)}(s, \theta) \psi(\mathbf{1}_{i \in J}, x) \\
&\quad + \mathbf{1}_{\tau(A_J)}(s, -\theta) e^x \psi(\mathbf{1}_{i \in J}, -x) ds \nu(d\theta \times dx).
\end{aligned}$$

Thus, putting the parts together, we have

$$\begin{aligned}
&\log \mathbb{E} \exp \left\{ \sum_{i=1}^n u_i \left(I_{\Lambda}(t_i) - \log \mathbb{E} e^{I_{\Lambda}(t_i)} \right) \right\} \\
&= \int_{S_+^{d-1} \times \mathbb{R}_0} \sum_{J \in \mathcal{N}_n} \int_{\mathbb{R}_+} \mathbf{1}_{\tau(A_J)}(s, \theta) \left[\psi(u_J, x) - \sum_{i=1}^n \psi(\mathbf{1}_{i \in J}, x) \right] \\
&\quad + \mathbf{1}_{\tau(A_J)}(s, -\theta) e^x \left[\psi(u_J, -x) - \sum_{i=1}^n \psi(\mathbf{1}_{i \in J}, -x) \right] ds \nu(d\theta \times dx) \\
&= \int_{S_+^{d-1} \times \mathbb{R}_0} \log \mathbb{E} \exp \left\{ \sum_{i=1}^n u_i \left(I_{\Lambda^{\alpha, z}}(t_i) - \log \mathbb{E} e^{I_{\Lambda^{\alpha, z}}(t_i)} \right) \right\} \nu(d\theta \times dx). \tag{4.21}
\end{aligned}$$

From the Brown-Resnick stationarity of $\zeta^{\alpha, z}$ it follows that for $h \in \mathbb{R}^d$

$$\log \mathbb{E} \exp \left\{ \sum_{i=1}^n u_i \zeta^{\alpha, z}(t_i + h) \right\} = \log \mathbb{E} \exp \left\{ \sum_{i=1}^n u_i \zeta^{\alpha, z}(t_i) \right\}.$$

This together with Proposition 6 in [Kabluchko et al. \(2009\)](#) and (4.21) shows that also ζ is Brown-Resnick stationary. □

4.4 Estimation

When modeling extreme events via max-stable processes it is important to have estimation methods to determine the model parameters. As seen in Chapter 3 inference for max-stable processes is a challenging task as multivariate densities are usually unknown. Fortunately, in many cases already the bivariate distributions contain all necessary information. For instance, the composite likelihood approach in [Padoan et al. \(2010\)](#) uses this fact to construct estimators for max-stable processes.

In this section we show how inference for the max-stable, stationary “Lévy-Brown-Resnick” processes η in (4.7) can be carried out. In fact, it turns out that by suitably conditioning the data, the problem simplifies to estimation of the parameters of the underlying Lévy process. Similar to the results in Chapter 3 and in [Engelke et al. \(2012d\)](#), the first approach uses the incremental representation of η to derive the asymptotic distribution of extremal increments. The second approach is based on spectral densities as in [Coles & Tawn \(1991\)](#) and is applicable if the Lévy process possesses a density.

Since spectral measures are defined for Fréchet margins, in this section we consider the transformed “Lévy-Brown-Resnick” process

$$\eta^*(t) = \max_{i \in \mathbb{N}} Y_i V_i, \quad t \in \mathbb{R}, \quad (4.22)$$

where $V_i = e^{Z_i(t)}$ and the $Z_i, i \in \mathbb{N}$ are as in Theorem 5 and $\sum_{i=1}^{\infty} \delta_{Y_i}$ is a PPP with intensity $1/y^2 dy$. In particular, η^* is stationary with unit Fréchet margins.

4.4.1 Extremal increments

By construction, the process η^* admits the *incremental representation* (4.22) (cf. [Engelke et al. \(2012c,d\)](#)). Similarly to Theorem 4 in Chapter 3, the following result gives the asymptotic distribution of suitably conditioned increments of processes in the MDA of η^* .

Theorem 10. *Suppose that the stochastic process $\{Q(t) : t \in \mathbb{R}\}$ is in the MDA of η^* with standard Pareto margins, i.e., for any $t_0, \dots, t_k \in \mathbb{R}, k \in \mathbb{N}$, we have the weak convergence*

$$\left(\frac{1}{n} \max_{i=1, \dots, n} Q_i(t_0), \dots, \frac{1}{n} \max_{i=1, \dots, n} Q_i(t_k) \right) \xrightarrow{d} (\eta^*(t_0), \dots, \eta^*(t_k)), \quad n \rightarrow \infty,$$

where $(Q_i)_{i \in \mathbb{N}}$ are independent copies of Q . Further, let U be Pareto distributed and independent of $V = \exp(Z)$. Let $a(n)$ be any sequence of positive numbers such that

$\lim_{n \rightarrow \infty} a(n) = \infty$. Then, as n tends to ∞ , we have the convergence in distribution on \mathbb{R}^{k+1}

$$\left(\frac{Q(t_0)}{a(n)}, \frac{Q(t_1)}{Q(t_0)}, \dots, \frac{Q(t_k)}{Q(t_0)} \mid Q(t_0) > a(n) \right) \xrightarrow{d} (U, V(t_1), \dots, V(t_k)).$$

Proof. The assertion of this theorem is a consequence of the more general statement of Theorem 2.1 in [Engelke et al. \(2012d\)](#). \square

Remark 12. For statistical applications the sequence $a(n)$, $n \in N$, should be chosen in such a way that $\lim_{n \rightarrow \infty} a(n)/n = 0$. In this case, the number of observations that exceed the threshold $a(n)$, i.e., the cardinality of the index set $I(n)$ of extremal observations

$$I(n) = \{i \in \{1, \dots, n\} : Q_i(t_0) > a(n)\},$$

converges to ∞ with probability 1, as n tends to ∞ . Thus, the set of extremal observations

$$\left\{ \left(\frac{Q_i(t_1)}{Q_i(t_0)}, \dots, \frac{Q_i(t_k)}{Q_i(t_0)} \right) : i \in I(n) \right\},$$

is approximately a sample of i.i.d. realizations of the random vector $(V(t_1), \dots, V(t_k))$.

4.4.2 Spectral densities

Recall the definition of the exponent measure and the spectral density from Section 3.2.2 in Chapter 3.

Suppose that the Lévy process $\{Z(t) : t \geq 0\}$ possesses the density f_t at each point $t \geq 0$. For the bivariate distributions $(\eta^*(0), \eta^*(t))$ of the process η^* we compute the spectral density s_t of the exponent measure. To this end, note that for arbitrary $t \geq 0$ we have

$$-\log \mathbb{P}(\eta^*(0) \leq x_0, \eta^*(t) \leq x_1) = \mathbb{E} \max \left\{ \frac{1}{x_0}, \frac{1}{x_1} e^{Z(t)} \right\}.$$

Thus, we compute

$$\mathbb{E} \max \left\{ \frac{1}{x_0}, \frac{1}{x_1} e^{Z(t)} \right\} = \frac{1}{x_1} \int_{\log \frac{x_1}{x_0}}^{\infty} e^y f_t(y) dy + \frac{1}{x_0} \int_{-\infty}^{\log \frac{x_1}{x_0}} f_t(y) dy. \quad (4.23)$$

In order to obtain the spectral density s_t on $S_+ = \{x, y \geq 0 : x + y = 1\}$, we observe that by Theorem 1 in [Coles & Tawn \(1991\)](#) for $\omega \in [0, 1]$, $r > 0$

$$r^{-3} s_t(\omega) = \left(-\frac{\partial^2}{\partial x_0 \partial x_1} [-\log \mathbb{P}(\eta^*(0) \leq x_0, \eta^*(t) \leq x_1)] \right)_{x_0=r\omega, x_1=r(1-\omega)}.$$

Hence, by equation (4.23)

$$\begin{aligned} \frac{\partial}{\partial x_0} \mathbb{E} \max \left\{ \frac{1}{x_0}, \frac{1}{x_1} e^{Z(t)} \right\} &= \frac{1}{x_0^2} f_t \left(\log \frac{x_1}{x_0} \right) - \frac{1}{x_0^2} \int_{-\infty}^{\log \frac{x_1}{x_0}} f_t(y) dy - \frac{1}{x_0^2} f_t \left(\log \frac{x_1}{x_0} \right) \\ &= -\frac{1}{x_0^2} \int_{-\infty}^{\log \frac{x_1}{x_0}} f_t(y) dy, \end{aligned}$$

and consequently

$$-\frac{\partial^2}{\partial x_0 \partial x_1} \mathbb{E} \max \left\{ \frac{1}{x_0}, \frac{1}{x_1} e^{Z(t)} \right\} = \frac{1}{x_0^2 x_1} f_t \left(\log \frac{x_1}{x_0} \right),$$

such that

$$s_t(\omega) = \frac{1}{\omega^2(1-\omega)} f_t(\log[(1-\omega)/\omega]), \quad \omega \in [0,1].$$

For $\Theta \subset \mathbb{R}^d$, let $\{f_{t,\theta} : \theta \in \Theta\}$ be a parametric model for the densities of Z and suppose that $\{Q_i\}_{1 \leq i \leq n}$ is an independent sequence of copies of a process $\{Q(t) : t \in \mathbb{R}\}$ with standard Pareto margins in the MDA of η^* . Further, for fixed $t > 0$ put $\mathbf{Q}_i = (Q_i(0), Q_i(t))$. In this case the spectral densities can be used to estimate the model parameters θ . In fact, after a further transformation of Q_1, \dots, Q_n to pseudo polar coordinates $(r_1, \omega_1), \dots, (r_n, \omega_n)$, for a high threshold $r_0 > 0$, the set $\{(r_i, \omega_i) : i \in I_0\}$, $I_0 = \{1 \leq i \leq n : r_i > r_0\}$, is fitted to a Poisson point process on $[0, \infty) \times S_+$ with intensity $1/r^2 dr \times s_{t,\theta}(\omega) d\omega$. The approximate likelihood is given by

$$L(\theta, \{(r_i, \omega_i), i \in I_0\}) = \exp(-r_0^{-1}) \prod_{i \in I_0} r_i^{-2} s_{t,\theta}(\omega_i).$$

Note that it suffices to maximize the reduced likelihood

$$\tilde{L}(\theta, \{\omega_i, i \in I_0\}) = \prod_{i \in I_0} f_{t,\theta}(\log[(1-\omega_i)/\omega_i]), \quad (4.24)$$

to obtain an estimator for the parameter θ , which fully characterizes the Lévy process $\{Z(t) : t \geq 0\}$ and thus also the ‘‘Lévy-Brown-Resnick’’ process η^* .

4.4.3 Examples

For most Lévy processes the densities either do not exist or are unknown in explicit form. However, there are several important examples where the densities f_t do exist and where we can write down explicitly a parametric model for the spectral density s_t .

Gamma process

Let $\{T(t; \gamma, \delta), t \geq 0\}$, $\gamma, \delta > 0$, be the Gamma Lévy process, i.e. a subordinator with density

$$f_t(x) = \frac{\delta^\gamma}{\Gamma(\gamma t)} x^{\gamma t - 1} e^{-x\delta}, \quad x > 0,$$

where Γ denotes the Gamma function. Its Laplace transform is given by

$$\mathbb{E}e^{\lambda T(s; \gamma, \delta)} = (1 - \lambda/\delta)^{-\gamma s}, \quad \lambda < \delta,$$

and it thus satisfies the moment condition (4.3) if $\delta > 1$. Taking for instance $\{-T(t; \gamma, \delta) : t \geq 0\}$ for any $\gamma, \delta > 0$ or $\{T(t; \gamma, \delta) : t \geq 0\}$ for $\gamma > 0, \delta > 1$ for L^+ in the definition (4.6) of Z , we can estimate γ and δ as above by observing data in the MDA of the corresponding ‘‘Lévy-Brown-Resnick’’ process η^* in (4.22).

Inverse Gaussian

Consider the inverse Gaussian Lévy process $\{T(t; a, b) : t \geq 0\}$, which is a subordinator and defined as follows. Let $\{W(t) : t \geq 0\}$ be a standard Brownian motion and $a, b > 0$. Then T is the first passage time

$$T(t; a, b) = \inf\{u > 0 : W(u) + bu = at\}, \quad t \geq 0.$$

It has an inverse Gaussian distribution with density

$$f_t(x) = \frac{at}{\sqrt{2\pi}} x^{-3/2} e^{abt} e^{-\frac{1}{2}(a^2 t^2 x^{-1} + b^2 x)}, \quad x > 0,$$

and Laplace transform

$$\mathbb{E}e^{-\lambda T(t; a, b)} = \exp\left\{tab \left(1 - \sqrt{1 + 2\lambda/b^2}\right)\right\}, \quad \lambda > -b^2/2.$$

Taking $\{-T(t; a, b) : t \geq 0\}$ in the definition (4.6) of Z , we obtain a max-stable, stationary process $\eta_{a,b}^*$, which depends on the two parameters $a, b > 0$. Note that the density of $Z(t)$ for $t > 0$ is given by $\tilde{f}_t(x) = f_t(x + tab(1 - \sqrt{1 + 2/b^2}))$. Hence, according to (4.24), we need to maximize

$$\log \tilde{L}((a, b), \{\omega_i, i \in I_0\}) \propto |I_0| \log a + |I_0| abt - \frac{1}{2} a^2 t^2 \sum_{i \in I_0} x_i^{-1} - \frac{1}{2} b^2 \sum_{i \in I_0} x_i,$$

where $x_i = \log[(1 - \omega_i)/\omega_i] + tab(1 - \sqrt{1 + 2/b^2})$. This problem can be solved numerically.

Normal inverse Gaussian

Another prominent example for a Lévy process is the normal inverse Gaussian process $\{R(t; a, \alpha, \beta) : t \geq 0\}$ with parameters $a, \alpha, \beta > 0$ and $\alpha > \beta$ (cf. [Barndorff-Nielsen \(1997\)](#)). Let T be the inverse Gaussian subordinator with parameters $a > 0$ and $b = \sqrt{\alpha^2 - \beta^2}$ as above and define a drifted Brownian motion $\{W^{(\beta)}(t) = W(t) + \beta t : t \geq 0\}$, $\beta \in \mathbb{R}$, where $\{W(t) : t \geq 0\}$ is an independent standard Brownian motion. The normal inverse Gaussian process is defined as the time-changed process $\{R(t) = W^{(\beta)}(T(t)), t \geq 0\}$ and has density

$$f_t(x) = \frac{\alpha a t K_1\left(\alpha \sqrt{a^2 t^2 + x^2}\right)}{\pi \sqrt{a^2 t^2 + x^2}} e^{at\sqrt{\alpha^2 - \beta^2} + \beta x}, \quad x \in \mathbb{R},$$

where K_1 denotes the modified Bessel function of the second kind. It has Laplace transform

$$\mathbb{E}e^{\lambda R(t)} = \exp\left(ta\left(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + \lambda)^2}\right)\right), \quad -\alpha - \beta < \lambda < \alpha - \beta.$$

As above, this can be used to fit the corresponding “Lévy-Brown-Resnick” process η^* .

5 Maxima of independent, non-identically distributed Gaussian vectors

Recall from Section 3.1.1 that for a triangular array with i.i.d. entries of d -variate zero-mean, unit-variance normal distributions with covariance matrix Σ_n in the n -th row satisfying

$$\lim_{n \rightarrow \infty} b_n^2(\mathbf{1}\mathbf{1}^\top - \Sigma_n)/2 = A \in [0, \infty)^{d \times d}, \quad (5.1)$$

the row-wise maxima converge to the d -variate, max-stable Hüsler-Reiss distribution whose dependence structure is fully characterized by the matrix A . Note that condition (5.1) implies that all off-diagonal entries of Σ_n converge to 1 as $n \rightarrow \infty$. A slightly more general representation is given in Kabluchko (2011a) in terms of Poisson point processes and negative definite kernels.

In fact, it turns out that these distributions not only attract Gaussian arrays but also classes of related distributions. For instance, Hashorva (2005) shows, that the convergence of maxima holds for triangular arrays of general bivariate elliptical distributions, if the random radius is in the domain of attraction of the Gumbel distribution. The generalization to multivariate elliptical distributions can be found in Hashorva (2006). Moreover, Hashorva *et al.* (2012) prove that also non-elliptical distributions are in the domain of attraction of the Hüsler-Reiss distribution, for instance multivariate χ^2 -distributions.

Apart from being one of the few known parametric families of multivariate extreme value distributions, the Hüsler-Reiss distributions play a prominent role in modeling spatial extremes since they are the finite dimensional distributions of Brown-Resnick processes (Brown & Resnick, 1977; Kabluchko *et al.*, 2009).

Recently, Hashorva & Weng (2013) analyzed maxima of stationary Gaussian triangular arrays where the variables in each row are identically distributed but not necessarily independent. They show that weak dependence is asymptotically negligible, whereas stronger dependence may influence the max-limit distribution.

In this chapter we consider independent triangular arrays $\mathbf{X}_{i,n} = (X_{i,n}^{(1)}, \dots, X_{i,n}^{(d)})$, $n \in \mathbb{N}$ and $1 \leq i \leq n$, where $\mathbf{X}_{i,n}$ is a zero-mean Gaussian random vector with covariance matrix $\Sigma_{i,n}$. Thus, in each row the random variables are independent, but may have different dependence structures. Letting $\mathbf{M}_n = (M_n^{(1)}, \dots, M_n^{(d)})$ denote the vector consisting of the componentwise maxima $M_n^{(j)} = \max_{i=1, \dots, n} X_{i,n}^{(j)}$, $j \in \{1, \dots, d\}$, we are interested in the convergence of the rescaled, row-wise maximum

$$b_n(\mathbf{M}_n - b_n), \quad (5.2)$$

as $n \rightarrow \infty$, and the respective limit distributions.

In Section 5.1 we start with bivariate triangular arrays. For this purpose, we introduce a sequence of counting measures which capture the dependence structure in each row and which is used to state necessary and sufficient conditions for the convergence of (5.2). Moreover, the limits turn out to be new max-stable distributions that generalize (3.5). The results on triangular arrays are used to completely characterize the max-limits of independent sequences of bivariate Gaussian vectors. Explicit examples for the bivariate limit distributions are given at the end of this section. The multivariate case is treated in Section 5.2, giving rise to a class of d -dimensional max-stable distributions. In Section 5.3, we show how these distributions arise as the finite dimensional margins of suitably normalized and randomly rescaled maxima of n independent Gaussian processes. In fact, the limit processes are max-stable, stationary random fields which can be seen as max-mixtures of Brown-Resnick processes. Furthermore, it is shown that these processes offer a large variety of extremal correlation functions which makes them interesting for modeling dependencies in spatial extremes. Finally, Section 5.4 comprises the proofs of the main theorems of this chapter.

5.1 The bivariate case

In order to state the main results in the bivariate case, we need probability measures on the extended positive half-line $[0, \infty]$. To this end, let $([0, \infty], d)$ be a compact metric space such that a function $g : [0, \infty] \rightarrow \mathbb{R}$ is continuous iff it is continuous in the usual topology on $[0, \infty)$ and the limit $\lim_{x \rightarrow \infty} g(x)$ exists and equals $g(\infty)$.

5.1.1 Limit theorems

Consider a triangular array of independent bivariate Gaussian random vectors $\mathbf{X}_{i,n} = (X_{i,n}^{(1)}, X_{i,n}^{(2)})$, $n \in \mathbb{N}$ and $1 \leq i \leq n$, with zero expectation and covariance matrix

$$\text{Cov}(\mathbf{X}_{i,n}) = \begin{pmatrix} \sigma_{i,n,1}^2 & \sigma_{i,n,1,2} \\ \sigma_{i,n,1,2} & \sigma_{i,n,2}^2 \end{pmatrix}.$$

Further, denote by $\rho_{i,n} = \sigma_{i,n,1,2} / (\sigma_{i,n,1} \sigma_{i,n,2})$ the correlation of $\mathbf{X}_{i,n}$. For $n \in \mathbb{N}$, we define a probability measure η_n on $[0, \infty] \times \mathbb{R}^2$ by

$$\eta_n = \frac{1}{n} \sum_{i=1}^n \delta_{(\sqrt{b_n^2(1-\rho_{i,n})}/2, b_n^2(1-\sigma_{i,n,1}), b_n^2(1-\sigma_{i,n,2}))} \quad (5.3)$$

which encodes the suitably normalized variances and correlations in the n -th row. Here, for any measurable space (S, \mathcal{S}) and $a \in S$, δ_a denotes the Dirac measure on the point a . In this general situation, the next theorem gives a sufficient condition in terms of η_n for the convergence of row-wise maxima of this triangular array.

Theorem 11. For $n \in \mathbb{N}$ and $1 \leq i \leq n$, let $\mathbf{X}_{i,n}$ and η_n be defined as above. Further suppose that for some $\epsilon > 0$ the measures $(\eta_n)_{n \in \mathbb{N}}$ satisfy the uniform integrability condition

$$\lim_{K \rightarrow \infty} \sup_{n \in \mathbb{N}} \int_{[0, \infty] \times (K, \infty) \times \mathbb{R}} e^{\theta(1+\epsilon)} \eta_n(d(\lambda, \theta, \gamma)) + \int_{[0, \infty] \times \mathbb{R} \times (K, \infty)} e^{\gamma(1+\epsilon)} \eta_n(d(\lambda, \theta, \gamma)) = 0. \quad (5.4)$$

If for $n \rightarrow \infty$, η_n converges weakly to some probability measure η on $[0, \infty] \times \mathbb{R}^2$, i.e. $\eta_n \Rightarrow \eta$, then

$$\max_{i=1, \dots, n} b_n(\mathbf{X}_{i,n} - b_n) \quad (5.5)$$

converges in distribution to a random vector with distribution function F_η given by

$$\begin{aligned} -\log F_\eta(x, y) = & \int_{[0, \infty] \times \mathbb{R}^2} \Phi \left(\lambda + \frac{y - x + \theta - \gamma}{2\lambda} \right) e^{-(x-\theta)} \\ & + \Phi \left(\lambda - \frac{y - x + \theta - \gamma}{2\lambda} \right) e^{-(y-\gamma)} \eta(d(\lambda, \theta, \gamma)), \end{aligned} \quad (5.6)$$

for $x, y \in \mathbb{R}$.

Remark 13. An equivalent condition for the uniform integrability in (5.4) is that for some $\tau > 0$

$$\sup_{n \in \mathbb{N}} \int_{[0, \infty] \times \mathbb{R}^2} e^{\theta(1+\tau)} + e^{\gamma(1+\tau)} \eta_n(d(\lambda, \theta, \gamma)) < \infty.$$

Note in particular that this implies

$$\sup_{n \in \mathbb{N}, 1 \leq i \leq n} \frac{1}{n} \left(e^{b_n^2(1-1/\sigma_{i,n,1})(1+\tau)} + e^{b_n^2(1-1/\sigma_{i,n,2})(1+\tau)} \right) < \infty.$$

Since $b_n^2 \sim 2 \log n$ for n large, it follows that the variances of both components are uniformly bounded. Thus, the single random variables in each row satisfy the uniform asymptotical negligibility condition (see for instance [Balkema & Resnick \(1977\)](#))

$$\max_{i=1, \dots, n} \mathbb{P}(b_n(X_{i,n}^{(j)} - b_n) > x) \rightarrow 0, \quad n \rightarrow \infty, \quad (5.7)$$

for $j = 1, 2$ and any $x \in \mathbb{R}$.

Remark 14. In fact, one can extend the distribution F_η to mixture measures η taking infinite mass at negative infinity. The only condition which needs to be satisfied is

$$\int_{[0, \infty] \times \mathbb{R}^2} e^\theta + e^\gamma \eta(d(\lambda, \theta, \gamma)) < \infty.$$

Note that the idea of constructing new extreme value distributions as in (5.6) is not new.

Indeed, it is well-known that any mixture of spectral measures is again a spectral measure. In our case, however, these mixture distributions also arise naturally as the max-limits of independent Gaussian triangular arrays.

Note that the one-dimensional marginals of F_η are Gumbel distributed with different location parameters, for instance

$$-\log F_\eta(x, \infty) = \exp \left[-x + \log \int_{[0, \infty] \times \mathbb{R}^2} e^\theta \eta(d(\lambda, \theta, \gamma)) \right].$$

Moreover, F_η is a max-stable distribution since

$$F_\eta^n(x + \log n, y + \log n) = F_\eta(x, y),$$

for all $n \in \mathbb{N}$. This is a remarkable fact, since in general limits of row-wise maxima of triangular arrays are not max-stable, not even if the random variables in each row are identically distributed.

In order to obtain a necessary condition and to simplify the sufficient condition, we need to impose stronger assumptions on the univariate margins. We denote by $\mathcal{M}_1([0, \infty])$ the space of all probability measures on $[0, \infty]$. By Helly's theorem this space is sequentially compact.

Theorem 12. *Consider a triangular array of independent bivariate Gaussian random vectors $\mathbf{X}_{i,n} = (X_{i,n}^{(1)}, X_{i,n}^{(2)})$, $n \in \mathbb{N}$ and $1 \leq i \leq n$, where $X_{i,n}^{(1)}$ and $X_{i,n}^{(2)}$ are standard normal random variables. Denote by $\rho_{i,n}$ the correlation of $\mathbf{X}_{i,n}$. Let*

$$\nu_n = \frac{1}{n} \sum_{i=1}^n \delta_{\sqrt{b_n^2(1-\rho_{i,n})}/2} \quad (5.8)$$

be a probability measure on $[0, \infty]$. For $n \rightarrow \infty$,

$$\max_{i=1, \dots, n} b_n(\mathbf{X}_{i,n} - b_n) \quad (5.9)$$

converges in distribution if and only if ν_n converges weakly to some probability measure ν on $[0, \infty]$, i.e. $\nu_n \Rightarrow \nu$. In this case, the limit of (5.9) has distribution function F_ν given by

$$-\log F_\nu(x, y) = \int_0^\infty \Phi \left(\lambda + \frac{y-x}{2\lambda} \right) e^{-x} + \Phi \left(\lambda + \frac{x-y}{2\lambda} \right) e^{-y} \nu(d\lambda), \quad (5.10)$$

$x, y \in \mathbb{R}$. The distribution in (5.10) uniquely determines the measure ν , i.e., for two probability measures $\nu, \tilde{\nu} \in \mathcal{M}_1([0, \infty])$ with $\nu \neq \tilde{\nu}$ it follows that $F_\nu \neq F_{\tilde{\nu}}$. Furthermore, F_ν depends continuously on ν , in the sense that if $\nu_n \Rightarrow \nu$, as $n \rightarrow \infty$, and $\nu_n, \nu \in \mathcal{M}_1([0, \infty])$, then F_{ν_n} converges pointwise to F_ν .

Remark 15. *For an arbitrary probability measure $\nu \in \mathcal{M}_1([0, \infty])$, let $(R_i)_{i \in \mathbb{N}}$ be a sequence*

of i.i.d. samples of ν . Putting $\rho_{i,n} = \max(1 - 2R_i^2/b_n^2, -1)$ in Theorem 12 yields

$$\nu_n = \frac{1}{n} \sum_{i=1}^n \delta_{\min(R_i, b_n)} \Rightarrow \nu, \quad \text{a.s.},$$

by the law of large numbers. Hence, (5.9) converges a.s. in distribution to F_ν .

Remark 16. If ν is a probability measure on $[0, \infty)$, an alternative construction of the distribution F_ν is the following (Kabluchko, 2011a, Section 3): Let $\sum_{i=1}^{\infty} \delta_{U_i}$ be a Poisson point process on \mathbb{R} with intensity $e^{-u} du$ and suppose that B has the normal distribution $N(-2S^2, 4S^2)$ with random mean and variance, where S is ν -distributed. Then, for a sequence $(B_i)_{i \in \mathbb{N}}$ of i.i.d. copies of B , the bivariate random vector $\max_{i \in \mathbb{N}}(U_i, U_i + B_i)$ has distribution F_ν .

The above theorem can be applied to completely characterize the maxima of a sequence of independent bivariate Gaussian random vectors with unit variance.

Corollary 1. Suppose that $\mathbf{X}_i = (X_i^{(1)}, X_i^{(2)})$, $n \in \mathbb{N}$ and $1 \leq i \leq n$, is a sequence of independent bivariate Gaussian random vectors where $X_i^{(1)}$ and $X_i^{(2)}$ are standard normal random variables. Denote by ρ_i the correlation of \mathbf{X}_i and let

$$\nu_n = \frac{1}{n} \sum_{i=1}^n \delta_{\sqrt{b_n^2(1-\rho_i)/2}}$$

be a probability measure on $[0, \infty]$. For $n \rightarrow \infty$,

$$\max_{i=1, \dots, n} b_n(\mathbf{X}_i - b_n) \tag{5.11}$$

converges in distribution if and only if ν_n converges weakly to some probability measure ν on $[0, \infty]$. In this case, the limit of (5.11) has distribution function F_ν as in (5.10). Furthermore, for all $\nu \in \mathcal{M}_1([0, \infty])$, F_ν is attained as a limit of (5.11) for a suitable sequence $(\mathbf{X}_i)_{i \in \mathbb{N}}$.

Remark 17. It is worthwhile to note that in general, the class of max-selfdecomposable distributions, i.e. the max-limits of sequences of independent (not necessarily identically distributed) random variables, is a proper subclass of max-infinitely-divisible distributions, i.e. the max-limits of triangular arrays with i.i.d. random variables in each row. The latter coincides with the class of max-limits of triangular arrays, where the rows are merely independent but not identically distributed (Balkema & Resnick, 1977; Gerttse, 1986). In the (bivariate) Gaussian case the above shows that the max-limits of i.i.d. triangular arrays, namely the Hüsler-Reiss distributions in (3.5), are a proper subclass of max-limits of independent triangular arrays, namely the distributions in (5.10), which, on the other hand, coincide with the max-limits of independent sequences.

In multivariate and spatial extreme value theory it is important to have flexible and tractable models for spatial dependence of extremal events. On this account, in this section we show how the mixtures of Hüsler-Reiss distributions give rise to new models for bivariate spectral densities and, in the spatial domain, to new classes of extremal correlation functions.

5.1.2 Examples

The max-stable distributions F_ν in Theorem 12 for $\nu \in \mathcal{M}_1([0, \infty])$ are max-mixtures of Hüsler-Reiss distributions with different dependency parameters. They constitute a large class of new bivariate max-stable distributions. We derive two of them explicitly by evaluating the integral in (5.10).

Example 1 (Rayleigh distributed ν). *The Rayleigh distribution has density*

$$f_\sigma(\lambda) = \frac{\lambda}{\sigma^2} e^{-\frac{\lambda^2}{2\sigma^2}}, \quad \lambda \geq 0, \quad (5.12)$$

for $\sigma > 0$. Choosing the dependence parameter λ according to the Rayleigh distribution ν_σ , we obtain the bivariate distribution function

$$-\log F_{\nu_\sigma}(x, y) = \int_0^\infty \left[\Phi\left(\lambda + \frac{y-x}{2\lambda}\right) e^{-x} + \Phi\left(\lambda + \frac{x-y}{2\lambda}\right) e^{-y} \right] \frac{\lambda}{\sigma^2} e^{-\frac{\lambda^2}{2\sigma^2}} d\lambda, \quad (5.13)$$

for $x, y \in \mathbb{R}$. In order to evaluate this integral, we apply partial integration and use formulae 3.471.9 and 3.472.3 in [Gradshteyn & Ryzhik \(2007\)](#). Equation (5.13) then simplifies to

$$F_{\nu_\sigma}(x, y) = \exp \left[-e^{-\min(x, y)} - \frac{1}{\eta} e^{-\frac{y+x}{2}} e^{-\frac{|y-x|\eta}{2}} \right], \quad x, y \in \mathbb{R}, \quad (5.14)$$

where $\eta = \sqrt{1 + 1/\sigma^2} \in (1, \infty)$. Note that σ parameterizes the dependence of F_{ν_σ} . As σ goes to 0 (i.e., η goes to ∞), then the margins become equal. On the other hand, as σ goes to ∞ (i.e., η goes to 1), then the margins become completely independent.

Example 2 (Type-2 Gumbel distributed ν). *The Type-2 Gumbel distribution has density*

$$f_b(\lambda) = 2b\lambda^{-3} e^{-\frac{b}{\lambda^2}}, \quad \lambda \geq 0, \quad (5.15)$$

for $b > 0$. With similar arguments as for the Rayleigh distribution the distribution function F_{ν_b} , where ν_b has density f_b , is given by

$$F_{\nu_b}(x, y) = \exp \left[-e^{-x} - e^{-y} + e^{-\frac{y+x}{2}} e^{-\sqrt{\left(\frac{y-x}{2}\right)^2 + 2b}} \right], \quad x, y \in \mathbb{R}. \quad (5.16)$$

Also in this case, the parameter $b \in (0, \infty)$ interpolates between complete independence and complete dependence of the bivariate distribution. In particular, if $b \rightarrow 0$, then the margins are equal and, on the other hand, if $b \rightarrow \infty$ then the margins are independent.

Every multivariate max-stable distribution admits a spectral representation ([Resnick, 2008](#), Chapter 5), where the spectral measure contains all information about the extremal dependence. Recently, [Cooley et al. \(2010\)](#) and [Ballani & Schlather \(2011\)](#) constructed new parametric models for spectral measures. For the bivariate Hüsler-Reiss distribution, [de Haan & Pereira \(2006\)](#) give an explicit form of its spectral density on the positive sphere

$S_+^1 = \{(x_1, x_2) \in [0, \infty)^2, x_1^2 + x_2^2 = 1\}$. More precisely, they show that for $\lambda \in (0, \infty)$

$$-\log F_\lambda(x, y) = \int_0^{\pi/2} \max\{e^{-x} \sin \theta, e^{-y} \cos \theta\} s_\lambda(\theta) d\theta, \quad x, y \in \mathbb{R},$$

and give a rather complicated expression for s_λ . Using the equation

$$\phi\left(\lambda - \frac{\log \tan \theta}{2\lambda}\right) = \frac{\sin \theta}{\cos \theta} \phi\left(\lambda + \frac{\log \tan \theta}{2\lambda}\right), \quad \lambda \in (0, \infty), \theta \in [0, \pi/2],$$

their expression can be considerably simplified and the spectral density becomes

$$s_\lambda(\theta) = \frac{1}{2\lambda \sin \theta \cos^2 \theta} \phi\left(\lambda + \frac{\log(\tan \theta)}{2\lambda}\right), \quad \theta \in [0, \pi/2].$$

For the spectral density s_ν of the Hüsler-Reiss mixture distribution F_ν as in (5.10), where ν does neither have an atom at 0 nor at ∞ , we have the relation

$$s_\nu(\theta) = \int_0^\infty s_\lambda(\theta) \nu(d\lambda), \quad \theta \in [0, \pi/2].$$

For the two examples above we can compute the corresponding spectral densities.

Proposition 4. *For the Rayleigh distribution with parameter $\sigma > 0$, s_{ν_σ} is given by*

$$s_{\nu_\sigma}(\theta) = \frac{e^{-\frac{1}{\sqrt{2}}|\log \tan \theta| \sqrt{1+1/\sigma^2}}}{4\sqrt{\sigma^4 + \sigma^2} (\sin \theta \cos \theta)^{3/2}}, \quad \theta \in [0, \pi/2].$$

Similarly, for the Type-2 Gumbel distribution with parameter $b > 0$, the spectral density has the form

$$s_{\nu_b}(\theta) = \frac{e^{-u_b(\theta)}}{4(\sin \theta \cos \theta)^{3/2}} \left(1 - \frac{(\log \tan \theta)^2}{4u_b(\theta)^2}\right) \left(1 + \frac{1}{u_b(\theta)}\right), \quad \theta \in [0, \pi/2],$$

with $u_b(\theta) = \sqrt{(\log \tan \theta)^2 / 4 + 2b}$.

Figure 5.1 illustrates how these spectral measures interpolate between complete independence and complete dependence for different parameters.

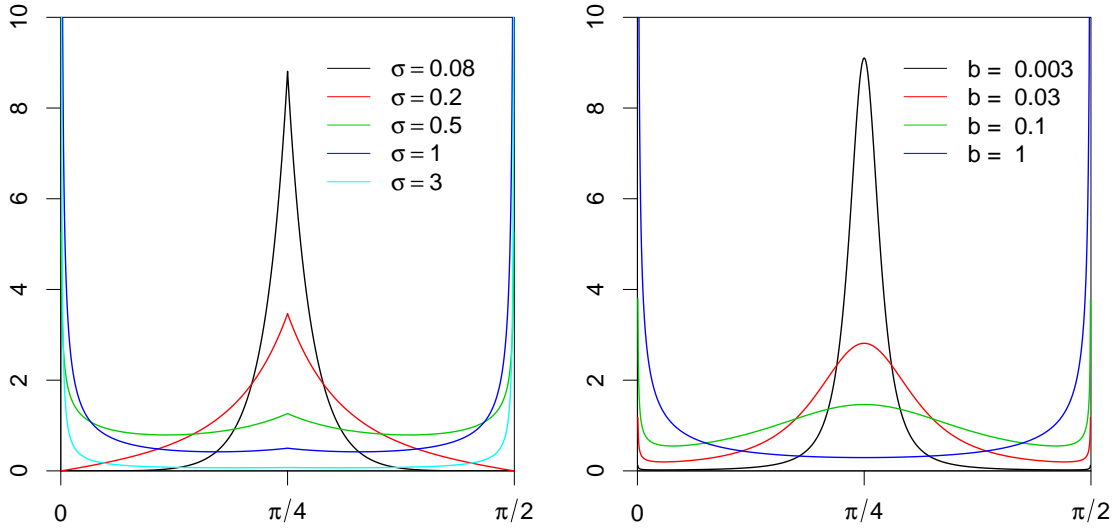


Figure 5.1: Spectral densities of the Rayleigh (left) and Type-2 Gumbel (right) mixture distribution for different parameters σ and b , respectively

5.2 The multivariate case

Similarly as in [Hüsler & Reiss \(1989\)](#), the results for standard bivariate Gaussian random vectors can be generalized to d -dimensional random vectors. To this end, define a triangular array of independent d -dimensional, non-degenerate (i.e. with positive definite covariance matrix) Gaussian random vectors $\mathbf{X}_{i,n} = (X_{i,n}^{(1)}, \dots, X_{i,n}^{(d)})$, $n, d \in \mathbb{N}$ and $1 \leq i \leq n$, where $X_{i,n}^{(j)}$, $j \in \{1, \dots, d\}$, are standard normal random variables. Denote by $\Sigma_{i,n} = (\rho_{j,k}(i,n))_{1 \leq j,k \leq d}$ the correlation matrix of $\mathbf{X}_{i,n}$. Let $\mathbf{1} = (1, \dots, 1)^\top \in \mathbb{R}^d$ and

$$\eta_n = \frac{1}{n} \sum_{i=1}^n \delta_{\sqrt{b_n^2(\mathbf{1}\mathbf{1}^\top - \Sigma_{i,n})/2}} \quad (5.17)$$

be a probability measure on the metric space $[0, \infty)^{d \times d}$, equipped with the product metric. Throughout this chapter, squares and square roots of matrices are to be understood component-wise. For a measure τ on $[0, \infty)^{d \times d}$ we will denote by τ^2 the image measure of τ under the transformation $[0, \infty)^{d \times d} \rightarrow [0, \infty)^{d \times d}$, $\Lambda \mapsto \Lambda^2$. Further, recall from [Section 3.1.1](#) the definition of the space of strictly (conditionally) negative definite matrices D and the d -dimensional Hüsler-Reiss distribution H_Λ . In particular, note that D is a suitable subspace for the measures η_n^2 since $\eta_n^2(D) = 1$ for all $n \in \mathbb{N}$. For $\Lambda = (\lambda_{j,k})_{1 \leq j,k \leq d} \in [0, \infty)^{d \times d}$, define a family of transformed matrices by

$$\Gamma_{l,m}(\Lambda) = 2 \left(\lambda_{m_j, m_l}^2 + \lambda_{m_k, m_l}^2 - \lambda_{m_j, m_k}^2 \right)_{1 \leq j, k \leq l-1},$$

where $2 \leq l \leq d$ and $m = (m_1, \dots, m_l)$ with $1 \leq m_1 < \dots < m_l \leq d$. It follows from the proof of Lemma 2.1 in Berg *et al.* (1984) that if $\Lambda \in D$, then $\Gamma_{l,m}(\sqrt{\Lambda})$ is a (strictly) positive definite matrix.

With this notation we are now in a position to state the following theorem.

Theorem 13. *Consider a triangular array of independent d -dimensional Gaussian random vectors as above. If for $n \rightarrow \infty$ the measure η_n in (5.17) converges weakly to some probability measure η on $[0, \infty)^{d \times d}$, i.e. $\eta_n \Rightarrow \eta$, s.t. $\eta^2(D) = 1$, then*

$$\max_{i=1, \dots, n} b_n(\mathbf{X}_{i,n} - b_n)$$

converges in distribution to a random vector with distribution function H_η given by

$$H_\eta(x_1, \dots, x_d) = \exp \left(\int_{[0, \infty)^{d \times d}} \log H_\Lambda(x) \eta(d\Lambda) \right), \quad x \in \mathbb{R}^d. \quad (5.18)$$

Remark 18. *Similarly to Remark 16, we can give an alternative construction of the distribution H_η in terms of Poisson point processes. Let $\sum_{i=1}^{\infty} \delta_{U_i}$ be a Poisson point process on \mathbb{R} with intensity $e^{-u} du$ and suppose that \mathbf{B} has the multivariate normal distribution $N(-\text{diag}(\Gamma_{d,(1, \dots, d)}(\Lambda))/2, \Gamma_{d,(1, \dots, d)}(\Lambda))$ with random mean and variance, where Λ is η -distributed. Then, for a sequence $(\mathbf{B}_i)_{i \in \mathbb{N}}$ of i.i.d. copies of \mathbf{B} , the bivariate random vector $\max_{i \in \mathbb{N}} (U_i, U_i + \mathbf{B}_i)$ has distribution H_η .*

Remark 19. *We believe that the above theorem also holds in the case when η has positive measure on non-strictly conditionally negative definite matrices, i.e. $\eta^2(D) < 1$. Our proof of this theorem however breaks down in this situation such that another technique might be necessary.*

5.3 Application to Brown-Resnick processes

The d -dimensional Hüsler-Reiss distributions arise in the theory of maxima of Gaussian random fields as the finite dimensional distributions of Brown-Resnick processes (Brown & Resnick, 1977) and its generalizations (Kablichko *et al.*, 2009). In this section we introduce a new class of max-stable processes with finite dimensional distributions given by (5.18) for suitable measures η . In fact, they are the max-limits of Gaussian random fields with random scaling in space.

Recall the definition and properties of the Brown-Resnick process in (3.6). Theorem 17 in Kablichko *et al.* (2009) implies that the subclass belonging to variograms of fractional Brownian motions on \mathbb{R}^d , i.e. $\gamma(t) = \|t\|^\alpha$, $\alpha \in (0, 2)$, are the max-limits of suitably rescaled Gaussian processes whose covariance functions satisfy the following regular variation condition. Let $\{X(t), t \in \mathbb{R}^d\}$ be a zero-mean, unit variance Gaussian process with covariance function $C(t_1, t_2) = \mathbb{E}[X(t_1)X(t_2)]$. Assume that

$$\lim_{\epsilon \rightarrow 0} \frac{1 - C(\epsilon t_1, \epsilon t_2)}{L(\epsilon) \epsilon^\alpha} = 2\gamma(t_1 - t_2) \quad (5.19)$$

holds uniformly for bounded $t_1, t_2 \in \mathbb{R}^d$, where L is continuous and slowly varying at 0 and $\gamma(t) = \|t\|^\alpha$, $\alpha \in (0, 2)$, is a continuous variogram. Further, define normalizing sequences b_n as above and

$$s_n = \inf\{s > 0 : L(s)s^\alpha = b_n^{-2}\}. \quad (5.20)$$

In the next theorem we prove, that by introducing a random space scaling in Theorem 17 in [Kabluchko et al. \(2009\)](#), we obtain new stochastic processes which have finite dimensional distributions given by (5.18) for a certain measure η .

Theorem 14. *Let X_i , $i \in \mathbb{N}$, be independent copies of the process X above, satisfying the regular variation assumption (5.19). Further, let S_i , $i \in \mathbb{N}$, be independent random variables distributed according to a probability measure ν on $(0, \infty)$. Then the finite dimensional distributions of the process*

$$Y_n(t) = \max_{i=1, \dots, n} b_n(X_i(S_i^{2/\alpha} s_n t) - b_n), \quad t \in \mathbb{R}^d,$$

converge to the distribution in (5.18). More precisely, for $t_1, \dots, t_m \in \mathbb{R}^d$, the respective measure η is concentrated on the subspace $\{\lambda \Lambda_0, \lambda > 0\}$ with $\Lambda_0 = \left(\sqrt{\gamma(t_j - t_k)}\right)_{1 \leq j, k \leq m}$ and is given by

$$\eta(d\lambda \Lambda_0) = \nu(d\lambda), \quad (5.21)$$

and by 0 everywhere else.

Remark 20. *Note that, by Kolmogorov's extension theorem, the family of finite dimensional limit distributions in the above theorem gives rise to a new stationary, max-stable stochastic process $\{Y(t), t \in \mathbb{R}^d\}$ on \mathbb{R}^d .*

In fact, it is possible to define this max-stable processes via another construction which allows for a broader class of variograms than those in (5.19). Let

$$V_d = \left\{ \gamma : \mathbb{R}^d \rightarrow [0, \infty) : \gamma(0) = 0, \gamma \text{ conditionally negative definite} \right\}$$

denote the space of all variograms on \mathbb{R}^d , equipped with the product σ -algebra. Further, let \mathbb{Q} be an arbitrary probability measure on this space and γ_i , $i \in \mathbb{N}$, be an i.i.d. sequence of random variables with distribution \mathbb{Q} . For each $i \in \mathbb{N}$, let ξ_i be a Brown-Resnick process as in (3.6) with variogram $4\gamma_i$. Consider the max-mixtures κ_n of the processes ξ_i , given by

$$\kappa_n(t) = \max_{i=1, \dots, n} \xi_i(t) - \log n, \quad t \in \mathbb{R}^d,$$

for $n \in \mathbb{N}$. It can be shown that, as $n \rightarrow \infty$, the finite dimensional distributions of the processes κ_n converge to those of a max-stable, stationary process κ . The latter are given by the distribution in (5.18) with η induced by \mathbb{Q} .

Recall from Chapter 1 that the extremal correlation function of the process in (3.6) is

given by

$$\rho_\gamma(h) = 2 \left(1 - \Phi \left(\sqrt{\gamma(h)}/2 \right) \right), \quad h \in \mathbb{R}^d.$$

Remark 20 defines new max-stable and stationary processes and thus also extremal correlation functions. Moreover, from the construction it is obvious that processes with this dependence structure can be simulated easily as max-mixtures of Brown-Resnick processes. In fact, for an arbitrary variogram γ and mixture measure ν on $(0, \infty)$, we let the measure \mathbb{Q} in this remark be the law of $S^2\gamma$, where S is ν -distributed. The corresponding process κ possesses the extremal correlation function

$$\rho_{\gamma, \nu}(h) = \int_0^\infty 2 \left(1 - \Phi \left(s\sqrt{\gamma(h)} \right) \right) \nu(ds), \quad h \in \mathbb{R}^d. \quad (5.22)$$

Gneiting (1999) analyzes this kind of scale mixtures of the complementary error function in a more general framework. The following corollary is a consequence of Theorem 3.7 and 3.8 therein.

Corollary 2. *For a fixed variogram γ the class of extremal correlation functions in (5.22) is given by all functions $\varphi(\sqrt{\gamma(h)})$, $h \in \mathbb{R}^d$, where $\varphi : [0, \infty) \rightarrow \mathbb{R}$ is a continuous function with $\varphi(0) = 1$, $\lim_{h \rightarrow \infty} \varphi(h) = 0$, and*

$$(-1)^k \frac{d^k}{dh^k} [-\varphi'(\sqrt{h})] \quad (5.23)$$

is nonnegative for infinitely many positive integers k .

For instance, if ν_1 is the Rayleigh distribution (5.12) with density f_1 , we immediately obtain

$$\rho_{\gamma, \nu_1}(h) = 2 \left(1 - \int_0^\infty \Phi(\lambda) f_{\sqrt{\gamma(h)}}(\lambda) d\lambda \right) = 1 - \left(\frac{\gamma(h)}{\gamma(h) + 1} \right)^{1/2}, \quad h \in \mathbb{R}^d,$$

from equation (5.14). In fact, $\rho_{\gamma, \nu_1}(h) = \psi(\gamma(h))$, with $\psi(x) = 1 - (x/(x+1))^{1/2}$ being a completely monotone member of the Dagum family (Berg et al., 2008). However, it is interesting to note, that when writing $\rho_{\gamma, \nu_1}(h) = \varphi(\sqrt{\gamma(h)})$ with $\varphi(x) = 1 - (x^2/(x^2+1))^{1/2}$ as in Proposition 2, the function φ merely satisfies (5.23) but is not completely monotone. Similarly, for the Type-2 Gumbel distribution with $b = 1$, the extremal correlation function is given by $\rho(h) = \exp(-\sqrt{2\gamma(h)})$. In particular, it follows that for any variogram γ and any $r > 0$ the function

$$\rho(h) = \exp \left(-r\sqrt{\gamma(h)} \right), \quad h \in \mathbb{R}^d,$$

is an extremal correlation function. Since this class of extremal correlation functions is closed under the operation of mixing with respect to probability measures, this implies that for any

measure $\mu \in \mathcal{M}_1((0, \infty))$ the Laplace transform $\mathcal{L}\mu$ yields an extremal correlation function

$$\rho_\mu(h) = \mathcal{L}\mu(\sqrt{\gamma(h)}) = \int_0^\infty e^{-r\sqrt{\gamma(h)}} \mu(dr), \quad h \in \mathbb{R}^d.$$

Equivalently, for a completely monotone function ψ with $\psi(0) = 1$, the function $\psi(\sqrt{\gamma(h)})$ is an extremal correlation function. A corresponding max-stable, stationary random field is given by a max-mixture of Brown-Resnick processes with suitable $\nu \in \mathcal{M}_1((0, \infty))$.

5.4 Proofs

Proof of Theorem 11. Let $x, y \in \mathbb{R}$ and put $u_n(z) = b_n + z/b_n$, for $z \in \mathbb{R}$.

$$\begin{aligned} & \log \mathbb{P} \left(\max_{i=1, \dots, n} X_{i,n}^{(1)} \leq u_n(x), \max_{i=1, \dots, n} X_{i,n}^{(2)} \leq u_n(y) \right) \\ &= \sum_{i=1}^n \log \left(1 - \left[\mathbb{P}(X_{i,n}^{(1)} > u_n(x)) + \mathbb{P}(X_{i,n}^{(2)} > u_n(y)) - \mathbb{P}(X_{i,n}^{(1)} > u_n(x), X_{i,n}^{(2)} > u_n(y)) \right] \right) \\ &= - \sum_{i=1}^n \mathbb{P}(X_{i,n}^{(1)} > u_n(x)) - \sum_{i=1}^n \mathbb{P}(X_{i,n}^{(2)} > u_n(y)) \\ & \quad + \sum_{i=1}^n \mathbb{P}(X_{i,n}^{(1)} > u_n(x), X_{i,n}^{(2)} > u_n(y)) + R_n \end{aligned} \quad (5.24)$$

where R_n is a remainder term from the Taylor expansion of $\log(1 - z) = -z - z^2/2 + o(z^2)$, as $z \rightarrow 0$. Thus, by (5.7) there is an $n_0 \in \mathbb{N}$ s.t. for all $n \geq n_0$ we have

$$\begin{aligned} |R_n| &\leq \sum_{i=1}^n \left[\mathbb{P}(X_{i,n}^{(1)} > u_n(x)) + \mathbb{P}(X_{i,n}^{(2)} > u_n(y)) \right]^2 \\ &\leq \max_{i=1, \dots, n} \left[\mathbb{P}(X_{i,n}^{(1)} > u_n(x)) + \mathbb{P}(X_{i,n}^{(2)} > u_n(y)) \right] \\ & \quad \cdot \sum_{i=1}^n \left[\mathbb{P}(X_{i,n}^{(1)} > u_n(x)) + \mathbb{P}(X_{i,n}^{(2)} > u_n(y)) \right]. \end{aligned} \quad (5.25)$$

For the one-dimensional margins we observe

$$\begin{aligned} - \sum_{i=1}^n \mathbb{P}(X_{i,n}^{(1)} > u_n(x)) &= - \sum_{i=1}^n \int_{u_n(x)/\sigma_{i,n,1}}^\infty \phi(z) dz \\ &= - \sum_{i=1}^n \int_{x/\sigma_{i,n,1} - b_n^2(1-1/\sigma_{i,n,1})}^\infty \frac{1}{b_n} \phi(u_n(z)) dz \\ &= - \int_{[0, \infty] \times \mathbb{R}^2} \int_{(1-\theta/b_n^2)x-\theta}^\infty e^{-z-z^2/(2b_n^2)} dz \eta_n(d(\lambda, \theta, \gamma)), \end{aligned}$$

where for the last equation, we used $b_n = n\phi(b_n)$ and the definition of the measure η_n in (5.3) to replace the sum by the integral. For $n \in \mathbb{N}$, let

$$h_n(\theta) = \int_{(1-\theta/b_n^2)x-\theta}^{\infty} e^{-z-z^2/(2b_n^2)} dz, \quad \theta \in \mathbb{R}.$$

Clearly, as $n \rightarrow \infty$, h_n converges uniformly on compact sets to the function $h(\theta) = \exp(\theta - x)$. Note that h and h_n are continuous functions on \mathbb{R} . Put $\omega = (\lambda, \theta, \gamma)$ and observe for $K > 0$ that

$$\begin{aligned} & \left| \int_{[0, \infty] \times \mathbb{R}^2} h_n(\theta) \eta_n(d\omega) - \int_{[0, \infty] \times \mathbb{R}^2} h(\theta) \eta(d\omega) \right| \\ & \leq \left| \int_{[0, \infty] \times \mathbb{R}^2} h_n(\theta) \mathbf{1}_{h_n > K} \eta_n(d\omega) - \int_{[0, \infty] \times \mathbb{R}^2} h(\theta) \mathbf{1}_{h > K} \eta(d\omega) \right| \\ & \quad + \left| \int_{[0, \infty] \times \mathbb{R}^2} h_n(\theta) \mathbf{1}_{h_n < K} \eta_n(d\omega) - \int_{[0, \infty] \times \mathbb{R}^2} h(\theta) \mathbf{1}_{h < K} \eta(d\omega) \right|. \end{aligned} \quad (5.26)$$

By Theorem 5.5 in Billingsley (1968) (see also the remark after the theorem), $\eta_n h_n^{-1}$ converges weakly to ηh^{-1} . Moreover, since $h \mathbf{1}_{h < K}$ and the $h_n \mathbf{1}_{h_n < K}$ are uniformly bounded in n , the second summand in (5.26) converges to 0 as $n \rightarrow \infty$, for arbitrary $K > 0$. By the uniform integrability condition (5.4) and Fatou's Lemma we have $\int_{[0, \infty] \times \mathbb{R}^2} h(\theta) \eta(d\omega) < \infty$ and hence, also the first summand in (5.26) tends to zero as $K, n \rightarrow \infty$. Consequently,

$$-\sum_{i=1}^n \mathbb{P} \left(X_{i,n}^{(1)} > u_n(x) \right) \rightarrow -\int_{[0, \infty] \times \mathbb{R}^2} \exp[-(x - \theta)] \eta(d\omega), \quad (5.27)$$

Similarly, we get

$$-\sum_{i=1}^n \mathbb{P} \left(X_{i,n}^{(2)} > u_n(y) \right) \rightarrow -\int_{[0, \infty] \times \mathbb{R}^2} \exp[-(y - \gamma)] \eta(d\omega). \quad (5.28)$$

It now also follows from (5.7), (5.25), (5.27) and (5.28) that the remainder term R_n converges to zero as $n \rightarrow \infty$.

We now turn to the third term in (5.24).

$$\begin{aligned}
& \sum_{i=1}^n \mathbb{P}(X_{i,n}^{(1)}/\sigma_{i,n,1} > u_n(x)/\sigma_{i,n,1}, X_{i,n}^{(2)}/\sigma_{i,n,2} > u_n(y)/\sigma_{i,n,2}) \\
&= \sum_{i=1}^n \int_{u_n(y)/\sigma_{i,n,2}}^{\infty} \left[1 - \Phi \left(\frac{u_n(x)/\sigma_{i,n,1} - \rho_{i,n}z}{(1 - \rho_{i,n}^2)^{1/2}} \right) \right] \phi(z) dz \\
&= \frac{1}{n} \sum_{i=1}^n \int_{y/\sigma_{i,n,2} - b_n^2(1-1/\sigma_{i,n,2})}^{\infty} \left[1 - \Phi \left(\frac{u_n(x)/\sigma_{i,n,1} - \rho_{i,n}u_n(z)}{(1 - \rho_{i,n}^2)^{1/2}} \right) \right] e^{-z-z^2/(2b_n^2)} dz \\
&= \int_{[0,\infty] \times \mathbb{R}^2} \int_{(1-\gamma/b_n^2)y-\gamma}^{\infty} [1 - \Phi(s_n(\lambda, \theta, z, x))] e^{-z-z^2/(2b_n^2)} dz \eta_n(d\omega),
\end{aligned}$$

where we used $b_n = n\phi(b_n)$ for the second last equation and s_n is defined by

$$s_n(\lambda, \theta, z, x) := \frac{\lambda}{(1 - \lambda^2/b_n^2)^{1/2}} + \frac{(1 - \theta/b_n^2)x - z - \theta}{(1 - \lambda^2/b_n^2)^{1/2}2\lambda} + \frac{\lambda z}{(1 - \lambda^2/b_n^2)^{1/2}b_n^2}.$$

For the last equation, we replaced the sum by the integral w.r.t. the empirical measure η_n in (5.3). Note that for $i \in \{1, \dots, n\}$, in fact a short computation yields

$$s_n \left(\sqrt{b_n^2(1 - \rho_{i,n})/2}, b_n^2(1 - 1/\sigma_{i,n,1}), z, x \right) = \frac{u_n(x)/\sigma_{i,n,1} - \rho_{i,n}u_n(z)}{(1 - \rho_{i,n}^2)^{1/2}}.$$

For $n \in \mathbb{N}$, let

$$g_n(\lambda, \theta, \gamma) = \mathbf{1}_{\lambda \leq b_n} \int_{(1-\gamma/b_n^2)y-\gamma}^{\infty} [1 - \Phi(s_n(\lambda, \theta, z, x))] e^{-z-z^2/(2b_n^2)} dz$$

be a measurable function on $[0, \infty] \times \mathbb{R}^2$. It is easy to see, that as $n \rightarrow \infty$, g_n converges pointwise to the function

$$g(\lambda, \theta, \gamma) = \int_{y-\gamma}^{\infty} [1 - \Phi(s(\lambda, \theta, z, x))] e^{-z} dz,$$

with

$$s(\lambda, \theta, z, x) := \lambda + \frac{x - z - \theta}{2\lambda}.$$

Note that g is a continuous function on $[0, \infty] \times \mathbb{R}^2$ and

$$g(0, \theta, \gamma) = g_n(0, \theta, \gamma) = \exp(-\max(x - \theta, y - \gamma))$$

and $g(\infty, \theta, \gamma) = g_n(\infty, \theta, \gamma) = 0$, for any $(\theta, \gamma) \in \mathbb{R}^2$ and n sufficiently large. Here, the values are understood as the limits as $\lambda \rightarrow 0$ and $\lambda \rightarrow \infty$ (using dominated convergence), respectively, e.g., $\lim_{\lambda \rightarrow 0} g(\lambda, \theta, \gamma) = \int_{y-\gamma}^{\infty} \mathbf{1}_{z > x-\theta} e^{-z} dz = \exp(-\max(x-\theta, y-\gamma))$. In order to establish the weak convergence $\eta_n g_n^{-1} \Rightarrow \eta g^{-1}$, we show that g_n converges uniformly on compact sets to g as $n \rightarrow \infty$. To this end, let $C = [0, \infty] \times [\theta_0, \theta_1] \times [\gamma_0, \gamma_1]$ be an arbitrary compact set in $[0, \infty] \times \mathbb{R}^2$ and let $\epsilon > 0$ be given. First, note that instead of g_n it suffices to consider the function \tilde{g}_n , defined as

$$\tilde{g}_n(\lambda, \theta, \gamma) = \mathbf{1}_{\lambda \leq b_n} \int_{(1-\gamma/b_n^2)y-\gamma}^{\infty} [1 - \Phi(s_n(\lambda, \theta, z, x))] e^{-z} dz,$$

since for n large enough

$$\sup_{(\lambda, \theta, \gamma) \in C} |g_n(\lambda, \theta, \gamma) - \tilde{g}_n(\lambda, \theta, \gamma)| \leq \mathbf{1}_{\lambda \leq b_n} \int_{-2|y|-\gamma_1}^{\infty} e^{-z} (1 - e^{-z^2/(2b_n^2)}) dz \rightarrow 0,$$

as $n \rightarrow \infty$, by dominated convergence. Further, for any $\epsilon > 0$, let $z_1 > -\log \epsilon$ which implies $\int_{z_1}^{\infty} e^{-z} dz < \epsilon$. We note that for n large enough

$$\begin{aligned} s_n(\lambda, \theta, z, x) &\geq \left(1 - \lambda^2/b_n^2\right)^{-1/2} \left(\lambda \left(1 + \frac{-2|y| - \gamma_1}{b_n^2}\right) + \frac{-2|x| - z_1 - \theta_1}{2\lambda}\right) \\ &\geq \left(\frac{\lambda}{2} + \frac{-2|x| - z_1 - \theta_1}{2\lambda}\right), \end{aligned}$$

for all $\lambda \leq b_n$, $-2|y| - \gamma_1 \leq z \leq z_1$ and $(\lambda, \theta, \gamma) \in C$, independently of $n \in \mathbb{N}$. Hence, there is a $\lambda_1 > 0$ s.t. for all $\lambda_1 \leq \lambda \leq b_n$

$$1 - \Phi(s_n(\lambda, \theta, z, x)) < \epsilon e^{-2|y|-\gamma_1}.$$

Thus, for all $n \in \mathbb{N}$ large enough,

$$\sup_{(\lambda, \theta, \gamma) \in C, \lambda \geq \lambda_1} \tilde{g}_n(\lambda, \theta, \gamma) \leq \mathbf{1}_{\lambda \leq b_n} \left(\int_{-2|y|-\gamma_1}^{z_1} \epsilon e^{-2|y|-\gamma_1} e^{-z} dz + \int_{z_1}^{\infty} e^{-z} dz \right) \leq 2\epsilon,$$

and in the same manner, $\sup_{(\lambda, \theta, \gamma) \in C, \lambda \geq \lambda_1} g(\lambda, \theta, \gamma) \leq 2\epsilon$. Furthermore, we observe

$$\lim_{\lambda \rightarrow 0} \Phi(s_n(\lambda, \theta, z, x)) = \mathbf{1}_{z < (1-\theta/b_n^2)x-\theta} \quad \text{and} \quad \lim_{\lambda \rightarrow 0} \Phi(s(\lambda, \theta, z, x)) = \mathbf{1}_{z < x-\theta}.$$

Choose $n_0 \in \mathbb{N}$ such that for all $n > n_0$ and all $\theta \in [\theta_0, \theta_1]$ we find an open interval (a_θ, b_θ) of size $\epsilon/2$ that contains $\{(1-\theta/b_n^2)x-\theta, x-\theta\}$. Put $I_\theta = (a_\theta - \epsilon/4, b_\theta + \epsilon/4)$, then we find a $\lambda_0 > 0$, s.t. for all $(\lambda, \theta, \gamma) \in C, \lambda \leq \lambda_0$, $z \in I_\theta$ and $n > n_0$, we have

$|\Phi(s_n(\lambda, \theta, z, x)) - \Phi(s(\lambda, \theta, z, x))| \leq \epsilon$. Consequently,

$$\begin{aligned} & \sup_{(\lambda, \theta, \gamma) \in C, \lambda \leq \lambda_0} |\tilde{g}_n(\lambda, \theta, \gamma) - g(\lambda, \theta, \gamma)| \\ & \leq \sup_{(\lambda, \theta, \gamma) \in C, \lambda \leq \lambda_0} \int_{-2|y|-\gamma_1}^{\infty} \left(\mathbf{1}_{z \in I_\theta} + \epsilon \mathbf{1}_{z \in \mathbb{R} \setminus I_\theta} \right) e^{-z} dz \leq 2\epsilon e^{2|y|+\gamma_1}. \end{aligned}$$

Choose $n_1 \in \mathbb{N}$, s.t. $b_{n_1} > \lambda_1$. For $\lambda_0 \leq \lambda \leq \lambda_1$ and $n > n_1$,

$$\begin{aligned} & |s_n(\lambda, \theta, z, x) - s(\lambda, \theta, z, x)| \tag{5.29} \\ & = \left| \left(\lambda + \frac{x - z - \theta}{2\lambda} \right) \left(1 - \frac{1}{(1 - \lambda_1^2/b_n^2)^{1/2}} \right) - \frac{\lambda^2 z - \theta}{(1 - \lambda_1^2/b_n^2)^{1/2} b_n^2 2\lambda} \right| \\ & \leq M_1 \left| 1 - \frac{1}{(1 - \lambda_0^2/b_n^2)^{1/2}} \right| + \frac{M_2}{(1 - \lambda_1^2/b_n^2)^{1/2} b_n^2} \rightarrow 0 \end{aligned}$$

for $n \rightarrow \infty$, uniformly in $z \in [-2|y| - \gamma_1, z_1]$ and $(\lambda, \theta, \gamma) \in C$ with $\lambda_0 \leq \lambda \leq \lambda_1$. Here, M_1 and M_2 are positive constants that only depend on $x, y, \lambda_0, \lambda_1, \theta_0, \theta_1, \gamma_1$. Let $n_2 \in \mathbb{N}$, s.t. for all $n > \max(n_1, n_2)$ the difference in (5.29) is less than or equal to $\epsilon e^{-2|y|-\gamma_1}$. By the Lipschitz continuity of Φ , we obtain for all $\lambda_0 \leq \lambda \leq \lambda_1$ and $(\lambda, \theta, \gamma) \in C$,

$$\begin{aligned} & \int_{-2|y|-\gamma_1}^{\infty} |\Phi(s_n(\lambda, \theta, z, x)) - \Phi(s(\lambda, \theta, z, x))| e^{-z} dz \\ & \leq \int_{-2|y|-\gamma_1}^{z_1} |s_n(\lambda, \theta, z, x) - s(\lambda, \theta, z, x)| e^{-z} dz + \int_{z_1}^{\infty} e^{-z} dz \\ & \leq \int_{-2|y|-\gamma_1}^{z_1} \epsilon e^{-2|y|-\gamma_1} e^{-z} dz + \int_{z_1}^{\infty} e^{-z} dz \leq 2\epsilon. \end{aligned}$$

Putting the parts together yields

$$\lim_{n \rightarrow \infty} \sup_{(\lambda, \theta, \gamma) \in C} |\tilde{g}_n(\lambda, \theta, \gamma) - g(\lambda, \theta, \gamma)| = 0.$$

The assumptions of Theorem 5.5 in Billingsley (1968) are satisfied and therefore $\eta_n g_n^{-1}$ converges weakly to ηg^{-1} . By a similar argument as in (5.26) together with the uniform integrability condition (5.4) we obtain for $n \rightarrow \infty$

$$\sum_{i=1}^n \mathbb{P}(X_{i,n}^{(1)} > u_n(x), X_{i,n}^{(2)} > u_n(y)) \rightarrow \int_{[0, \infty] \times \mathbb{R}^2} g(\lambda, \theta, \gamma) \eta(d(\lambda, \theta, \gamma)).$$

Finally, partial integration gives

$$\begin{aligned} g(\lambda, \theta, \gamma) &= e^{-(y-\gamma)} + e^{-(x-\theta)} - \Phi\left(\lambda + \frac{y-x+\theta-\gamma}{2\lambda}\right) e^{-(x-\theta)} \\ &\quad - \Phi\left(\lambda - \frac{y-x+\theta-\gamma}{2\lambda}\right) e^{-(y-\gamma)}. \end{aligned}$$

Together with (5.24), (5.27), (5.28) and the fact that R_n converges to zero, this implies the desired result. \square

Proof of Theorem 12. The sufficient part is a simple consequence of Theorem 11, where the covariance matrix of $\mathbf{X}_{i,n}$ is given by

$$\begin{pmatrix} 1 & \rho_{i,n} \\ \rho_{i,n} & 1 \end{pmatrix}.$$

For the necessary part, suppose that the sequence $(\max_{i=1,\dots,n} b_n(\mathbf{X}_{i,n} - b_n))_{n \in \mathbb{N}}$ of bivariate random vectors converges in distribution to some random vector Y . Let the ν_n , $n \in \mathbb{N}$, be defined as in (5.8) and assume that the sequence $(\nu_n)_{n \in \mathbb{N}} \subset \mathcal{M}_1([0, \infty])$ does not converge. Then, by sequential compactness, it has at least two different accumulation points $\nu, \tilde{\nu} \in \mathcal{M}_1([0, \infty])$. By the first part of this theorem, $(\max_{i=1,\dots,n} b_n(\mathbf{X}_{i,n} - b_n))_{n \in \mathbb{N}}$ converges in distribution to $F_\nu \equiv F_{\tilde{\nu}}$. It now suffices to show that $F_\nu \equiv F_{\tilde{\nu}}$ implies $\nu \equiv \tilde{\nu}$ to conclude that $(\nu_n)_{n \in \mathbb{N}} \subset \mathcal{M}_1([0, \infty])$ converges to some measure ν and that Y has distribution F_ν .

The fact that there is a one-to-one correspondence between Hüsler-Reiss distributions F_λ and the dependence parameter $\lambda \in [0, \infty]$ is straightforward (Kabluchko *et al.*, 2009). Showing a similar result in our case, however, requires more effort.

To this end, for two measures $\nu_1, \nu_2 \in \mathcal{M}_1([0, \infty])$ define random variables Y_1 and Y_2 with distribution F_{ν_1} and F_{ν_2} , respectively. First, suppose that $\nu_1(\{\infty\}) = \nu_2(\{\infty\}) = 0$. For $j = 1, 2$, by Remark 16 we have the stochastic representation $Y_j = \max_{i \in \mathbb{N}} (U_{i,j}, U_{i,j} + B_{i,j})$, where $\sum_{i=1}^{\infty} \delta_{U_{i,j}}$ are Poisson point process on \mathbb{R} with intensity $e^{-u} du$ and the $(B_{i,j})_{i \in \mathbb{N}}$ are i.i.d. copies of the random variable B_j with normal distribution $N(-2S_j^2, 4S_j^2)$, where S_j is ν_j -distributed. Assume that

$$F_{\nu_1}(x, y) = F_{\nu_2}(x, y), \quad \text{for all } x, y \in \mathbb{R}, \quad (5.30)$$

i.e. the max-stable distributions of Y_1 and Y_2 are equal. Since a Poisson point process is determined by its intensity on a generating system of the σ -algebra, it follows that the point processes $\Pi_1 = \sum_{i=1}^{\infty} \delta_{(U_{i,1}, U_{i,1} + B_{i,1})}$ and $\Pi_2 = \sum_{i=1}^{\infty} \delta_{(U_{i,2}, U_{i,2} + B_{i,2})}$ are equal in distribution. Therefore, the measurable mapping

$$h : \mathbb{R}^2 \rightarrow \mathbb{R}^2, (x_1, x_2) \mapsto (x_1, x_2 - x_1)$$

induces two Poisson point processes $h(\Pi_1)$ and $h(\Pi_2)$ on \mathbb{R}^2 with coinciding intensity measures $e^{-u} du \mathbb{P}_{B_1}(dx)$ and $e^{-u} du \mathbb{P}_{B_2}(dx)$, respectively. Hence, B_1 and B_2 have the same distribution. Denote by ψ_j the Laplace transform of the Gaussian mixture B_j , $j = 1, 2$. A straightforward calculation yields for $u \in (0, 1)$

$$\psi_j(u) = \mathbb{E} \exp(uB_j) = \int_{[0, \infty)} \exp(-2\lambda^2(u - u^2)) \nu_j(d\lambda), \quad j = 1, 2.$$

By Lemma 7 in Kabluchko *et al.* (2009) this implies the equality of measures $\nu_1^2(d\lambda) = \nu_2^2(d\lambda)$, where ν_j^2 is the image measure of ν_j under the transformation $[0, \infty] \rightarrow [0, \infty]$, $\lambda \mapsto \lambda^2$, for

$j = 1, 2$. Hence, it also holds that $\nu_1 \equiv \nu_2$.

For arbitrary $\nu_1, \nu_2 \in \mathcal{M}_1([0, \infty])$, we first need to show that $\nu_1(\{\infty\}) = \nu_2(\{\infty\})$. For $j = 1, 2$, observe that for $n \in \mathbb{N}$

$$\begin{aligned} & -\log F_{\nu_j}(-n, 0) + \log F_{\nu_j}(-n, n) \\ &= \int_{[0, \infty)} \Phi\left(\lambda + \frac{n}{2\lambda}\right) e^n + \Phi\left(\lambda - \frac{n}{2\lambda}\right) - \Phi\left(\lambda + \frac{n}{\lambda}\right) e^n - \Phi\left(\lambda - \frac{n}{\lambda}\right) e^{-n} \nu_j(d\lambda) \\ & \quad + (1 - e^{-n}) \nu_j(\{\infty\}). \end{aligned}$$

Since the curvature of Φ is negative on the positive real line, we have the estimate

$$e^n \left| \Phi\left(\lambda + \frac{n}{2\lambda}\right) - \Phi\left(\lambda + \frac{n}{\lambda}\right) \right| \leq \frac{n}{2\lambda\sqrt{2\pi}} e^n e^{-(\lambda+n/(2\lambda))^2/2},$$

where the latter term converges pointwise to zero as $n \rightarrow \infty$. Moreover, it is uniformly bounded in $n \in \mathbb{N}$ and $\lambda \in [0, \infty)$ by a constant and hence, by dominated convergence

$$\lim_{n \rightarrow \infty} -\log F_{\nu_j}(-n, 0) + \log F_{\nu_j}(-n, n) = \nu_j(\{\infty\}), \quad j = 1, 2.$$

It therefore follows from (5.30) that $\nu_1(\{\infty\}) = \nu_2(\{\infty\})$. If $\nu_1(\{\infty\}) < 1$ we apply the above to the restricted probability measures $\nu_j(\cdot \cap [0, \infty)) / (1 - \nu_j(\{\infty\}))$ on $[0, \infty)$, $j = 1, 2$, to obtain $\nu_1 \equiv \nu_2$.

The last claim of the theorem follows from the fact, that the integrand in (5.10) is bounded and continuous in λ for fixed $x, y \in \mathbb{R}$, and hence, for $\nu, \nu_n \in \mathcal{M}_1([0, \infty])$, $n \in \mathbb{N}$, weak convergence of ν_n to ν ensures the pointwise convergence of the distribution functions. \square

Proof of Corollary 1. The first statement is a consequence of Theorem 12, because every sequence of random vectors can be understood as a triangular array where the columns contain equal random vectors.

For the second claim, let $\nu \in \mathcal{M}_1([0, \infty])$ be an arbitrary probability measure. Similarly as in Remark 15, define an i.i.d. sequence $(R_i)_{i \in \mathbb{N}}$ of samples of ν . Choosing $\rho_i = \max(1 - 2R_i^2/b_i^2, -1)$ as correlation of \mathbf{X}_i yields

$$\nu_n = \frac{1}{n} \sum_{i=1}^n \delta_{\mathbf{1}_{R_i < b_i} R_i b_n / b_i + \mathbf{1}_{R_i > b_i} b_n}.$$

First, consider the measures $\tilde{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{R_i b_n / b_i}$, for $n \in \mathbb{N}$. For $y \in [0, \infty]$ with $\nu(\{y\}) = 0$ we observe

$$\tilde{\nu}_n([0, y]) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[0, y]}(R_i b_n / b_i). \quad (5.31)$$

Fix $\epsilon > 0$ and recall from (3.2) that $b_n / \sqrt{2 \log n} \rightarrow 1$ as $n \rightarrow \infty$. Hence, choose n large enough such that $i > n^{1/(1+\epsilon)^2}$ implies $b_n / b_i < 1 + \epsilon$. Let n_ϵ denote the smallest integer

which is strictly larger than $n^{1/(1+\epsilon)^2}$, then (5.31) yields

$$\begin{aligned} \left| \tilde{\nu}_n([0,y]) - \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[0,y]}(R_i) \right| &\leq \frac{n_\epsilon}{n} + \frac{1}{n} \left| \sum_{i=n_\epsilon}^n \mathbf{1}_{[0,y]}(R_i b_n/b_i) - \sum_{i=n_\epsilon}^n \mathbf{1}_{[0,y]}(R_i) \right| \\ &\leq \frac{n_\epsilon}{n} + \frac{1}{n} \sum_{i=n_\epsilon}^n \mathbf{1}_{(y/(1+\epsilon),y]}(R_i). \end{aligned}$$

Letting $n \rightarrow \infty$ gives

$$\lim_{n \rightarrow \infty} \left| \tilde{\nu}_n([0,y]) - \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[0,y]}(R_i) \right| \leq \nu((y/(1+\epsilon),y]), \quad \text{a.s.}$$

Since ϵ was arbitrary and $\nu(\{y\}) = 0$, it follows from the law of large numbers that $\tilde{\nu}_n$ converges a.s. weakly to ν , as $n \rightarrow \infty$. Similarly, one can see that the sequence $(\nu_n)_{n \in \mathbb{N}}$ has a.s. the same limit as $(\tilde{\nu}_n)_{n \in \mathbb{N}}$, as $n \rightarrow \infty$. \square

Proof of Theorem 13. Let $u_n(z) = b_n + z/b_n$ for $z \in \mathbb{R}$, $u_n(\mathbf{x}) = (u_n(x_1), \dots, u_n(x_d))^\top$ for $\mathbf{x} \in \mathbb{R}^d$ and for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ write $\mathbf{x} > \mathbf{y}$ if $x_i > y_i$ for all $1 \leq i \leq d$.

Let $\mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$ be a fixed vector and $A_{i,n}^l = \{X_{i,n}^{(l)} \leq u_n(x_l)\}$ for $n \in \mathbb{N}$, $1 \leq i \leq n$ and $1 \leq l \leq d$.

$$\begin{aligned} \log \mathbb{P} \left(\max_{i=1, \dots, n} X_{i,n}^{(1)} \leq u_n(x_1), \dots, \max_{i=1, \dots, n} X_{i,n}^{(d)} \leq u_n(x_d) \right) \\ = \sum_{i=1}^n \log \mathbb{P} \left[\bigcap_{l=1}^d A_{i,n}^l \right] = - \sum_{i=1}^n \mathbb{P} \left[\bigcup_{l=1}^d (A_{i,n}^l)^C \right] + R_n \end{aligned} \quad (5.32)$$

where R_n is a remainder term from the Taylor expansion of \log . Using the same arguments as for the remainder term in (5.25), we conclude that R_n converges to zero as $n \rightarrow \infty$. By the additivity formula we have

$$- \mathbb{P} \left[\bigcup_{l=1}^d (A_{i,n}^l)^C \right] = \sum_{l=1}^d (-1)^l \sum_{m: 1 \leq m_1 < \dots < m_l \leq d} \mathbb{P} \left[\bigcap_{k=1}^l (A_{i,n}^{m_k})^C \right]. \quad (5.33)$$

Consequently, by (5.32) and (5.33) it suffices to show that

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(\mathbf{X}_{i,n} > u_n(\mathbf{x})) = \int_{[0, \infty)^{d \times d}} h_{d, (1, \dots, d), \Lambda}(x_1, \dots, x_d) \eta(d\Lambda). \quad (5.34)$$

Let $\mathbf{Z} = (Z_1, \dots, Z_d)$ be a standard normal random vector with independent margins and let $K = \{1, \dots, d-1\}$. For a vector $\mathbf{x} \in \mathbb{R}^d$ let $\mathbf{x}_K = (x_1, \dots, x_{d-1})$. If $A = (a_{j,k})_{1 \leq j, k \leq d} \in \mathbb{R}^{d \times d}$ is a matrix, let $A_{d,K} = (a_{d,1}, \dots, a_{d,d-1})$, $A_{K,d} = (a_{1,d}, \dots, a_{d-1,d})$ and $A_{K,K} = (a_{j,k})_{j, k \in K}$. Similarly as in the proof of Theorem 1.1 in Hashorva *et al.* (2012), we define a new matrix

$B_{i,n} \in \mathbb{R}^{(d-1) \times (d-1)}$ by

$$B_{i,n} B_{i,n}^\top = (\Sigma_{i,n})_{K,K} - \boldsymbol{\sigma}_{i,n} \boldsymbol{\sigma}_{i,n}^\top, \quad \boldsymbol{\sigma}_{i,n} = (\Sigma_{i,n})_{K,d}, \quad (5.35)$$

which is well-defined since $(\Sigma_{i,n})_{K,K} - \boldsymbol{\sigma}_{i,n} \boldsymbol{\sigma}_{i,n}^\top$ is positive definite as the Schur complement of $(\Sigma_{i,n})_{d,d}$ in the positive definite matrix $\Sigma_{i,n}$. This enables us to write the vector $\mathbf{X}_{i,n}$ as the joint stochastic representation

$$\left(X_{i,n}^{(1)}, \dots, X_{i,n}^{(d-1)} \right) \stackrel{d}{=} B_{i,n} \mathbf{Z}_K + Z_d \boldsymbol{\sigma}_{i,n}, \quad X_{i,n}^{(d)} \stackrel{d}{=} Z_d.$$

Therefore, since Z_d is independent of \mathbf{Z}_K ,

$$\begin{aligned} \mathbb{P}(\mathbf{X}_{i,n} > u_n(\mathbf{x})) &= \mathbb{P}(B_{i,n} \mathbf{Z}_K + Z_d \boldsymbol{\sigma}_{i,n} > u_n(\mathbf{x}_K), Z_d > u_n(x_d)) \\ &= \int_{x_d}^{\infty} \mathbb{P}(B_{i,n} \mathbf{Z}_K + u_n(s) \boldsymbol{\sigma}_{i,n} > u_n(\mathbf{x}_K)) b_n^{-1} \phi(b_n) e^{-s-s^2/(2b_n^2)} ds \\ &= \frac{1}{n} \int_{x_d}^{\infty} S \left(\left(b_n^2 (\mathbf{1}\mathbf{1}^\top - \Sigma_{i,n}) \right)_{K,d} + x_K - s \mathbf{1} + s b_n^{-2} \left(b_n^2 (\mathbf{1}\mathbf{1}^\top - \Sigma_{i,n}) \right)_{K,d} \middle| b_n^2 B_{i,n} B_{i,n}^\top \right) \\ &\quad \cdot e^{-s-s^2/(2b_n^2)} ds. \end{aligned} \quad (5.36)$$

It follows from the definition of $B_{i,n}$ in equation (5.35) that

$$\begin{aligned} B_{i,n} B_{i,n}^\top &= (\mathbf{1}\mathbf{1}^\top - \Sigma_{i,n})_{K,d} \mathbf{1}^\top + \mathbf{1} (\mathbf{1}\mathbf{1}^\top - \Sigma_{i,n})_{d,K} - (\mathbf{1}\mathbf{1}^\top - \Sigma_{i,n})_{K,K} \\ &\quad - (\mathbf{1}\mathbf{1}^\top - \Sigma_{i,n})_{K,d} (\mathbf{1}\mathbf{1}^\top - \Sigma_{i,n})_{d,K}. \end{aligned}$$

Together with (5.36) and the definition of η_n this yields

$$\sum_{i=1}^n \mathbb{P}(\mathbf{X}_{i,n} > u_n(\mathbf{x})) = \int_D p_n(A) \eta_n^2(dA),$$

where p_n is a measurable function from D to $[0, \infty)$ given by

$$\begin{aligned} p_n(A) &= \int_{x_d}^{\infty} S \left(2A_{K,d} + x_K - s \mathbf{1} + 2b_n^{-2} s A_{K,d} \middle| \Gamma_{d,(1,\dots,d)}(\sqrt{A}) - 4b_n^{-2} A_{K,d} A_{d,K} \right) \\ &\quad e^{-s-s^2/(2b_n^2)} ds. \end{aligned}$$

Further, let p be the measurable function from D to $[0, \infty)$

$$p(A) = \int_{x_d}^{\infty} S \left(2A_{K,d} + x_K - s \mathbf{1} \middle| \Gamma_{d,(1,\dots,d)}(\sqrt{A}) \right) e^{-s} ds.$$

Note that $\eta_n \Rightarrow \eta$ if and only if $\eta_n^2 \Rightarrow \eta^2$. In view of (5.34) it suffices to show that

$$\lim_{n \rightarrow \infty} \int_D p_n(A) \eta_n^2(dA) = \int_D p(A) \eta^2(dA). \quad (5.37)$$

To this end, let $A_0 \in D$ and $\{A_n, n \in \mathbb{N}\}$ be a sequence in D that converges to A_0 . We will show that $p_n(A_n) \rightarrow p(A_0)$ as $n \rightarrow \infty$. By dominated convergence it is sufficient to show the convergence of the survivor functions. Since A_0 is in D , recall that $\Gamma_{d,(1,\dots,d)}(\sqrt{A_0})$ is in the space $\mathcal{M}_{(d-1)}$ of $(d-1)$ -dimensional, non-degenerate covariance matrices. Moreover, since $\mathcal{M}_{(d-1)} \subset \mathbb{R}^{(d-1) \times (d-1)}$ is open and $\Gamma_{d,(1,\dots,d)}(\sqrt{A_n}) - b_n^{-2}4(A_n)_{K,d}(A_n)_{d,K}$ converges to $\Gamma_{d,(1,\dots,d)}(\sqrt{A_0})$, there is an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ we have $\Gamma_{d,(1,\dots,d)}(\sqrt{A_n}) - b_n^{-2}4(A_n)_{K,d}(A_n)_{d,K} \in \mathcal{M}_{(d-1)}$. Since also $2(A_n)_{K,d} + x_K - s\mathbf{1} + b_n^{-2}s2(A_n)_{K,d}$ converges to $2(A_0)_{K,d} + x_K - s\mathbf{1}$ as $n \rightarrow \infty$, we conclude that the survivor functions converge and consequently $p_n(A_n) \rightarrow p(A_0)$. Applying Theorem 5.5 in Billingsley (1968) yields (5.37) and therefore concludes the proof. \square

The following lemma is needed for the proof of Theorem 14 below.

Lemma 1. *For $q \in \mathbb{N}$, let Z be random vector in \mathbb{R}^q and $V_i, i \in \mathbb{N}$, a sequence of i.i.d. copies of V . Further, suppose $\epsilon_{i,n}, n \in \mathbb{N}$ and $1 \leq i \leq n$, is a triangular array of errors such that $\|\epsilon_{i,n}\|$ goes to 0 a.s. uniformly for bounded V_i , as $n \rightarrow \infty$. For $f \in C_b(\mathbb{R}^q)$ the following convergence holds almost surely,*

$$\frac{1}{n} \sum_{i=1}^n f(V_i + \epsilon_{i,n}) \rightarrow \mathbb{E}f(V), \quad \text{as } n \rightarrow \infty.$$

Proof. For arbitrary $N > 0$ note that

$$\frac{1}{n} \sum_{i=1}^n f(V_i + \epsilon_{i,n}) = \frac{1}{n} \sum_{i=1}^n f(V_i + \epsilon_{i,n}) \mathbf{1}_{\|V_i\| \leq N} + \frac{1}{n} \sum_{i=1}^n f(V_i + \epsilon_{i,n}) \mathbf{1}_{\|V_i\| > N}. \quad (5.38)$$

Since for arbitrary $\delta > 0$ we find $n_0 \in \mathbb{N}$ such that $\epsilon_{i,n} < \delta$ a.s. for all $n > n_0$ and $1 \leq i \leq n$ with $\|V_i\| \leq N$, it follows from the uniform continuity of f on compact sets that almost surely

$$\frac{1}{n} \sum_{i=1}^n |f(V_i + \epsilon_{i,n}) - f(V_i)| \mathbf{1}_{\|V_i\| \leq N} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (5.39)$$

Further, observe that as $n \rightarrow \infty$, the absolute value of the second sum in (5.38) is a.s. bounded by $\mathbb{P}(\|V\| > N) \sup_{x \in \mathbb{R}^q} |f(x)|$, which converges to zero as $N \rightarrow \infty$. Thus, the assertion follows from the law of large numbers, the triangle inequality, (5.38) and (5.39). \square

Proof of Theorem 14. First, note that s_n defined in (5.20) goes to 0 for $n \rightarrow \infty$. For $m \in \mathbb{N}$ and pairwise different $t_1, \dots, t_m \in \mathbb{R}^d$ we interpret the random vector $(Y(t_1), \dots, Y(t_m))$, conditionally on $\{S_i\}_{i \in \mathbb{N}}$, as the normalized maximum of n independent, yet differently distributed, m -variate normal random vectors

$$\mathbf{Z}_{i,n} = \left(X_i(S_i^{2/\alpha} s_n t_1), \dots, X_i(S_i^{2/\alpha} s_n t_m) \right), \quad n \in \mathbb{N}, i \in \{1, \dots, n\}.$$

In view of (5.19), their covariance matrices are specified by

$$\begin{aligned} (1/2)b_n^2(\mathbf{1}\mathbf{1}^\top - \text{Cov}(\mathbf{Z}_{i,n})) &= \left(b_n^2 \gamma(S_i^{2/\alpha}(t_j - t_k)) L(s_n) s_n^\alpha + b_n^2 o(L(s_n) s_n^\alpha) \right)_{1 \leq j, k \leq m} \\ &= \left(\gamma(S_i^{2/\alpha}(t_j - t_k)) + o(1) \right)_{1 \leq j, k \leq m}, \end{aligned}$$

because of (5.20) and the continuity of L . The remainder term goes to 0 for $n \rightarrow \infty$ uniformly in $S_i^{2/\alpha} t_j, S_i^{2/\alpha} t_k \in \mathbb{R}^d$, as long as they stay bounded. In order to apply Theorem 13 we define η_n as in (5.17) by

$$\eta_n = \frac{1}{n} \sum_{i=1}^n \delta \left(\sqrt{\gamma(S_i^{2/\alpha}(t_j - t_k)) + o(1)} \right)_{1 \leq j, k \leq m}.$$

For a continuous, bounded function $f \in C_b([0, \infty)^{m \times m})$, we apply Lemma 1 with $V = \left(\sqrt{\gamma(S_i^{2/\alpha}(t_j - t_k))} \right)_{1 \leq j, k \leq m}$ to obtain

$$\int_{[0, \infty)^{m \times m}} f(\Lambda) \eta_n(d\Lambda) \rightarrow \int_{[0, \infty)^{m \times m}} f(\Lambda) \eta(d\Lambda), \quad \text{as } n \rightarrow \infty,$$

where η is as in (5.21). Thus, η_n converges weakly to η . It remains to check the condition $\eta^2(D) = 1$, which is fulfilled if $\{\lambda A_0^2, \lambda > 0\} \subset D$. Since γ is a variogram, A_0^2 is conditionally negative definite for $\lambda > 0$. Suppose that there is an $\mathbf{x} \in \mathbb{R}^m \setminus \{0\}$ s.t. $\sum_{i=1}^m x_i = 0$ and $\mathbf{x}^\top A_0^2 \mathbf{x} = 0$. For some $t_0 \notin \{t_1, \dots, t_m\}$ this gives

$$0 = - \sum_{i,j=1}^m x_i x_j \gamma(t_i - t_j) = \sum_{i,j=1}^m x_i x_j [\gamma(t_i - t_0) + \gamma(t_j - t_0) - \gamma(t_i - t_j)] = \mathbf{x}^\top \mathbf{C} \mathbf{x}, \quad (5.40)$$

where $\mathbf{C} = (C(t_i, t_j))_{1 \leq i, j \leq m}$ is a covariance matrix of a fractional Brownian motion $\{X(t) : t \in \mathbb{R}^d\}$ with $X(t_0) = 0$ and covariance function $C(s, t) = \|s - t_0\|^\alpha + \|t - t_0\|^\alpha - \|s - t\|^\alpha$, for $s, t \in \mathbb{R}^d$. Thus, (5.40) implies that the Gaussian random vector $(X(t_1), \dots, X(t_m))$ is degenerate. However, this contradicts the properties of fractional Brownian motion stated in Lemma 7.1 in Pitt (1978) and therefore $\lambda A_0^2 \in D$ for all $\lambda > 0$.

Hence, conditionally on $\{S_i\}_{i \in \mathbb{N}}$, the distribution function of $(Y(t_1), \dots, Y(t_m))$ converges by Theorem 13 pointwise to the distribution function H_η . Finally, it suffices to note that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{E} [\mathbb{P}(Y(t_1) \leq y_1, \dots, Y(t_m) \leq y_m | \{S_i\}_{i \in \mathbb{N}})] \\ &= \mathbb{E} \left[\lim_{n \rightarrow \infty} \mathbb{P}(Y(t_1) \leq y_1, \dots, Y(t_m) \leq y_m | \{S_i\}_{i \in \mathbb{N}}) \right] \\ &= \mathbb{E} H_\eta(y_1, \dots, y_m) = H_\eta(y_1, \dots, y_m), \end{aligned}$$

where the first equation follows by the dominated convergence theorem. \square

Proof of Proposition 4. By the construction of the spectral measure (cf. (de Haan & Resnick,

1977)), the spectral density s_ν can be computed by

$$s_\nu(\theta) = r^3 \left(\frac{\partial^2}{\partial x \partial y} \log F_\nu(x, y) \right)_{x=r \cos(\theta), y=r \sin(\theta)}, \quad \theta \in [0, \pi/2].$$

Applying this to $F_\nu = F_{\nu_\sigma}$ and $F_\nu = F_{\nu_b}$ from (5.14) and (5.16), respectively, yields the desired result. \square

Bibliography

- Albin, J. M. P. (1993). Extremes of totally skewed stable motion. *Statist Probab Lett* **16**:219–224.
- Bacro, J.-N. & Gaetan, C. (2012). Estimation of spatial max-stable models using threshold exceedances. Available from <http://arxiv.org/abs/1205.1107>.
- Balkema, A. A. & Resnick, S. I. (1977). Max-infinite divisibility. *J Appl Probability* **14**:309–319.
- Ballani, F. & Schlather, M. (2011). A construction principle for multivariate extreme value distributions. *Biometrika* **98**:633–645.
- Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand J Statist* **24**:1–13.
- Berg, C., Christensen, J. P. R. & Ressel, P. (1984). *Harmonic Analysis on Semigroups*, vol. 100 of *Graduate Texts in Mathematics*. Springer-Verlag, New York. Theory of positive definite and related functions.
- Berg, C., Mateu, J. & Porcu, E. (2008). The Dagum family of isotropic correlation functions. *Bernoulli* **14**:1134 – 1149.
- Bertoin, J. (1996). *Lévy Processes*. Cambridge University Press, Cambridge.
- Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons Inc., New York.
- Blanchet, J. & Davison, A. C. (2011). Spatial modeling of extreme snow depth. *Ann Appl Stat* **5**:1699–1725.
- Boldi, M.-O. & Davison, A. C. (2007). A mixture model for multivariate extremes. *J R Stat Soc Ser B Stat Methodol* **69**:217–229.
- Borodin, A. N. & Salminen, P. (1996). *Handbook of Brownian motion—facts and formulae*. Probability and its Applications. Birkhäuser Verlag, Basel.
- Breiman, L. (1968). A delicate law of the iterated logarithm for non-decreasing stable processes. *Ann Math Statist* **39**:1818–1824.
- Brown, B. M. & Resnick, S. I. (1977). Extreme values of independent stochastic processes. *J Appl Probab* **14**:732–739.

- Brown, M. (1970). A property of Poisson processes and its application to macroscopic equilibrium of particle systems. *Ann Math Statist* **41**:1935–1941.
- Coles, S. G. & Tawn, J. A. (1991). Modelling extreme multivariate events. *J R Stat Soc Ser B Stat Methodol* **53**:377–392.
- Cooley, D., Naveau, P. & Poncet, P. (2006). Variograms for spatial max-stable random fields. In *Dependence in Probability and Statistics* (eds. P. Bertail, P. Soulier & P. Doukhan), vol. 187 of *Lecture Notes in Statistics*, chap. 17, pp. 373–390. Springer, New York.
- Cooley, D., Davis, R. A. & Naveau, P. (2010). The pairwise beta distribution: a flexible parametric multivariate model for extremes. *J Multivariate Anal* **101**:2103–2117.
- Daley, D. J. & Vere-Jones, D. (2003). *An introduction to the theory of point processes. Vol. I. Probability and its Applications* (New York). Springer-Verlag, New York, second edn. Elementary theory and methods.
- Das, B. & Hashorva, E. (2012). Extremal behaviour of non-Gaussian processes attracted by Brown-Resnick process. Preprint.
- Davison, A. C. & Gholamrezaee, M. M. (2012). Geostatistics of extremes. *Proc R Soc Lond Ser A Math Phys Eng Sci* **468**:581–608.
- Davison, A. C. & Smith, R. L. (1990). Models for exceedances over high thresholds. *J R Stat Soc Ser B Stat Methodol* **52**:393–442.
- Davison, A. C., Padoan, S. A. & Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statist Sci* **27**:161–186.
- de Haan, L. (1984). A spectral representation for max-stable processes. *Ann Probab* **12**(4):1194–1204.
- de Haan, L. & Ferreira, A. (2006). *Extreme Value Theory*. Springer, New York.
- de Haan, L. & Pereira, T. T. (2006). Spatial extremes: models for the stationary case. *Ann Statist* **34**:146–168.
- de Haan, L. & Pickands, J., III (1986). Stationary min-stable stochastic processes. *Probab Theory Relat Fields* **72**(4):477–492.
- de Haan, L. & Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Probability Theory and Related Fields* **40**:317–337.
- Dombry, C., Éyi-Minko, F. & Ribatet, M. (2011). Conditional simulations of Brown-Resnick processes. Available from <http://arxiv.org/abs/1112.3891>.
- Dombry, C., Éyi-Minko, F. & Ribatet, M. (2012). Conditional simulation of max-stable processes. Available from <http://arxiv.org/abs/1208.5376>.

- Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events: for Insurance and Finance*. Springer, London.
- Engelke, S. & Schlather, M. (2011). Environmental and Ecological Statistics with R. Song S. Qian (2010). Boca Raton, FL, USA: Chapman & Hall/CRC. ISBN: 978-1-4200-6206-9. *Biom J* **53**:867–867.
- Engelke, S. & Woerner, J. H. C. (2013). A unifying approach to fractional Lévy processes. *Stoch Dyn* **13**.
- Engelke, S., Kabluchko, Z. & Schlather, M. (2011). An equivalent representation of the Brown-Resnick process. *Statist Probab Lett* **81**:1150 – 1154.
- Engelke, S., Kabluchko, Z. & Schlather, M. (2012a). A generalization of the Brown-Resnick process in terms of Lévy processes. Unpublished manuscript.
- Engelke, S., Kabluchko, Z. & Schlather, M. (2012b). Maxima of independent, non-identically distributed Gaussian vectors. Available from <http://arxiv.org/abs/1205.0947>.
- Engelke, S., Malinowski, A., Kabluchko, Z. & Schlather, M. (2012c). Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. Available from <http://arxiv.org/abs/1207.6886>.
- Engelke, S., Malinowski, A., Oesting, M. & Schlather, M. (2012d). Representations of max-stable processes based on single extreme events. Available from <http://arxiv.org/abs/1209.2303>.
- Ethier, S. N. & Kurtz, T. G. (1986). *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York. Characterization and convergence.
- Fisher, R. A. & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc Cambridge Philos Soc* **24**:180–190.
- Genton, M. G., Ma, Y. & Sang, H. (2011). On the likelihood function of Gaussian max-stable processes. *Biometrika* **98**:481–488.
- Gerritse, G. (1986). Supremum self-decomposable random vectors. *Probab Theory Relat Fields* **72**:17–33.
- Giné, E., Hahn, M. G. & Vatan, P. (1990). Max-infinitely divisible and max-stable sample continuous processes. *Probab Theory Related Fields* **87**:139–165.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Ann of Math (2)* **44**:423–453.
- Gneiting, T. (1999). Radial positive definite functions generated by euclid’s hat. *Journal of Multivariate Analysis* **69**:88 – 119.

- Gradshteyn, I. S. & Ryzhik, I. M. (2007). *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, seventh edn.
- Hashorva, E. (2005). Elliptical triangular arrays in the max-domain of attraction of Hüsler-Reiss distribution. *Statist Probab Lett* **72**:125–135.
- Hashorva, E. (2006). On the multivariate Hüsler-Reiss distribution attracting the maxima of elliptical triangular arrays. *Statist Probab Lett* **76**:2027–2035.
- Hashorva, E. & Weng, Z. (2013). Limit laws for extremes of dependent stationary Gaussian arrays. *Statist Probab Lett* **83**:320–330.
- Hashorva, E., Kabluchko, Z. & Wübker, A. (2012). Extremes of independent chi-square random vectors. *Extremes* **15**:35–42.
- Huser, R. & Davison, A. C. (2012). Space-time modelling of extreme events. Available from <http://arxiv.org/abs/1201.3245>.
- Hüsler, J. & Reiss, R.-D. (1989). Maxima of normal random vectors: between independence and complete dependence. *Statist Probab Lett* **7**:283–286.
- Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statist Probab Lett* **9**:75–81.
- Kabluchko, Z. (2009a). Functional limit theorems for sums of independent geometric Lévy processes. Available from <http://arxiv.org/abs/0911.4139v1>.
- Kabluchko, Z. (2009b). Spectral representations of sum- and max-stable processes. *Extremes* **12**:401–424.
- Kabluchko, Z. (2011a). Extremes of independent Gaussian processes. *Extremes* **14**:285–310.
- Kabluchko, Z. (2011b). Functional limit theorems for sums of independent geometric Lévy processes. *Bernoulli* **17**:942–968.
- Kabluchko, Z. & Schlather, M. (2010). Ergodic properties of max-infinitely divisible processes. *Stochastic Process Appl* **120**:281–295.
- Kabluchko, Z., Schlather, M. & de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann Probab* **37**:2042–2065.
- Kallenberg, O. (2002). *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edn.
- Karatzas, I. & Shreve, S. E. (1991). *Brownian motion and stochastic calculus*, vol. 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edn.
- Kotz, S. & Nadarajah, S. (2000). *Extreme Value Distributions*. Imperial College Press, London. Theory and applications.

- Leadbetter, M. R., Lindgren, G. & Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York.
- Mori, T. (1992). Representation of linearly additive random fields. *Probab The Rel Fields* **92**:91–115.
- Oesting, M., Kabluchko, Z. & Schlather, M. (2012). Simulation of Brown-Resnick processes. *Extremes* **15**:89–107.
- Padoan, S. A., Ribatet, M. & Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *J Amer Statist Assoc* **105**:263–277.
- Pitt, L. D. (1978). Local times for Gaussian vector fields. *Indiana Univ Math J* **27**:309–330.
- Qian, S. S. (2010). *Environmental and Ecological Statistics with R*. Chapman and Hall/CRC, Boca Raton, FL, Cambridge.
- Resnick, S. I. (2008). *Extreme Values, Regular Variation and Point Processes*. Springer, New York.
- Rogers, L. C. G. & Pitman, J. W. (1981). Markov functions. *Ann Probab* **9**:573–582.
- Rogers, L. C. G. & Williams, D. (2000). *Diffusions, Markov Processes, and Martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge. Itô calculus.
- Rootzén, H. & Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli* **12**:917–930.
- Samorodnitsky, G. & Taqqu, M. S. (1994). *Stable non-Gaussian random processes*. Stochastic Modeling. Chapman & Hall, New York.
- Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge.
- Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes* **5**:33–44.
- Schlather, M. & Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika* **90**:139–156.
- Sibuya, M. (1960). Bivariate extreme statistics. I. *Ann Inst Statist Math Tokyo* **11**:195–210.
- Smith, R. (1990). Max-stable processes and spatial extremes. Unpublished manuscript.
- Stoev, S. A. (2008). On the ergodicity and mixing of max-stable processes. *Stochastic Process Appl* **118**:1679–1705.
- Stoev, S. A. & Taqqu, M. S. (2005). Extremal stochastic integrals: a parallel between max-stable processes and α -stable processes. *Extremes* **8**:237–266.

-
- Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika* **77**:245–253.
- Turkman, K. F., Turkman, M. A. A. & Pereira, J. M. (2010). Asymptotic models and inference for extremes of spatio-temporal data. *Extremes* **13**(4):375–397.
- Wang, Y. & Stoev, S. A. (2010). On the structure and representations of max-stable processes. *Adv Appl Probab* **42**:855–877.
- Wang, Y. & Stoev, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Adv in Appl Probab* **43**:461–483.

