ANALYSIS AND SIMULATION OF MULTIVARIATE AND SPATIAL EXTREMES

HABILITATIONSSCHRIFT

vorgelegt von
Dr. rer. nat. Marco Oesting

aus
Georgsmarienhütte

eingereicht bei der Naturwissenschaftlich-Technischen Fakultät
der Universität Siegen

Siegen 2019

gedruckt auf alterungsbeständigem holz- und säurefreiem Papier

Preface

This thesis is a compilation of parts of the work I did during my postdoc phase at University of Mannheim, AgroParisTech/INRA, University of Twente, University of Siegen and Ulm University. During this time, I have also partially been funded by VolkswagenStiftung within the project "Mesoscale Weather Extremes – Theory, Spatial Modeling and Prediction (WEX-MOP)" and by the ANR project McSim. The financial support is gratefuly acknowledged.

The completion of this work would not have been possible without the support of numerous people whom I would like to thank:

- The professors at my home instuitutions, in particular Liliane Bel, Alfred Müller, Hans-Peter Scheffler, Martin Schlather, Alexander Schnurr, Evgeny Spodarev and Alfred Stein, for their scientific and personal advice, for sharing their knowledge and opening new perspectives.
- All the co-authors of my publications and submitted manuscripts so far, that is, Liliane Bel, Clément Dombry, Sebastian Engelke, Raphaël de Fondeville, Petra Friederichs, Zakhar Kabluchko, Christian Lantuéjoul, Alexander Malinowski, Peter Menck, Mathieu Ribatet, Claudia Schillings, Martin Schlather, Alexander Schnurr, Alfred Stein, Kirstin Strokorb and Chen Zhou, for sharing their ideas and time for fruitful discussions.
- All the colleagues at my various home institutions for their help, enjoyable times and having made me feel at home at different places.
- My family and friends for their constant support and companionship on my varied journey through life.

Thank you! *Danke!* *Merci!* *Dank jullie wel!*

Contents

1 Overview

1.1 Introduction to Univariate and Spatial Extremes

Extreme events have a strong influence on various aspects of human life. While financial crises and stock market crashes can lead to large financial losses for individuals, companies and economies, natural hazards such as severe storms, heavy precipitation, floods or droughts often cause personal injury, death and infrastructure damage.
In order to estimate the risks associated with extreme events and take appropriate countermeasures, the probabilities of exceeding extreme, possibly previously unobserved, thresholds must be estimated from data or, conversely, the corresponding thresholds for specified excess probabilities need to be determined – two of the key questions of extreme value theory and statistics.

Mathematically speaking, one is interested in the tail behavior of some random variable X describing the "impact" (e.g. the financial loss or the amount of precipitation) of some extreme event. There are two main approaches that are closely related to each other according to the Pickands–Balkema–de Haan Theorem (Balkema and de Haan, 1974; Pickands, 1975): the block maxima and the peaks-over-threshold (POT) approach.
In this thesis, we will mainly focus on the block maxima approach, that is, we consider the behavior of the random variable

$$M_n = \max_{i=1,\dots,n} X_i$$

for large n where $X_1, X_2, \dots$ are independent copies of the random variable X. Now assume that $\{a_n\} \subset (0, \infty)$ and $\{b_n\} \subset \mathbb{R}$ are normalizing sequences such that $a_n^{-1}(M_n - b_n)$ converges in distribution to a non-degenerate random variable Z as $n \to \infty$. Then, by the Fisher–Tippett–Gnedenko Theorem (Fisher and Tippett, 1928; Gnedenko, 1943), one of the main results in extreme value theory, the limiting variable Z necessarily follows a generalized extreme value (GEV) distribution

$$\mathbb{P}(Z \leq z) = G_{\xi,\mu,\sigma}(z) := \begin{cases} \exp\left(-\left(1+\xi\frac{z-\mu}{\sigma}\right)_+^{-1/\xi}\right), & \xi \neq 0, \\ \exp\left(-\exp\left(-z\right)\right), & \xi = 0, \end{cases} \qquad z \in \mathbb{R},$$

with shape parameter $\xi \in \mathbb{R}$, location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$. The class of GEV distributions encompasses three types of distributions: the class of Fréchet distributions with heavy tails, the light-tailed Gumbel distribution and the class of Weibull distributions with finite upper endpoint.
All of these distributions are stable with respect to the maximum operation: For every $n \in \mathbb{N}$, there are sequences $\{c_n\} \subset (0, \infty)$ and $\{d_n\} \subset \mathbb{R}$ such that

$$\max_{i=1,\dots,n} \frac{Z_i - d_n}{c_n} =_d X,$$

where $Z_1, Z_2, \dots$ are independent copies of Z and "$=_d$" denotes equality in distribution. Such a random variable Z and its distribution, respectively, are called *max-stable*. It can

be shown that any max-stable distribution is either degenerate or a GEV distribution. More details on the extremal behavior of single random variables, so-called *univariate extreme value theory*, can be found in the textbooks of Resnick (1987), Embrechts et al. (1997) and de Haan and Ferreira (2006).

In many practical examples, however, extreme events cannot be fully described by a single random variable such as the price of a single stock or the amount of precipitation at a single site. Instead, the joint extremal behavior of various components of random vectors or even stochastic processes has to be considered resulting in *multivariate extreme value theory* and *extreme value theory of stochastic processes*, respectively. Most of this thesis focuses on the general case of stochastic processes on some domain $S \subset \mathbb{R}^d$, having spatial processes in environmental applications in mind.
There is an analogous result to the Fisher–Tippett–Gnedenko theorem in this setting: Let $X_i = \{X_i(s),\ s \in S\}$, $i \in \mathbb{N}$, be independently and identically distributed sample-continuous stochastic processes. Furthermore, assume the existence of sequences of continuous functions $a_n : S \to (0, \infty)$ and $b_n : S \to \mathbb{R}$ and a sample-continuous process $\{Z(s),\ s \in S\}$ with non-degenerate marginal distributions such that

$$\left\{ \max_{i=1,\dots,n} \frac{X_i(s) - b_n(s)}{a_n(s)},\ s \in S \right\} \overset{n\to\infty}{\longrightarrow} \{Z(s),\ s \in S\} \tag{1.1}$$

weakly in the space $C(S)$ of continuous functions on S. Then, it can be shown that the limit process Z is a *max-stable process*, i.e., for every $n \in \mathbb{N}$, there are sequences of continuous functions $c_n : S \to (0, \infty)$ and $d_n : S \to \mathbb{R}$ such that

$$\left\{ \max_{i=1,\dots,n} \frac{Z_i(s) - d_n(s)}{c_n(s)},\ s \in S \right\} =_d \{Z(s),\ s \in S\},$$

where $Z_1, Z_2, \dots$ are independent copies of the process Z (cf. Chapter 9 in de Haan and Ferreira, 2006).
Necessarily, by the Fisher–Tippett–Gnedenko theorem, for each $s \in S$, the random variable $Z(s)$ follows a GEV distribution. As max-stability is preserved under marginal transformations within the class of GEV distributions (cf. Resnick 1987, Prop. 5.10, and de Haan and Ferreira 2006, Thm. 9.2.1, for instance), one often focuses on the class of *simple* max-stable processes, i.e. max-stable processes with unit Fréchet margins.

Max-stable processes and their finite-dimensional counterparts, multivariate extreme value distributions, form the main object we study in this thesis. The thesis consists of eight articles that have been published recently in various scientific journals. It contains contributions to the theory and statistics of max-stable processes including their representation (Chapter 2), likelihood-based inference (Chapter 3), unconditional simulation (Chapters 4, 5 and 6) and conditional simulation with applications to downscaling and statistical post-processing (Chapters 7, 8 and 9). Brief overviews over the main results in these areas and the author's contributions will be given in the following sections of this introductory chapter.

1.2 Spectral Representation of Max-Stable Processes

The class of simple max-stable processes can be characterized by the following spectral representation (cf. de Haan, 1984; Giné et al., 1990; Penrose, 1992, among others):

$$Z(s) = \max_{i \in \mathbb{N}} \zeta_i W_i(s), \qquad s \in S, \tag{1.2}$$

where $\{\zeta_i\}_{i\in\mathbb{N}}$ are the points of a Poisson point process on $(0,\infty)$ with intensity measure $\zeta^{-2}\mathrm{d}\zeta$, and, independently of the ζ_i, the W_i are independent copies of a sample-continuous nonnegative stochastic process $W = \{W(s),\ s \in S\}$, the so-called *spectral process*, satisfying $\mathbb{E}W(s) = 1$ for all $s \in S$.

The spectral representation and the specific choice of W play an important role in extreme value theory and statistics as many procedures for the estimation and simulation of max-stable processes are based on an appropriate choice of the spectral process W which is not unique, cf. Engelke et al. (2014, 2015); Dieker and Mikosch (2015); Dombry et al. (2016a); Oesting et al. (2018b); Oesting and Strokorb (2018), for instance. Popular examples include the following processes:

- Often a the spectral process of the form

$$W(s) = \Gamma\left(\tfrac{\nu+1}{2}\right)^{-1}\sqrt{\tfrac{2^{2-\nu}\pi}{\mathrm{Var}(G(s))^\nu}}\max\{0, G(s)\}^\nu, \quad s \in S,$$

  is considered where G is a centered Gaussian process and $\nu > 0$. This results in the max-stable *extremal-t process* (Opitz, 2013). For $\nu = 1$, one obtains the so-called *extremal Gaussian process* (Schlather, 2002).

- If the spectral process is a log-Gaussian process, i.e.

$$W(s) = \exp\left(G(s) - \tfrac{1}{2}\mathrm{Var}(G(s))\right), \quad s \in S,$$

  where, again, G is a centered Gaussian process, we obtain a Brown–Resnick process (Kabluchko et al., 2009) for Z.

If $S = \mathbb{R}^d$ (or an additive subgroup), special attention is given to stationary processes. The extremal-t process and, thus, also the extremal Gaussian process, is stationary if the underlying Gaussian process G is stationary. For the Brown–Resnick process, stationarity of the increments of G already implies stationarity of Z.
An important subclass of stationary simple max-stable processes is given by the class of simple max-stable processes allowing for a *mixed moving maxima representation* (cf. Schlather, 2002; Stoev and Taqqu, 2005; Stoev, 2008, for instance)

$$Z(s) = \max_{i\in\mathbb{N}} U_i F_i(s - S_i), \qquad s \in \mathbb{R}^d, \tag{1.3}$$

where $\{(U_i, S_i)\}_{i\in\mathbb{N}}$ are the points of a Poisson point process on $(0,\infty)\times\mathbb{R}^d$ with intensity measure $u^{-2}\mathrm{d}u\,\mathrm{d}s$, and, independently of the (U_i, S_i), the F_i are independent copies of a sample-continuous process F satisfying $\mathbb{E}\left(\int_{\mathbb{R}^d} F(s)\,\mathrm{d}s\right) = 1$. A popular example of a max-stable process with such a representation is the Gaussian extreme value process (Smith, 1990) where F is deterministic and equals the density of a multivariate Gaussian distribution. For further examples and the existence of such a representation, see Kabluchko (2009), Strokorb et al. (2015) and Kabluchko and Stoev (2016), among others. By construction, the mixed moving maxima representation (1.3) is of a different type than the general spectral representation (1.2), but can be transformed in several ways (cf. Engelke et al., 2014; Oesting et al., 2018b, for instance).

Motivated by practical applications where discontinuities are present because of measurement errors or physical phenomena, one often drops the assumption of sample-continuity and considers the broader class of processes arising as limits in (1.1) with respect to finite-dimensional distributions. In case of a countable index set $S \subset \mathbb{R}^d$, de Haan's (1984)

representation (1.2) is still valid with a discontinuous spectral process W. A more specific popular max-stable model is the Reich and Shaby (2012) model

$$Z(s) = \varepsilon^{(1/\alpha)}(s) \cdot \Big[\sum\nolimits_{l=1}^{L} B_l w_l(s)^{1/\alpha} \Big]^{\alpha}, \qquad s \in S, \tag{1.4}$$

where $B_1, \ldots, B_L$ are independent α-stable random variables whose distribution is given by their Laplace transform $\mathbb{E}(\exp(-tB_l)) = \exp(-t^\alpha)$, $t > 0$, for some $\alpha \in (0,1)$, and, independently of the B_l, ε is a noise process on S with $(1/\alpha)$-Fréchet marginal distributions. The functions $w_1, \ldots, w_L : S \to [0,1]$ are deterministic weight functions such that $\sum_{l=1}^{L} w_l(s) = 1$ for all $s \in S$. If S is a dense subset of $\mathbb{R}^d$, the sample path properties of Z are of interest. Due to the noise process $\varepsilon^{(1/\alpha)}$, which can be perceived as a multiplicative nugget effect, the sample paths of Z cannot be extended to continuous functions on $\mathbb{R}^d$. This feature makes the Reich and Shaby (2012) model (1.4) different from the models discussed above, but also attractive for environmental applications.

The question of the relation between the Reich–Shaby model (1.4) and the general spectral representation (1.2) has been the motivation for the article Oesting (2018) which is presented in **Chapter 2**. In this article, a general framework of representations of max-stable processes via ℓ^p norms, $p \in (1, \infty]$, is established, namely

$$Z(s) = \frac{\varepsilon^{(p)}(s)}{\Gamma(1 - p^{-1})} \Big[\sum\nolimits_{i \in \mathbb{N}} (\zeta_i W_i(s))^p \Big]^{1/p}, \qquad s \in S, \tag{1.5}$$

where $\varepsilon^{(p)}$, $\{\zeta_i\}_{i \in \mathbb{N}}$ and $\{W_i\}_{i \in \mathbb{N}}$ are all independent and defined as above. This representation covers both the representation (1.4) of the Reich–Shaby model as a finite sum and the spectral representation (1.2) by de Haan (1984) as the maximum over an infinite number of processes. Thus, on the one hand, representation (1.5) allows to extend the Reich–Shaby model to a class of processes whose finite-dimensional distributions are generalized logistic mixtures (cf. Fougères et al., 2009, 2013). On the other hand, it provides a way to include multiplicative nugget effects into well-known max-stable models.
Various results for max-stable processes with ℓ^p norm based representation (1.5) are presented. It is demonstrated that each process with an ℓ^p norm based representation also possesses an ℓ^q norm based representation for $q > p$ with the special case $q = \infty$ corresponding to de Haan's (1984) representation (1.2). Explicit formulae to switch from one representation to another equivalent one are provided. In particular, this includes the transformation from representation (1.5) for $p < \infty$ to representation (1.2). The converse transformation, however, is not always possible. Here, necessary and sufficient conditions for the existence of ℓ^p norm based representations are given. Interestingly, these conditions are closely connected to the theory of negative definite functions on semigroups.
Furthermore, mixing and ergodicity properties of stationary processes with ℓ^p norm based representation (1.5) are discussed. It is shown that the properties are the same as for the "denoised" analogues, which have been studied in works by Stoev (2008) and Kabluchko and Schlather (2010), for instance.

1.3 Likelihood-Based Inference for Multivariate Extreme Value Distributions and Max-Stable Processes

In view of the large variety of max-stable models, the question of inference for these models arises. More precisely, we assume that a number of independent observations of the same k-dimensional random vector Z, e.g. observations at pairwise disjoint sites $s_1, \ldots, s_k \in S$, are

available. In the classical setting introduced in Section 1.1, these data correspond to block maxima and are assumed to approximately follow a multivariate max-stable distribution or, equivalently, a max-stable process on a finite index set $S = \{s_1, \dots, s_k\}$. There is a large variety of methods for parametric, semi-parametric and non-parametric inference for max-stable distributions. Here, for the sake of brevity, we will focus on parametric approaches based on likelihoods.

Due to the representation (1.4) of the Reich and Shaby (2012) model as a finite sum, its likelihood becomes tractable within a Bayesian hierarchical model. Conditional on $B_1, \dots, B_L$, the random variables $Z(s_1), \dots, Z(s_k)$ are independent Fréchet variables. Even though there is no closed-form expression for the likelihood of the stable variables B_i, it can be processed by introducing auxiliary variables (Stephenson, 2009).
For a general simple max-stable model with representation (1.2), the computation of the likelihood is more intricate. In this case, we have that the distribution function F of Z is of the form $F(z) = \mathbb{P}(Z \leq z) = \exp(-V(z))$ where

$$V(z) = \mathbb{E}\left(\max_{i=1}^{k} \frac{W(s_i)}{z_i}\right), \qquad z = (z_1, \dots, z_k)^\top \in (0, \infty)^k,$$

denotes the so-called exponent function. If V is continuously differentiable, then the corresponding density f exists (see Dombry et al., 2017a, for more details on the existence of f) and satisfies

$$f(z) = \sum\nolimits_{\tau \in \mathcal{P}_k} f(z, \tau) := \sum\nolimits_{\tau \in \mathcal{P}_k} \exp\{-V(z)\} \prod\nolimits_{j=1}^{|\tau|} \{-\partial_{\tau_j} V(z)\}, \tag{1.6}$$

where $\mathcal{P}_k$ is the set of all partitions $\tau = \{\tau_1, \dots, \tau_{|\tau|}\}$ of $\{1, \dots, k\}$ and $\partial_{\tau_j} V(\cdot; \theta)$ denotes the partial derivative of V with respect to the arguments z_i, $i \in \tau_j$. As the number of summands in (1.6) grows super-exponentially in k, maximum likelihood estimators can be directly implemented for small dimension k only.
Instead, often, a composite pairwise likelihood built from bivariate sublikelihoods only is considered. The resulting maximum composite likelihood estimator (Padoan et al., 2010) is known to be usually less efficient than the maximum likelihood estimator based on full likelihoods, an effect that might be mitigated by including likelihood terms corresponding to higher-dimensional subsets (cf. Genton et al., 2011; Huser and Davison, 2013; Castruccio et al., 2016, for instance).
Alternative approaches are based on the fact that representation (1.2) naturally suggests a partition τ of the vector Z with j_1 and j_2 being in the same subset of $\{1, \dots, k\}$ if and only if the maxima $Z(s_{j_1})$ and $Z(s_{j_2})$ in (1.2) are attained by the same function $\zeta_i W_i(\cdot)$. The term $f(z, \tau)$ in (1.6) then provides the joint density of the vector Z and the underlying random partition τ, which is integrated out in the sum (1.6). If, additionally to the block maxima, the occurrence times of the maxima within the blocks are available for each component, these occurrence times can be used to form an observed partition $\hat{\tau}$ in an analogous way. For this case, Stephenson and Tawn (2005) suggest to consider the joint likelihood of the data and the observed partition, that is, only one single summand in (1.6), for maximum likelihood estimation. Due to the approximation of the underlying partition by its empirical counterpart, however, the resulting estimator often suffers from a significant bias. A correction is proposed in Wadsworth (2015). More details on these likelihood-based estimators for max-stable models, alternative approaches based on threshold exceedances and a comparison of different estimators are given in Huser et al. (2016).

Several recent works put the focus on different ways to treat the unknown random partition, allowing for inference based on full likelihoods. For instance, Huser et al. (2019)

suggest a stochastic expectation-maximization algorithm perceiving the underlying random partitions as missing observations. As an alternative, Thibaud et al. (2016) propose to use the random partition as latent variable in a Bayesian framework to fit a Brown–Resnick process to extreme low temperature data. In Dombry et al. (2017b), which is **Chapter 3** in this thesis, we generalize and further investigate this approach. More precisely, we consider N independent data $z^{(1)}, \dots, z^{(N)}$ from a max-stable distribution belonging to a parametric family $\{F_\theta,\ \theta \in \Theta\}$ equipped with an appropriate prior distribution and also account for the unknown partitions $T^{(1)}, \dots, T^{(N)}$ corresponding to each data vector, thus allowing for the use of the Bayesian full likelihood. Sampling from the joint posterior distribution of $(\theta, T^{(1)}, \dots, T^{(N)})$ can then be performed via Markov chain Monte Carlo (MCMC) methods. Similarly to Thibaud et al. (2016), we update each variable separately by the means of a Metropolis-Hastings sampler for θ and the Gibbs sampler proposed in Dombry et al. (2013) for the random partitions.
We show that, under appropriate conditions, the median of the posterior distribution of θ is an asymptotically normal estimator of the true parameter θ_0 as $N \to \infty$. Furthermore, it possesses the same asymptotic variance as the maximum likelihood estimator. The most important conditions for this result are differentiability in quadratic mean and the existence of uniformly consistent tests. While conditions for differentiability of simple max-stable models, i.e. max-stable models with unit Fréchet margins, are given in Dombry et al. (2017a), we provide a sufficient condition for the existence of uniformly consistent tests in terms of pairwise extremal coefficients. Making use of further results of Dombry et al. (2017a), the validity of all of these conditions is verified for a variety of max-stable process models and multivariate extreme value distributions. For the logistic model, simulation studies are conducted to compare the finite-sample performance of the Bayesian full likelihood estimator to various other estimators such as the pairwise likelihood estimator (Padoan et al., 2010), the independence likelihood estimator for marginal parameters, the Stephenson-Tawn estimator (Stephenson and Tawn, 2005) and its bias corrected version (Wadsworth, 2015). Furthermore, genuinely Bayesian techniques such as Bayesian model comparison are discussed.

1.4 Simulation of Max-Stable Processes

Besides its use for the construction of parametric models, the spectral representation of max-stable processes is of particular importance for their simulation. A first approach to this problem has been considered in Schlather (2002). It relies on the intuition that the product $\zeta_i W_i(\cdot)$ is more likely to contribute to the maximum in (1.2) if ζ_i is large, and on the fact that the Poisson points $\{\zeta_i\}_{i\in\mathbb{N}}$ can be generated subsequently in a descending order, i.e., such that $\zeta_1 > \zeta_2 > \dots$ a.s. Hence, for $n \in \mathbb{N}$, Schlather (2002) proposes the finite approximation

$$Z^{(n)}(s) = \max_{i=1,\dots,n} \zeta_i W_i(s), \quad s \in K, \tag{1.7}$$

to the exact process $\{Z^{(\infty)}(s),\ s \in K\} \overset{d}{=} \{Z(s),\ s \in K\}$ on a compact simulation domain $K \subset S$. Provided that the spectral process W is uniformly bounded on K by some constant $C > 0$, a stopping rule can be used to determine a finite, but random number T such that $\{Z^{(T)}(s),\ s \in K\} = \{Z^{(\infty)}(s),\ s \in K\}$ with probability one. Thus, the max–stable process Z can be simulated exactly in finite time. A similar procedure also works in case of a mixed moving maxima representation (1.3) provided that the corresponding process F is uniformly bounded and has compact support.

If, in contrast, the spectral process is not uniformly bounded, one might use an approximation $Z^{(N)}$ where N is determined according to the above mentioned stopping rule for sufficiently large C. In practice, such an approximation provides satisfactory results in many cases where the spectral process is stationary with light tails and rather small variance such as for extremal-t process with small $\nu > 0$. For Brown–Resnick processes, however, approximations often turn out to be highly inaccurate. To overcome these issues, several approximate simulation algorithms based on other equivalent representations are considered in Oesting et al. (2012), which has been part of the author's PhD thesis.
In recent years, also exact simulation procedures have been developed starting from the work by Dieker and Mikosch (2015) for the case of Brown–Resnick processes on a finite index set. Their approach is based on a measure transform resulting in spectral functions that are normalized w.r.t. the L^1 norm.

The use of a family of more general measure transforms for the exact simulation of arbitrary max-stable processes is investigated in the article Oesting et al. (2018b) which is presented in **Chapter 4**. In this article, we start from the general representation

$$Z(s) = \max_{i\in\mathbb{N}} U_i V_i(s), \qquad s \in K, \tag{1.8}$$

where $\{(U_i, V_i)\}_{i\in\mathbb{N}}$ are the points of a Poisson point process on $(0,\infty) \times C_+(K)$ with intensity measure $u^{-2}\,\mathrm{d}u\,H(\mathrm{d}v)$ for some Radon measure H satisfying $\int v(s)\,H(\mathrm{d}v) = 1$ for all $s \in K$, and $C_+(K)$ denotes the space of nonnegative continuous functions on K. This representation (1.8) comprises both de Haan's (1984) representation (1.2) and the mixed moving maxima representation (1.3). Then, introducing some probability density g on $C_+(K)$ w.r.t. H, it can be shown that

$$Z(s) =_d \max_{i\in\mathbb{N}} U_i \frac{\tilde{V}_i(s)}{g(\tilde{V}_i)}, \qquad s \in K, \tag{1.9}$$

where $\{(U_i, \tilde{V}_i)\}_{i\in\mathbb{N}}$ are the points of a Poisson point process on $(0,\infty) \times C_+(K)$ with intensity measure $u^{-2}\,\mathrm{d}u\,g(v)H(\mathrm{d}v)$, i.e. the process allows for an equivalent spectral representation of type (1.2) with spectral processes $W_i(s) = \tilde{V}_i(s)/g(\tilde{V}_i)$. Provided that g is chosen such that these processes are uniformly bounded with probability one, again, a stopping rule can be applied such that for some finite but random number T the approximation $Z^{(T)}$ as defined in (1.7) provides an exact sample from the law of Z.
The article Oesting et al. (2018b) aims at minimizing the expected number $\mathbb{E}T$ of functions to be involved in the finite approximation among all probability densities g. As the resulting optimization problem cannot be solved directly, two surrogates for the original problem are stated, both of them being solved by the same probability density $g^*(h) = c^{-1}\sup_{s\in K} h(s)$, $h \in C_+(K)$, where $c > 0$ is an appropriate normalization constant. The corresponding spectral functions $W_i(s) = c\tilde{V}_i(s)/\sup_{t\in K}\tilde{V}_i(t)$, $i \in \mathbb{N}$, consequently satisfy $\sup_{s\in K} W_i(s) = c$ a.s., which is why the resulting spectral representation is called normalized spectral representation. While the existence of such a representation has already been verified in Giné et al. (1990), the article Oesting et al. (2018b) provides the underlying transformation to obtain such a representation from any other equivalent representation and proposes its use for exact simulation. For the evaluation of the computational costs of the simulation, an expression for the expectation $\mathbb{E}T$ is given. Explicit formulae for the spectral functions and the normalizing constant c are available in case of moving maxima processes with radial symmetric and non-increasing shapes. More specifically, for the extremal Gaussian process (Smith, 1990), the performance of the exact

simulation algorithm based on the normalized spectral representation is demonstrated and compared to the approach by Schlather (2002).

In the general case of a max-stable process with representation (1.2), i.e. representation (1.8) with H being equal to the distribution of W, sampling of normalized spectral functions or, equivalently, the underlying processes $\tilde{V}_i$ in (1.9), is not straightforward due to the underlying measure transform

$$\mathbb{P}(\tilde{V}_i \in \mathrm{d}v) = c^{-1} \sup_{s\in K} v(s) \cdot \mathbb{P}(W \in \mathrm{d}v), \qquad v \in C(K),\ v \geq 0. \tag{1.10}$$

While Oesting et al. (2018b) suggest an approach by MCMC techniques, de Fondeville and Davison (2018) propose to simulate normalized spectral functions via rejection sampling in the context of Pareto processes. For the important case of Brown–Resnick processes where W is a log-Gaussian process, both the MCMC algorithm and the rejection sampling approach are refined in Oesting et al. (2019), i.e. **Chapter 5** of this thesis. More precisely, for a finite simulation domain $K = \{s_1, \ldots, s_n\}$, we propose a Metropolis-Hastings algorithm using a mixture of the type

$$f_{\mathrm{prop}}(v) = \sum\nolimits_{i=1}^{N} p_i f_i(v) = \sum\nolimits_{i=1}^{N} p_i v(s_i) \mathbb{P}(W \in \mathrm{d}v), \quad v \in (0,\infty)^N,$$

as proposal density, where $p_i \geq 0$ are weights such that $\sum_{i=1}^{N} p_i = 1$. In case of a Brown–Resnick process, each modified density f_i is also a log-Gaussian density (cf. Dombry et al., 2016a) and, thus, allows for sampling in an efficient way. Specifying the acceptance probability in an appropriate way, the distribution of the resulting Markov chain can be shown to converge to the desired distribution in total variation norm. In order to obtain good mixing properties, we recommend to choose the weights p_i, $i = 1, \ldots, N$, such that the relative deviation between f_{prop} and the target density is minimized, resulting in a quadratic optimization problem.
As an alternative to approximate sampling from the desired distribution, in Oesting et al. (2019), we also consider exact simulation via rejection sampling. Here, using the same type of proposal density f_{prop} as before with weights $p_1 = \ldots = p_N = \frac{1}{N}$ results in an procedure equivalent to the one proposed by de Fondeville and Davison (2018). Then, the expected computational complexity is the same as for the algorithm developed by Dieker and Mikosch (2015) (see Oesting and Strokorb, 2019). To reduce the computational effort, we propose a more flexible approach with proposal density of the type $f_{\mathrm{prop}} = \sum_{i=1}^{N} p_i g_{i,\varepsilon}$, where $g_{i,\varepsilon}$, $\varepsilon \in [0,1)$, are log-Gaussian densities with an increased variance. Ideally, one would choose the weights $p_1, \ldots, p_N$ and the additional parameter ε such that the average acceptance rate is maximal. As the solution of this optimization problem cannot be calculated directly, we present lower bounds for the average acceptance rate and present strategies for their maximization. An example demonstrates the potential of our modifications to improve the original approaches.
Thus, the work presented in Chapter 5 complements the results from Chapter 4 by showing that, apart from the overall normalization constant c, also Brown–Resnick processes can be simulated exactly and efficiently via the normalized spectral representation.

In addition to simulation via the normalized spectral representation, there are (at least) two other general exact simulation methods both of which are presented in **Chapter 6**. This chapter corresponds to the publication Dombry et al. (2016a). Our first approach is based on the concept of so-called extremal functions which has been developed by Dombry and Éyi-Minko (2012, 2013). More precisely, it can be shown that, for each $s \in S$, there is one and only one function $\phi_i = \zeta_i W_i$ such that $Z(s) = \phi_i(s)$ in (1.2). Such a function, also

denoted by ϕ_s^+, is called extremal function at point $s \in S$. In order to simulate Z exactly on a finite simulation domain $K = \{s_1, \dots, s_N\}$, we propose to subsequently simulate the extremal functions $\phi_{s_1}^+, \phi_{s_2}^+, \dots, \phi_{s_N}^+$. By definition, the process

$$\tilde{Z}^{(N)}(s) = \max_{i=1,\dots,N} \phi_{s_i}^+(s), \quad s \in S,$$

satisfies $\tilde{Z}^{(N)}(s) = Z(s)$ for all $s \in K$. An important ingredient for the simulation of the extremal functions $\phi_{s_i}^+$ is the probability measure P_{s_i}. This measure is constructed from the original spectral distribution of W by a transformation and is concentrated on those functions $f \in C(K)$ that satisfy $f(s_i) = 1$. In Dombry et al. (2016a), we give explicit formulae for P_{s_i} for the most popular max-stable process models and various multivariate extreme value distributions. In all of these cases, the distributions P_{s_i} allow for simulation in an efficient way.

Our second approach generalizes the procedure developed by Dieker and Mikosch (2015) for the simulation of Brown–Resnick processes on a finite domain $K = \{s_1, \dots, s_N\}$ to more general max-stable processes. It is based on a transformation of the spectral vector such that it becomes normalized w.r.t. the L^1 norm. As this also implies a uniform bound for the spectral vector, exact simulation can be performed as suggested in Schlather (2002). This procedure is closely connected to the other two procedures discussed in Chapter 4 and Chapter 6. First, the resulting representation is a special case of the transformed representation (1.9) with $g(f) = N^{-1} \sum_{i=1}^N f(s_i)$. Second, the distribution of the vectors $\tilde{V}_i$, $i \in \mathbb{N}$, is a mixture of the distributions P_{s_j}, $j = 1, \dots, N$, occurring in the simulation procedure via extremal functions. Thus, using the above results on the P_{s_j}, exact simulation can be performed for the most popular max-stable models.
For both exact approaches discussed in Dombry et al. (2016a), expressions for the computational complexity defined as the expected number of spectral vectors to be simulated to obtain one max-stable realization are given. Here, the extremal functions approach turns out to be always preferable to the second approach. There is no clear ordering, however, between the extremal functions approach and the procedure based on normalized spectral functions, see Dombry et al. (2016a) and Oesting et al. (2018b) for examples.

In addition to the generic algorithms discussed above, there are also approaches tailored specifically to Brown–Resnick processes. For instance, Liu et al. (2019+) propose an algorithm for their exact simulation based on record breakers. In contrast, Oesting and Strokorb (2018) revisit non-exact simulation procedures. Allowing for some small simulation error, they suggest to reduce the computational effort by minimizing the maximal variance of the underlying log-Gaussian spectral processes. Interestingly, in the non-exact setting, the approach by Dieker and Mikosch (2015), or, more generally, the second approach in Dombry et al. (2016a), become competitive again even though not being efficient for exact simulation, see Oesting and Strokorb (2019) for more details.

1.5 Conditional Simulation of Max-Stable Processes and its Applications in Environmental Sciences

In many practical examples, additional information on the realization of a stochastic process is given. For instance, its value at specific sites might be obtained from station measurements or average values over certain grid cells might be available from satellite data or numerical models. In order to get more insight in the distribution of the process given these information, conditional simulations are necessary.

For a max-stable process Z on S, the problem of conditional sampling given values $Z(s_1), \dots, Z(s_N)$ for $s_1, \dots, s_N \in S$, has been addressed in a series of papers starting from Wang and Stoev (2011) who consider spectrally discrete max-linear models. In Dombry and Éyi-Minko (2013), general results for max-stable processes, and, even more generally, max-infinitely divisible processes, are provided. For three popular regular cases, the Brown–Resnick, extremal Gaussian and extremal-t processes, these are explicitly calculated and applied in Dombry et al. (2013) and Ribatet (2013). Their results are complemented by the work of Oesting and Schlather (2014), which has been part of the author's PhD thesis and examines the irregular cases of mixed moving maxima processes.
Even though the works mentioned above use different techniques in the proofs, their algorithms all rely on the concepts of extremal functions (cf. Section 1.4) and the partition of the index set formed by the extremal functions (cf. Section 1.3) which is also called hitting scenario in this context. Based on these ideas, a three-step algorithm is proposed. First, a hitting scenario is sampled conditionally on $Z(s_1), \dots, Z(s_N)$. Second, extremal functions corresponding to this hitting scenario are simulated. Third, the spectral functions that do not contribute to the maxima $Z(s_1), \dots, Z(s_N)$, but might contribute to $Z(s)$ for other $s \in S$ are sampled. Here, the last step is closely related to the problem of unconditional simulation, see the review article Dombry et al. (2016b) for more details.

Besides values of the process at specific sites, other information such as functionals of the process might be given. Two types of such functionals are considered in Oesting (2015) and Oesting et al. (2018a), which correspond to Chapter 7 and Chapter 8 of this thesis, respectively. In **Chapter 7**, we focus on the problem of conditional simulation of a simple max-stable process Z on some compact domain K given the values of continuous max-linear functionals $L_1, \dots, L_n : C_+(K) \to [0, \infty)$, i.e. continuous functionals satisfying

$$L_j(\max\{a_1 f_1, a_2 f_2\}) = \max\{a_1 L_j(f_1), a_2 L_j(f_2)\}, \quad a_1, a_2 \geq 0,\ f_1, f_2 \in C_+(K).$$

Examples of such functionals are suprema over subsets of K. From the spectral representation (1.8) and the max-linearity of the L_j, we obtain

$$L_j(Z) = \max_{i \in \mathbb{N}} \zeta_i L_j(W_i), \quad j = 1, \dots, n.$$

Thus, one can consider the extended process $Z_L = (\{Z(s),\ s \in K\}, L_1(Z), \dots, L_n(Z))$. By the above considerations, this process is then a max-stable process with values in $C_+(K) \times [0, \infty)^n$ and possesses a spectral representation of the form (1.2) with spectral process $W_L = (\{W(s),\ s \in K\}, L_1(W), \dots, L_n(W))$. Hence, one can adapt the general methodology developed in Dombry and Éyi-Minko (2013) for conditional simulation leading to a three-step procedure similar to the one described above. In Oesting (2015), following Dombry and Éyi-Minko (2013), general formula for the conditional distributions involved in each step are given. Particular emphasis is put on the case that only one max-linear condition is given. Then, the distribution of corresponding extremal function is closely related to the distribution of a spectral process with is normalized w.r.t. the max-linear functional. Similarly to Chapter 4, an explicit formula can be obtained for the Gaussian extreme value process, i.e. a moving maxima process. In the case of multiple max-linear conditions, the conditional distributions can be rewritten in terms of the Lebesgue density of the random vector $(L_1(W), \dots, L_n(W))$ provided that it exists.

In **Chapter 8**, we consider the case of one single condition given by a positively homogeneous functional, i.e. a functional $\ell : C_+(K) \to [0, \infty)$ such that

$$\ell(af) = a\ell(f), \quad a > 0,\ f \in C_+(K).$$

For instance, one could consider weighted averages, suprema or minima over subsets of K. We start with the spectrally discrete case, i.e. a max-linear model

$$Z(s) = \max_{j=1,\dots,n} a_j(s) Z_j, \qquad s \in K, \tag{1.11}$$

where $Z_1, \dots, Z_n$ are independent α-Fréchet variables and $a_1, \dots, a_n \in C_+(K)$ are deterministic weight functions (cf. Wang and Stoev, 2011). Due to the homogeneity of ℓ, the random variables $Z_1, \dots, Z_n$ and, thus, also the process Z, are uniquely determined by the value of $\ell(Z)$ and the ratios Z_j/Z_1, $j = 2, \dots, n$. Here, the density of the vector $(Z_j/Z_1)_{j=2,\dots,n}$ conditionally on $\ell(Z)$ can be stated explicitly up to some normalizing constant. Thus, we propose a Metropolis–Hastings algorithm with independence sampler to sample from the conditional distribution. This idea can be extended to the case of conditionally max-linear models even further to general max-stable models. To this end, we use similar ideas as discussed in Section 1.4, and exploit the fact that a max-stable process can be represented as

$$Z(s) =_d Z^{(T)}(s) = \max_{i=1,\dots,T} \zeta_i W_i(s), \quad s \in K,$$

with a finite, but random number T provided the spectral functions are uniformly bounded. Here, the random spectral functions W_i, $i \in \mathbb{N}$, can be interpreted analogously to the weight functions $a_1, \dots, a_n$ in (1.11), while the Poisson points ζ_i, $i \in \mathbb{N}$, take a similar role as the α-Fréchet variables $Z_1, \dots, Z_n$. Thus, the Metropolis-Hastings algorithm above can be adapted to sample from the joint distribution of T, $\{W_i\}_{i=1,\dots,T}$ and $\{\zeta_i\}_{i=1,\dots,T}$ conditional on $\ell(Z)$, using the building blocks obtained from unconditional simulations of the max-stable process Z as proposals. The performances of the algorithms for the max-linear and the general max-stable case are demonstrated in simulation studies. Furthermore, the behavior of the Markov chains produced by these algorithms are analyzed.

Conditions of the same type as in Chapter 8 play an important role in environmental sciences. Climate models, for instance, often provide output on a rather coarse grid. In many cases, the value corresponding to a grid cell can be interpreted as a spatial average over the cell. As averaging typically tends to conceal localized extremes, one is interested in the behavior of the process at fine scale, i.e. the problem of downscaling arises. A natural approach to tackle this question is via conditional simulation. Focusing on extremes, Bechler et al. (2015), for instance, propose to use transfer functions in order to assign the value for each grid cell to a single site inside the cell. Then, conditional simulation of an appropriate max-stable process given its values at these sites can be performed. In the application part of Oesting et al. (2018a), which has been omitted in this thesis, the methodology developed in our work, is applied to downscale precipitation data in the southeast of France.

Often, the aggregated values that are available do not correspond to an aggregation of block maxima, i.e. the max-stable process Z itself, but an aggregation of a single event, i.e. a process X in the max-domain of attraction of Z. This broader setting is considered in Engelke et al. (2018), which is part of Raphaël de Fondeville's PhD thesis. In this paper, we develop methods to estimate the parameters of the underlying max-stable process Z from aggregated data given by homogeneous functionals $\ell_1(X), \dots, \ell_L(X)$. These are based on limit results for the joint distribution of the aggregated data. Conditional simulations from the limit model can then be used for downscaling.

Downscaling can be seen as a specific type of statistical post-processing, which is required because of the phenomenon that output of numerical models such as a weather prediction

models often does not reflect extremes accurately. Such a post-processing procedure is presented in Oesting et al. (2017), which is **Chapter 9** of this thesis. Here, we develop a bivariate spatial process model $(Z^{(1)}, Z^{(2)})^\top$ where the first component $Z^{(1)}$ corresponds to standardized observed block maxima and the second component $Z^{(2)}$ to the forecast analogues. More precisely, we propose a bivariate Brown–Resnick model defined via

$$\begin{pmatrix} Z^{(1)}(s) \\ Z^{(2)}(s) \end{pmatrix} = \begin{pmatrix} \max_{i\in\mathbb{N}} \zeta_i \exp(G_i^{(1)}(s) - \mathrm{Var}(G_i^{(1)}(s))/2) \\ \max_{i\in\mathbb{N}} \zeta_i \exp(G_i^{(2)}(s) - \mathrm{Var}(G_i^{(2)}(s))/2) \end{pmatrix}, \qquad s \in \mathbb{R}^d, \tag{1.12}$$

where $\{\zeta_i\}_{i\in\mathbb{N}}$ are the points of a Poisson point process on $(0,\infty)$ with intensity measure $\zeta^{-2}\mathrm{d}\zeta$ and, independently of the Poisson points, $G_i = (G_i^{(1)}, G_i^{(2)})$, $i \in \mathbb{N}$, are independent copies of a centered bivariate Gaussian process $G = \{(G^{(1)}(s), G^{(2)}(s))^\top,\ s \in \mathbb{R}^d\}$. The bivariate Brown–Resnick process Z is stationary if and only if the pseudo cross-variogram

$$2\gamma(h) = (2\gamma_{ij}(h))_{1\le i,j\le 2} = \big(\mathrm{Var}(G^{(i)}(s+h) - G^{(j)}(s))\big)_{1\le i,j\le 2},$$

does not depend on $s \in \mathbb{R}^d$ (cf. Genton et al., 2015; Molchanov and Stucki, 2013). We further show that stationarity of Z implies that the components of $(\sqrt{\gamma_{ij}(h)})_{1\le i,j\le 2}$ differ by bounded functions only, i.e. the asymptotic behavior of all the components of γ is essentially the same provided that at least one component tends to infinity. Based on the bivariate Matérn model (Gneiting et al., 2010), we develop a flexible bivariate unbounded cross-variogram model that allows for different smoothness properties of the two component. The model can be fitted by a weighted least squares fit of the corresponding extremal coefficients. Statistical post-processing can then be performed by simulating from the bivariate random field model Z conditionally on its second component $Z^{(2)}$ at various sites. The procedure is demonstrated for daily maxima of station data and forecasts for wind gusts in Northern Germany. Model verification by the means of scores reveals that, in contrast to a univariate max-stable model, the post-processing procedure based on the bivariate model leads to convincing improvements of the forecasts.

Contributions to the Chapters of the Thesis:

In the final part of the introductory chapter, I would like to comment on my contributions to the eight published articles that form Chapters 2–9 of this thesis.

- Oesting (2018) and Oesting (2015), i.e. Chapter 2 and Chapter 8 of this thesis, are single author papers.
- Dombry et al. (2017b) and Dombry et al. (2016a), that is, Chapter 3 and Chapter 6, are joint work with Clément Dombry and Sebastian Engelke. Each of the three authors contributed in equal parts to both articles.
- Martin Schlather, Chen Zhou and me contributed equally to Oesting et al. (2018b), i.e. Chapter 4 of this thesis.
- I wrote the majority of Oesting et al. (2019), i.e. Chapter 5. Martin Schlather contributed with a few ideas and the time relevant programming part while Claudia Schillings made contributions to one of the optimization procedures.
- As first author, I contributed in a leading manner to the three-author articles Oesting et al. (2018a) and Oesting et al. (2017), which correspond to Chapter 7 and Chapter 9, respectively. In this thesis, the application part of Oesting et al. (2018a) is left out as it has mainly been implemented by Liliane Bel.

2 Equivalent Representations of Max-Stable Processes via ℓ^p Norms

Up to minor modifications and corrections, this chapter is a reprint of the article Oesting (2018) which has appeared in the *Journal of Applied Probability*.

While max-stable processes are typically written as pointwise maxima over an infinite number of stochastic processes, in this chapter, we consider a family of representations based on ℓ^p norms. This family includes both the construction of the Reich–Shaby model and the classical spectral representation by de Haan as special cases. As the representation of a max-stable process is not unique, we present formulae to switch between different equivalent representations. We further provide a necessary and sufficient condition for the existence of a ℓ^p norm based representation in terms of the stable tail dependence function of a max-stable process. Finally, we discuss several properties of the represented processes such as ergodicity or mixing.

2.1 Introduction

Arising as limits of rescaled maxima of stochastic processes, max-stable processes play an important role in spatial and spatio-temporal extremes. Here, a stochastic process $X = \{X(s),\, s \in S\}$ on a countable index set S is called max-stable if there exist sequences $\{a_n(\cdot)\}_{n\in\mathbb{N}}$ and $\{b_n(\cdot)\}_{n\in\mathbb{N}}$ of functions $a_n : S \to (0,\infty]$ and $b_n : S \to \mathbb{R}$ such that, for all $n \in \mathbb{N}$,

$$\mathcal{L}(X) = \mathcal{L}\bigg(\max_{i=1}^{n} \frac{X_i - b_n}{a_n} \bigg),$$

where X_i, $i \in \mathbb{N}$, are independent copies of X and the maximum is taken pointwise. From univariate extreme value theory, it is well-known that the marginal distributions of X, if non-degenerate, are necessarily Generalized Extreme Value (GEV) distributions, i.e.

$$\mathbb{P}(X(s) \le x) = \exp\bigg(-\bigg(1 + \xi(s)\frac{x-\mu(s)}{\sigma(s)}\bigg)^{-1/\xi(s)} \bigg), \quad 1 + \xi(s)\frac{x-\mu(s)}{\sigma(s)} > 0,$$

with $\xi(s) \in \mathbb{R}$, $\mu(s) \in \mathbb{R}$ and $\sigma(s) > 0$ for $s \in S$. As max-stability is preserved by marginal transformations, it is common practice in extreme value theory to consider only one type of marginal distributions, e.g. the case that the shape parameter ξ is positive. In this case, the marginal distributions are of α-Fréchet type, i.e., up to affine transformations, the marginal distribution functions are of the form

$$\Phi_\alpha(x) = \exp\left(-x^{-\alpha}\right), \quad x > 0,$$

for some $\alpha > 0$. Here, we will focus on the case of max-stable processes with unit Fréchet margins, i.e. $X(s) \sim \Phi_1$ for all $s \in S$. In this case, X is called a simple max-stable process.

By de Haan (1984), the class of simple max-stable processes on S can be fully characterized: A stochastic process $\{X(s),\, s \in S\}$ is simple max-stable if and only if it possesses the spectral representation

$$X(s) = \max_{i \in \mathbb{N}} A_i V_i(s), \quad s \in S, \tag{2.1}$$

where $\sum_{i\in\mathbb{N}} \delta_{A_i}$ is a Poisson point process on $(0,\infty)$ with intensity measure $a^{-2}\mathrm{d}a$ and $V_i = \{V_i(s),\, s \in S\}$ are independent copies of a stochastic process V such that $\mathbb{E}(V(s)) = 1$ for all $s \in S$ (Giné et al., 1990; Penrose, 1992, see also). It is important to note that this representation is not unique. As different representations of the same max-stable process might be convenient for different purposes such as estimation (see Engelke et al., 2014, 2015, among others) or simulation (cf. Oesting et al., 2012; Dieker and Mikosch, 2015; Oesting et al., 2018b, for instance), finding novel representations is of interest.

Recently, Reich and Shaby (2012) came up with a class of max-stable processes written as a product

$$X(s) = U^{(p)}(s) \cdot \left[\sum_{l=1}^{L} B_l w_l(s)^p\right]^{1/p}, \qquad s \in S, \tag{2.2}$$

where $\{U^{(p)}(s)\}_{s\in S}$ is a noise process with $U^{(p)}(s) \sim_{iid} \Phi_p$, the functions $w_l : S \to [0,\infty)$, $l = 1,\dots,L$, are deterministic weight functions such that $\sum_{l=1}^{L} w_l(s) = 1$ for all $s \in S$ and, independently from $\{U^{(p)}(s)\}_{s\in S}$, the independent random variables B_l, $l = 1,\dots,L$, follow a stable law given by the Laplace transform

$$\mathbb{E}\{\exp(-t \cdot B_l)\} = \exp\left(-t^{1/p}\right), \qquad t > 0.$$

The parameter $p \in (1,\infty)$ determines the strength of the effect of the noise process which – analogously to the terminology in geostatistics – is also called a nugget effect. In Reich and Shaby (2012), the weight functions w_l are chosen as shifted and appropriately rescaled Gaussian density functions yielding an approximation of the well-known Gaussian extreme value process (Smith, 1990) joined with a nugget effect. Similarly, Reich and Shaby (2012) propose analogues to popular max-stable processes such as extremal Gaussian processes (Schlather, 2002) and Brown-Resnick processes (Kabluchko et al., 2009) by choosing appropriately rescaled realizations of Gaussian and log-Gaussian processes, respectively, as weight functions. Due to the flexibility in modeling the strength of the nugget by the additional parameter p and the tractability of the likelihood which allows to embed the model in a hierarchical Bayesian model, the Reich–Shaby model (2.2) has found its way into several applications (cf. Shaby and Reich, 2012; Reich et al., 2014; Stephenson et al., 2015; Sebille et al., 2017, for instance).

While a simple max-stable process in the spectral representation (2.1) is written as the pointwise supremum of an infinite number of processes, i.e. the pointwise ℓ_∞ norm of the random sequence $\{A_i \cdot V_i(s)\}_{i\in\mathbb{N}}$, the Reich–Shaby model (2.2) is represented as the pointwise p norm of the finite random vector $(B_l^{1/p} \cdot w_l(s))_{l=1,\dots,L}$. In this chapter, we will present a more general class of representations of max-stable processes by writing them as pointwise ℓ^p norms of sequences of stochastic processes, including, for instance, both de Haan's representation and the Reich–Shaby model as special cases. The finite-dimensional distributions of the resulting processes will turn out to be generalized logistic mixtures introduced by Fougères et al. (2009) and Fougères et al. (2013).

This chapter is structured as follows: In Section 2.2, we will introduce the spectral representation based on ℓ^p norms. As a single max-stable process might allow for equivalent ℓ^p

norm based representations for different $p \in (1, \infty]$, we give formulae to switch between them in Section 2.3. Section 2.4 provides a full characterization of the resulting class of processes whose properties are finally discussed in Section 2.5.

2.2 Generalization of the Spectral Representation

Denoting by

$$\|\boldsymbol{A} \circ \boldsymbol{V}(s)\|_p = \begin{cases} \left[\sum_{i\in\mathbb{N}} (A_i \cdot V_i(s))^p\right]^{1/p}, & p \in (1,\infty), \\ \max_{i\in\mathbb{N}} A_i \cdot V_i(s), & p = \infty, \end{cases}$$

the ℓ^p norm of the Hadamard product of the two random sequences $\boldsymbol{A} = \{A_i\}_{i\in\mathbb{N}}$ and $\boldsymbol{V}(s) = \{V_i(s)\}_{i\in\mathbb{N}}$, $s \in S$, the spectral representation (2.1) can be rewritten as

$$X(s) = \|\boldsymbol{A} \circ \boldsymbol{V}(s)\|_\infty, \quad s \in S.$$

We present a more general representation replacing the ℓ^∞ norm by some general ℓ^p norm, $p \in (1, \infty]$, and multiplication by an independent noise process with Φ_p marginal distributions. Here, we use the convention that Φ_∞ denotes the weak limit of Φ_p as $p \to \infty$, i.e. $\Phi_\infty(x) = \mathbf{1}_{[1,\infty)}(x)$ is a degenerate distribution function.

Theorem 2.1. *Let $p \in (1, \infty]$ and $\{U^{(p)}(s)\}_{s\in S}$ be a collection of independent Φ_p random variables. Further, let $\sum_{i\in\mathbb{N}} \delta_{A_i}$ be a Poisson process on $(0, \infty)$ with intensity $a^{-2}\mathrm{d}a$ and, independently of $\sum_{i\in\mathbb{N}} \delta_{A_i}$, let $W_i^{(p)}$, $i \in \mathbb{N}$, be independent copies of a nonnegative stochastic process $\{W^{(p)}(s), s \in S\}$ with $\mathbb{E}\{W^{(p)}(s)\} = 1$ for all $s \in S$. Then, the process X, defined by*

$$X(s) = \frac{U^{(p)}(s)}{\Gamma(1 - p^{-1})} \|\boldsymbol{A} \circ \boldsymbol{W}^{(p)}(s)\|_p, \qquad s \in S, \tag{2.3}$$

is simple max-stable.

Proof. For $p = \infty$, we have $U^{(p)}(s) = 1$ a.s. and, thus, representation (2.3) is of the same form as representation (2.1). Consequently, max-stability follows from de Haan (1984). For $p \in (1, \infty)$, we first show that $\|\boldsymbol{A} \circ \boldsymbol{W}^{(p)}(s)\|_p < \infty$ a.s. According to Campbell's Theorem (cf. Kingman, 1993, p. 28), this holds true if and only if

$$\mathbb{E}\left(\int_0^\infty \min\{|aW^{(p)}(s)|^p, 1\}\, a^{-2}\, \mathrm{d}a\right) < \infty. \tag{2.4}$$

Substituting $v = aW(s)$, we can easily see that the left-hand side of (2.4) equals

$$\mathbb{E}\left(W^{(p)}(s)\right) \cdot \int_0^\infty \min\{|v|^p, 1\} v^{-2}\, \mathrm{d}v = 1 + \frac{1}{p-1}.$$

Thus, $\|\boldsymbol{A} \circ \boldsymbol{W}^{(p)}(s)\|_p < \infty$ a.s. Then, for $s_1, \ldots, s_n \in S$, $x_1, \ldots, x_n > 0$, $n \in \mathbb{N}$, we obtain

$$\begin{aligned} &\mathbb{P}(X(s_i) \le x_i,\, i = 1, \ldots, n) \\ &= \mathbb{E}\left(\mathbb{P}\left(U^{(p)}(s_i) \le \frac{\Gamma(1 - p^{-1}) x_i}{\|\boldsymbol{A} \circ \boldsymbol{W}^{(p)}(s_i)\|_p},\, i = 1, \ldots, n \;\middle|\; \boldsymbol{A}, \boldsymbol{W}^{(p)}\right)\right) \\ &= \mathbb{E}\left(\exp\left(-\sum_{i=1}^n \left(\frac{\Gamma(1 - p^{-1}) x_i}{\|\boldsymbol{A} \circ \boldsymbol{W}^{(p)}(s_i)\|_p}\right)^{-p}\right)\right). \end{aligned}$$

Using well-known results on the Laplace functional of Poisson point processes, this yields

$$\begin{aligned}
&\mathbb{P}(X(s_i) \le x_i,\, i = 1, \dots, n) \\
&= \exp\left(\mathbb{E}\Big(\int_0^\infty \Big\{ \exp\Big(-\sum_{i=1}^n \Big(\frac{aW^{(p)}(s_i)}{\Gamma(1-p^{-1})x_i}\Big)^p\Big) - 1\Big\} a^{-2}\,\mathrm{d}a\Big)\right) \\
&= \exp\left(\mathbb{E}\left(\left\|\left(\frac{W^{(p)}(s_i)}{x_i}\right)_{i=1}^n\right\|_p\right) \cdot \frac{1}{p\Gamma(1-p^{-1})} \cdot \int_0^\infty \left(e^{-a} - 1\right) a^{-1-p^{-1}}\,\mathrm{d}a\right) \\
&= \exp\left(-\mathbb{E}\left(\left\|\left(\frac{W^{(p)}(s_i)}{x_i}\right)_{i=1}^n\right\|_p\right)\right) \qquad (2.5)
\end{aligned}$$

where we used Formula 3.478.2 in Gradshteyn and Ryzhik (2007). Thus, for m independent copies $X_1, \dots, X_m$ of X, $m \in \mathbb{N}$, the homogeneity of the ℓ^p norm yields

$$\mathbb{P}\left(\frac{1}{m}\max_{j=1,\dots,m} X_j(s_i) \le x_i,\ i = 1, \dots, n\right) = \mathbb{P}\left(X(s_i) \le x_i,\, i = 1, \dots, n\right),$$

i.e. Z is simple max-stable. □

Remark 2.2. Theorem 2.1 could alternatively be verified by observing that the process $T(s) = \|\boldsymbol{A} \circ \boldsymbol{W}^{(p)}(s)\|_p^p$, $s \in S$, is α-stable with $\alpha = 1/p$ (see also the proof of Theorem 2.5). Thus, all the finite-dimensional distributions of X are generalized logistic mixtures (cf. Fougères et al., 2009, 2013) and, consequently, are max-stable distributions.

Noting that the finite-dimensional distributions of the Reich–Shaby model (2.2) are given by

$$\mathbb{P}\left(X(s_i) \le x_i,\, i = 1, \dots, n\right) = \exp\left(-\sum_{j=1}^L \left\| \left(\frac{w_j(s_i)}{x_i}\right)_{i=1}^n \right\|_p\right),$$

it can be easily seen that (2.2) is a special case of representation (2.3) where W follows the discrete distribution $\mathbb{P}(W = Lw_i) = 1/L$, $i = 1, \dots, L$. Further, the classical spectral representation (2.1) by de Haan (1984) can be recovered from representation (2.3) with $p = \infty$.

Analogously to the law of the spectral processes $\{V_i(s),\, s \in S\}_{i\in\mathbb{N}}$ in representation (2.1), the law of the processes $\{W_i^{(p)}(s),\, s \in S\}_{i\in\mathbb{N}}$ in the ℓ^p norm based representation of a given process $\{X(s),\, i \in S\}$ is not unique: Let Y_i, $i \in \mathbb{N}$, be independently and identically distributed random variables with $\mathbb{E}(Y_i) = 1$ which are independent from $\sum_{i\in\mathbb{N}} \delta_{A_i}$ and $\{W^{(p)}(s),\, s \in S\}$. Then, the processes $\{U^{(p)}(s)/\Gamma(1-p^{-1}) \cdot \|\boldsymbol{A} \circ \boldsymbol{W}^{(p)}(s)\|_p,\, s \in S\}$ and $\{U^{(p)}(s)/\Gamma(1-p^{-1}) \cdot \|\boldsymbol{A} \circ \boldsymbol{Y} \circ \boldsymbol{W}^{(p)}(s)\|_p,\, s \in S\}$ are equal in distribution.
Consequently, even for some fixed $p \in (1, \infty]$ representation (2.3) of a simple max-stable process X is not unique. Furthermore, there might be representations of type (2.3) with different p for the same process X. Such equivalent representations are discussed in the following section.

2.3 Equivalent Representations

By de Haan (1984), the class of simple max-stable processes is fully covered by the class of processes which allow for the spectral representation (2.1), i.e. representation (2.3) with $p = \infty$. Thus, any ℓ^p norm based representation (2.3) with $p < \infty$ of a simple max-stable process can be transformed to an equivalent representation of type (2.1). This transformation is presented in the following proposition. Even more generally, it is shown how a ℓ^q norm based representation can be derived from a ℓ^p norm based representation with $p < q < \infty$.

Proposition 2.3. *Let X be a simple max-stable process with representation (2.3) for some $p \in (1, \infty)$. Then, the following holds:*

1. *The process X allows for the spectral representation (2.1) with*

$$V(\cdot) =_d \frac{U^{(p)}(\cdot)}{\Gamma(1-p^{-1})} W^{(p)}(\cdot). \tag{2.6}$$

2. *For $q \in (p, \infty)$, the process X satisfies*

$$X(\cdot) =_d \frac{U^{(q)}(\cdot)}{\Gamma(1-q^{-1})} \|\boldsymbol{A} \circ \boldsymbol{W}^{(q)}(\cdot)\|_q, \tag{2.7}$$

where $\{U^{(q)}(s)\}_{s\in S}$ is a collection of independent Φ_q random variables and $W_i^{(q)}$, $i \in \mathbb{N}$, are independent copies of a stochastic process $\{W^{(q)}(s),\, s \in S\}$ given by

$$W^{(q)}(s) = \frac{\Gamma(1-q^{-1})}{\Gamma(1-p^{-1})} (T_{(p/q)}(s))^{p/q} \cdot W^{(p)}(s), \quad s \in S.$$

Here, independently from the process $W^{(p)}$, the collection $\{T_{(p/q)}(s)\}_{s\in S}$ consists of independent stable random variables whose law is given by the Laplace transform

$$\mathbb{E}\left(e^{-tT_{(p/q)}(s)}\right) = e^{-t^{p/q}}, \quad t \geq 0.$$

Proof. 1. By comparing the finite-dimensional distributions of the processes defined via (2.1) and (2.3), it suffices to show that

$$\frac{1}{\Gamma(1-p^{-1})} \mathbb{E}\left(\left\|\left(\frac{U^{(p)}(s_i)W^{(p)}(s_i)}{x_i}\right)_{i=1}^n\right\|_\infty\right) = \mathbb{E}\left(\left\|\left(\frac{W^{(p)}(s_i)}{x_i}\right)_{i=1}^n\right\|_p\right), \tag{2.8}$$

for all $s_1, \ldots, s_n \in S$, $x_1, \ldots, x_n > 0$, $n \in \mathbb{N}$. To this end, we first note that, for $y > 0$,

$$\mathbb{P}\left(\left\|\left(\frac{U^{(p)}(s_i)W^{(p)}(s_i)}{x_i}\right)_{i=1}^n\right\|_\infty \leq y \;\middle|\; \boldsymbol{W}^{(p)}\right) = \exp\left(-\frac{1}{y^p}\sum_{i=1}^n \left(\frac{W^{(p)}(s_i)}{x_i}\right)^p\right),$$

that is, conditionally on the vector $\boldsymbol{W}^{(p)}$, the norm $\|(U^{(p)}(s_i)W^{(p)}(s_i)/x_i)_{i=1}^n\|_\infty$ follows a p-Fréchet distribution with scale parameter $\|(W^{(p)}(s_i)/x_i)_{i=1}^n\|_p$. Thus,

$$\begin{aligned}\mathbb{E}\left(\left\|\left(\frac{U^{(p)}(s_i)W^{(p)}(s_i)}{x_i}\right)_{i=1}^n\right\|_\infty\right) &= \mathbb{E}_W\left\{\mathbb{E}\left(\left\|\left(\frac{U^{(p)}(s_i)W^{(p)}(s_i)}{x_i}\right)_{i=1}^n\right\|_\infty \;\middle|\; \boldsymbol{W}^{(p)}\right)\right\} \\ &= \mathbb{E}_W\left\{\Gamma(1-p^{-1})\left\|\left(\frac{W^{(p)}(s_i)}{x_i}\right)_{i=1}^n\right\|_p\right\},\end{aligned}$$

i.e. Equation (2.8).

2. From the first part of the proposition, it follows that the right-hand side of (2.7) allows for a spectral representation (2.1) where the spectral functions are independent copies of the process $\widetilde{V}$ given by

$$\widetilde{V}(\cdot) = \frac{U^{(q)}(\cdot) \cdot (T_{p/q}(\cdot))^{1/q}}{\Gamma(1-p^{-1})} \cdot W^{(p)}(\cdot),$$

while the spectral functions of the process X on the left-hand side of (2.7) are independent copies of the process V given in (2.6). Conditioning on the value of the stable random variable $T_{(p/q)}(s)$, it can be shown that the product $U^{(q)}(s) \cdot T_{(p/q)}(s)$ has the distribution function Φ_p for all $s \in S$ (cf. Fougères et al., 2009) and, thus, $\widetilde{V}(\cdot) =_d V(\cdot)$.

□

Remark 2.4. Even though the transformation in the second part of the proposition requires $p < q < \infty$, the two cases $p = q$ and $q = \infty$ can be regarded as limiting cases.
As $q \searrow p$, we obtain that $U^{(q)}(\cdot) \to_d U^{(p)}(\cdot)$ and $\{T_{(p/q)}(s)\}_{s\in S}$ converges in distribution to a collection of random variables which equal 1 a.s. Thus, in the limit $p = q$, there is no transformation.
As $q \to \infty$, we have that $\Gamma(1-q^{-1}) \to 1$ and each $U^{(q)}(s)$, $s \in S$, converges to 1 a.s. Further, by Theorem 1.4.5 in Samorodnitsky and Taqqu (1994), for each $s \in S$, the random variable $T_{(p/q)}(s)$ can be represented as

$$T_{(p/q)}(s) = \frac{1}{\Gamma(1-p/q)} \sum_{i\in\mathbb{N}} (\tilde{A}_i Y_i)^{q/p}$$

where $\{\tilde{A}_i\}_{i\in\mathbb{N}}$ are the points of a Poisson point process on $(0,\infty)$ with intensity $\tilde{a}^{-2}\mathrm{d}\tilde{a}$ and Y_i, $i \in \mathbb{N}$, are independently and identically distributed non-negative random variables with expectation 1. Thus, as $q \to \infty$,

$$\big(T_{(p/q)}(s)\big)^{1/q} =_d \left(\frac{1}{\Gamma(1-p/q)} \sum_{i\in\mathbb{N}} (\tilde{A}_i Y_i)^{q/p}\right)^{1/q} \longrightarrow_d \max_{i\in\mathbb{N}} (\tilde{A}_i Y_i)^{1/p}$$

which has the distribution function Φ_p. Consequently, $\big(T_{(p/q)}(\cdot)\big)^{1/q} \to_d U^{(p)}(\cdot)$.

Denoting by $\mathcal{MS}$ the class of all simple max-stable processes and by $\mathcal{MS}_p$ the class of simple max-stable processes allowing for a ℓ^p norm based spectral representation (2.3), Proposition 2.3 yields

$$\mathcal{MS}_p \subset \mathcal{MS}_q \subset \mathcal{MS}_\infty = \mathcal{MS}, \quad 1 < p < q < \infty.$$

A full characterization of the class $\mathcal{MS}_p$ is given in the following section.

2.4 Existence of ℓ^p Norm Based Representations

In the following, we will present a necessary and sufficient criterion for the existence of a ℓ^p norm based representation of a simple max-stable process X in terms of the stable

tail dependence functions of its finite-dimensional distributions. For a simple max-stable distribution $(X(s_1), \dots, X(s_n))^\top$, its stable tail dependence function $l_{s_1,\dots,s_n}$ is defined via

$$\begin{aligned} l_{s_1,\dots,s_n} &: [0,\infty)^n \to [0,\infty) \\ &(x_1,\dots,x_n) \mapsto -\log\left\{\mathbb{P}\left(X(s_1) \le \frac{1}{x_1}, \dots, X(s_n) \le \frac{1}{x_n}\right)\right\}. \end{aligned}$$

From the spectral representation (2.1), we obtain the form

$$l_{s_1,\dots,s_n}(x) = \mathbb{E}\left(\max_{i=1,\dots,n} x_i W(s_i)\right), \quad x \in [0,\infty)^n. \tag{2.9}$$

The stable tail dependence function is homogeneous and convex (cf. Beirlant et al., 2004, among others). Further, from Equation (2.9) together with dominated convergence, we can deduce that the stable tail dependence function is continuous.

Theorem 2.5. *Let $\{X(s),\ s \in S\}$ a simple max-stable process and $p \in (1,\infty)$. Then, the following statements are equivalent:*

(i) X possesses a ℓ^p norm based representation (2.3)*.*

(ii) For all pairwise distinct $s_1, \dots, s_n \in S$ and $n \in \mathbb{N}$, the function $f^{(p)}_{s_1,\dots,s_n}$, defined by

$$f^{(p)}_{s_1,\dots,s_n}(x) = l_{s_1,\dots,s_n}\left(x_1^{1/p}, \dots, x_n^{1/p}\right), \quad x = (x_1,\dots,x_n) \in [0,\infty)^n,$$

is conditionally negative definite on the additive semigroup $[0,\infty)^n$, that is, for all $x^{(1)}, \dots, x^{(m)} \in [0,\infty)^n$ and $a_1, \dots, a_m \in \mathbb{R}$ such that $\sum_{i=1}^m a_i = 0$, we have

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j f^{(p)}_{s_1,\dots,s_n}(x^{(i)} + x^{(j)}) \le 0. \tag{2.10}$$

Proof. Firstly, we show that (i) implies (ii). To this end, let X be a simple max-stable process with representation (2.3). Then, from (2.5), we obtain that

$$\begin{aligned} f^{(p)}_{s_1,\dots,s_n}(x) = & -\log\left\{\mathbb{P}\left(X(s_1) \le \frac{1}{x_1^{1/p}}, \dots, X(s_n) \le \frac{1}{x_n^{1/p}}\right)\right\} \\ &= \mathbb{E}\left\{\left(\sum_{i=1}^n x_i W^{(p)}(s_i)^p\right)^{1/p}\right\}, \qquad x = (x_1,\dots,x_n) \in [0,\infty)^n. \end{aligned}$$

Now, let $w(s_1), \dots, w(s_n) \ge 0$ be fixed. Then, by a straightforward computation, it can be seen that the function $x \mapsto \sum_{k=1}^n x_k w(s_k)^p$ is conditionally negative definite on $[0,\infty)^n$. As the function $y \mapsto y^{1/p}$ is a Bernstein function and the composition of a conditionally negative function and a Bernstein function yields a conditionally negative definite function (Berg et al., 1984, Theorem 3.2.9), the function $x \mapsto (\sum_{k=1}^n x_k w(s_k)^p)^{1/p}$ is conditionally negative definite, as well. Being a mixture, the same is true for $f^{(p)}_{s_1,\dots,s_n}$.

Secondly, we show that (ii) implies (i). From the conditionally negative definiteness of $f^{(p)}_{s_1,\dots,s_n}$, it follows that $e^{-f^{(p)}_{s_1,\dots,s_n}}$ is positive definite on $[0,\infty)^n$ (Berg et al., 1984, Theorem 3.2.2). As $l_{s_1,\dots,s_n}$ is non-negative and continuous, $e^{-f^{(p)}_{s_1,\dots,s_n}}$ is further bounded by 1

and continuous. Thus, by Theorem 4.4.7 in Berg et al. (1984), there exists a unique finite measure $\mu_{s_1,\dots,s_n}$ on $[0,\infty)^n$ with Laplace transform

$$\mathcal{L}\mu_{s_1,\dots,s_n}(x) = \int_{[0,\infty)^n} \exp\left(-\langle x, a\rangle\right) \mu(\mathrm{d}a) = \exp\left(-f_{s_1,\dots,s_n}(x)\right), \quad x \in [0,\infty)^n. \tag{2.11}$$

Because of $\mu_{s_1,\dots,s_n}([0,\infty)^n) = \exp(-l_{s_1,\dots,s_n}(0,\dots,0)) = 1$, $\mu_{s_1,\dots,s_n}$ is a probability measure. Further,

$$\begin{aligned} &l_{s_1,\dots,s_n}(x_1,\dots,x_{i-1},0,x_{i+1},\dots,x_n) \\ &= l_{s_1,\dots,s_{i-1},s_{i+1},\dots,s_n}(x_1,\dots,x_{i-1},x_{i+1},\dots,x_n) \end{aligned} \tag{2.12}$$

for all $x = (x_1,\dots,x_n) \in [0,\infty)^n$ and $i \in \{1,\dots,n\}$ implies that

$$\begin{aligned} &\mu_{s_1,\dots,s_n}(A_1 \times \dots \times A_{i-1} \times [0,\infty) \times A_{i+1} \times \dots \times A_n) \\ &= \mu_{s_1,\dots,s_{i-1},s_{i+1},\dots,s_n}(A_1 \times \dots \times A_{i-1} \times A_{i+1} \times \dots \times A_n) \end{aligned}$$

for all Borel sets $A_1,\dots,A_n \subset [0,\infty)$ and $i \in \{1,\dots,n\}$. Consequently, the family $\{\mu_{s_1,\dots,s_n} : s_1,\dots,s_n \in S,\ n \in \mathbb{N}\}$ of probability measures satisfies the consistency conditions from Kolmogorov's existence theorem. Thus, there exists a stochastic process $\{T(s),\ s \in S\}$ with finite-dimensional distributions $\mu_{\cdot}$.
Now, let $\{U^{(p)}(s)\}_{s\in S}$ be a collection of independent Φ_p random variables and

$$\tilde{X}(s) = U^{(p)}(s)T(s)^{1/p}, \quad s \in S.$$

Then, for all pairwise distinct $s_1,\dots,s_n \in S$ and $x_1,\dots,x_n > 0$, we have

$$\begin{aligned} &\mathbb{P}\left(\tilde{X}(s_1) \le x_1,\dots,\tilde{X}(s_n) \le x_n\right) \\ &= \mathbb{E}\left\{\mathbb{P}\left(U^{(p)}(s_1) \le \frac{x_1}{T^{1/p}(s_1)},\dots,U^{(p)}(s_n) \le \frac{x_n}{T^{1/p}(s_n)} \;\middle|\; T(s_1),\dots,T(s_n)\right)\right\} \\ &= \mathbb{E}\left\{\exp\left(-\sum_{i=1}^n \frac{T(s_i)}{x_i^p}\right)\right\}, \end{aligned}$$

By Equation (2.11), we obtain

$$\begin{aligned} \mathbb{P}(\tilde{X}(s_1) \le x_1,\dots,\tilde{X}(s_n) \le x_n) &= \exp\left(-f^{(p)}_{s_1,\dots,s_n}(x_1^{-p},\dots,x_n^{-p})\right) \\ &= \mathbb{P}\left(X(s_1) \le x_1,\dots,X(s_n) \le x_n\right). \end{aligned}$$

Thus, X allows for the spectral representation

$$X(s) = U^{(p)}(s)T^{1/p}(s), \qquad s \in S. \tag{2.13}$$

Now, let $T^{(1)},\dots,T^{(m)}$ be m independent copies of T, $m \in \mathbb{N}$. Then, for all $s_1,\dots,s_n \in S$ and $x = (x_1,\dots,x_n) \in [0,\infty)^n$, we have

$$\begin{aligned} &\mathbb{E}\left\{\exp\left(-\left\langle x, \left(\sum_{k=1}^m T^{(k)}(s_i)\right)_{i=1}^n \right\rangle\right)\right\} = \left[\mathbb{E}\left\{\exp(-\langle x, (T(s_i))_{i=1}^n\rangle)\right\}\right]^m \\ &= \exp(-m \cdot l_{s_1,\dots,s_n}(x_1^{1/p},\dots,x_n^{1/p})) = \exp(-l_{s_1,\dots,s_n}((m^p x_1)^{1/p},\dots,(m^p x_n)^{1/p})) \\ &= \mathbb{E}\left\{\exp\left(\langle x, m^p (T(s_i))_{i=1}^n\rangle\right)\right\}, \end{aligned}$$

where we used the homogeneity of the stable tail dependence function. Hence, for all $s_1, \dots, s_n \in S$, the vectors $\left(\sum_{k=1}^m T^{(k)}(s_i)\right)_{i=1}^n$ and $m^p(T(s_i))_{i=1}^n$ have the same distribution, i.e. $\{T(s),\, s \in S\}$ is an α-stable process with $\alpha = 1/p$. Thus, from Theorem 13.1.2 and Theorem 3.10.1 in Samorodnitsky and Taqqu (1994), we can deduce that $\{T(s),\, s \in S\}$ allows for the representation

$$T(s) = \frac{1}{\Gamma(1-p^{-1})^p} \sum_{i \in \mathbb{N}} A_i^p \tilde{W}_i(s), \quad s \in S, \tag{2.14}$$

where $\{A_i\}_{i\in\mathbb{N}}$ are the points of a Poisson point process on $[0,\infty)$ with intensity $a^{-2}\mathrm{d}a$ and $\{\tilde{W}_i(s),\, s \in S\}$ are independent and identically distributed stochastic processes which are independent from $\{A_i\}_{i\in\mathbb{N}}$ and satisfy $\mathbb{E}(\tilde{W}_i(s)^{1/p}) = l_s(1) = 1$ for all $s \in S$. Defining $W_i^{(p)}(s) = \tilde{W}_i(s)^{1/p}$, $s \in S$, $i \in \mathbb{N}$, and plugging Equation (2.14) into Equation (2.13), we obtain Equation (2.3). □

Remark 2.6. Note that Theorem 2.5 assumes that, for each $s_1, \dots, s_n \in S$, $\ell_{s_1,\dots,s_n}$ is the stable tail dependence function of the simple max-stable vector $(X(s_1), \dots, X(s_n))^\top$. The conditional negative definiteness of the function $f^{(p)}_{s_1,\dots,s_n}$ is an additional condition. In particular, it is always satisfied for $p = \infty$ – i.e. any simple max-stable process allows for de Haan's (1984) spectral representation (2.1) – as $f^{(\infty)}_{s_1,\dots,s_n} \equiv l_{s_1,\dots,s_n}(1,\dots,1)$ is always conditionally negative definite.
In order to check whether a function $l_{s_1,\dots,s_n}$ is the stable tail dependence function of some process X with an ℓ^p norm based representation, we first need to ensure that $l_{s_1,\dots,s_n}$ is a valid stable tail dependence function. This can be done by checking necessary and sufficient conditions given in Molchanov (2008) and Ressel (2013), for instance.

Using an integral representation of continuous conditionally negative definite functions on $[0,\infty)^n$ (cf. Paragraph 4.4.6 in Berg et al., 1984) condition (ii) in Theorem 2.5 can be reformulated yielding the following corollary.

Corollary 2.7. *For a simple max-stable process $\{X(s),\, s \in S\}$ and $p \in (1,\infty)$, the following statements are equivalent:*

(i) X possesses a ℓ^p norm based representation (2.3).

(ii) For all pairwise distinct $s_1, \dots, s_n \in S$ and $n \in \mathbb{N}$, there exist a vector

$$c(s_1,\dots,s_n) = (c_1(s_1,\dots,s_n), \dots, c_n(s_1,\dots,s_n))^\top \in [0,\infty)^n$$

and a Radon measure $\mu_{s_1,\dots,s_n}$ on $[0,\infty)^n$ such that the stable tail dependence function $l_{s_1,\dots,s_n}$ satisfies

$$l_{s_1,\dots,s_n}(x) = \sum_{i=1}^n c_i(s_1,\dots,s_n) \cdot x_i^p + \int_{[0,\infty)^n} \left\{1 - \exp\left(-\sum_{i=1}^n a_i x_i^p\right)\right\} \mu_{s_1,\dots,s_n}(\mathrm{d}a),$$

for all $x = (x_1,\dots,x_n)^\top \in [0,\infty)^n$.

From the characterization given in Theorem 2.5, we can deduce necessary conditions on the dependence structure of a max-stable process with ℓ^p norm based representation (2.3) in terms of its extremal coefficients: For a general simple max-stable process $\{X(s),\, s \in S\}$ and a finite set $\tilde{S} = \{s_1, \ldots, s_n\} \subset S$, let the extremal coefficient $\theta(\tilde{S})$ be defined via

$$\mathbb{P}\left(\max_{s\in S} X(s) \leq x\right) = \exp\left(-\frac{\theta(\tilde{S})}{x}\right), \quad x > 0.$$

Then, we necessarily have $\theta(\tilde{S}) \in [1, n]$ where $\theta(\tilde{S}) = n$ if and only if $X(s_1)$, …, $X(s_n)$ are independent and $\theta(\tilde{S}) = 1$ if and only if $X(s_1) = X(s_2) = \ldots = X(s_n)$ a.s. The extremal coefficient is closely connected to the stable tail dependence function via the relation

$$\theta(\{s_1, \ldots, s_n\}) = l_{s_1,\ldots,s_n}(1, \ldots, 1).$$

If X further allows for an ℓ^p norm based representation (2.3), we obtain the following condition.

Proposition 2.8. *Let $\{X(s),\, s \in S\}$ be a simple max-stable process with representation (2.3) and $S_1, S_2 \subset S$ be finite and disjoint. Then, we have*

$$\theta(S_1 \cup S_2) \geq 2^{1/p}\frac{\theta(S_1) + \theta(S_2)}{2}.$$

Proof. Let $S_1 = \{s_1, s_2, \ldots, s_{k_1}\}$ and $S_2 = \{s_{k_1+1}, \ldots, s_{k_1+k_2}\}$ be disjoint. Furthermore, let $\{e_1, \ldots, e_{k_1+k_2}\}$ denote the standard basis in $\mathbb{R}^{k_1+k_2}$. As the function

$$(x_1, \ldots, x_{k_1+k_2}) \mapsto l_{s_1,\ldots,s_{k_1+k_2}}\left(x_1^{1/p}, \ldots, x_{k_1+k_2}^{1/p}\right)$$

is conditionally negative definite by Theorem 2.5, inequality (2.10) particularly holds true for $m = 2$, $a_1 = 1$, $a_2 = -1$, $x^{(1)} = \sum_{i=1}^{k_1} e_i$ and $x^{(2)} = \sum_{i=k_1+1}^{k_1+k_2} e_i$, i.e.

$$l_{s_1,\ldots,s_{k_1+k_2}}\left(2^{\frac{1}{p}}\sum_{i=1}^{k_1} e_i\right) + l_{s_1,\ldots,s_{k_1+k_2}}\left(2^{\frac{1}{p}}\sum_{i=k_1+1}^{k_1+k_2} e_i\right) - 2l_{s_1,\ldots,s_{k_1+k_2}}\left(\sum_{i=1}^{k_1+k_2} e_i\right) \leq 0.$$

Using the homogeneity and property (2.12) of the stable tail dependence function, we obtain

$$2^{1/p} l_{s_1,\ldots,s_{k_1}}(1, \ldots, 1) + 2^{1/p} l_{s_{k_1+1},\ldots,s_{k_1+k_2}}(1, \ldots, 1) - 2l_{s_1,\ldots,s_{k_1+k_2}}(1, \ldots, 1) \leq 0.$$

As $\theta(\tilde{S}) = l_{\tilde{S}}(1, \ldots, 1)$ for any finite $\tilde{S} \subset S$, this yields the assertion. □

Of particular interest in extreme value analysis is the case of the pairwise extremal coefficient function (cf. Smith, 1990; Schlather and Tawn, 2003) where $\tilde{S} = \{s_1, s_2\}$. Then, Proposition 2.8 provides the lower bound

$$\theta(\{s_1, s_2\}) \geq 2^{1/p} \qquad \text{for all } s_1 \neq s_2 \in S. \tag{2.15}$$

For the particular case of model (2.2), this bound has already been found by Reich and Shaby (2012) motivating their interpretation of model (2.2) as a max-stable process with nugget effect in analogy to the Gaussian case.

The bound (2.15) and the characterization of simple max-stable processes with a ℓ^p norm based representation given in Theorem 2.5 can be used to show the existence of a *minimal* ℓ^p norm based representation of a simple max-stable process X, i.e. the existence of some $p_{\min}(X)$ such that $X \in \mathcal{MS}_p$ if and only if $p \geq p_{\min}(X)$.

Corollary 2.9. *Let $\{X(s),\, s \in S\}$ be a simple max-stable process such that not all random variables $\{X(s)\}_{s\in S}$ are independent. Then, there exists a number $p_{\min}(X) \in (1,\infty]$ such that $X \in \mathcal{MS}_p$ if and only if $p \geq p_{\min}(X)$.*

Proof. By de Haan (1984), any simple max-stable process X satisfies $X \in \mathcal{MS}_\infty$. Thus, the assertion follows directly if

$$p_{\min}(X) = \inf\{p > 1 :\, X \in \mathcal{MS}_p\} = \infty.$$

Thus, we restrict ourselves to the case that $p_{\min}(X) < \infty$. As not all the random variables $\{X(s)\}_{s\in S}$ are independent, there exist $s_1, s_2 \in S$ and $\varepsilon > 0$ such that

$$\theta(\{s_1, s_2\}) < 2^{1/(1+\varepsilon)}.$$

Hence, by Equation (2.15), we obtain that $p_{\min}(X) \geq 1 + \varepsilon$. Now, using the fact that $\mathcal{MS}_p \subset \mathcal{MS}_q$ for $p < q$, it remains to show that $X \in \mathcal{MS}_{p_{\min}(X)}$. By Theorem 2.5, for all pairwise distinct $s_1, \ldots, s_n \in S$, $n \in \mathbb{N}$, $a_1, \ldots, a_m \in \mathbb{R}$ such that $\sum_{i=1}^m a_i = 0$, $x^{(1)}, \ldots, x^{(m)} \in [0,\infty)^n$ and $m \in \mathbb{N}$ we have that

$$\sum_{i=1}^m \sum_{j=1}^n a_i a_j l_{s_1,\ldots,s_n}\left((x_1^{(i)} + x_1^{(j)})^{1/p}, \ldots, (x_n^{(i)} + x_n^{(j)})^{1/p}\right) \leq 0$$

for all $p > p_{\min}(X)$. By the continuity of $l_{s_1,\ldots,s_m}$, the same holds true for $p = p_{\min}(X)$, and, thus, by Theorem 2.5, $X \in \mathcal{MS}_{p_{\min}(X)}$. □

For any $p \in (1,\infty]$, we now give an example for a simple max-stable process $X^{(p)}$ such that $p_{\min}(X^{(p)}) = p$. Thus, we will also see that

$$\mathcal{MS}_p \subsetneq \mathcal{MS}_q \subsetneq \mathcal{MS}_\infty = \mathcal{MS}, \quad 1 < p < q < \infty.$$

We consider the process $X_{\log}^{(p)} \in \mathcal{MS}_p$ which possesses an ℓ^p norm based representation (2.3) with $W(s) = 1$ a.s. for all $s \in S$. From Equation (2.5), for pairwise distinct $s_1, \ldots, s_n \in S$, we obtain the finite-dimensional distributions

$$\mathbb{P}\left(X_{\log}^{(p)}(s_i) \leq x_i,\ 1 \leq i \leq n\right) = \exp\left\{-\left(\sum_{i=1}^n x_i^{-p}\right)^{1/p}\right\}, \qquad x_1, \ldots, x_n > 0,$$

i.e. all the multivariate distributions are multivariate logistic distributions (Gumbel, 1960). Thus, the process $X_{\log}^{(p)}$ has pairwise extremal coefficients $\theta(s,t) = 2^{1/p}$ for all $s, t \in S$, $s \neq t$. From Equation (2.15), it follows that $X_{\log}^{(p)} \notin \mathcal{MS}_{p'}$ for $p' < p$. Consequently, we have $p_{\min}(X_{\log}^{(p)}) = p$.

While we have $\theta(s,t) = 2^{1/p_{\min}(X)}$ for the process $X = X_{\log}^{(p)}$, the connection between $p_{\min}(X)$ and the pairwise extremal coefficients $\theta(s,t)$ is more involved in general. To see this, we consider the case $S = \{s_1, s_2\}$. In this case, for a process $X \in \mathcal{MS}_p$, the condition $\theta(s_1, s_2) = 2^{1/p}$ implies $W^{(p)}(s_1) = W^{(p)}(s_2)$ a.s., i.e. X necessarily follows a bivariate logistic distribution. For any other bivariate simple max-stable distribution, we have $\theta(s_1, s_2) > 2^{1/p_{\min}(X)}$.

2.5 Properties of Processes with ℓ^p Norm Based Representation

In this section, we will analyze several properties of simple max-stable processes with an ℓ^p norm based representation in more detail. We will particularly focus on properties related to the dependence structure of the process such as stationarity, ergodicity and mixing. A characteristic feature of a process X with ℓ^p norm based representation (2.3) is the additional noise introduced via the process $\{U^{(p)}(s),\ s \in S\}$. Thus, we will compare the process X to a "denoised" reference process

$$X^*(s) = \max_{i\in\mathbb{N}} A_i W_i^{(p)}(s), \quad s \in S,$$

i.e. the simple max-stable process constructed via the same spectral functions used in the original (ℓ^∞ norm based) spectral representation (2.1). As the processes X and X^* just differ by the Fréchet noise process $U^{(p)}$, we will call X^* the denoised max-stable process associated to X.
The following proposition relates the extremal coefficients $\theta(\{s_1, s_2\})$, $s_1, s_2 \in S$, of X to the extremal coefficients $\theta^*(\{s_1, s_2\}) = \mathbb{E}\left(\max\{W^{(p)}(s_1), W^{(p)}(s_2)\}\right)$ of the associated denoised process X^*. We obtain that extremal dependence of the process X is always weaker than dependence of the associated denoised process – as expected.

Proposition 2.10. *Let $\{X(s),\ s \in S\}$ be a simple max-stable process with ℓ^p norm based representation (2.3) with $p \in (1, \infty]$. Then, for the pairwise extremal coefficients $\theta(\{s_1, s_2\})$, we obtain the bounds*

$$\theta^*(\{s_1, s_2\}) \le \theta(\{s_1, s_2\}) \le 2^{1/p} \cdot \theta^*(\{s_1, s_2\})^{1-\frac{1}{p}},$$

where $\theta^*(\{s_1, s_2\})$ are the pairwise extremal coefficients of the associated denoised process X^*.

Proof. In the case $p = \infty$, we have

$$\theta(\{s_1, s_2\}) = \mathbb{E}\left(\max\{W^{(p)}(s_1), W^{(p)}(s_2)\}\right) = \theta^*(\{s_1, s_2\}),$$

which equals both the lower and the upper bound given in the assertion.
Now, let $p \in (1, \infty)$. Then, we have the lower bound

$$\begin{aligned}\theta(\{s_1, s_2\}) &= \mathbb{E}\left\{\left(W^{(p)}(s_1)^p + W^{(p)}(s_2)^p\right)^{1/p}\right\} \\ &\ge \mathbb{E}\left(\max\{W^{(p)}(s_1), W^{(p)}(s_2)\}\right) = \theta^*(\{s_1, s_2\}).\end{aligned}$$

Further, for any $p < r < \infty$ and $\boldsymbol{w} \in [0, \infty)^2$, we obtain

$$\|\boldsymbol{w}\|_p^p \le \|\boldsymbol{w}\|_1^{\frac{r-p}{r-1}} \cdot \|\boldsymbol{w}\|_r^{r\frac{p-1}{r-1}}$$

(cf. Theorem 18 in Hardy et al., 1952), or equivalently

$$\|\boldsymbol{w}\|_p \le \|\boldsymbol{w}\|_1^{\frac{1}{p}\frac{r-p}{r-1}} \cdot \|\boldsymbol{w}\|_r^{\frac{1-p^{-1}}{1-r^{-1}}}.$$

As $r \to \infty$, this yields

$$\|\boldsymbol{w}\|_p \le \|\boldsymbol{w}\|_1^{1/p} \cdot \|\boldsymbol{w}\|_\infty^{1-p^{-1}}.$$

Taking the expectation of $\boldsymbol{w}$ with respect to the joint distribution of $W^{(p)}(s_1)$ and $W^{(p)}(s_2)$ and applying Hölder's inequality, we obtain the upper bound

$$\begin{aligned}
&\theta(\{s_1, s_2\}) = \mathbb{E}\left\{\left(W^{(p)}(s_1)^p + W^{(p)}(s_2)^p\right)^{1/p}\right\} \\
&\leq \mathbb{E}\left\{\left(W^{(p)}(s_1) + W^{(p)}(s_2)\right)^{1/p} \cdot \max\{W^{(p)}(s_1), W^{(p)}(s_2)\}^{1-p^{-1}}\right\} \\
&\leq \left[\mathbb{E}\left\{W^{(p)}(s_1) + W^{(p)}(s_2)\right\}\right]^{1/p} \left[\mathbb{E}\left(\max\{W^{(p)}(s_1), W^{(p)}(s_2)\}\right)\right]^{1-p^{-1}}.
\end{aligned}$$

The assertion follows from $\mathbb{E}\{W^{(p)}(s_1)\} = \mathbb{E}\{W^{(p)}(s_2)\} = 1$. □

In the following, we will consider the case that $S = \mathbb{Z}$. In this case, properties such as stationarity, ergodicity or mixing are of interest. For a simple max-stable process $\{X(s),\, s \in \mathbb{Z}\}$ with representation (2.1), necessary and sufficient conditions for these properties can be expressed in terms of the distribution of the spectral function V: By Kabluchko et al. (2009), X is stationary if and only if

$$\mathbb{E}\left\{V(s_1)^{u_1} \cdot \ldots \cdot V(s_n)^{u_n}\right\} = \mathbb{E}\left\{V(s_1 + s)^{u_1} \cdot \ldots \cdot V(s_n + s)^{u_n}\right\} \tag{2.16}$$

for all $n \in \mathbb{N}$, $s, s_1, \ldots, s_n \in \mathbb{Z}$ and $u_1, \ldots, u_n \in [0, 1]$ such that $\sum_{i=1}^n u_i = 1$. For stationary simple max-stable processes, Kabluchko and Schlather (2010) give necessary and sufficient conditions for ergodicity and mixing in terms of the pairwise extremal coefficients $\theta(\{s_1, s_2\}) = \mathbb{E}(\max\{V(s_1), V(s_2)\})$, stating that X is mixing if and only

$$\lim_{r \to \infty} \theta(\{0, r\}) = 2, \tag{2.17}$$

and X is ergodic if and only if

$$\lim_{r \to \infty} \frac{1}{r} \sum_{k=1}^{r} \theta(\{0, k\}) = 2, \tag{2.18}$$

respectively.
Now, we transfer these results to a max-stable process X with ℓ^p norm based representation (2.3) giving necessary and sufficient conditions in terms of $W^{(p)}$. For the associated denoised process X^*, Equations (2.16)–(2.18) depend on the distribution $W^{(p)} = V$ only, while the structure of the process X is more difficult according to the relation $V(\cdot) = [\Gamma(1 - p^{-1})]^{-1} U^{(p)}(\cdot) W^{(p)}(\cdot)$ (cf. Proposition 2.3). The following result, however, shows that those conditions simplify to the conditions for the associated denoised process X^*.

Proposition 2.11. *Let $\{X(s),\, s \in \mathbb{Z}\}$ be a simple max-stable process with ℓ^p norm based representation* (2.3) *and let X^* be the denoised process associated to X. Then, the following holds:*

1. *X is stationary if and only if X^* is stationary.*

If X is stationary, we further have

2. *X is mixing if and only if X^* is mixing.*
3. *X is ergodic if and only if X^* is ergodic.*

Proof. 1. By Kabluchko et al. (2009) and Proposition 2.3, the process X is stationary if and only if (2.16) holds for $V(\cdot) = [\Gamma(1-p^{-1})]^{-1} U^{(p)}(\cdot) W^{(p)}(\cdot)$. The left-hand side of (2.16) equals

$$\mathbb{E}\left\{V(s_1)^{u_1} \cdot \ldots \cdot V(s_n)^{u_n}\right\} = \frac{1}{\Gamma(1-p^{-1})} \mathbb{E}\left\{\prod_{i=1}^n U^{(p)}(s_i)^{u_i} W^{(p)}(s_i)^{u_i}\right\}$$
$$= \frac{1}{\Gamma(1-p^{-1})} \mathbb{E}\left\{\prod_{i=1}^n U^{(p)}(s_i)^{u_i}\right\} \mathbb{E}\left\{\prod_{i=1}^n W^{(p)}(s_i)^{u_i}\right\}$$
$$= \frac{\prod_{i=1}^n \Gamma(1-u_i p^{-1})}{\Gamma(1-p^{-1})} \mathbb{E}\left\{\prod_{i=1}^n W^{(p)}(s_i)^{u_i}\right\},$$

where we used the fact that $U^{(p)}(s_i)^{u_i}$, $i = 1, \ldots, n$, are independent Φ_{p/u_i} random variables. Thus, X is stationary if and only if Equation (2.16) holds for $V = W^{(p)}$, i.e. if and only if X^* is stationary.

2. By Kabluchko and Schlather (2010), the process X is mixing if and only if Equation (2.17) holds where θ denotes the pairwise extremal coefficient of X. Proposition 2.10 yields the bounds

$$\lim_{r\to\infty} \theta^*(\{0,r\})) \leq \lim_{r\to\infty} \theta(\{0,r\}) \leq 2^{1/p} \lim_{r\to\infty} \theta^*(\{0,r\})^{1-p^{-1}} \leq 2.$$

Thus, $\lim_{r\to\infty} \theta(\{0,r\}) = 2$ if and only if $\lim_{r\to\infty} \theta^*(\{0,r\}) = 2$ which is equivalent to X^* being mixing.

3. The proof runs analogously to the proof of the second assertion. The process X is ergodic if and only if Equation (2.18) holds. From Proposition 2.10 and Jensen's inequality, we obtain

$$\lim_{r\to\infty} \frac{1}{r} \sum_{k=1}^r \theta^*(\{0,k\}) \leq \lim_{r\to\infty} \frac{1}{r} \sum_{k=1}^r \theta(\{0,k\})$$
$$\leq 2^{1/p} \lim_{r\to\infty} \frac{1}{r} \sum_{k=1}^r \theta^*(\{0,k\})^{1-p^{-1}} \leq 2^{1/p} \lim_{r\to\infty} \left[\frac{1}{r} \sum_{k=1}^r \theta^*(\{0,k\})\right]^{1-p^{-1}} \leq 2.$$

Consequently, we have that $\lim_{r\to\infty} r^{-1} \sum_{k=1}^r \theta(\{0,k\}) = 2$ holds true if and only if $\lim_{r\to\infty} r^{-1} \sum_{k=1}^r \theta^*(\{0,k\}) = 2$

□

Remark 2.12. The mixing properties of a stochastic process $\{X(s),\, s \in S\}$ are described more precisely by its mixing coefficients. For two subsets $S_1, S_2 \subset S$, the β-mixing coefficient $\beta(S_1, S_2)$ is defined by

$$\beta(S_1, S_2) = \sup\{|\mathcal{P}_{S_1 \cup S_2}(C) - \mathcal{P}_{S_1} \otimes \mathcal{P}_{S_2}(C)|,\, C \in \mathcal{C}_{S_1 \cup S_2}\},$$

where, for each $\tilde{S} \subset S$, the probability measure $\mathcal{P}_{\tilde{S}}$ denotes the distribution of the restricted process $\{X(s),\, s \in \tilde{S}\}$ on the space of non-negative functions on $\tilde{S}$ endowed with the Borel-σ algebra $\mathcal{C}_{\tilde{S}}$.

For the case of a max-stable process, Dombry and Éyi-Minko (2012) provide the upper bound

$$\beta(S_1, S_2) \leq 4 \sum_{s_1 \in S_1} \sum_{s_2 \in S_2} [2 - \theta(s_1, s_2)].$$

Applying Proposition 2.10 , we obtain

$$\beta(S_1, S_2) \leq 4 \sum_{s_1 \in S_1} \sum_{s_2 \in S_2} [2 - \theta(s_1, s_2)] \leq 4 \sum_{s_1 \in S_1} \sum_{s_2 \in S_2} [2 - \theta^*(s_1, s_2)],$$

i.e. the upper bound for a process with ℓ^p norm based representation (2.3) is lower than the bound for the associated denoised process.

As Proposition 2.11 states, a max-stable process with ℓ^p norm based representation (2.3) shares properties such as stationarity, ergodicity and mixing with the associated denoised process. In particular, the "noisy" analogues of well-studied max-stable processes might be used without changing any of these properties.

Acknowledgements

The author is grateful to Prof. Stilian Stoev and Dr. Kirstin Strokorb for pointing out some connections to other work as well as to Christopher Dörr for pointing out some typos. The helpful comments and suggestions by an anonymous referee are gratefully acknowledged.

3 Bayesian Inference for Multivariate Extreme Value Distributions

joint work with Clément Dombry and Sebastian Engelke

This chapter is based on the research article Dombry et al. (2017b) which has appeared in the *Electronic Journal of Statistics*. Besides some minor modifications, also some changes in the structure of the chapter have been made: all the proofs have been shifted from the appendix to appropriate places in the main body of the chapter.

Statistical modeling of multivariate and spatial extreme events has attracted broad attention in various areas of science. Max-stable distributions and processes are the natural class of models for this purpose, and many parametric families have been developed and successfully applied. Due to complicated likelihoods, the efficient statistical inference is still an active area of research, and usually composite likelihood methods based on bivariate densities only are used. Thibaud et al. (2016) use a Bayesian approach to fit a Brown–Resnick process to extreme temperatures. In this chapter, we extend this idea to a methodology that is applicable to general max-stable distributions and that uses full likelihoods. We further provide simple conditions for the asymptotic normality of the median of the posterior distribution and verify them for the commonly used models in multivariate and spatial extreme value statistics. A simulation study shows that this point estimator is considerably more efficient than the composite likelihood estimator in a frequentist framework. From a Bayesian perspective, our approach opens the way for new techniques such as Bayesian model comparison in multivariate and spatial extremes.

3.1 Introduction

Extremes and the impacts of rare events have been brought into public focus in the context of climate change or financial crises. The temporal or spatial concurrence of several such events has often shown to be most catastrophic. Arising naturally as limits of rescaled componentwise maxima of random vectors, max-stable distributions are frequently used to describe this joint behavior of extremes. The generalization to continuous domains gives rise to max-stable processes that have become popular models in spatial extreme value statistics (e.g., Davison and Gholamrezaee, 2012), and are applied in various fields such as meteorology (Buishand et al., 2008; Engelke et al., 2015; Einmahl et al., 2016) and hydrology (Asadi et al., 2015).

For a k-dimensional max-stable random vector $Z = (Z_1, \ldots, Z_k)$ with unit Fréchet margins, there exists an exponent function V describing the dependence between the components of Z such that $\mathbb{P}[Z \leq z] = \exp\{-V(z)\}$, $z \in (0, \infty)^k$. Many parametric models $\{F_\theta,\ \theta \in \Theta\}$ for the distribution function of Z have been proposed (cf. Schlather, 2002; Boldi and Davison, 2007; Kabluchko et al., 2009; Opitz, 2013), but likelihood-based inference remains challenging. The main reason is the lack of simple forms of the likelihood

$L(z;\theta)$ in these models, which, by Faá di Bruno's formula, is given by

$$L(z;\theta) = \sum_{\tau\in\mathcal{P}_k} L(z,\tau;\theta) = \sum_{\tau\in\mathcal{P}_k} \exp\{-V(z)\} \prod_{j=1}^{|\tau|} \{-\partial_{\tau_j} V(z)\}, \tag{3.1}$$

where $\mathcal{P}_k$ is the set of all partitions $\tau = \{\tau_1, \dots, \tau_{|\tau|}\}$ of $\{1, \dots, k\}$ and $\partial_{\tau_j} V(\cdot;\theta)$ denotes the partial derivative of the exponent function $V = V_\theta$ of F_θ with respect to the variables z_i, $i \in \tau_j$. The fact that the cardinality of $\mathcal{P}_k$ is the kth Bell number that grows super-exponentially in the dimension k inhibits the use of the maximum likelihood methods based on $L(z;\theta)$ in (3.1).

The most common way to avoid this problem is to maximize the composite pairwise likelihood that relies only on the information in bivariate sub-vectors of Z (Padoan et al., 2010). Apart from the fact that this likelihood is misspecified, there might also be considerable losses in efficiency by using the composition of bivariate likelihoods instead of the full likelihood $L(z;\theta)$. To reduce this efficiency loss, higher order composite likelihood has been considered (Genton et al., 2011; Huser and Davison, 2013; Castruccio et al., 2016).

In practice, to obtain observations from the random variable Z, the data, typically a multivariate time series, is split into disjoint blocks and a max-stable distribution is fitted to the componentwise maxima within each block. To increase the efficiency, not only the block maxima but additional information from the time series can be exploited. The componentwise occurrence times of the maxima within each block lead to a partition τ of $\{1, \dots, k\}$ with indices belonging to the same subset if and only if the maxima in this component occurred at the same time. The knowledge of this partition makes inference much easier, as a single summand $L(z,\tau;\theta)$ in the full likelihood $L(z;\theta)$ given in (3.1) corresponds to the likelihood contribution of the specific partition τ. This joint likelihood $L(z,\tau;\theta)$ was introduced in Stephenson and Tawn (2005) and is tractable for many extreme value models and, consequently, can be used for inference if occurrence times are available. In real data applications, however, the distribution of the block maxima is only approximated by a max-stable distribution and the distribution of the observed partitions of occurrence times are only approximations to the limiting distribution (as the block size tends to infinity) given by the likelihood $L(z,\tau;\theta)$. This approximation introduces a significant bias in the Stephenson–Tawn estimator and a bias correction has been proposed in Wadsworth (2015).

In many cases, only observations $z^{(1)}, \dots, z^{(N)} \in \mathbb{R}^k$ of the random max-stable vector Z are available, but there is no information about the corresponding partitions $\tau^{(1)}, \dots, \tau^{(N)}$. In this case, the Stephenson–Tawn likelihood cannot be used since the partition information is missing. In the context of conditional simulation of max-stable processes, Dombry et al. (2013) proposed a Gibbs sampler to obtain conditional samples of $\tau^{(l)}$ given the observation $z^{(l)}$, $l = 1, \dots, N$. Thibaud et al. (2016) use this approach to treat the missing partitions as latent variables in a Bayesian framework to estimate the parameters of a Brown–Resnick model (cf., Kabluchko et al., 2009) for extreme temperature. They obtain samples from the posterior distribution

$$L\left(\theta, \{\tau^{(l)}\}_{l=1}^N \mid \{z^{(l)}\}_{l=1}^N\right) \propto \pi_\theta(\theta) \prod_{l=1}^N L(z^{(l)}, \tau^{(l)};\theta), \tag{3.2}$$

via a Markov chain Monte Carlo algorithm, where π_θ is the prior distribution on Θ.

In this chapter, we extend the Bayesian approach to general max-stable distributions and provide various examples of parametric models F_θ where it can be applied. The first focus is to study the statistical efficiency of the point estimators obtained as the median of the

posterior distribution (3.2). This frequentist perspective allows to compare the efficiency of the Bayesian estimator that uses the full likelihoods to other frequentist estimators. A simulation study shows a substantial improvement of the estimation error when using full likelihoods rather than the commonly used pairwise likelihood estimator of Padoan et al. (2010).

From the Bayesian perspective, this approach opens up many new possibilities for Bayesian techniques in multivariate extreme value statistics. Besides readily available credible intervals, we discuss how Bayesian model comparison can be implemented. Thanks to the full, well-specified likelihoods in our approach, no adjustment of the posterior distribution as in the composite pairwise likelihood methods (Ribatet et al., 2012) is required.

Finally, we note that Huser et al. (2019) follow a complementary approach to ours where they apply an expectation-maximization algorithm to use full likelihoods $L(z;\theta)$ in the frequentist framework. The large sample asymptotic behavior of the frequentist and Bayesian estimators are the same (see Section 3.3 below) but the Monte-Carlo Markov Chain computation of the Bayesian estimator offers better convergence guarantees than the expectation-maximization computation of the maximum likelihood estimator. Moreover, alternatively to the perspective of max-stability and block maxima, inference can be based on threshold exceedances (Engelke et al., 2014; Wadsworth and Tawn, 2014; Thibaud and Opitz, 2015) and the corresponding multivariate Pareto distributions (Rootzén and Tajvidi, 2006; Rootzén et al., 2018).

The chapter is organized as follows. In Section 3.2 we provide some background on max-stable distributions and their likelihoods, and we present the general methodology for the Bayesian full-likelihood approach. Section 3.3 develops an asymptotic theory for the resulting estimator. We show in Section 3.4 that our method and the asymptotic theory are applicable for the popular models in multivariate and spatial extremes, including the Brown–Resnick and extremal-t processes. The simulation studies in Section 3.5 quantify the finite-sample efficiency gains of the Bayesian approach when used as a frequentist point estimator of the extremal dependence parameters. Interestingly, this advantage persists when the dependence is a nuisance parameter and one is only interested in estimating marginal parameters (Section 3.5.3), at least in the case of a well-specified model. The posterior distribution and genuinely Bayesian techniques are studied in Section 3.6, with a focus on Bayesian model comparison. Section 3.7 concludes the paper with a discussion on computational aspects.

3.2 Methodology

In Section 3.2.1 we review some facts on max-stable distributions and their likelihoods. We describe the general setup of our approach and review the Markov chain Monte Carlo algorithm from Thibaud et al. (2016) and the Gibbs sampler from Dombry et al. (2013) in Section 3.2.2.

3.2.1 Max-Stable Distributions, Partitions and Joint Likelihoods

Let us assume from now on that the max-stable vector Z belongs to a parametric family $\{F_\theta, \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^p$ is the parameter space, and that it admits a density f_θ. The exponent function of F_θ is $V_\theta(z) = -\log F_\theta(z)$. If there is no confusion we might omit the dependence on θ for simplicity.

Recall that if Z has standard Fréchet margins, it can be represented as the componentwise

maximum

$$Z_i = \max_{j\in\mathbb{N}} \psi_i^{(j)}, \quad i = 1, \dots, k, \tag{3.3}$$

where $\{\psi^{(j)} : j \in \mathbb{N}\}$ is a Poisson point process on $E = [0,\infty)^k \setminus \{0\}$ with intensity measure Λ such that $\Lambda(E \setminus [0, z]) = V(z)$. For more details and an exact simulation method of Z via this representation, we refer to Dombry et al. (2016a).
Analogously to the occurrence times in case of block maxima, the Poisson point process induces a random limit partition T of the index set $\{1, \dots, k\}$, where two indices $i_1 \neq i_2$ belong to the same subset if and only if $Z_{i_1} = \psi_{i_1}^{(j)}$ and $Z_{i_2} = \psi_{i_2}^{(j)}$ for the same $j \in \mathbb{N}$ (Dombry and Éyi-Minko, 2013). The joint likelihood of the max-stable vector Z and the limit partition T under the model F_θ satisfies

$$L(z, \tau; \theta) = \exp\{-V(z)\} \prod_{j=1}^{|\tau|} \{-\partial_{\tau_j} V(z)\}, \quad z \in (0,\infty)^k,\ \tau \in \mathcal{P}_k, \tag{3.4}$$

and it equals the likelihood introduced in Stephenson and Tawn (2005). This fact provides another interpretation of Equation (3.1), namely that the likelihood of Z is the integrated joint likelihood of Z and T.
In Dombry et al. (2017a) it has been shown that the existence of a density for the simple max-stable random vector Z with exponent measure Λ is equivalent to the existence of a density λ_I for the restrictions of Λ to the different faces $E_I \subset E$ defined by

$$E_I = \{z \in E;\ z_i > 0 \text{ for } i \in I \text{ and } z_i = 0 \text{ for } i \notin I\}, \quad \emptyset \neq I \subset \{1, \dots, k\},$$

that is,

$$\Lambda(A) = \sum_{\emptyset \neq I \subset \{1,\dots,k\}} \int_{\{z_I :\, z \in A \cap E_I\}} \lambda_I(z_I) \mu_I(\mathrm{d}z_I),$$

Thus, the Stephenson–Tawn likelihood $L(z, \tau; \theta)$ can be rewritten as

$$L(z, \tau; \theta) = \exp\{-V(z)\} \prod_{j=1}^{\ell} \omega(\tau_j, z), \tag{3.5}$$

where

$$\omega(\tau_j, z) = \sum_{\tau_j \subset I \subset \{1,\dots,k\}} \int_{(0, z_{\tau_j^c \cap I})} \lambda_I(z_{\tau_j}, u_j) \mathrm{d}u_j, \tag{3.6}$$

and $\tau_1, \dots, \tau_\ell$ denote the $\ell = |\tau|$ different blocks of the partition τ, and z_{τ_j} and $z_{\tau_j^c}$ are the restrictions of z to τ_j and $\tau_j^c = \{1, \dots, k\} \setminus \tau_j$, respectively.
Equation (3.5) provides a formula for the joint likelihood of max-stable distributions with unit Fréchet margins and its partition. From this we can deduce a formula for the joint likelihood of a general max-stable distribution that admits a density. More precisely, let $\bar{Z}$ be a k-dimensional max-stable random vector whose ith component, $i = 1, \dots, k$, has a generalized extreme value distribution with parameters $(\mu_i, \sigma_i, \xi_i) \in \mathbb{R} \times (0,\infty) \times \mathbb{R}$, i.e.

$$\mathbb{P}(\bar{Z}_i \leq z_i) = \exp\left\{ -\left(1 + \xi_i \frac{z_i - \mu_i}{\sigma_i}\right)_+^{-1/\xi_i} \right\}, \quad z_i \in \mathbb{R}.$$

Then, $U_i(\bar{Z}_i)$ has unit Fréchet distribution where U_i denotes the marginal transformation

$$U_i(x) = \left(1 + \xi_i \frac{x - \mu_i}{\sigma_i}\right)^{1/\xi_i}, \quad 1 + \xi_i \frac{x - \mu_i}{\sigma_i} > 0.$$

For a vector $z_I = (z_i)_{i\in I}$ with $I \subset \{1,\dots,k\}$, we define $U(z_I) = (U_i(z_i))_{i\in I}$, such that $\mathbb{P}(\bar{Z} \leq z) = \exp\{-V[U(z)]\}$, where V is the exponent measure of the normalized max-stable distribution $U(\bar{Z})$. Consequently, the joint density of the general max-stable vector $\bar{Z}$ and the limit partition T is

$$L(z,\tau;\theta) = \exp\{-V[U(z)]\} \cdot \left(\prod_{j=1}^{\ell} \omega(\tau_j, U(z))\right) \cdot \left(\prod_{i=1}^{k} \frac{1}{\sigma_i} U_i(z_i)^{1-\xi_i}\right), \tag{3.7}$$

for $z \in (0,\infty)^k$ such that $1 + \xi_i(z_i - \mu_i)/\sigma_i > 0,\ i = 1,\dots,k$ and $\tau = \{\tau_1,\dots,\tau_\ell\} \in \mathcal{P}_k$.

3.2.2 Bayesian Inference and Markov Chain Monte Carlo

Extreme value statistics is concerned with the estimation and uncertainty quantification of the parameter vector $\theta \in \Theta$. Here, θ might include both marginal and dependence parameters of the max-stable model. In a Bayesian setup we introduce a prior $\pi_\theta(\theta)$ on the parameter space Θ. Given independent data $z^{(1)},\dots,z^{(N)} \in \mathbb{R}^k$ from the max-stable distribution $Z \sim F_\theta$, we are interested in the posterior distribution of the parameter θ conditional on the data. As explained in Section 3.1, the complex structure of the full likelihood $L(\{z^{(l)}\};\theta) = \prod_{l=1}^{N} L(z^{(l)};\theta)$ prevents a direct assessment of the posterior distribution, which is proportional to the product of $L(\{z^{(l)}\};\theta)$ and the prior density $\pi_\theta(\theta)$. Instead, we introduce the corresponding limit partitions $T^{(1)},\dots,T^{(N)}$ as latent variables and sample from the joint distribution of $(\theta, T^{(1)},\dots,T^{(N)})$ conditional on the data $z^{(1)},\dots,z^{(N)}$, which is given in Equation (3.2).

It is customary to use Monte Carlo Markov Chain methods to sample from a target distribution which is known up to a multiplicative constant only. The aim is to construct a Markov chain which possesses the target distribution as stationary distribution and has good mixing properties. To this end, in each step of the Markov chain, the parameter vector θ and the partitions $T^{(1)},\dots,T^{(N)}$ are updated separately by the Metropolis–Hastings algorithm and a Gibbs sampler, respectively (cf., Thibaud et al., 2016).

For fixed partitions $T^{(l)} = \tau^{(l)}$, $l = 1,\dots,N$, and the current state θ for the parameter vector, we propose a new state θ^* according to a probability density $q(\theta,\cdot)$ which satisfies $q(\theta_1,\theta_2) > 0$ if and only if $q(\theta_2,\theta_1) > 0$ for $\theta_1,\theta_2 \in \Theta$. The proposal is accepted, that is, θ is updated to θ^*, with probability

$$a(\theta,\theta^*) = \min\left\{\frac{\prod_{l=1}^{N} L(z^{(l)},\tau^{(l)};\theta^*)\pi_\theta(\theta^*)q(\theta^*,\theta)}{\prod_{l=1}^{N} L(z^{(l)},\tau^{(l)};\theta)\pi_\theta(\theta)q(\theta,\theta^*)}, 1\right\} \tag{3.8}$$

where $L(z,\tau;\theta)$ is given by (3.7). In general, there are various ways of choosing an appropriate proposal density q. For instance, it might be advisable to update the vector θ component by component. It has to be ensured that any state θ_2 with positive posterior density can be reached from any other state θ_1 with positive posterior density in a finite number of steps, that is, that the Markov chain is irreducible. The convergence of the Markov chain to its stationary distribution (3.2) is then guaranteed. Note that the framework described above enables estimation of marginal and dependence parameters simultaneously. In particular, it allows for response surface methodology such as (log-)linear models for the marginal parameters.

For a fixed parameter vector $\theta \in \Theta$ we use the Gibbs sampler in Dombry et al. (2013) to update the current states of the partitions $\tau^{(1)},\dots,\tau^{(N)} \in \mathcal{P}_k$ conditional on the data $z^{(1)},\dots,z^{(N)}$. Thanks to independence, for each $l = 1,\dots,N$, we can update $\tau = \tau^{(l)}$

conditional on $z = z^{(l)}$ separately, where the conditional distribution is

$$L(\tau \mid z; \theta) = \frac{L(z, \tau; \theta)}{\sum_{\tau' \in \mathcal{P}_k} L(z, \tau'; \theta)} = \frac{1}{C_z} \prod_{j=1}^{\ell} \omega\{\tau_j, U(z)\}, \tag{3.9}$$

with C_z the normalization constant

$$C_z = \sum\nolimits_{\tau \in \mathcal{P}_k} \prod_{j=1}^{\ell} \omega\{\tau_j, U(z)\}.$$

For $i \in \{1, \ldots, k\}$, let τ_{-i} be the restriction of τ to the set $\{1, \ldots, k\} \setminus \{i\}$. As usual with Gibbs samplers, our goal is to simulate from

$$\mathbb{P}_\theta \left(T = \cdot \mid T_{-i} = \tau_{-i},\, Z = z\right), \tag{3.10}$$

where τ is the current state of the Markov chain and $\mathbb{P}_\theta$ denotes the probability under the assumption that Z follows the law F_θ. It is easy to see that the number of possible updates according to (3.10) is always less than k, so that a combinatorial explosion is avoided. Indeed, the index i can be reallocated to any of the components of τ_{-i} or to a new component with a single point: the number of possible updates $\tau^* \in \mathcal{P}_k$ such that $\tau^*_{-i} = \tau_{-i}$ equals ℓ if $\{i\}$ is a partitioning set of τ, and $\ell + 1$ otherwise.
The distribution (3.10) has nice properties. From (3.9), we obtain that

$$\mathbb{P}_\theta(T = \tau^* \mid T_{-i} = \tau_{-i}, Z = z) = \frac{L(z, \tau^*)}{\sum_{\tau' \in \mathcal{P}_k} L(z, \tau') \mathbf{1}_{\{\tau'_{-i} = \tau_{-i}\}}} \propto \frac{\prod_{j=1}^{|\tau^*|} w\{\tau^*_j, U(z)\}}{\prod_{j=1}^{|\tau|} w\{\tau_j, U(z)\}}. \tag{3.11}$$

for all $\tau^* \in \mathcal{P}_k$ with $\tau^*_{-i} = \tau_{-i}$. Since τ and τ^* share many components, all the factors in the right-hand side of (3.11) cancel out except at most four of them. This makes the Gibbs sampler particularly convenient.
We suggest a random scan implementation of the Gibbs sampler, meaning that one iteration of the Gibbs sampler selects randomly an element $i \in \{1, \ldots, k\}$ and then updates the current state τ according to the proposal distribution (3.10). For the sake of simplicity, we use the uniform random scan, i.e., i is selected according to the uniform distribution on $\{1, \ldots, k\}$.

3.3 Asymptotic Results

In the previous section, we presented a procedure that allows to sample from the posterior distribution of the parameter θ of a parametric model $\{f_\theta, \theta \in \Theta\}$ given a sample of N observations. In this section, we will discuss the asymptotic properties of the posterior distribution as the sample size N tends to ∞.
The asymptotic analysis of Bayes procedures usually relies on the Bernstein–von Mises theorem which allows for an asymptotic normal approximation of the posterior distribution of $\sqrt{N}(\theta - \theta_0)$, given the observations $z^{(1)}, \ldots, z^{(N)}$ from the parametric model f_{θ_0}. The theorem then implies the asymptotic normality and efficiency of Bayesian point estimators such as the posterior mean or posterior median with the same asymptotic variance as the maximum likelihood estimator.
A key assumption is that for every $\varepsilon > 0$ there exists a sequence of uniformly consistent (non-randomized) tests $\phi_N = \phi_N(z^{(1)}, \ldots, z^{(N)}) \in \{0, 1\}$ for testing the null hypothesis

$H_0 : \theta = \theta_0$ against $H_1 : \|\theta - \theta_0\|_\infty \geq \varepsilon$, where H_0 is not rejected if and only if $\phi_N = 0$. The uniformity means that

$$\mathbb{P}_{\theta_0}(\phi_N = 1) \to 0 \quad \text{and} \quad \sup_{\|\theta-\theta_0\|_\infty \geq \varepsilon} \mathbb{P}_\theta(\phi_N = 0) \to 0 \quad \text{as } N \to \infty. \tag{3.12}$$

where $\mathbb{P}_\theta$ denotes the probability measure induced by N independent copies of $Z \sim f_\theta$.

Theorem 3.1 (Bernstein-von Mises, Theorems 10.1 and 10.8 in van der Vaart (1998))**.** *Let the parametric model $\{f_\theta, \theta \in \Theta\}$ be differentiable in quadratic mean at θ_0 with non-singular Fisher information matrix I_{θ_0}, and assume that the mapping $\theta \mapsto \sqrt{f_\theta(z)}$ is differentiable at θ_0 for f_{θ_0}-almost every z. For every $\varepsilon > 0$, suppose there exists a sequence of uniformly consistent tests ϕ_N as in* (3.12)*. Suppose further that the prior distribution $\pi_{\text{prior}}(\mathrm{d}\theta)$ is absolutely continuous in a neighborhood of θ_0 with a continuous positive density at θ_0. Then, under the distribution f_{θ_0}, the posterior distribution satisfies*

$$\left\| \pi_{\text{post}}(\mathrm{d}\theta \mid z^{(1)}, \dots, z^{(N)}) - \mathcal{N}\left(\theta_0 + N^{-1/2}\Delta_{N,\theta_0}, N^{-1} I_{\theta_0}^{-1}\right) \right\|_{TV} \xrightarrow{d} 0 \quad \text{as } N \to \infty,$$

where $\Delta_{N,\theta_0} = N^{-1/2} \sum_{i=1}^N I_{\theta_0}^{-1} \partial_\theta \log f_{\theta_0}(z^{(i)})$ and $\|\cdot\|_{TV}$ is the total variation distance. As a consequence, if the prior distribution $\pi_{\text{prior}}(\mathrm{d}\theta)$ has a finite mean, the posterior median $\hat{\theta}_n^{Bayes}$ is asymptotically normal and efficient, that is, it satisfies

$$\sqrt{N}(\hat{\theta}_N^{Bayes} - \theta_0) \xrightarrow{d} \mathcal{N}(0, I_{\theta_0}^{-1}), \quad \text{as } N \to \infty.$$

In order to apply this theorem to max-stable distributions, two main assumptions are required: the differentiability in quadratic mean of the statistical model and the existence of uniformly consistent tests satisfying (3.12). Differentiability in quadratic mean is a technical condition that imposes a certain regularity on the likelihood f_{θ_0}. For the case of multivariate max-stable models this property has been considered in detail in Dombry et al. (2017a), where equivalent conditions on the exponent function and the spectral density are given.

We now discuss the existence of uniformly consistent tests and propose a criterion based on pairwise extremal coefficients. This criterion turns out to be simple and general enough since it applies for most of the standard models in extreme value theory. Indeed, in many cases, pairwise extremal coefficients can be explicitly computed and allow for identifying the parameter θ.

For a max-stable vector Z with unit Fréchet margins, the pairwise extremal coefficient $\eta_{i_1,i_2} \in [1, 2]$ between margins $1 \leq i_1 < i_2 \leq k$ is defined by

$$\mathbb{P}(Z_{i_1} \leq z, Z_{i_2} \leq z) = \exp\left\{-\frac{\eta_{i_1,i_2}}{z}\right\}, \quad z > 0.$$

It is the scale exponent of the unit Fréchet variable $Z_{i_1} \vee Z_{i_2}$ and hence satisfies

$$\eta_{i_1,i_2} = \left(\mathbb{E}\left[\frac{1}{Z_{i_1} \vee Z_{i_2}}\right]\right)^{-1}.$$

In the case that Z follows the distribution f_θ, we write $\eta_{i_1,i_2}(\theta)$ for the associated pairwise extremal coefficient.

Proposition 3.2. *Let $\theta_0 \in \Theta$ and $\varepsilon > 0$. Assume that*

$$\inf_{\|\theta-\theta_0\|_\infty \geq \varepsilon} \max_{1 \leq i_1 < i_2 \leq k} |\eta_{i_1,i_2}(\theta) - \eta_{i_1,i_2}(\theta_0)| > 0. \tag{3.13}$$

Then there exists a uniformly consistent sequence of tests ϕ_N satisfying (3.12)*.*

Remark 3.3. The identifiability of the model parameters $\theta \in \Theta$ through the pairwise extremal coefficients $\eta_{i_1,i_2}(\theta)$, $1 \leq i_1 < i_2 \leq k$ is a direct consequence of Equation (3.13) provided that the equation holds for every $\varepsilon > 0$.

Remark 3.4. If $\theta = (\theta_1, \ldots, \theta_p) \in \Theta$, and for any $1 \leq j \leq p$ there exists $1 \leq i_1 < i_2 \leq k$, such that $\eta_{i_1,i_2}(\theta)$ depends only on θ_j and it is strictly monotone with respect to this component, then Equation (3.13) is satisfied.

Proof of Proposition 3.2. For a random vector Z with distribution f_θ and $1 \leq i_1 < i_2 \leq k$, the random variable $1/(Z_{i_1} \vee Z_{i_2})$ follows an exponential distribution with parameter $\eta_{i_1,i_2}(\theta) \in [1,2]$ and variance $\eta^{-2}_{i_1,i_2}(\theta) \in [1/4, 1]$. Hence,

$$T^{-1}_{i_1,i_2} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{Z^{(i)}_{i_1} \vee Z^{(i)}_{i_2}}$$

is an unbiased estimator of $\eta^{-1}_{i_1,i_2}(\theta)$ with variance less than or equal to $1/N$. Chebychev's inequality entails

$$\mathbb{P}_\theta\big(\big|T^{-1}_{i_1,i_2} - \eta^{-1}_{i_1,i_2}(\theta)\big| > \delta\big) \leq \frac{1}{N\delta^2}, \quad \text{for all } \delta > 0. \tag{3.14}$$

Define the test ϕ_N by

$$\phi_N = \begin{cases} 0 & \text{if } \max_{1\leq i_1<i_2\leq k} \big|T^{-1}_{i_1,i_2} - \eta^{-1}_{i_1,i_2}(\theta_0)\big| \leq \delta, \\ 1 & \text{otherwise.} \end{cases}$$

We prove below that, for $\delta > 0$ small enough, the sequence ϕ_N satisfies (3.12). For $\theta = \theta_0$, the union bound together with Equation (3.14) yield

$$\mathbb{P}_{\theta_0}(\phi_N = 1) \leq \sum_{1\leq i_1<i_2\leq k} \mathbb{P}_{\theta_0}\big(\big|T^{-1}_{i_1,i_2} - \eta^{-1}_{i_1,i_2}(\theta_0)\big| > \delta\big) \leq \frac{k(k-1)}{2N\delta^2} \longrightarrow 0,$$

as $N \to \infty$. On the other hand, Equation (3.13) implies that there is some $\gamma > 0$ such that

$$\max_{1\leq i_1<i_2\leq k} \big|\eta^{-1}_{i_1,i_2}(\theta) - \eta^{-1}_{i_1,i_2}(\theta_0)\big| \geq \gamma, \quad \text{for all } \|\theta - \theta_0\|_\infty \geq \varepsilon.$$

Let $\theta \in \Theta$ be such that $\|\theta - \theta_0\|_\infty \geq \varepsilon$ and consider $1 \leq i_1 < i_2 \leq k$ realizing the maximum in the above equation. By the triangle inequality,

$$\big|T^{-1}_{i_1,i_2} - \eta^{-1}_{i_1,i_2}(\theta)\big| \geq \big|\eta^{-1}_{i_1,i_2}(\theta) - \eta^{-1}_{i_1,i_2}(\theta_0)\big| - \big|T^{-1}_{i_1,i_2} - \eta^{-1}_{i_1,i_2}(\theta_0)\big| \geq \gamma - \big|T^{-1}_{i_1,i_2} - \eta^{-1}_{i_1,i_2}(\theta_0)\big|,$$

so that, on the event $\{\phi_N = 0\} \subset \big\{\big|T^{-1}_{i_1,i_2} - \eta^{-1}_{i_1,i_2}(\theta_0)\big| \leq \delta\big\}$, we have

$$\big|T^{-1}_{i_1,i_2} - \eta^{-1}_{i_1,i_2}(\theta)\big| \geq \gamma - \delta.$$

Applying Equation (3.14) again, we deduce, for $\delta \in (0, \gamma)$,

$$\mathbb{P}_\theta(\phi_N = 0) \leq \mathbb{P}_\theta\big(\big|T^{-1}_{i_1,i_2} - \eta^{-1}_{i_1,i_2}(\theta)\big| \geq \gamma - \delta\big) \leq \frac{1}{N(\gamma-\delta)^2}.$$

Since the upper bound goes to 0 uniformly in θ with $\|\theta - \theta_0\|_\infty \geq \varepsilon$, as $N \to \infty$, this proves Equation (3.12). □

3.4 Examples

In the previous sections we discussed the implementation of a Markov chain Monte Carlo algorithm to obtain samples from the posterior distribution $L\left(\theta, \{\tau^{(l)}\}_{l=1}^N \mid \{z^{(l)}\}_{l=1}^N\right)$ and its asymptotic behavior as $N \to \infty$. The only model-specific quantity needed to run the algorithm are the weights $\omega(\tau_j, z)$ in (3.6). In this section, we provide explicit formulas for these weights for several classes of popular max-stable models and prove that the models satisfy the assumptions of the Bernstein–von Mises theorem; see Theorem 3.1. It follows that the posterior median $\hat{\theta}_N^{Bayes}$ is asymptotically normal and efficient for these models. For the calculation of the weights $\omega(\tau_j, z)$, we first note that all of the examples in this section admit densities as a simple consequence of Proposition 2.1 in Dombry et al. (2017a). We further note that, for the models considered in Subsections 3.4.1-3.4.4, we have $\lambda_I \equiv 0$ for all $I \subsetneq \{1, \dots, k\}$, i.e.

$$\Lambda(A) = \int_A \lambda(z)\, \mu(\mathrm{d}z), \quad A \subset [0,\infty)^k \setminus \{0\},$$

and, consequently, Equation (3.6) simplifies to

$$\omega(\tau_j, z) = \int_{(0, z_{\tau_j^c})} \lambda(z_{\tau_j}, u_j)\, \mathrm{d}u_j.$$

For the posterior median $\hat{\theta}_n^{Bayes}$, in the sequel, we will always assume that the prior distribution is absolutely continuous with strictly positive density in a neighborhood of θ_0, and that it has finite mean. Given the differentiability in quadratic mean of the model, it suffices to verify condition (3.13) in Proposition 3.2. This implies the existence of a uniformly consistent sequence of tests and, by Remark 3.3, the identifiability of the model. Theorem 3.1 then ensures asymptotic normality of the posterior median.
Henceforth, analogously to the notation $z_I = (z_i)_{i \in I}$ for a vector $z \in \mathbb{R}^k$ and an index set $\emptyset \neq I \subset \{1, \dots, k\}$, we write $A_{I,J} = (A_{ij})_{i \in I, j \in J}$ for a matrix $A = (A_{ij})_{1 \le i,j \le k}$ and index sets $\emptyset \neq I, J \subset \{1, \dots, k\}$.

3.4.1 The Logistic Model

One of the simplest multivariate extreme value distributions is the logistic model where

$$V(z) = \left(z_1^{-1/\theta} + \cdots + z_k^{-1/\theta}\right)^{\theta}, \quad \theta \in (0,1). \tag{3.15}$$

The logistic model is symmetric in its variables and interpolates between independence as $\theta \uparrow 1$ and complete dependence as $\theta \downarrow 0$.

Proposition 3.5. *Let $\tau = (\tau_1, \dots, \tau_\ell) \in \mathcal{P}_k$ and $z \in E$. The weights $\omega(\tau_j, z)$ in (3.6) for the logistic model with exponent measure (3.15) are*

$$\omega(\tau_j, z) = \theta^{-|\tau_j|+1} \frac{\Gamma(|\tau_j| - \theta)}{\Gamma(1-\theta)} \left\langle \Big(\sum_{i=1}^k z_i^{-1/\theta}\Big)^{\theta - |\tau_j|} \prod_{i \in \tau_j} z_i^{-1-1/\theta}. \right. \tag{3.16}$$

Proof. Taking partial derivative of the exponent function (3.15) we obtain

$$-\partial_{\tau_j} V_\theta(z) = \prod_{i=1}^{|\tau_j|-1} \left(\frac{i}{\theta} - 1\right) \left(\sum_{i=1}^k z_i^{-1/\theta}\right)^{\theta - |\tau_j|} \prod_{i \in \tau_j} z_i^{-1/\theta - 1}.$$

We note that

$$\frac{\Gamma(|\tau_j| - \theta)}{\Gamma(1-\theta)} = \prod_{i=1}^{|\tau_j|-1} (i - \theta).$$

Using this, Equation (3.5) becomes for the logistic model

$$L(\tau, z; \theta) = \exp\{-V(z)\} \prod_{j=1}^{\ell} \omega(\tau_j, z)$$

with

$$\omega(\tau_j, z) = \theta^{-1|\tau_j|+1} \frac{\Gamma(|\tau_j| - \theta)}{\Gamma(1-\theta)} \left(\sum_{i=1}^{k} z_i^{-1/\theta} \right)^{\theta - |\tau_j|} \prod_{i \in \tau_j} z_i^{-1-1/\theta}.$$

□

Remark 3.6. From (3.16), it can be seen that we can also write

$$L(\tau, z) = \exp(-V(z)) \left(\prod_{i=1}^{k} z_i^{-1-1/\theta} \right) \left(\sum_{i=1}^{k} z_i^{-1/\theta} \right)^{-k} \theta^{-k} \prod_{j=1}^{\ell} \tilde{\omega}(\tau_j, z)$$

with

$$\tilde{\omega}(\tau_j, z) = \theta \frac{\Gamma(|\tau_j| - \theta)}{\Gamma(1-\theta)} \left(\sum_{i=1}^{k} z_i^{-1/\theta} \right)^{\theta}.$$

This suggests to use the simplified weights $\tilde{\omega}$ for the Gibbs sampler.

Proposition 3.7. *For the logistic model with $\theta_0 \in (0,1)$, the posterior median $\hat{\theta}_N^{Bayes}$ is asymptotically normal and efficient as $N \to \infty$.*

Proof. From Proposition 4.1 in Dombry et al. (2017a) it follows that the model is differentiable in quadratic mean. For any $1 \leq i_1 < i_2 \leq k$, the pairwise extremal coefficient of the logistic model with parameter $\theta \in (0,1)$ is $\eta_{i_1,i_2}(\theta) = 2^{\theta}$, a strictly increasing function in θ. The assertion of the proposition follows by Remark 3.4. □

3.4.2 The Dirichlet Model

The Dirichlet model (Coles and Tawn, 1991) is defined by its spectral density h on the simplex $S^{k-1} = \{w \in [0,\infty)^k \ : \ w_1 + \cdots + w_k = 1\}$. For parameters $\alpha_1, \ldots, \alpha_k > 0$, it is given by

$$h(w) = \frac{1}{k} \frac{\Gamma(1 + \sum_{i=1}^{k} \alpha_i)}{(\sum_{i=1}^{k} \alpha_i w_i)^{k+1}} \prod_{i=1}^{k} \frac{\alpha_i}{\Gamma(\alpha_i)} \left(\frac{\alpha_i w_i}{\sum_{j=1}^{k} \alpha_j w_j} \right)^{\alpha_i - 1}, \quad w \in S^{k-1}, \tag{3.17}$$

and it has no mass on lower-dimensional faces of S^{k-1} (Coles and Tawn, 1991). Equivalently, the exponent function of the Dirichlet model is given by

$$V(z) = k\mathbb{E}\left[\max_{i=1,\ldots,k} \frac{W_i}{z_i} \right],$$

where W is a random vector with density $h(w)$.

Proposition 3.8. *Let $\tau = (\tau_1, \dots, \tau_\ell) \in \mathcal{P}_k$ and $z \in E$. The weights $\omega(\tau_j, z)$ in (3.6) for the Dirichlet model with spectral density (3.17) are*

$$\omega(\tau_j, z) = \prod_{i \in \tau_j} \frac{\alpha_i^{\alpha_i} z_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} \int_0^\infty e^{-\frac{1}{r}\sum_{i \in \tau_j} \alpha_i z_i} \left(\prod_{i \in \tau_j^c} F_{\alpha_i}(\alpha_i z_i / r) \right) r^{-2-\sum_{i=1}^k \alpha_i} \, \mathrm{d}r, \qquad (3.18)$$

where

$$F_\alpha(x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} \, \mathrm{d}t$$

is the distribution function of a Gamma variable with shape $\alpha > 0$.

Proposition 3.9. *Consider the Dirichlet model with $\theta_0 = (\alpha_1, \dots, \alpha_k) \in \Theta = (0, \infty)^k$. For $k \geq 3$ and almost every $\theta_0 \in \Theta$, the posterior median $\hat{\theta}_N^{Bayes}$ is asymptotically normal and efficient as $N \to \infty$.*

Both the proof of Proposition 3.8 and Proposition 3.9 rely on the following lemma.

Lemma 3.10. *Let $Y(\alpha_1), \dots, Y(\alpha_k)$ be independent random variables such that $Y(\alpha)$ has a Gamma distribution with shape parameter $\alpha > 0$ and scale 1.*

(i) Let $U_1 > U_2 > \dots$ be the points of a Poisson point process on $(0, \infty)$ with intensity $u^{-2} \, \mathrm{d}u$ and, independently of the U_i, let $\tilde{Y}^{(1)}, \tilde{Y}^{(2)}, \dots$ independent copies of the random vector $\tilde{Y} = (Y(\alpha_i)/\alpha_i)_{1 \leq i \leq k}$. Then the simple max-stable random vector $Z = \bigvee_{i \geq 1} U_i \tilde{Y}^{(i)}$ has angular density (3.17).

(ii) In the Dirichlet max-stable model with angular density (3.17), the pair extremal coefficient η_{i_1, i_2}, $1 \leq i_1 < i_2 \leq k$, is given by

$$\eta_{i_1, i_2} = \eta(\alpha_{i_1}, \alpha_{i_2}) = \mathbb{E}\left[\frac{Y(\alpha_{i_1})}{\alpha_{i_1}} \vee \frac{Y(\alpha_{i_2})}{\alpha_{i_2}} \right].$$

Furthermore, $\eta : (0, \infty)^2 \to [1, 2]$ is continuously differentiable and strictly decreasing in both components.

Proof. For the proof of the first part, we note that the intensity of the spectral measure is given by

$$\lambda(z) = \int_0^\infty f_{\tilde{Y}}(z/u) u^{-k-2} \, \mathrm{d}u, \quad z \in (0, \infty)^k,$$

where

$$f_{\tilde{Y}}(\tilde{y}) = \prod_{i=1}^k \frac{\alpha_i^{\alpha_i}}{\Gamma(\alpha_i)} \tilde{y}_i^{\alpha_i - 1} e^{-\alpha_i \tilde{y}_i}, \quad \tilde{y} \in (0, \infty)^k,$$

is the density of the random vector $\tilde{Y}$. A direct computation yields

$$\lambda(z) = \frac{\Gamma(1 + \sum_{i=1}^d \alpha_i)}{(\sum_{i=1}^d \alpha_i z_i)^{1 + \sum_{i=1}^d \alpha_i}} \prod_{i=1}^d \frac{\alpha_i^{\alpha_i} z_i^{\alpha_i - 1}}{\Gamma(\alpha_i)}.$$

We see that the restriction of λ to the simplex S^{k-1} is equal to h which proves the claim. The first statement of the second part, is a direct consequence of the first part since

$$\eta(\alpha_1, \alpha_2) = -\log \mathbb{P}(Z_1 \leq 1, Z_2 \leq 1) = \mathbb{E}\left[\frac{Y(\alpha_1)}{\alpha_1} \vee \frac{Y(\alpha_2)}{\alpha_2} \right].$$

The proof of the strict monotonicity relies on the notion of convex order, see Chapter 3.4 in Denuit et al. (2005). For two real-valued random variables X_1, X_2 we say that X_1 is lower than X_2 in convex order if $\mathbb{E}[\varphi(X_1)] \leq \mathbb{E}[\varphi(X_1)]$ for all convex functions $\varphi : \mathbb{R} \to \mathbb{R}$ such that the expectations exist. It is known that the family of random variables $(Y(\alpha)/\alpha)_{\alpha>0}$ is non-increasing in convex order (Ramos et al., 2000, Section 4.3), and, in this case, the Lorenz order is equivalent to the convex order (Denuit et al., 2005, Property 3.4.41). We show below that this implies that $\eta(\alpha_1, \alpha_2)$ is strictly decreasing in its arguments.
Let $\alpha_1' > \alpha_1 > 0$ and $\alpha_2 > 0$ and let us prove that $\eta(\alpha_1', \alpha_2) < \eta(\alpha_1, \alpha_2)$. For independent random variables $Y(\alpha_1)$, $Y(\alpha_1')$ and $Y(\alpha_2)$, we have

$$\eta(\alpha_1', \alpha_2) = \mathbb{E}\left[\frac{Y(\alpha_1')}{\alpha_1'} \vee \frac{Y(\alpha_2)}{\alpha_2}\right] \quad \text{and} \quad \eta(\alpha_1, \alpha_2) = \mathbb{E}\left[\frac{Y(\alpha_1)}{\alpha_1} \vee \frac{Y(\alpha_2)}{\alpha_2}\right].$$

Using that $Y(\alpha_1')/\alpha_1'$ is lower than $Y(\alpha_1)/\alpha_1$ in convex order, we obtain

$$\mathbb{E}\left[\frac{Y(\alpha_1')}{\alpha_1'} \vee \frac{y_2}{\alpha_2}\right] \leq \mathbb{E}\left[\frac{Y(\alpha_1)}{\alpha_1} \vee \frac{y_2}{\alpha_2}\right] \quad \text{for all } y_2 > 0, \tag{3.19}$$

because the map $u \mapsto u \vee (y_2/\alpha_2)$ is convex. Replacing y_2 by $Y(\alpha_2)$ and integrating, we get $\eta(\alpha_1', \alpha_2) \leq \eta(\alpha_1, \alpha_2)$. The equality $\eta(\alpha_1', \alpha_2) = \eta(\alpha_1, \alpha_2)$ would imply that for almost every $y_2 > 0$ the equality holds in (3.19) which is true if and only if $Y(\alpha_1)/\alpha_1$ and $Y(\alpha_1')/\alpha_1'$ have the same distribution. Since this is not the case, $\eta(\alpha_1', \alpha_2) < \eta(\alpha_1, \alpha_2)$ and η is strictly decreasing in α_1. By symmetry, η is also strictly decreasing in α_2.
Finally, the fact that $(\alpha_1, \alpha_2) \mapsto \eta(\alpha_1, \alpha_2)$ is continuously differentiable follows from the integral representation

$$\eta(\alpha_1, \alpha_2) = \int_0^\infty \int_0^\infty \frac{y_1}{\alpha_1} \vee \frac{y_2}{\alpha_2} \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1 - 1} y_2^{\alpha_2 - 1} e^{-y_1} e^{-y_2} \, \mathrm{d}y_1 \, \mathrm{d}y_2, \tag{3.20}$$

for $\alpha_1, \alpha_2 > 0$, and standard theorems for integrals depending on a parameter. □

Proof of Proposition 3.8. From the construction given in the first part of Lemma 3.10, we obtain

$$\begin{aligned}
\lambda(z) &= \int_0^\infty \left(\prod_{i=1}^k \frac{\alpha_i^{\alpha_i}}{\Gamma(\alpha_i)} \left(\frac{z_i}{r}\right)^{\alpha_i - 1} e^{-(\alpha_i z_i / r)} \right) r^{-2-k} \, \mathrm{d}r \\
&= \prod_{i=1}^k \frac{\alpha_i^{\alpha_i} z_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} \int_0^\infty e^{-\frac{1}{r}\sum_{i=1}^k \alpha_i z_i} r^{-2-\sum_{i=1}^k \alpha_i} \, \mathrm{d}r
\end{aligned}$$

and, consequently,

$$\begin{aligned}
&\int_{u_j < z_{\tau_j^c}} \lambda(z_{\tau_j}, u_j) \, \mathrm{d}u_j \\
&= \prod_{i \in \tau_j} \frac{\alpha_i^{\alpha_i} z_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} \int_0^\infty \int_{u_j < z_{\tau_j^c}} \prod_{i \in \tau_j^c} \left(\frac{\alpha_i^{\alpha_i} z_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} e^{-(\alpha_i z_i / r)} \right) \mathrm{d}u_j \cdot e^{-\frac{1}{r}\sum_{i \in \tau_j} \alpha_i z_i} r^{-2-\sum_{i=1}^k \alpha_i} \, \mathrm{d}r \\
&= \prod_{i \in \tau_j} \frac{\alpha_i^{\alpha_i} z_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} \int_0^\infty e^{-\frac{1}{r}\sum_{i \in \tau_j} \alpha_i z_i} \left(\prod_{i \in \tau_j^c} F_{\alpha_i}(\alpha_i z_i / r) \right) r^{-2-\sum_{i \in \tau_j} \alpha_i} \, \mathrm{d}r,
\end{aligned}$$

where

$$F_\alpha(x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha - 1} e^{-t} \, \mathrm{d}t,$$

is the distribution function of a Gamma variable with shape $\alpha > 0$. □

Proof of Proposition 3.9. Proposition 4.2 in Dombry et al. (2017a) implies that the model is differentiable in quadratic mean. In order to verify Equation (3.13) for the Dirichlet model, we consider the mapping

$$\Psi:\ (0,\infty)^k \to [1,2]^k,\ \theta=(\alpha_1,\dots,\alpha_k)\mapsto(\eta_{1,2},\eta_{2,3},\eta_{1,3},\eta_{1,4},\dots,\eta_{1,k}).$$

We first show that Ψ is injective. To this end, let $\theta^{(1)}\neq\theta^{(2)}\in\Theta$ where $\psi_i=(\alpha_1^{(i)},\dots,\alpha_k^{(i)})$, $i=1,2$. We distinguish between two cases.
First, we assume that $\theta^{(1)}$ and $\theta^{(2)}$ share at least one common component. Then, there is a pair $(i,j)\in\{(1,2),(2,3),(1,3),(1,4),\dots,(1,k)\}$ such that $(\alpha_i^{(1)},\alpha_j^{(1)})$ and $(\alpha_i^{(2)},\alpha_j^{(2)})$ differ in exactly one component. As $\eta_{i,j}=\eta(\alpha_i,\alpha_j)$ is strictly decreasing both in α_i and α_j, by Lemma 3.10, we have that $\eta(\alpha_i^{(1)},\alpha_j^{(1)})\neq\eta(\alpha_i^{(2)},\alpha_j^{(2)})$.
Secondly, we consider the case that $\theta^{(1)}$ and $\theta^{(2)}$ do not share any common component. Then, there is a pair $(i,j)\in\{(1,2),(2,3),(1,3)\}$ such that both components of the vector $(\alpha_i^{(1)}-\alpha_i^{(2)},\alpha_j^{(1)}-\alpha_j^{(2)})$ have the same sign and, again, by the strict monotonicity of $\eta(\alpha_i,\alpha_j)$, it follows that $\eta(\alpha_i^{(1)},\alpha_j^{(1)})\neq\eta(\alpha_i^{(2)},\alpha_j^{(2)})$.
Hence, in both cases, $\Psi(\theta^{(1)})\neq\Psi(\theta^{(2)})$, that is, Ψ is injective and there exists a unique inverse function $\Psi^{-1}:\Psi((0,\infty)^k)\to(0,\infty)^k$.
Consider the set

$$\Theta'=\left\{(\alpha_1,\dots,\alpha_k)\in(0,\infty)^k:\ \partial_{\alpha_i}\eta(\alpha_i,\alpha_j)<0,\ \partial_{\alpha_j}\eta(\alpha_i,\alpha_j)<0\ \forall 1\le i<j\le k\right\}.$$

Note that since η is continuously differentiable and strictly decreasing in its argument, $\Theta\setminus\Theta'$ has Lebesgue measure 0. For all $\theta_0=(\alpha_1,\dots,\alpha_k)\in\Theta'$, the Jacobian $D\Psi$ satisfies

$$\begin{aligned}\det\{D\Psi(\theta_0)\}=&\{\partial_{\alpha_1}\eta(\alpha_1,\alpha_2)\cdot\partial_{\alpha_2}\eta(\alpha_2,\alpha_3)\cdot\partial_{\alpha_3}\eta(\alpha_1,\alpha_3)\\&+\partial_{\alpha_2}\eta(\alpha_1,\alpha_2)\cdot\partial_{\alpha_3}\eta(\alpha_2,\alpha_3)\cdot\partial_{\alpha_1}\eta(\alpha_1,\alpha_3)\}\cdot\prod_{j=4}^{k}\partial_{\alpha_j}\eta(\alpha_1,\alpha_j)\neq 0.\end{aligned}$$

The inverse function theorem then implies that Ψ^{-1} is continuously differentiable at $\Psi(\theta_0)$, that is, for every $\varepsilon>0$, there exists $\delta>0$ such that $\|\Psi(\theta_0)-\Psi_\infty(\theta)\|_\infty<\delta$ implies $\|\theta_0-\theta\|_\infty<\varepsilon$. In particular, we obtain

$$\inf_{\|\theta_0-\theta\|_\infty>\varepsilon}\|\Psi(\theta_0)-\Psi(\theta)\|_\infty\ge\delta,$$

that is, Equation (3.13), and the asymptotic normality and efficiency of the posterior median for $\theta_0\in\Theta'$ follow from Proposition 3.2. Finally, we note that each extremal coefficient η is continuously differentiable and strictly decreasing with respect to both components by Lemma 3.10. Thus, $\partial_{\alpha_1}\eta(\alpha_1,\alpha_2)<0$ and $\partial_{\alpha_2}\eta(\alpha_1,\alpha_2)<0$ for almost every $\theta\in\Theta$. □

Remark 3.11. We believe that the result for the posterior median holds true even for every $\theta_0\in\Theta$. In the proof of 3.9, we need the partial derivatives of $(\alpha_1,\alpha_2)\mapsto\eta(\alpha_1,\alpha_2)$ to be negative, but this can only be concluded almost everywhere.

3.4.3 The Extremal-t Model and the Schlather Process

The extremal-t model (Nikoloulopoulos et al., 2009; Opitz, 2013) is given by an exponent measure of the form

$$V(z)=c_\nu\mathbb{E}\left[\max_{i=1,\dots,k}\frac{\max\{0,W_i\}^\nu}{z_i}\right],\tag{3.21}$$

where $(W_1, \dots, W_k)^\top$ is a standardized Gaussian vector with correlation matrix Σ, the constant c_ν is given by $c_\nu = \sqrt{\pi} 2^{-(\nu-2)/2} \Gamma\{(\nu+1)/2\}^{-1}$ and $\nu > 0$.

Proposition 3.12 (Thibaud and Opitz, 2015)**.** *Let $\tau = (\tau_1, \dots, \tau_\ell) \in \mathcal{P}_k$ and $z \in E$. The weights $\omega(\tau_j, z)$ in* (3.6) *for the extremal-t model with exponent function* (3.21) *are*

$$\begin{aligned}\omega(\tau_j, z) = T_{|\tau_j|+\nu}\big(z_{\tau_j^c}^{1/\nu} - \tilde{\mu}, \tilde{\Sigma}\big) \cdot \nu^{1-|\tau_j|} \cdot \pi^{(1-|\tau_j|)/2} \cdot \det(\Sigma_{\tau_j,\tau_j})^{-1/2} \\ \cdot \frac{\Gamma\{(\nu+|\tau_j|)/2\}}{\Gamma\{(\nu+1)/2\}} \cdot \prod_{i \in \tau_j} |z_i|^{1/\nu - 1} \cdot \Big\{ \big(z_{\tau_j}^{1/\nu}\big)^\top \Sigma_{\tau_j,\tau_j}^{-1} z_{\tau_j}^{1/\nu} \Big\}^{-(\nu+|\tau_j|)/2}\end{aligned} \tag{3.22}$$

where $\tilde{\mu} = \Sigma_{\tau_j^c,\tau_j} \Sigma_{\tau_j}^{-1} z_{\tau^j}^{1/\nu}$,

$$\tilde{\Sigma} = (|\tau_j| + \nu)^{-1} \big(z_{\tau^j}^{1/\nu}\big)^\top \Sigma_{\tau_j}^{-1} z_{\tau_j} \big(\Sigma_{\tau_j^c} - \Sigma_{\tau_j^c,\tau_j} \Sigma_{\tau_j,\tau_j}^{-1} \Sigma_{\tau_j,\tau_j^c}\big)$$

and $T_k(\cdot; \Sigma)$ denotes a multivariate Student distribution function with k degrees of freedom and scale matrix Σ.

Proposition 3.13. *Consider the extremal-t model with $\theta_0 = (\Sigma, \nu)$ where Σ is a positive definite correlation matrix and $\nu > 0$. Then, for fixed $\nu > 0$ the posterior median $\hat{\theta}_N^{Bayes}$ is asymptotically normal and efficient as $N \to \infty$.*

Proof. By Proposition 4.3 in Dombry et al. (2017a), the model is differentiable in quadratic mean (even if $\nu > 0$ is not fixed). For any $1 \le i_1 < i_2 \le k$, and fixed $\nu > 0$, the pairwise extremal coefficient of the extremal-t model with parameter matrix $\Sigma = \{\rho_{ij}\}_{1 \le i,j \le k}$ is

$$\eta_{i_1,i_2}(\Sigma) = 2T_{\nu+1}\left(\sqrt{(\nu+1)\frac{1-\rho_{i_1 i_2}}{1+\rho_{i_1 i_2}}}\right), \tag{3.23}$$

where $T_{\nu+1}$ denotes the distribution function of a t-distribution with $\nu + 1$ degrees of freedom. Therefore, $\eta_{i_1,i_2}(\Sigma)$ as a function of $\rho_{i_1 i_2} \in [-1, 1]$ is strictly decreasing and the claim follows by Remark 3.4 together with Proposition 3.2. □

Remark 3.14. If ν is not fixed, then the parameter $\theta = (\Sigma, \nu)$ cannot be identified from the pairwise extremal coefficients and Equation (3.13) is not satisfied. The identifiability can still be shown by considering the behavior of the bivariate angular measure at the origin (Engelke and Ivanovs, 2017, Section A.3.3).

A popular model in spatial extremes is the extremal-t process (Opitz, 2013), a max-stable process $\{Z(x),\ x \in \mathbb{R}^d\}$ whose finite-dimensional distributions $(Z(x_1), \dots, Z(x_k))^\top$, $x_1, \dots, x_k \in \mathbb{R}^d$ have an exponent function of the form (3.21) where the Gaussian vector is replaced by a standardized stationary Gaussian process $\{W(x),\ x \in \mathbb{R}^d\}$ evaluated at $x_1, \dots, x_k$. The correlation matrix Σ then has the form

$$\Sigma = \{\rho(x_i - x_j)\}_{1 \le i,j \le k},$$

where $\rho : \mathbb{R}^d \to [-1, 1]$ is the correlation function of the Gaussian process W. The special case $\nu = 1$ corresponds to the extremal Gaussian process (Schlather, 2002), also called Schlather process.

Corollary 3.15. *Let Z be the Schlather process on $\mathbb{R}^d$ with correlation function ρ coming from the parametric family*

$$\rho(h) = \exp\left(-\frac{\|h\|_2^\alpha}{s}\right), \quad (s,\alpha) \in \Theta = (0,\infty) \times (0,2].$$

Suppose that Z is observed at pairwise distinct locations $t_1, \ldots, t_k \in \mathbb{R}^d$ such that not all pairs of locations have the same Euclidean distance. Then, the posterior median of $\theta = (s,\alpha)$ is asymptotically normal.

Proof. Analogously to the proof of Corollary 4.4 in Dombry et al. (2017a), it can be shown that the model is differentiable in quadratic mean. Suppose that $\|t_1 - t_2\|_2 \neq \|t_2 - t_3\|_2$ and observe that the mapping

$$\Psi : \Theta \to \Psi(\Theta), \quad \theta = (s,\alpha) \mapsto \{\rho_{ij}\}_{1\leq i,j\leq k} = \left\{\exp\left(-\frac{\|t_i - t_j\|_2^\alpha}{s}\right)\right\}_{1\leq i,j\leq k}$$

is continuously differentiable. Since

$$\alpha = \frac{\log\{\log \rho_{12}\} - \log\{\log \rho_{23}\}}{\log \|t_1 - t_2\|_2 - \log \|t_2 - t_3\|_2}, \qquad s = -\frac{\|t_1 - t_2\|_2^\alpha}{\log \rho_{12}},$$

the same holds true for the inverse mapping Ψ^{-1}.
Further, from the continuity of Ψ^{-1} at $\Psi(\theta_0)$ at any $\theta_0 \in \Theta$, we obtain that for every $\varepsilon > 0$ there is some $\delta > 0$ such that for all $\Psi(\theta) \in \Psi(\Theta)$ with $\|\Psi(\theta_0) - \Psi(\theta)\|_\infty < \delta$ we have $\|\theta_0 - \theta\|_\infty < \varepsilon$. Consequently,

$$\inf_{\|\theta_0 - \theta\|_\infty > \varepsilon} \|\Psi(\theta_0) - \Psi(\theta)\|_\infty \geq \delta.$$

From the proof of Proposition 3.13 and with the same notation as in (3.23), we obtain that $\|\Psi(\theta_0) - \Psi(\theta)\|_\infty \geq \delta$ implies $\max_{1\leq i_1<i_2\leq k} |\eta_{i_1,i_2}\{\Psi(\theta_0)\} - \eta_{i_1,i_2}\{\Psi(\theta)\}| > \delta'$ for some $\delta' > 0$, that is, Equation (3.13) holds. The assertion follows then from Proposition 3.2. □

3.4.4 The Hüsler–Reiss Model and the Brown–Resnick Model

The Hüsler–Reiss distribution (cf., Hüsler and Reiss, 1989; Kabluchko et al., 2009) can be characterized by its exponent function

$$V(z) = \mathbb{E}\left[\max_{i=1,\ldots,k} \frac{\exp\{W_i - \Sigma_{ii}/2\}}{z_i}\right], \tag{3.24}$$

where $W = (W_1, \ldots, W_k)^\top$ is a Gaussian vector with expectation 0 and covariance matrix Σ. It can be shown that the exponent function can be parameterized by the matrix

$$\Lambda = \{\lambda_{i,j}^2\}_{1\leq i,j\leq k} = \left\{\frac{1}{4}\mathbb{E}(W_i - W_j)^2\right\}_{1\leq i,j\leq k}$$

as we have the equality

$$V(z) = \sum_{p=1}^{k} z_p^{-1} \Phi_{k-1}\left(2\lambda_{p,-p}^2 + \log(z_{-p}/z_p); \Sigma^{(p)}\right), \quad z \in (0,\infty)^k, \tag{3.25}$$

(cf. Nikoloulopoulos et al., 2009), where for each $p = 1, \ldots, k$, the matrix $\Sigma^{(p)}$ has (i,j)th entry $2(\lambda_{p,i}^2 + \lambda_{p,j}^2 - \lambda_{i,j}^2)$, $i, j \neq p$ and $\Phi_{k-1}(\cdot, \Sigma^{(p)})$ denotes the $(k-1)$-dimensional normal distribution function with covariance matrix $\Sigma^{(p)}$.
Note that the positive definiteness of the matrices $\Sigma^{(p)}$, $p = 1, \ldots, k$, follows from the fact that Λ is conditionally negative definite, i.e.

$$\sum_{1 \leq i,j \leq k} a_i a_j \lambda_{i,j}^2 \leq 0 \tag{3.26}$$

for all $a_1, \ldots, a_k \in \mathbb{R}$ summing up to 0 (cf. Berg et al., 1984, Lemma 3.2.1). In the following, we will assume that Λ is even strictly positive definite, i.e. equality in (3.26) holds true if and only if $a_1 = \ldots = a_k = 0$. Then, all the matrices $\Sigma_{I,I}^{(p)}$ with $p \in \{1, \ldots, k\}$ and $\emptyset \neq I \subset \{1, \ldots, k\}$ are strictly positive definite.

Proposition 3.16 (see also Wadsworth and Tawn, 2014; Asadi et al., 2015, for instance)**.** *Let $\tau = (\tau_1, \ldots, \tau_\ell) \in \mathcal{P}_k$ and $z \in E$. For $j \in \{1, \ldots, k\}$, choose any $p \in \tau_j$ and let $\tilde{\tau} = \tau_j \setminus \{p\}$, $\tilde{\tau}^c = \{1, \ldots, k\} \setminus \tau_j$. The weights $\omega(\tau_j, z)$ in* (3.6) *for the Hüsler–Reiss distribution with exponent function* (3.25) *are*

$$\omega(\tau_j, z) = \frac{1}{z_p^2 \prod_{i \in \tilde{\tau}} z_i} \varphi_{|\tilde{\tau}|} \left\{ z_{\tilde{\tau}}^*; \Sigma_{\tilde{\tau},\tilde{\tau}}^{(p)} \right\} \Phi_{|\tilde{\tau}^c|} \left\{ z_{\tilde{\tau}^c}^* - \Sigma_{\tilde{\tau}^c,\tilde{\tau}}^{(p)} (\Sigma_{\tilde{\tau},\tilde{\tau}}^{(p)})^{-1} z_{\tilde{\tau}}^*; \hat{\Sigma}^{(p)} \right\}, \tag{3.27}$$

where

$$z^* = \left\{ \log\left(\frac{z_i}{z_p}\right) + \frac{\Gamma(x_i, x_p)}{2} \right\}_{i=1,\ldots,k} \text{ and } \hat{\Sigma}^{(p)} = \Sigma_{\tilde{\tau}^c,\tilde{\tau}^c}^{(p)} - \Sigma_{\tilde{\tau}^c,\tilde{\tau}}^{(p)} (\Sigma_{\tilde{\tau},\tilde{\tau}}^{(p)})^{-1} \Sigma_{\tilde{\tau},\tilde{\tau}^c}^{(p)}.$$

Here $\Phi_k(\cdot; \Sigma)$ denotes a k-dimensional Gaussian distribution function with mean 0 and covariance matrix Σ, and $\varphi_k(\cdot; \Sigma)$ its density. The functions Φ_0 and φ_0 are set to be constant 1.

Proposition 3.17. *Consider the Hüsler–Reiss model with $\theta_0 = \Lambda$ being a strictly conditionally negative definite matrix. Then, the posterior median $\hat{\theta}_N^{Bayes}$ is asymptotically normal and efficient as $N \to \infty$.*

Proof. From Proposition 4.5 in Dombry et al. (2017a), it follows that the model is differentiable in quadratic mean. For any $1 \leq i_1 < i_2 \leq k$, the pairwise extremal coefficient of the Hüsler–Reiss model with parameter matrix $\Lambda = \{\lambda_{i,j}^2\}_{1 \leq i,j \leq k}$ is

$$\eta_{i_1,i_2}(\Lambda) = 2\Phi_1\left\{\sqrt{\lambda_{i_1,i_2}^2}\right\},$$

which is a strictly increasing function in $\lambda_{i_1,i_2}^2 > 0$, and the claim follows by Remark 3.4 together with Proposition 3.2. □

Hüsler–Reiss distributions also appear as the finite dimensional distributions of the max-stable Brown–Resnick process, a popular class in spatial extreme value statistics. Here, the Gaussian vectors $(W_1, \ldots, W_k)^\top$ in (3.24) are the finite-dimensional distributions of a centered Gaussian process $\{W(x), x \in \mathbb{R}^d\}$ which is parameterized via a conditionally negative definite variogram $\gamma : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty)$, $\gamma(x_1, x_2) = \mathbb{E}(W(x_1) - W(x_2))^2$. If W has stationary increments, we have that $\gamma(x_1, x_2) = \gamma(x_1 - x_2, 0) =: \gamma(x_1 - x_2)$ and the resulting Brown–Resnick process is stationary (Brown and Resnick, 1977; Kabluchko et al., 2009). The most common parametric class of variograms belonging to Gaussian processes with stationary increments is the class of fractional variograms, which we consider in the following corollary.

Corollary 3.18. *Consider a Brown–Resnick process on $\mathbb{R}^d$ with variogram coming from the parametric family*

$$\gamma(h) = \|h\|_2^\alpha / s, \quad (s, \alpha) \in \Theta = (0, \infty) \times (0, 2).$$

Suppose that the process is observed on a finite set of locations $t_1, \dots, t_m \in \mathbb{R}^d$ such that the pairwise Euclidean distances are not all equal. Then the posterior median of $\theta = (s, \alpha)$ is asymptotically normal.

Proof. Analogously to Corollary 4.6 in Dombry et al. (2017a), the model can be shown to be differentiable in quadratic mean. Suppose that $\|t_1 - t_2\|_2 \neq \|t_2 - t_3\|_2$. As the mapping $\Psi : \Theta \to \Psi(\Theta)$, $\theta = (\lambda, \alpha) \mapsto \{\lambda_{ij}^2\}_{1 \leq i,j \leq k} = \left\{ \frac{\|t_i - t_j\|_2^\alpha}{4s} \right\}_{1 \leq i,j \leq k}$ is continuously differentiable and

$$\alpha = \frac{\log \gamma_{12} - \log \gamma_{23}}{\log \|t_1 - t_2\|_2 - \log \|t_2 - t_3\|_2}, \qquad s = \frac{\|t_i - t_j\|_2^\alpha}{4\lambda_{ij}^2},$$

the inverse mapping Ψ^{-1} is continuously differentiable, as well. The same arguments as in the proof of Corollary 3.15 together with the proof of Proposition 3.17 yield the assertion. □

3.5 Simulation Study

Let $z^{(l)} = (z_1^{(l)}, \dots, z_k^{(l)})$, $l = 1, \dots, N$, be N realizations of a k-dimensional max-stable vector Z whose distribution belongs to some parametric family $\{F_\theta,\ \theta \in \Theta\}$. As described in Section 3.2, including the partition $\tau^{(l)}$ associated to a realization $z^{(l)}$ in a Bayesian framework allows to obtain samples from the posterior distribution $L(\theta \mid z^{(1)}, \dots, z^{(N)})$ of θ given the data. This procedure uses the full dependence information of the multivariate distribution Z. This is in contrast to frequentist maximum likelihood estimation for the max-stable vector Z, where even in moderate dimensions the likelihoods are to complicated for practical applications. Instead, at the price of likelihood misspecification, it is common practice to use only pairwise likelihoods which are assumed to be mutually independent. The maximum pairwise likelihood estimator (Padoan et al., 2010) is then

$$\hat{\theta}_{\mathrm{PL}} = \operatorname*{argmax}_{\theta \in \Theta} \sum_{l=1}^{N} \sum_{1 \leq i < j \leq k} \log f_{\theta;i,j}(z_i^{(l)}, z_j^{(l)}), \tag{3.28}$$

where $f_{\theta;i,j}$ denotes the joint density of the ith and jth component of Z under the model F_θ. Using only bivariate information on the dependence results in efficiency losses.
In this section, we analyze the performance of our proposed Bayesian estimator and compare it to $\hat{\theta}_{\mathrm{PL}}$ and other existing methods. Since the latter are all frequentist approaches, for a Markov chain whose stationary distribution is the posterior, we obtain a point estimator $\hat{\theta}_{\mathrm{Bayes}}$ of θ as the posterior median, i.e.,

$$\hat{\theta}_{\mathrm{Bayes}} = \mathrm{median}\big\{L(\theta | z^{(1)}, \dots, z^{(N)})\big\}.$$

As the parametric model we choose the logistic distribution introduced in Subsection 3.4.1 with parameter space $\Theta = (0, 1)$ and uniform prior. This choice covers a range of situations from strong to very weak dependence. Other choices of parametric models will result in different efficiency gains but the general observations in the next sections should remain the same.

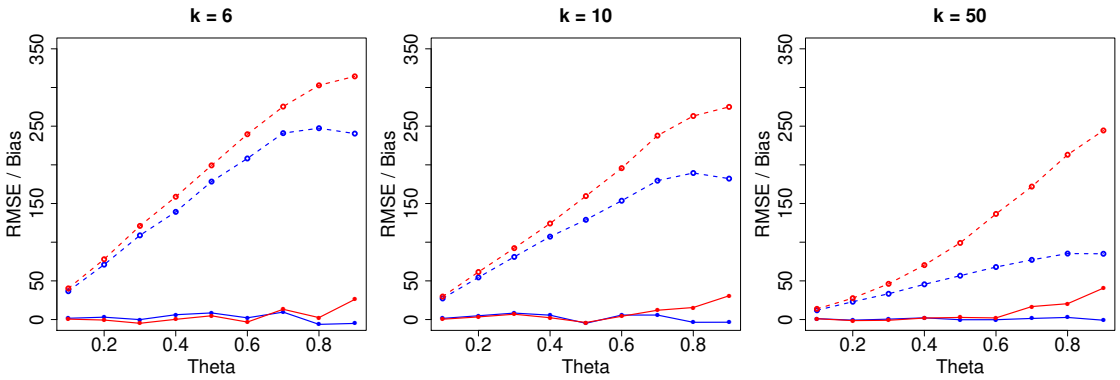

Figure 3.1: Root mean squared errors (dashed) and biases (solid) of $\hat{\theta}_{\text{Bayes}}$ (blue) and $\hat{\theta}_{\text{PL}}$ (red) for different dimensions k and different parameters θ. Values have been multiplied by 10000.

We note that other functionals of the posterior distribution can be used to obtain point estimators. Simulations based on the posterior mean, for instance, gave very similar results, and we therefore restrict to the posterior median in the sequel. Similarly, changing the prior distributions does not have a strong effect on the posterior distribution for the sample sizes we consider; see also Section 3.6.1.

3.5.1 Max-Stable Data

We first take the marginal parameters to be fixed and known and quantify the efficiency gains of $\hat{\theta}_{\text{Bayes}}$ compared to $\hat{\theta}_{\text{PL}}$. We simulate $N = 100$ samples $z^{(1)}, \dots, z^{(N)}$ from the logistic distribution for different dimensions $k \in \{6, 10, 50\}$ and different dependence parameters $\theta = 0.1 \times i$, $i = 1, \dots, 9$. For each combination of dimension k and parameter θ we then run a Markov chain with length 1500, where we discard the first 500 steps as the burn-in time. The empirical median of the remaining 1000 elements gives $\hat{\theta}_{\text{Bayes}}$. The chain is sufficiently long to reliably estimate the posterior median; see also the mixing properties in Section 3.6.1. The maximum pairwise likelihood estimator $\hat{\theta}_{\text{PL}}$ is obtained according to (3.28). The whole procedure is repeated 1500 times to compute the corresponding root mean squared errors and biases shown in Figure 3.1.

As expected, the use of full dependence information substantially decreases the root mean squared errors and thus increases the efficiency of the estimates. In extreme value statistics, where typically only small data sets are available, this allows to reduce uncertainty due to parameter estimation. The advantage of this additional information becomes stronger for both higher dimensions and weaker dependence, analogously to the observations in Huser et al. (2016). This behavior can to some extent be understood by the results in Shi (1995) on the Fisher information of the logistic distribution for different dimensions and dependence parameters. When $\theta \downarrow 0$, pairwise likelihood performs just as well as full likelihood, which is sensible since, up to a multiplicative constant, the pairwise likelihood equals the full likelihood when $\theta = 0$.

It is interesting to note that the estimates $\hat{\theta}_{\text{Bayes}}$ appear to be unbiased in almost all cases, whereas the pairwise estimator has a finite sample bias for θ close to 1.

3.5.2 Data in the Max-Domain of Attraction

In applications, the max-stable distribution Z might not be observed exactly but only as an approximation by componentwise block maxima of data vectors $X^{(1)}, \dots, X^{(b)}$ in its max-domain of attraction with standard Fréchet margins, where $b \in \mathbb{N}$ is the block size. Indeed, the random vector

$$\tilde{Z} = \frac{1}{b}\left(\max_{l=1,\dots,b} X_1^{(l)}, \dots, \max_{l=1,\dots,b} X_k^{(l)}\right),$$

approximates the distribution of Z, where the approximation improves for increasing b. In this situation we can associate to $\tilde{Z}$ the partition of occurrence times of the maxima, say $\tilde{\tau}$. Stephenson and Tawn (2005) proposed to use this information on the partition to simplify the likelihood of the max-stable distribution. For N observations $\tilde{z}^{(1)}, \dots, \tilde{z}^{(N)}$ of $\tilde{Z}$ with partitions $\tilde{\tau}^{(1)}, \dots, \tilde{\tau}^{(N)}$ they defined the estimator

$$\hat{\theta}_{\mathrm{ST}} = \operatorname*{argmax}_{\theta\in\Theta} \sum_{l=1}^{N} \log L(\tilde{z}^{(l)}, \tilde{\tau}^{(l)}; \theta).$$

This estimator suffers from two kinds of misspecification biases. Firstly, the $\tilde{z}^{(l)}$ are only approximately Z distributed and, secondly, the partitions $\tilde{\tau}^{(l)}$ are only finite sample approximations to the true distribution of the limit partition T. For the latter, Wadsworth (2015) proposed a bias reduction method for moderate dimensions and showed in a simulation study that it significantly decreases the bias of the Stephenson–Tawn estimator in the case where the $X^{(k)}$ and thus also $\tilde{Z}$ follow exactly a max-stable logistic distribution. However, if the $X^{(k)}$ are samples from the outer power Clayton copula (cf. Hofert and Mächler, 2011) and thus only in the max-domain of attraction of the logistic distribution, then even the bias reduced estimator suffers from significant bias (cf., Wadsworth, 2015, Table 3).

We repeat the simulation study from Section 3.5.1 with the only difference that, instead of sampling from Z, we simulate $N = 100$ samples $\tilde{z}^{(1)}, \dots, \tilde{z}^{(N)}$ of $\tilde{Z}$, which is the rescaled maximum of $b = 50$ samples from the outer power Clayton copula for different parameters. Based on these data in the max-domain of attraction of the logistic distribution we estimate the dependence parameter θ using our Bayes estimator and compare it to the pairwise likelihood estimator. Both approaches ignore the additional information on the partitions $\tilde{\tau}^{(l)}$ that we have in this setup. On the other hand, we can also compute the Stephenson–Tawn estimator and its bias reduced version by Wadsworth (2015), which explicitly include the partition information.

Table 3.1 shows the root mean squared errors and biases of the four estimators. For all of them the bias plays a significant role for the overall estimation error and that is due to the model misspecification for only approximately max-stable data. This bias is however much stronger for $\hat{\theta}_{\mathrm{ST}}$ and $\hat{\theta}_{\mathrm{W}}$, which use the again misspecified partitions. In this case, the Bayes estimator that treats the partitions as unknown and samples from them automatically seems to be more robust and does not need a bias correction. At the same time it has a small variance and thus in many cases the smallest root mean squared error. Especially in higher dimensions (≥ 20) where the bias reduction of Wadsworth (2015) can no longer be used, the Bayes estimator still provides a robust and efficient method of inference. As one would expect, the pairwise likelihood estimator has the smallest bias since it is less sensitive to model misspecification, but it still a higher root mean squared error due to its higher variance.

	$\theta_0 = 0.1$		$\theta_0 = 0.4$		$\theta_0 = 0.7$		$\theta_0 = 0.9$	
k	6	10	6	10	6	10	6	10
RMSE(θ_{Bayes})	36	29	144	111	241	191	262	220
RMSE(θ_{PL})	40	32	159	127	279	235	311	286
RMSE(θ_{ST})	38	29	148	126	352	401	647	840
RMSE(θ_{W})	38	29	134	108	230	228	313	434
Bias(θ_{Bayes})	-9	-10	-44	-35	-57	-70	-96	-114
Bias(θ_{PL})	-10	-10	-47	-32	-44	-42	-46	-42
Bias(θ_{ST})	-12	-11	-90	-88	-315	-385	-634	-835
Bias(θ_{W})	-11	-11	-61	-58	-158	-194	-277	-422

Table 3.1: Root mean squared errors (top four rows) and biases (bottom four rows) of $\hat{\theta}_{\text{Bayes}}$, $\hat{\theta}_{\text{PL}}$, $\hat{\theta}_{\text{ST}}$ and $\hat{\theta}_{\text{W}}$, estimated from 1500 estimates; figures have been multiplied by 10000.

3.5.3 Estimation of Marginal Extreme Value Parameters

In spatial settings, the marginal extreme value parameters are often estimated by using the independence likelihood (Chandler and Bate, 2007), where all locations are assumed independent. This avoids to specify a dependence structure but can result in efficiency losses, even if only the marginal parameters are of interest.

We perform a simulation study to assess how using the full likelihoods in a Bayesian framework improves estimation of the marginal parameters. We fix the dimension $k = 10$ and set the marginal parameters to $\mu = 1$, $\sigma = 1$ and $\xi \in \{-0.2, 0.4, 1\}$, equal for all k margins. The dependence is logistic with unknown nuisance parameter $\theta_0 \in \{0.1, 0.4, 0.7, 0.9\}$.

Based on $N = 100$ independent samples from this model, we compare three different estimation procedures. The first one is our Bayesian approach using the full joint likelihood of the marginal parameters and the dependence parameter. We use a uniform prior for θ, and independent normal priors for μ, $\log\sigma$ and ξ with large standard deviations. For the univariate case, more sophisticated choices for the prior distributions are possible, including dependencies between the three extreme value parameters (e.g., Stephenson and Tawn, 2004; Northrop and Attalides, 2016).

The second procedure is the maximum pairwise likelihood estimator that only uses bivariate dependence, and the third is the maximum independence likelihood estimator that completely ignores dependence between different components. Each simulation and estimation is repeated 1500 times.

Table 3.2 contains the root mean squared errors of the marginal parameters for the three approaches. Interestingly, for the location and scale parameter we see only little difference between the three methods, meaning that they can be efficiently estimated without taking into account dependencies. For the shape parameter, however, there are substantial improvements in the estimation error by including the unknown dependence structure in the model and estimating it simultaneously. Since estimation of the shape is both the most difficult and the most important of the three extreme value parameters, the Bayesian approach is promising also for marginal tail estimation. Finally, we observe that there is already an efficiency gain for the shape parameter when only the pairwise dependence is considered, but it is even more remarkable in the Bayesian setting with full likelihoods. Table 3.2 also shows that these observations hold across different ranges for the shape parameter ξ. It should be noted that we considered the case of a well-specified model,

	$\theta_0 = 0.1$			$\theta_0 = 0.4$			$\theta_0 = 0.7$			$\theta_0 = 0.9$		
$\xi = -0.2$	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
Bayes full	105	75	22	97	58	21	71	37	20	51	28	19
Pairwise	106	73	31	98	57	28	72	39	24	52	30	21
Independence	111	75	67	101	58	52	73	39	35	52	30	24
$\xi = 0.4$	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
Bayes full	102	99	41	96	89	39	71	65	34	51	45	32
Pairwise	106	98	57	97	89	54	71	67	45	51	47	38
Independence	112	100	96	100	89	79	72	67	56	51	48	42
$\xi = 1$	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
Bayes full	109	155	85	100	144	76	77	111	59	53	75	46
Pairwise	106	146	94	96	135	90	74	108	72	52	75	53
Independence	110	146	127	98	135	107	74	109	81	52	76	57

Table 3.2: Root mean squared errors of (μ, σ, ξ) estimates with different values of ξ for the Bayesian approach, pairwise likelihoods and independence likelihoods, respectively, where θ is an unknown nuisance parameter; figures have been multiplied by 1000.

where the class of dependence structures is known. An additional model uncertainty might render the independence likelihood more favorable.

3.6 Applications in a Bayesian Framework

In the previous sections we discussed the efficiency gains of the Bayesian full likelihood approach in the frequentist framework of point estimates. The Markov chain from Section 3.2.2 however produces not only a point estimate but an estimate of the entire posterior distribution. For instance, this can directly be used to produce credible intervals for the parameter of interest. As a further application of our approach in the Bayesian framework, we will present Bayesian model comparison in this section.

3.6.1 The Posterior Distribution and Credible Intervals

As an illustration of the methodology we simulate a sample of $N = 15$ data $z^{(l)} = (z_1^{(l)}, \ldots, z_k^{(l)})$, $l = 1, \ldots, N$, from a k-dimensional max-stable vector Z whose distribution belongs to the parametric family of logistic distributions introduced in Subsection 3.4.1 with parameter space $\Theta = (0, 1)$. We run the Markov chain from Subsection 3.2.2. The left panel of Figure 3.2 shows the Markov chain for the parameter θ with simulated data from the logistic distribution in dimension $k = 10$ with $\theta_0 = 0.8$. The prior distribution is uniform, that is, $\pi_\theta = \text{Unif}(0, 1)$. The chain seems to have converged to its stationary distribution, namely the posterior distribution

$$L\left(\theta \mid \{z^{(l)}\}_{l=1}^N\right) \propto \pi_\theta(\theta) \prod_{l=1}^N L(z^{(l)}; \theta), \tag{3.29}$$

after a burn-in period of about 200 steps. The auto correlation of the Markov chain in Figure 3.3 suggests that there is serial dependence up to a lag of 30 steps. The parallel chain that updates the partitions is difficult to plot. The right panel of Figure 3.2 therefore shows

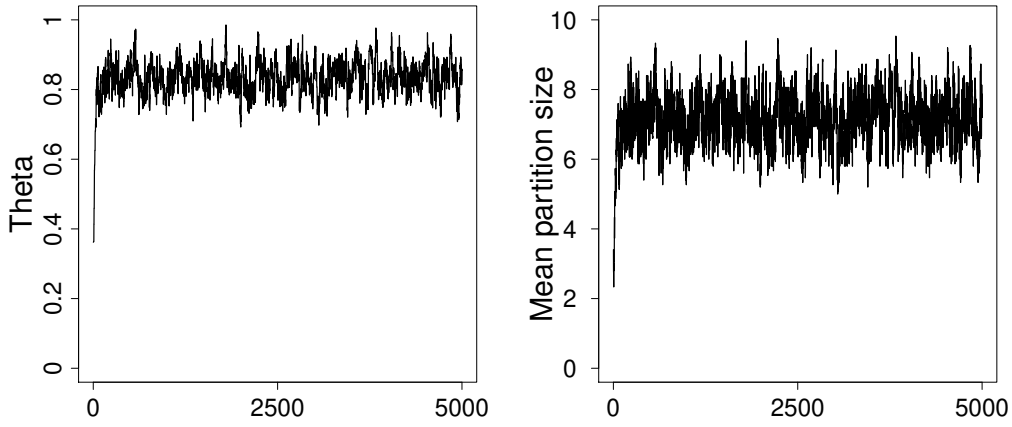

Figure 3.2: Markov chains for θ (left) and the mean partition size (right) with uniform prior.

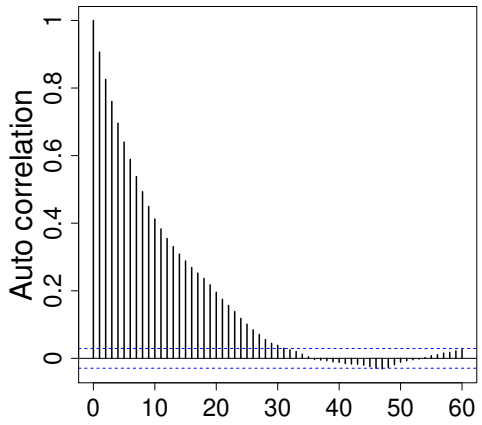

Figure 3.3: Auto correlation function of the Markov chain for the parameter θ.

in each step as a summary the mean number m of sets in the partitions $\tau^{(1)}, \ldots, \tau^{(N)}$, that is, $m = 1/N \sum_{l=1}^{N} |\tau^{(l)}|$. For complete independence ($\theta_0 = 1$) we must have $m = k = 10$, whereas for complete dependence ($\theta_0 = 0$) we have $m = 1$.

The left panel of Figure 3.4 shows a histogram and an approximated smooth version of the posterior distribution, together with the uniform prior. In order to assess the impact of the prior distribution on the posterior, the two other panels contain the corresponding plots for the same data set but for different priors, namely the beta distributions $\pi_\theta = \text{Beta}(0.5, 0.5)$ (center) and $\pi_\theta = \text{Beta}(4, 4)$ (right). Even for a relatively small amount of $N = 15$ data points, the influence of the prior is not very strong.

The Bayesian setup provides us with a whole distribution for the parameter instead of a point estimate only. From this we can readily deduce credible intervals for the parameter θ. This is an advantage compared to frequentist composite likelihood methods since the Fisher information matrix has a "sandwich" form adjusting for the misspecified likelihood, and confidence intervals are thus not easily computed (Padoan et al., 2010). When using composite likelihoods in a Bayesian setup, the posterior distributions are much too

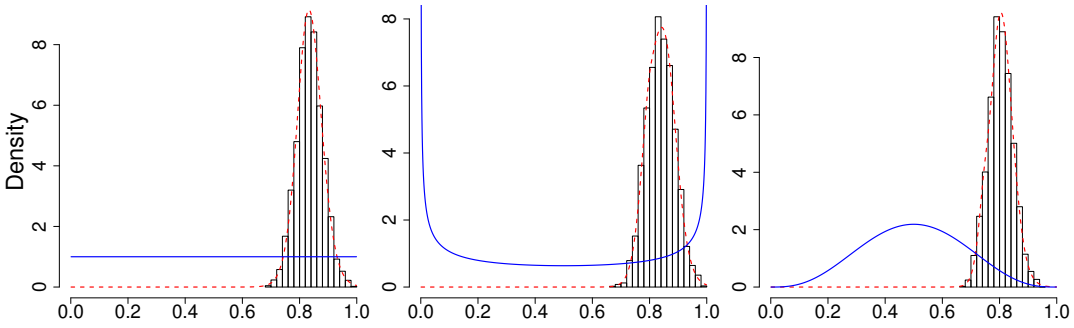

Figure 3.4: Histogram and smooth approximation of the posterior distribution (dotted red) for different priors (solid blue): Unif$(0,1)$ (left), Beta$(0.5,0.5)$ (center) and Beta$(4,4)$ (right).

concentrated and the empirical coverage rates are very small. Adjustments are necessary to obtain appropriate inference (Ribatet et al., 2012). Since our approach uses the full, correct likelihood, no adjustment is needed to obtain accurate empirical coverage rates. Indeed, in Table 3.3 we provide the coverage rates of the 95% credible intervals obtained in the simulation study in Section 3.5.1 for some values of θ_0.

3.6.2 Bayesian Model Comparison

Starting from data z from a family of max-stable distributions $\{F_\theta,\, \theta \in \Theta\}$, we consider two sub-models $M_1 : \theta \in \Theta_1$ and $M_2 : \theta \in \Theta_2$ for disjoint sets $\Theta_1, \Theta_2 \subset \Theta$. In Bayesian statistics, comparison of such models is often based on the Bayes factor $B_{1,2}$, which translates the prior odds into the posterior odds (e.g., Kass and Raftery, 1995), that is,

$$\frac{\pi_{\text{posterior}}(\Theta_1)}{\pi_{\text{posterior}}(\Theta_2)} = B_{1,2} \times \frac{\pi_{\text{prior}}(\Theta_1)}{\pi_{\text{prior}}(\Theta_2)}. \tag{3.30}$$

The Bayes factor can also be written as $B_{1,2} = L(z \mid M_1)/L(z \mid M_2)$, where

$$L(z \mid M_i) = \int_\Theta L(z;\theta)\pi(\theta \mid M_i)\,\mathrm{d}\theta, \quad i = 1, 2, \tag{3.31}$$

are the so-called marginal probabilities of the data and $\pi(\cdot \mid M_i)$ is the prior density of the parameter θ under the model M_i. Since the max-stable likelihood cannot be computed, the integral in (3.31) is computationally infeasible. However, we can use the estimation of the posterior probability (3.29) discussed in the previous subsection and estimate

$$B_{1,2} = \frac{\pi_\theta(\Theta_2)}{\pi_\theta(\Theta_1)} \times \frac{\int_{\Theta_1} L(\theta \mid \{z^{(l)}\}_{l=1}^N)\,\mathrm{d}\theta}{\int_{\Theta_2} L(\theta \mid \{z^{(l)}\}_{l=1}^N)\,\mathrm{d}\theta}. \tag{3.32}$$

	$\theta_0 = 0.1$			$\theta_0 = 0.4$			$\theta_0 = 0.7$			$\theta_0 = 0.9$		
k	6	10	50	6	10	50	6	10	50	6	10	50
Coverage (in %)	94	93	90	95	94	94	94	94	94	94	94	90

Table 3.3: Empirical coverage rates of 95% credible intervals obtained from the posterior distributions using full likelihood.

As an example, we consider a simple regression model

$$\xi_i = \alpha + i\beta, \quad i = 1, \dots, k, \tag{3.33}$$

for the marginal shape parameters $\xi_1, \dots, \xi_k$ of the k-dimensional max-stable distribution in dimension k. One might be interested in testing if there is a linear trend in the shape parameters, and, thus, in comparing the models $M_1 : \{\beta = 0\}$ and $M_2 : \{\beta \neq 0\}$. In order to compute the Bayes factor as the ratio of the posterior probabilities of the two models according to (3.32), the prior distribution π_β of β must be a mixture

$$p_{0,\pi} \times \delta_{\{0\}} + (1 - p_{0,\pi}) \times \pi_\beta^c$$

of a Dirac point mass $\delta_{\{0\}}$ on 0 and an appropriate continuous distribution π_β^c on $\mathbb{R}$ with mixture weight $p_{0,\pi} \in (0, 1)$. This ensures that we have a positive posterior probability on both sets $\{\beta = 0\}$ and $\{\beta \neq 0\}$ and the Bayes factor is well-defined. Here, it is important to note that the choice of the mixture weight $p_{0,\pi}$ does not have an effect on the Bayes factor $B_{1,2}$ as Equations (3.30)–(3.32) show.

Similarly as in Section 3.5.3, we simulate $N = 15$ data from a max-stable logistic distribution with dimension $k = 10$ and dependence parameter $\theta_0 = 0.5$, with marginal parameters $\mu_i = 1$, $\sigma_i = 1$ and ξ_i as in (3.33) with $\alpha = 1$ and different values for β, $i = 1, \dots, k$. The prior distributions for the dependence, location and scale parameters are chosen as in Section 3.5.3. The prior for α is standard normal and for the prior for β is a mixture of $0.5 \times \delta_{\{0\}} + 0.5 \times \pi_\beta^c$ of a point mass and a centered normal with standard deviation 0.5 as the continuous component π_β^c.

A Markov chain whose stationary distribution is the posterior distribution of the parameters given the data can be constructed analogously to Section 3.2.2. However, given the current state β of the Markov chain, the proposal β^* is not drawn from a continuous distribution with density q, but from a mixture

$$p_0(\beta)\delta_{\{0\}}(\cdot) + (1 - p_0(\beta))q^c(\beta, \cdot)$$

of a Dirac point mass on $\{0\}$ and a continuous distribution with density $q^c(\beta, \cdot)$, with mixture weight $p_0(\beta) \in (0, 1)$. To ensure convergence of the Markov Chain, the densities $q^c(\beta, \cdot)$ should be chosen such that $q^c(\beta, \beta^*) > 0$ if and only if $q^c(\beta^*, \beta) > 0$.

Figure 3.5 gives an illustration of the Bayes factors $B_{1,2}$ that compare the model without trend $M_1 : \{\beta = 0\}$ and the model with trend $M_2 : \{\beta \neq 0\}$ for the simulated data described above. The true trend varies from $\beta = 0$, in which case M_1 would be correct, over positive values up to $\beta = 0.08$ where M_2 is the correct model. As comparison, we implemented a Bayesian approach based on the independence likelihood (Chandler and Bate, 2007), which is the product of the marginal densities and ignores the dependence structure. The results show that using the full likelihood that takes the dependence into account and treats it as a nuisance parameter significantly facilitates the distinction between the two different models. The Bayes factors for the full likelihood show stronger support for M_1 if $\beta = 0$, and decrease more rapidly to 0 if $\beta > 0$ than the Bayes factors for independence likelihood.

Finally, we note that a similar approach has been proposed in the univariate setting for estimation of the shape parameter in Stephenson and Tawn (2004) in order to allow the Gumbel case $\xi = 0$ with positive probability.

3.7 Discussion

We present an approach that allows for inference of max-stable distributions based on full likelihoods by perceiving the underlying random partition of the data as latent variables

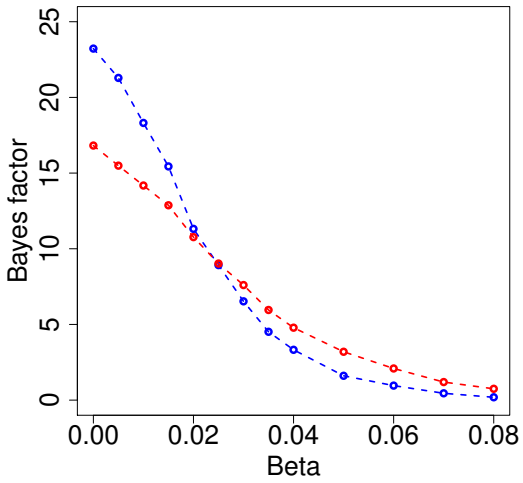

Figure 3.5: Bayes factors for different value of β for full likelihood (blue) and independence likelihood (red).

in a Bayesian framework. The formulas for $\omega(\tau_j, z)$ provided in Section 3.4 allow in principle to perform Bayesian inference based on full likelihoods for many popular max-stable distributions in any dimension. However, computational challenges arise for both the extremal-t and the Brown–Resnick model in higher dimensions since the corresponding $\omega(\tau_j, z)$ require the evaluation of a multivariate Student and Gaussian distribution functions, respectively, which have to be approximated numerically; see also Thibaud et al. (2016). The recent work de Fondeville and Davison (2018) on efficient computation of Gaussian distribution functions allows for even higher dimensions.

Making use of the weights $\omega(\tau_j, z)$, the posterior distribution of the parameters becomes numerically available by samples based on Markov chain Monte Carlo techniques. As the results in Section 3.6.1 indicate, the posterior distribution does not show strong influence of the prior distribution even in case of a rather small amount of data; cf., Figure 3.4. In most of the examples presented here, the proposal distributions for the model parameters in the Metropolis–Hastings algorithms were chosen to be centered around the current state of the Markov chain with an appropriate standard deviation, resulting in chains with satisfactory convergence and mixing properties; cf., Figures 3.2 and 3.3, for instance. Further improvements of these properties might be possible, e.g., by implementing an adaptive design of the Markov chain Monte Carlo algorithms.

In the frequentist framework, we propose to use the posterior median as a point estimator for the model parameters. As the simulation studies in Section 3.5 show, the use of full likelihoods considerably improves the estimation errors compared to the commonly used composite likelihood method even in the case of a rather small sample size. This complements our theoretical results on the asymptotic efficiency of the posterior median. Besides the point estimator in the frequentist setting, we can also make use of the posterior distribution in a Bayesian framework. In Section 3.6, we discuss the use of credible intervals and Bayesian model comparison for max-stable distributions. Further applications such as Bayesian prediction are possible.

Acknowledgments

We thank the editorial team and two referees for helpful comments. Financial support by the Swiss National Science Foundation (S. Engelke) and by the Bourgogne–Franche–Comté region (C. Dombry, grant OPE-2017-0068) is gratefully acknowledged.

4 Exact and Fast Simulation of Max-Stable Processes on a Compact Set Using the Normalized Spectral Representation

joint work with Martin Schlather and Chen Zhou

This chapter is based on the article Oesting et al. (2018b) which has appeared in *Bernoulli*. Only few changes have been made: Besides some minor modifications and a correction in Equation (4.47), the proof of Proposition 4.13 has been shifted from the appendix to the main body of the chapter.

The efficiency of simulation algorithms for max-stable processes relies on the choice of the spectral representation: different choices result in different sequences of finite approximations to the process. We propose a constructive approach yielding a normalized spectral representation that solves an optimization problem related to the efficiency of simulating max-stable processes. The simulation algorithm based on the normalized spectral representation can be regarded as max-importance sampling. Compared to other simulation algorithms hitherto, our approach has at least two advantages. First, it allows the exact simulation of a comprising class of max-stable processes. Second, the algorithm has a stopping time with finite expectation. In practice, our approach has the potential of considerably reducing the simulation time of max-stable processes.

4.1 Introduction

Max-stable processes have become a popular tool for modeling spatial extremes, particularly in environmental sciences, see, e.g. Coles (1993), Coles and Tawn (1996) and Padoan et al. (2010). A stochastic process $\{Z(y) : y \in K\}$ with standard Fréchet margins defined on an index set K, i.e. $\mathbb{P}(Z(y) \leq z) = \exp(-z^{-1})$, $z > 0$, for all $y \in K$, is called max-stable if

$$\frac{1}{n} \max_{i=1}^{n} Z_i =_d Z$$

for any $n \in \mathbb{N}$ and independent copies Z_i, $i = 1, \ldots, n$, of Z, where the maximum is taken pointwise. Max-stable processes occur naturally as limits of suitably normalized pointwise maxima of stochastic processes which motivates their usage in the context of spatial extremes.

Simulating max-stable processes is an important step in application for the following three reasons. First, while bivariate marginal distributions can be calculated frequently, higher dimensional marginal distributions do not have, in nearly all the cases, explicit formulae. Consequently, they can be addressed only by simulation. Second, most applications require the estimation of characteristics of max-stable processes that cannot be explicitly calculated. That leaves simulation as the only option, see, e.g. Buishand et al. (2008) and

Blanchet and Davison (2011). Finally, unconditional simulation appears as part of the conditional simulation of max-stable processes (Dombry et al., 2013; Oesting and Schlather, 2014).

Schlather (2002) suggested an algorithm to simulate max-stable processes. However, the simulation is exact only under substantial restrictions, namely when the shape function is bounded and has compact support. In this case, the algorithm ends in finite time almost surely. Compact support can be enforced by cutting the shape function, which introduces an approximation error. In some cases, such an error is not negligible. For example, consider moving maximum processes with monotone shape functions that have a pole at the origin. Strokorb et al. (2015) provided a derivation of such processes whose realizations have poles on a dense subset of the space. In particular, they are discontinuous everywhere with probability one. For such processes, any modification of the shape function towards a bounded shape function will alter dramatically the properties of the process. Consequently, the use of Schlather's (2002) algorithm becomes doubtful. By contrast, this chapter deals with the question of drawing random samples of max-stable processes in an exact and efficient way, including shape functions that are unbounded and do not have compact support.

Simulation of max-stable processes is based on their spectral representation (see de Haan, 1984; Giné et al., 1990; Kabluchko, 2009; Wang and Stoev, 2010, for instance): For any max-stable process Z with standard Fréchet margins defined on an index set K, there exists a spectral measure H defined on an appropriate set $\mathbb{H}$ of non-negative functions such that

$$Z(y) = \max_{(t,f)\in\Pi} tf(y), \qquad y \in K, \tag{4.1}$$

where Π is a Poisson point process on $(0,\infty)\times\mathbb{H}$ with intensity $t^{-2}\,\mathrm{d}t\,H(\mathrm{d}f)$ and

$$\int_{\mathbb{H}} f(y)\,H(\mathrm{d}f) = 1 \tag{4.2}$$

for all $y \in K$. The functions f in $\mathbb{H}$ are the *spectral functions* of the max-stable process Z. As any max-stable process can be obtained from a max-stable process with standard Fréchet marginals via marginal transformations, we will henceforth assume processes with standard Fréchet marginals, i.e. processes with representation (4.1) and (4.2).

According to the spectral representation (4.1), the construction of a max-stable process involves infinitely many points $(t,f)\in\Pi$. Nevertheless, since only the maximum over all functions tf counts, the number of points (t,f) that contribute to Z, i.e. $Z(y) = tf(y)$ for at least one point $y \in K$, is finite under mild conditions, see de Haan and Ferreira (2006), Corollary 9.4.4. However, their statement is a theoretical one that does not help for simulation purposes because one cannot determine ex ante which functions f will contribute.

Schlather's (2002) algorithm for simulating max-stable processes requires that K is compact, the shape functions $f : \mathbb{R}^d \to [0,\infty)$ are bounded by some $C \in (0,\infty)$ and the support is within a ball of radius r centered at the origin. For example, consider a stationary moving maxima processes with the measure H given by

$$H(\{f(\cdot - x) : f \in B, x \in A\}) = \Lambda(A)H_f(B) \tag{4.3}$$

for some probability measure H_f and the Lebesgue measure Λ. Let $K_r = \{x \in \mathbb{R}^d : \|x - y\| \le r \text{ for some } y \in K\}$. Under the aforementioned assumptions, Schlather (2002) showed that the right-hand side of (4.3) can be restricted to $\Lambda(A\cap K_r)H_f(B)$. Hence, the

intensity for $(t, f(\cdot - x))$ equals

$$t^{-2}\Lambda(K_r)\mathrm{d}t \cdot \frac{\Lambda(\mathrm{d}x \cap K_r)}{\Lambda(K_r)} H_f(\mathrm{d}f),$$

where $f(\cdot - x)$ can be reinterpreted as realization of a random function F with law $\Lambda(\mathrm{d}x \cap K_r)/\Lambda(K_r) \cdot H_f(\mathrm{d}f)$ that is independent of the Poisson point process with intensity $t^{-2}\Lambda(K_r)\mathrm{d}t$.
Schlather (2002) suggested to start with those points (t, f) that will contribute most likely to Z, i.e., with those points (t, f) that have the highest values of t. By ranking the points t in a descending order $t_1 > t_2, \ldots$, we have that $t_i =_d 1/(\sum_{j=1}^{i} E_j)$, where E_j are independent and identically distributed random variables with standard exponential distribution. Let $F_i \sim_{i.i.d.} F$ be independent of the E_j and

$$Z^{(m)}(y) = \max_{1 \le i \le m} \frac{1}{\sum_{j=1}^{i} E_j} F_i(y), \qquad y \in K,$$

be a finite approximation for Z. Then, $Z =_d Z^{(\infty)}$. Assume that for some m the inequality

$$Z^{(m)}(y) \ge \frac{1}{\sum_{j=1}^{m+1} E_j} \sup_{f \in \mathbb{H}} f(y) \quad \text{for all } y \in K, \tag{4.4}$$

holds. Then, obviously, $Z^{(n)}(y) = Z^{(\infty)}(y)$ for all $y \in K$ and all $n \ge m$. In other words, any spectral function f_i with $i > m$ cannot contribute to Z. This results in a *stopping rule* for a "m-step representation" of Z which can be used to construct an exact simulation procedure. Here, the minimal number of steps $M = \min\{m \in \mathbb{N} : m \text{ satisfies (4.4)}\}$ is a random integer. The properties of M depend on the choices of the spectral functions.
Schlather (2002) further proposed to replace inequality (4.4) by a stronger stopping rule $\inf_{y \in K} Z^{(m)}(y) \ge C/\sum_{j=1}^{m+1} E_j$. As a generalization of Schlather (2002), we consider stopping rule (4.4) with H being any locally finite measure, not necessarily of the form (4.3).
The random number of steps in the simulation of a max-stable process depends on the choice of the corresponding ensemble of spectral functions which is not unique (cf. de Haan and Ferreira, 2006, Remark 9.6.2). Some specific choices may bear severe disadvantages for the accuracy and speed of the simulation. For instance, finite approximations based on the original definition of the Brown–Resnick process are usually far from the actual process (Kabluchko et al., 2009; Oesting et al., 2012). The optimality of the choice of spectral functions with respect to the number of steps in simulating max-stable processes has not been discussed in literature yet. Here, we propose a choice of spectral functions, the normalized spectral representation and show that it is the solution to an optimization problem related to the number of steps in the simulation.
An illustrative example may clarify why the choice of spectral functions can have a significant impact on the distribution of the stochastic number M. Consider the simplest case where Z is univariate. Specializing (4.1) to $K = \{y_0\}$ and $f \equiv 1$, the random variable $Z(y_0)$ follows a univariate Fréchet distribution. It has a representation given by

$$Z(y_0) =_d \max_{t \in \Pi} t \tag{4.5}$$

where Π is the Poisson point process on $(0, \infty)$ with intensity $t^{-2}\,\mathrm{d}t$. By ranking the points t, we get that

$$Z^{(m)}(y_0) = \max_{1 \le i \le m} \frac{1}{\sum_{j=1}^{i} E_j} = \frac{1}{E_1} = Z^{(1)}(y_0).$$

In other words, $M \equiv 1$ based on the constant spectral function $f \equiv 1$. Differently, $Z(y_0)$ can also be constructed by (4.1) with $f(y_0)$ being a non-degenerate random variable satisfying $\mathbb{E}f(y_0) = 1$. In that case, the stochastic number M is greater than 1 with positive probability. Therefore, for simulating $Z(y_0)$, the spectral representation in (4.5) would be considered as optimal. This example illustrates the optimality we intend to achieve by the choice of spectral functions for an arbitrary max-stable process.
The very general optimality problem for general index sets K and arbitrary random functions f seems to be rather complicated. Therefore, we shall suggest a modified optimization problem and shall demonstrate that its solution is explicit and unique for each given max-stable process and index set K. The optimality is achieved by transforming any ensemble of spectral functions to a new ensemble of spectral functions satisfying $\sup_{y\in K} f(y) = c$, for all $f \in \mathbb{H}$. We call such a representation with all spectral functions sharing the same supremum the *normalized spectral representation*. This representation was initially used in constructing the spectral representation for max-stable processes with a continuous sample path on $K = [0, 1]$, see e.g. de Haan and Lin (2001) and de Haan and Ferreira (2006), Corollary 9.4.5. In this chapter, we provide a theoretical justification on the use of this normalized spectral representation in simulation algorithms.
This chapter is organized as follows. In Section 4.2, we revisit de Haan's (1984) spectral representation of max-stable processes and give the formula of spectral representation transformation, i.e. transforming one ensemble of spectral functions under a given spectral measure to another ensemble under a different spectral measure. Based on the transformed representation, we formulate a stopping rule that allows an exact simulation of max-stable processes. In Section 4.3, we pose an optimization problem for selecting the spectral representation that yields the most efficient simulation procedure. This problem is closely related to the problem of importance sampling. In addition, we give the explicit solution of a modified optimization problem. This results in the normalized spectral representation. Differences between the modified problem and the original optimization problem are evaluated in Section 4.4. Section 4.5 deals with the normalized spectral representation for moving maxima processes. For moving maxima processes and Brown–Resnick processes, the performance of the simulation procedure based on the normalized spectral representation is compared to other algorithms in Section 4.6. In Section 4.7, we summarize and discuss our results.

4.2 Transformation of Spectral Representations

Throughout the chapter we consider a max-stable process Z on some index set K that is assumed to be a compact Polish space. Further, we will assume that the spectral functions f from representation (4.1) lie in some Polish space $\mathbb{H} \subset [0, \infty)^K$ equipped with a σ-algebra $\mathcal{H}$ such that the mapping $f \mapsto \sup_{y\in K} f(y)$ is $(\mathcal{H}, \mathcal{B} \cap [0, \infty))$-measurable where $\mathcal{B}$ denotes the standard Borel σ-algebra on $\mathbb{R}$.

The subsequent proposition presents a general procedure to transform one spectral representation to another yielding the same max-stable process. It can be proved by standard arguments from extreme value theory, see Oesting et al. (2013) for details.

Proposition 4.1. *Let Z be a max-stable process with standard Fréchet margins defined as in* (4.1) *and* (4.2). *Suppose that H is a locally finite measure on $\mathbb{H}$. Let g be some probability density on $\mathbb{H}$ w.r.t. H, i.e. $g \geq 0$ and $\int_{\mathbb{H}} g(f)H(\mathrm{d}f) = 1$, such that*

$$H\left(\left\{f : g(f) = 0,\ \sup_{y\in K} f(y) > 0\right\}\right) = 0. \tag{4.6}$$

Then, by excluding all the functions $f \in \mathbb{H}$ with $g(f) = 0$, we get

$$Z(y) =_d \max_{(t,f)\in\tilde{\Pi}} t\frac{f(y)}{g(f)}, \qquad y \in K, \tag{4.7}$$

where $\tilde{\Pi}$ is a Poisson point process with intensity $t^{-2}\,\mathrm{d}t\, g(f)H(\mathrm{d}f)$.

Applying Proposition 4.1, a given ensemble of spectral functions $\{f\}_{(t,f)\in\Pi}$ can be transformed to a new ensemble $\{f/g(f)\}_{(t,f)\in\tilde{\Pi}}$, where f follows the transformed probability measure gH defined by

$$gH(A) = \int_A g(f)H(\mathrm{d}f)$$

for all measurable sets $A \subset \mathbb{H}$. For this transformed spectral representation (4.7), the stopping rule (4.4) can be formulated as follows. Denote

$$Z^{(m)}(y) = \max_{1\le i\le m} \frac{1}{\sum_{j=1}^{i} E_j} \cdot \frac{F_i(y)}{g(F_i)}, \qquad y \in K, \tag{4.8}$$

for standard exponentially distributed random variables E_j and $F_j \sim gH$, which are all independent. Let

$$Z^{(\infty)} = \lim_{m\to\infty} Z^{(m)}. \tag{4.9}$$

Then, $Z^{(\infty)} =_d Z$ and, for fixed $\omega \in \Omega$, we have $Z^{(m)} \equiv Z^{(\infty)}$ on K if

$$\operatorname*{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)Z^{(m)}(y)} \le \sum_{j=1}^{m+1} E_j, \tag{4.10}$$

where the essential supremum is taken w.r.t. the probability measure gH. Note that, by (4.6), up to a H null set, the set $\{f \in \mathbb{H} :\ g(f) = 0\}$ consists of functions $f \in \mathbb{H}$ with $f\,|_K \equiv 0$. Thus, for fixed g, we may exclude all the functions $f \in \mathbb{H}$ with $g(f) = 0$.
If the number $M_g := \min\{m \in \mathbb{N} :\ m \text{ satisfies (4.10)}\}$ is finite, the max-stable process Z can be simulated exactly in finite time via the following algorithm.

Algorithm 4.1: Exact simulation of a max-stable process via threshold stopping

Set $m = 0$ and $Z^{(0)}(y) = 0$ for all $y \in K$.
Simulate a standard exponentially distributed random variable E_1.
while (4.10) *is not satisfied* **do**
  Update m by $m + 1$.
  Sample F_m from gH.
  Set $Z^{(m)}(y) = \max\left\{Z^{(m-1)}(y), \frac{1}{\sum_{j=1}^{m} E_j}\frac{F_m(y)}{g(F_m)}\right\}$ for all $y \in K$.
  Simulate a standard exponentially distributed random variable E_{m+1}.
return $Z^{(m)}$

Algorithm 4.1 requires the evaluation of the stopping rule (4.10) which involves suprema with respect to $y \in K$ and $f \in \mathbb{H}$. While, in practice, the set K often is finite, i.e. the supremum w.r.t. y is a maximum, the evaluation of the essential supremum w.r.t. f is more difficult. However, with the normalized spectral representation (cf. Algorithm 4.2), it is possible to provide a bound for the essential supremum facilitating this evaluation; see Remark 4.12.
The exact simulation based on the stopping rule (4.10) requires the simulation of M_g processes from law gH and $M_g + 1$ exponentially distributed random variables. Therefore,

the time costs of Algorithm 4.1 depend on the stopping rule, which essentially depends on the choice of g. Consequently, the following question arises: which density g minimizes the random number of steps M_g? This optimization problem will be formulated in a more rigorous way in the next section.

4.3 The Optimization Problem

Henceforth, we will always assume that we are in the framework of Proposition 4.1. Further, the process Z is assumed to be almost surely strictly positive on K, i.e.

$$\mathbb{P}\left(\inf_{y\in K} Z(y) > 0\right) = 1. \tag{4.11}$$

Note that this assumption is valid for continuous sample path process.[1]

We are interested in minimizing the number of steps in Algorithm 4.1, M_g, i.e. the minimal number m such that (4.10) holds. Since M_g is a random variable, we aim at minimizing some of its mathematical characteristics. As many applications require a large number of simulations, a natural choice to start with is to minimize the mean Q_g of M_g by choosing a proper density g, i.e.,

$$Q_g = \mathbb{E}M_g, \qquad M_g = \min\left\{m \in \mathbb{N} : \operatorname*{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)Z^{(m)}(y)} \le \sum_{j=1}^{m+1} E_j\right\}. \tag{4.12}$$

In addition, we aim to determine at least one member of

$$\mathcal{G} = \arg\min_g Q_g.$$

We remark that the finiteness of Q_g is not ensured. As a first step, the following proposition provides a sufficient condition for M_g being finite almost surely. It follows directly from the definition of M_g.

Proposition 4.2. *Assume that* (4.11) *holds. Then M_g is finite a.s. if*

$$\operatorname*{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)} < \infty. \tag{4.13}$$

Proposition 4.18 below assures the finiteness of Q_g provided that condition (4.13) is fulfilled and Z is sample-continuous.
There is a close relation between our approach and the importance sampling (Hastings, 1970). We interpret (4.7) as "max-importance sampling" as follows. Importance sampling is targeted on calculating an integral $\int f(x)\lambda(\mathrm{d}x)$ in an efficient way for a function $f : \mathcal{X} \to \mathbb{R}$ and a measure λ on some space $\mathcal{X}$. It uses the fact that

$$\int f(x)\lambda(\mathrm{d}x) = \int \frac{f(x)}{g(x)}(g\lambda)(\mathrm{d}x), \tag{4.14}$$

[1] Assuming sample continuity, we even have

$$\mathbb{E}\left((\inf_{y\in K} Z(y))^{-1}\right) < \infty,$$

(Dombry and Éyi-Minko, 2012, Theorem 2.2). It can be shown that this also holds true under weaker assumptions than sample continuity (see Oesting et al., 2013).

where $g : \mathcal{X} \to (0, \infty)$ is a probability density w.r.t. λ and $(g\lambda)(A) = \int_A g(x)\lambda(\mathrm{d}x)$. Then the integral can be approximated by $\hat{I}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)}$, where $(X_i)_{i=1}^{n}$ is an i.i.d. sequence of random variables on $\mathcal{X}$ with law $g\lambda$. Importance sampling considers a practically advantageous choice of the density g such that $\mathrm{Var}(f(X_1)/g(X_1))$ and hence $\mathrm{Var}(\hat{I}^{(n)})$ are small. Exchanging the underlying space $\mathcal{X}$ with the function space $\mathbb{H}$, and replacing the integral with the max-integral (de Haan, 1984), we receive a similar description of our approach:

$$Z(y) = \int^{\vee} y(f) H^*(\mathrm{d}f) = \int^{\vee} \frac{y(f)}{g(f)} (gH^*)(\mathrm{d}f) \tag{4.15}$$

where $y : \mathbb{H} \to [0, \infty), f \mapsto f(y)$, is the punctual evaluation of a function f at y, H^* is a random discrete measure on $\mathbb{H}$ defined as $H^*(\{f\}) = \sup_{(t,f)\in\Pi} t$ and the integral $\int^{\vee}$ is understood as taking the maximum of the integrand weighted by the measure H^*.
Despite the formal correspondence between (4.14) and (4.15), there are some notable differences. First, in our approach, y, which is formally an element of the dual space of $\mathbb{H}$, is not fixed. Second, importance sampling is always an approximation to the integral value of interest. Under mild conditions, see Section 4.4, a finite approach to the right-hand side of equality (4.15) renders the exact value of the max-integral. Hence, we intend to choose g so that the required number of steps is minimized in order to eliminate the error of the finite approximation. In contrast, importance sampling searches for a g that leads to a high speed of convergence, i.e., a small variance of $\hat{I}^{(n)}$.
Whilst the theoretical optimum is well known in importance sampling, and the difficulty there is to find a numerically advantageous function g, the optimization problem (4.12) itself is difficult to solve, since both the numerator and the denominator of $f(y)/(g(f)Z^{(m)}(y))$ depend on y and the denominator is stochastic. To circumvent this difficulty, we modify the optimization problem in Subsection 4.3.1 and solve the modified problem in Subsection 4.3.2. The solution of the modified problem leads to the normalized spectral representation discussed in Section 4.3.3.

4.3.1 A modified optimization problem

We first motivate the modification of the optimization problem. Recall the stopping rule (4.10) as

$$\operatorname*{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)Z^{(m)}(y)} \le \sum_{j=1}^{m+1} E_j.$$

A stronger inequality that implies this stopping rule is

$$\operatorname*{esssup}_{f\in\mathbb{H}} \frac{\sup_{y\in K} f(y)}{g(f)\inf_{\tilde{y}\in K} Z^{(m)}(\tilde{y})} \le \sum_{j=1}^{m+1} E_j, \tag{4.16}$$

while a weaker inequality that is implied by the stopping rule is

$$\operatorname*{esssup}_{f\in\mathbb{H}} \frac{\sup_{y\in K} f(y)}{g(f)\sup_{\tilde{y}\in K} Z^{(m)}(\tilde{y})} \le \sum_{j=1}^{m+1} E_j. \tag{4.17}$$

The actual stopping time is larger than that under the weaker rule and smaller than that under the stronger rule. In other words, if we consider (4.16) and (4.17) as stopping rules, we might simulate too many or too few spectral functions, respectively. Next, we define

an ensemble of alternative stopping rules that also lie in between the rules (4.16) and (4.17). Suppose $T : [0,\infty)^K \to [0,\infty)$ is a functional that satisfies $T(\mathbf{1}) = 1$ and that is max-linear, i.e.

$$T(\max\{a_1 h_1, a_2 h_2\}) = \max\{a_1 T(h_1), a_2 T(h_2)\}$$

for all $a_1, a_2 \geq 0$ and $h_1, h_2 : K \to [0,\infty)$. Then, we have that $T(h) \leq T(g)$ for all $h \leq g$, which leads to

$$\inf_{y\in K} h(y) \leq T(h) \leq \sup_{y\in K} h(y) \tag{4.18}$$

for all $h : K \to [0,\infty)$. Therefore, any max-linear functional T delivers a stopping rule

$$\operatorname*{esssup}_{f\in\mathbb{H}} \frac{\sup_{y\in K} f(y)}{g(f)T(Z^{(m)})} \leq \sum_{j=1}^{m+1} E_j \tag{4.19}$$

that lies also in between the aforementioned stronger and weaker stopping rules.
We regard these new conditions (4.19) as surrogates for the actual stopping rule (4.10). The corresponding modified optimization problem is then

$$\mathcal{G}^* = \arg\min_g Q_g^*, \tag{4.20}$$

$$Q_g^* = \mathbb{E}M_g^* = \mathbb{E}\min\left\{ m \in \mathbb{N} : \operatorname*{esssup}_{f\in\mathbb{H}} \frac{\sup_{y\in K} f(y)}{g(f)T(Z^{(m)})} \leq \sum_{j=1}^{m+1} E_j \right\}.$$

Examples of T are $T(h) = \sup_{y\in K} h(y)$ and $T(h) = h(y_0)$ for some $y_0 \in K$. The corresponding modified problems based on these two specific T minimize the quantities

$$Q_g^{(1)} = \mathbb{E}\min\left\{ m \in \mathbb{N} : \operatorname*{esssup}_{f\in\mathbb{H}} \frac{\sup_{y\in K} f(y)}{g(f)\sup_{\tilde{y}\in K} Z^{(m)}(\tilde{y})} \leq \sum_{j=1}^{m+1} E_j \right\}, \tag{4.21}$$

and

$$Q_g^{(2)}(y_0) = \mathbb{E}\min\left\{ m \in \mathbb{N} : \operatorname*{esssup}_{f\in\mathbb{H}} \frac{\sup_{y\in K} f(y)}{g(f)Z^{(m)}(y_0)} \leq \sum_{j=1}^{m+1} E_j \right\}. \tag{4.22}$$

Note that the modified condition (4.19) does not correspond to the stopping rule under consideration (Algorithm 4.1).

4.3.2 The Solution of the Modified Optimization Problem

The following proposition provides a first but also a key step towards the solution of (4.20). It shows that the solution of (4.20) is independent of T.

Proposition 4.3. *The solution of the modified optimization problem in* (4.20) *satisfies*

$$\mathcal{G}^* = \arg\min_g \operatorname*{esssup}_{f\in\mathbb{H}} \frac{\sup_{y\in K} f(y)}{g(f)}.$$

Proof. If there exists some g such that Q_g^* is finite, then necessarily

$$\operatorname*{esssup}_{f\in H} \frac{\sup_{y\in K} f(y)}{g(f)} < \infty.$$

Thus, we can restrict ourselves to

$$g \in D = \left\{ g : \operatorname*{esssup}_{f\in\mathbb{H}} \frac{\sup_{y\in K} f(y)}{g(f)} < \infty \right\}$$

and assume w.l.o.g. that $D \neq \emptyset$. For $c = \int_{\mathbb{H}} \sup_{y\in K} f(y)\, H(\mathrm{d}f)$ and any $g \in D$, we have

$$c \leq \int_{\mathbb{H}} \operatorname*{esssup}_{h\in\mathbb{H}} \frac{\sup_{y\in K} h(y)}{g(h)} g(f)\, H(\mathrm{d}f) = \operatorname*{esssup}_{h\in\mathbb{H}} \frac{\sup_{y\in K} h(y)}{g(h)} < \infty. \tag{4.23}$$

Thus, by (4.18), for $c_T = \int_{\mathbb{H}} T(f)\, H(\mathrm{d}f)$, we obtain $c_T \leq c < \infty$.
Next, we prove $c_T > 0$ by contradiction. Assume that $c_T = 0$. This yields $T(f) = 0$ for H-a.e. $f \in \mathbb{H}$ which – by the max-linearity of Z – implies $T(Z) = 0$ a.s. in contradiction to $\inf_{y\in K} Z(y) > 0$ a.s. and (4.18). Thus, we conclude that $c_T \in (0, \infty)$.
Now, let $g \in D$. Using the max-linearity of T and the fact that

$$Q_g^* = \mathbb{E}M_g^* = \sum_{m=0}^{\infty} \mathbb{P}(M_g^* > m),$$

we have

$$\begin{aligned} Q_g^* - 1 &= \sum_{m=1}^{\infty} \mathbb{P}\left(\operatorname*{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)} \Big/ \sum_{j=1}^{m+1} E_j > \max_{1\leq k\leq m} \frac{1}{\sum_{j=1}^{k} E_j} \frac{T(f_k)}{g(f_k)} \right) \\ &= \sum_{m=1}^{\infty} \mathbb{P}\left(\frac{\sum_{j=1}^{k} E_j}{\sum_{j=1}^{m+1} E_j} > \frac{T(f_k)}{g(f_k)} \Big/ \operatorname*{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)},\ 1 \leq k \leq m \right). \end{aligned} \tag{4.24}$$

Note that

$$\frac{T(f_k)}{g(f_k)} \Big/ \operatorname*{esssup}_{f\in H} \sup_{y\in K} \frac{f(y)}{g(f)} \in [0, 1].$$

As the joint distribution of $(\sum_{j=1}^{k} E_j / \sum_{j=1}^{m+1} E_j)_{k=1,\ldots,m}$ equals the joint distribution of the order statistics of m independent random variables $U_1, \ldots, U_m$ which are uniformly distributed on $[0, 1]$, and as, by exchangeability,

$$\mathbb{P}(U_{(1)} > X_1, \ldots, U_{(m)} > X_m) = \mathbb{P}(U_1 > X_1, \ldots, U_m > X_m)$$

holds for i.i.d. $[0, 1]$-valued random variables $X_1, \ldots, X_m$, we obtain

$$\begin{aligned} Q_g^* &= 1 + \sum_{m=1}^{\infty} \left[1 - \mathbb{E}\left(\frac{T(f_1)}{g(f_1)} \right) \Big/ \operatorname*{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)} \right]^m \\ &= \operatorname*{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)} \Big/ \mathbb{E}\left(\frac{T(f_1)}{g(f_1)} \right) = \operatorname*{esssup}_{f\in\mathbb{H}} \frac{\sup_{y\in K} f(y)}{g(f)} \Big/ \int_{\mathbb{H}} T(f_1)\, H(\mathrm{d}f_1). \end{aligned} \tag{4.25}$$

This finishes the proof since $c_T \in (0, \infty)$. □

Remark 4.4. In the proof of Proposition 4.3, the distribution of the random variable M_g^* defined in (4.20) is calculated. We see that M_g^* follows a geometric distribution with parameter $\int_{\mathbb{H}} T(h)H(\mathrm{d}h) / \operatorname{esssup}_{f\in\mathbb{H}} \sup_{y\in K}(f(y)/g(f))$. Therefore, a density g that minimizes $\operatorname{esssup}_{f\in\mathbb{H}} \sup_{y\in K}(f(y)/g(f))$ does not only minimize the expectation of M_g^*, but also other characteristics. For example, the probability $\mathbb{P}(M_g^* > m_0)$ for a given $m_0 \in \mathbb{N}$, or the quantile of M_g^* at a given probability level. However, this property does not necessarily hold for the stochastic number M_g in the actual stopping rule (4.10).

We carry on to find the density g that minimizes $\operatorname{esssup}_{f\in\mathbb{H}} \frac{\sup_{y\in K} f(y)}{g(f)}$. Instead of considering the supremum in the numerator, we deal with a broader class of functionals in the following proposition.

Proposition 4.5. *Let $L : \mathbb{H} \to (0,\infty)$ be measurable and $c_L := \int_{\mathbb{H}} L(f)H(\mathrm{d}f) < \infty$. Then,*

$$g_L(f) = c_L^{-1} L(f) \tag{4.26}$$

is an element of

$$\mathcal{G}^{(L)} = \arg\min_g \operatorname*{esssup}_{f\in\mathbb{H}} \frac{L(f)}{g(f)}.$$

Furthermore, for every $g \in \mathcal{G}^{(L)}$, Equation (4.26) holds for H-a.e. $f \in \mathbb{H}$.

Proof. First, by contradiction, we show that the inequality

$$\operatorname*{esssup}_{f\in H} \frac{L(f)}{g(f)} \geq c_L \tag{4.27}$$

holds for all g. So, assume that (4.27) does not hold for some g considered in Proposition 4.1. Then some $\varepsilon > 0$ and some density g with $\int g(f)\, H(\mathrm{d}f) = 1$ exist such that, for almost all $f \in \mathbb{H}$, we have $L(f)/g(f) \leq c_L - \varepsilon$. Hence,

$$c_L = \int_{\mathbb{H}} L(f)\, H(\mathrm{d}f) \leq (c_L - \varepsilon) \int_{\mathbb{H}} g(f) H(\mathrm{d}f) < c_L$$

which is a contradiction. Hence, (4.27) is proved. Note that the choice $g(f) = c_L^{-1} L(f)$ implies equality in (4.27). The first assertion follows.
For the proof of the second assertion, assume that there is some $g \in \mathcal{G}^{(L)}$ such that (4.26) does not hold for H-a.e. $f \in \mathbb{H}$. Then, as

$$\int_{\mathbb{H}} g(f)\, H(\mathrm{d}f) = 1 = \int_{\mathbb{H}} c_L^{-1} L(f) H(\mathrm{d}f),$$

there is some set $A \subset \mathbb{H}$ with $H(A) > 0$ such that, for all $f \in A$, $g(f) < c_L^{-1} L(f)$, but $g(f) > 0$ by (4.6). This yields $gH(A) > 0$ and, hence, $\operatorname{esssup}_{f\in\mathbb{H}} L(f)/g(f) > c_L$, which is a contradiction to $g \in \mathcal{G}^{(L)}$. □

Remark 4.6. Let $\mathcal{L}_p(H)$ be the space of p-integrable functionals with respect to H. Then $c_L = \|L\|_{\mathcal{L}_1(H)}$ and $\operatorname{esssup}_{f\in\mathbb{H}} \frac{L(f)}{g(f)} = \|L/g\|_{\mathcal{L}_\infty(H)}$. Proposition 4.5 states a special case of Hölder's inequality for all density functions g, $c_L \leq \|L/g\|_{\mathcal{L}_\infty(H)} \|g\|_{\mathcal{L}_1(H)} = \|L/g\|_{\mathcal{L}_\infty(H)}$ and equality holds if and only if g is proportional to L (H-a.e.).
Similarly, in importance sampling, the second moment $\mathbb{E}(\tilde{f}(X_1)/g(X_1))^2 = \|\tilde{f}/\sqrt{g}\|_2^2$ is intended to be small. Now, $c_g := \|\tilde{f}\|_1 \leq \|\tilde{f}/\sqrt{g}\|_2 \, \|\sqrt{g}\|_2 = \mathbb{E}(\tilde{f}(X_1)/g(X_1))^2$ and equality holds if and only if g is proportional to $\tilde{f}$.
Both c_L and c_g can simply be regarded as a factor that normalize L and $\tilde{f}$, respectively.

The results stated above enable us to describe the solution of the optimization problem (4.20). To be rigorous, we first give a necessary and sufficient condition for the solvability of the problem. Here an optimization problem

$$\arg\min_{x\in A} h(x), \quad h : A \to \mathbb{R} \cup \{\infty\},$$

is called *solvable* if $\inf_{x\in A} h(x) \in (-\infty,\infty)$ and there exists some $x_0 \in A$ such that $h(x_0) = \inf_{x\in A} h(x)$. Our key theorem is given as follows.

Theorem 4.7. *The optimization problem* (4.20) *is solvable if and only if*

$$c = \int_{\mathbb{H}} \sup_{y \in K} f(y) H(\mathrm{d}f) < \infty.$$

Assuming $c < \infty$, the solution is given as

$$g^*(f) := c^{-1} \sup_{y \in K} f(y), \quad f \in \mathbb{H}. \tag{4.28}$$

The solution is unique H-a.s.

Proof. If $c = \infty$, Equation (4.23) and Proposition 4.3 yield that (4.20) is not solvable. For $c < \infty$, the solution and its uniqueness follow directly from Propositions 4.3 and 4.5 with $L(f) = \sup_{y \in K} f(y)$. □

Remark 4.8. It is obvious that g^* is also the H-a.s. solution for the two examples of the modified optimization problem in (4.21) and (4.22) and these problems are solvable if and only if $c < \infty$. Consequently, the original optimization problem (4.12) is not solvable if $c = \infty$, because $Q_g^{(1)} \leq Q_g$ for any g.

We close this subsection with analyzing under which conditions c is finite, i.e. the modified optimization problem is solvable. For instance, if Z is sample-continuous or if K is finite, $c < \infty$ follows from a result by Resnick and Roy (1991) (see also de Haan and Ferreira, 2006, Theorem 9.6.1) who showed the equivalence of the first and third assertion in the following proposition. Our result is more general as it shows the equivalence to further conditions and $\sup_{y \in K} f(y)$ is replaced by a general max-linear functional L.

Proposition 4.9. *Assume the framework of Proposition 4.1. Furthermore, assume that the functional $L : \mathbb{H} \to (0, \infty)$ is measurable and max-linear. Then the following conditions are equivalent:*

1. $c_L := \int_{\mathbb{H}} L(f)\, H(\mathrm{d}f) < \infty$
2. $\mathbb{P}(L(Z) \leq a) > 0$ *for some* $a > 0$
3. $\mathbb{P}(L(Z) < \infty) = 1$ *(or, equivalently,* $\mathbb{P}(L(Z) < \infty) > 0$*).*

If there is some stochastic process W such that

$$Z =_d \max_{t \in \Pi_0} t W_t, \tag{4.29}$$

where Π_0 is a Poisson point process on $(0, \infty)$ with intensity $t^{-2}\,\mathrm{d}t$ and W_t, $t \in \Pi_0$, are independent copies of W, we get another equivalent condition:

4. $\mathbb{E}L(W) < \infty$.

Proof. The assertion follows from the following continued equality:

$$\begin{aligned} \exp\left(-\frac{c_L}{a}\right) &= \exp\left(-\int_{\mathbb{H}} \int_{a/L(f)}^{\infty} t^{-2}\,\mathrm{d}t\, H(\mathrm{d}f)\right) = \mathbb{P}\left(L(Z) \leq a\right) \\ &= \exp\left(-\mathbb{E}_W\left(\int_{a/L(W)}^{\infty} u^{-2}\,\mathrm{d}u\right)\right) = \exp\left(-a^{-1}\mathbb{E}L(W)\right). \end{aligned}$$

for any $a > 0$. The equivalence to the third assertion can be seen from the relation $\mathbb{P}(L(Z) < \infty) = \lim_{a \to \infty} \mathbb{P}(L(Z) \leq a)$. □

4.3.3 The Normalized Spectral Representation

Plugging in the solution g^* of (4.20) given by (4.28) into (4.7), we obtain $Z =_d \tilde{Z}$ with

$$\tilde{Z}(y) = \max_{t \in \Pi_0} t \frac{cF_t(y)}{\sup_{\tilde{y} \in K} F_t(\tilde{y})}, \quad y \in K, \tag{4.30}$$

where Π_0 is a Poisson point process on $(0, \infty)$ with intensity $t^{-2}\mathrm{d}t$ and F_t, $t \in \Pi_0$, are independent random processes with density $g^*(f)\, H(\mathrm{d}f)$. It can be verified that the transformed spectral functions $\{cF_t / \sup_{y \in K} F_t(y)\}$ are independent copies of a stochastic process F^* with

$$\sup_{y \in K} F^*(y) \equiv c. \tag{4.31}$$

We define such a representation as the normalized spectral representation as follows.

Definition 4.10. Let Z be a max-stable process on K satisfying

$$Z =_d \max_{t \in \Pi_0} t F_t^*. \tag{4.32}$$

Here, Π_0 is a Poisson point process on $(0, \infty)$ with intensity $t^{-2}\mathrm{d}t$ and F_t^*, $t \in \Pi_0$, are independent copies of a stochastic process F^* satisfying (4.31) for some $c \in (0, \infty)$. Then, the right-hand side of (4.32) is called normalized spectral representation of Z.

Theorem 4.7 implies that the normalized spectral representation exists if and only if $c < \infty$. The following proposition implies that the constant c and the finite-dimensional distributions of the processes F_t^* in the normalized spectral representation are uniquely determined by the process Z.

Proposition 4.11. *Let Z be a max-stable process with a normalized spectral representation. Furthermore, let $Z^K := \sup_{y \in K} Z(y)$. Then, we have*

1. $c = -\log \mathbb{P}\left(Z^K \leq 1\right)$
2. *For any $y_1, \ldots, y_n \in K$, $v_1, \ldots, v_n > 0$, it holds*

$$\mathbb{P}(F^*(y_i) \leq v_i,\, 1 \leq i \leq n) = \lim_{z \to \infty} \mathbb{P}\left(\frac{Z(y_i)}{Z^K} \leq \frac{v_i}{c},\, 1 \leq i \leq n \,\middle|\, Z^K > z \right). \tag{4.33}$$

Proof. The first part is a consequence of the proof of Proposition 4.9. In order to prove the second part, we provide an lower and an upper bound for the probability of the event

$$A = \{Z(y_i) \leq v_i Z^K / c,\ i = 1, \ldots, n,\ Z^K > z\}.$$

To this end, let $\tilde{\Pi} = \{(t, F_t^*) :\ t \in \Pi_0\}$ and note that $Z^K > z$ if and only if there is some $(u, w) \in \tilde{\Pi}$ such that $Z^K = uc = u \sup_{y \in K} w(y) > z$. Now, suppose that further $Z(y_i) \leq v_i Z^K / c$ holds for all $i = 1, \ldots, n$. Then, this point (u, w) necessarily satisfies $w(y_i) \leq v_i$ for all $i = 1, \ldots, n$. This yields an upper bound for $\mathbb{P}(A)$. If, on the other hand, $Z(y_i) > v_i Z^K / c$ for some $i = 1, \ldots, n$, then one point $(u, w) \in \tilde{\Pi}$ with $uc > z$ satisfies $w(y_i) > v_i$ for some $i = 1, \ldots, n$ or there is some point $(u, w) \in \tilde{\Pi}$ with $uc \leq z$, but $uw(y_i)/z > v_i/c$ for some $i = 1, \ldots, n$. Considering the complementary probabilities, we obtain a lower bound. Summing up, we have

$$\mathbb{P}\left(\left| \tilde{\Pi} \cap \left\{ (u, w) :\ u > \frac{z}{c} \right\} \right| > 0,\ \left| \tilde{\Pi} \cap \left\{ (u, w) :\ u > \frac{z}{c},\ 1 > \min_{1 \leq i \leq n} \frac{v_i}{w(y_i)} \right\} \right| = 0, \right.$$

$$\left|\tilde{\Pi} \cap \left\{(u,w):\, u \le \frac{z}{c},\, \frac{u}{z} > \min_{1\le i\le n} \frac{v_i}{cw(y_i)}\right\}\right| = 0\Bigg) \quad \le \quad \mathbb{P}(A)$$
$$\le \mathbb{P}\Bigg(\left|\tilde{\Pi} \cap \left\{(u,w):\, u > \frac{z}{c},\, \max_{1\le i\le n} \frac{w(y_i)}{v_i} \le 1\right\}\right| > 0\Bigg). \tag{4.34}$$

The lower bound in (4.34) equals

$$\begin{aligned}
&\left(1 - \exp\left(-\frac{c}{z}\mathbb{P}(F^*(y_i) \le v_i,\, 1 \le i \le n)\right)\right) \\
&\cdot \exp\left(-\frac{c}{z}\mathbb{P}\left(\max_{1\le i\le n} \frac{F^*(y_i)}{v_i} > 1\right)\right) \exp\left(-\mathbb{E}\int_{\frac{z}{c}\wedge \min_{1\le i\le n} \frac{zv_i}{cF^*(y_i)}}^{\frac{z}{c}} u^{-2}\mathrm{d}u\right) \\
=\ &\left(1 - \exp\left(-\frac{c}{z}\mathbb{P}(F^*(y_i) \le v_i,\, 1 \le i \le n)\right)\right) \\
&\cdot \exp\left(-\frac{c}{z}\mathbb{P}\left(\max_{1\le i\le n} \frac{F^*(y_i)}{v_i} > 1\right)\right) \exp\left(-\frac{c}{z}\mathbb{E}\left(\max_{1\le i\le n} \frac{F^*(y_i)}{v_i} - 1\right)_+\right),
\end{aligned}$$

while the upper bound equals $1 - \exp\left(-\frac{c}{z}\mathbb{P}(F^*(y_i) \le v_i,\, 1 \le i \le n)\right)$. By using that $\mathbb{P}(Z^K > z) = 1 - e^{-c/z}$ and taking the limit $z \to \infty$, inequality (4.34) yields (4.33). □

Proposition 4.11 implies that the solution of the optimization problem (4.20) is unique in two different aspects. First, as stated in Theorem 4.7, any solution $g \in \mathcal{G}^*$ satisfies $g = g^*$ H-a.s. Second, the finite-dimensional distributions of the normalized spectral functions $\{f/g^*(f)\}$ do not depend on the initial choice of the spectral functions. In particular, the normalized spectral representation is unique if $\mathbb{H}$ is the space of continuous functions on K equipped with the Borel σ-algebra, i.e. the product σ-algebra on $[0,\infty)^K$.
Summarizing the results in this section, we suggest to make use of the normalized spectral representation for exact and efficient simulation as it is the unique solution to the modified optimization problem provided that $c < \infty$. Analogously to Algorithm 4.1, the algorithm for simulation via the normalized spectral representation is given as follows.

Algorithm 4.2: Exact simulation of a max-stable process Z via the normalized spectral representation

Set $m = 0$ and $Z^{(0)}(y) = 0$ for all $y \in K$.
Simulate a standard exponentially distributed random variable E_1.
while $\operatorname{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{cf(y)}{\sup_{\tilde{y}\in K} f(\tilde{y})Z^{(m)}(y)} > \sum_{j=1}^{m+1} E_j$ **do**
  Update m by $m+1$.
  Sample F_m with density $c^{-1}\sup_{y\in K} f(y)H(\,\mathrm{d}f)$.
  Set $Z^{(m)}(y) = \max\left\{Z^{(m-1)}(y), \frac{c}{\sum_{j=1}^m E_j} \frac{F_m(y)}{\sup_{\tilde{y}\in K} F_m(\tilde{y})}\right\}$ for all $y \in K$.
  Simulate a standard exponentially distributed random variable E_{m+1}.
return $Z^{(m)}$

Remark 4.12. In practice, with the normalized spectral representation, the criterion in the while-loop in Algorithm 4.2 might be replaced by the weaker criterion

$$\frac{c}{\inf_{y\in K} Z^{(m)}(y)} > \sum_{j=1}^{m+1} E_j.$$

Although the use of the weaker stopping rule based on this criterion may increase the number of iterations, it simplifies the evaluation of the essential supremum.

The constant c plays the role analogous to the bound of the quotient of the genuine density function and the proposed density function in rejection sampling. In particular, c controls the speed of the algorithm, while (4.10) guarantees the exactness.

Finally, we investigate the sampling of F_m in Algorithm 4.2. The simulation of the stochastic processes $F_m = \{F_m(y),\, y \in K\}$, $m \in \mathbb{N}$, according to the transformed measure g^*H may not be always straightforward. For some processes, such as the moving maxima processes (cf. Section 6.4), the distribution of F_m can be calculated explicitly, which allows for efficient sampling. For many other processes, such as Brown–Resnick or extremal t processes, there is no direct way to simulate the normalized spectral functions. One solution is to use the following Metropolis–Hastings algorithm.

Algorithm 4.3: Metropolis–Hastings algorithm for the simulation of normalized spectral functions

Simulate $f^{(1)}$ according to the law H.
for $k = 1, \ldots, n_{MCMC} - 1$ **do**
  Sample f^{prop} from H and set

$$f^{(k+1)} = \begin{cases} f^{\text{prop}} & \text{with probability } \min\left\{\frac{\sup_{y\in K} f^{\text{prop}}(y)}{\sup_{y\in K} f^{(k)}(y)}, 1\right\}, \\ f^{(k)} & \text{with probability } \max\left\{1 - \frac{\sup_{y\in K} f^{\text{prop}}(y)}{\sup_{y\in K} f^{(k)}(y)}, 0\right\}. \end{cases}$$

return $f^{(n_{MCMC})}$

If H is a probability measure, Algorithm 4.3 generates a Markov chain of length n_{MCMC} whose stationary distribution is g^*H based on simulations from H.

4.4 Evaluating the Modified Optimization Problem

In this section, we discuss how close the modified optimization problem and the original problem are. Considering the two examples (4.21) and (4.22) of the modified problem, we have that $Q_g^{(1)} \le Q_g^{(2)}(y_0) \le Q_g$ for all g under some mild condition on $\mathbb{H}$ (see the proof of Proposition 4.13). Therefore, the modified optimization problem is in fact minimizing a lower bound of $g \mapsto Q_g$. This section proceeds in two steps. First, we will improve the lower bound and show that the normalized spectral representation also minimizes the improved lower bound function. Second, we give a formula for calculating Q_g. In particular, this formula allows the calculation of Q_{g^*}, that is, the expected number of steps when simulating Z using the normalized spectral representation and the *original* stopping rule (4.10).
We start giving bounds for the expected number of iterations in the simulation algorithm.

Proposition 4.13. *Assume that the there exists a countable subset $K_0 \subset K$ such that $\sup_{y\in K} f(y) = \sup_{y\in K_0} f(y)$ for H-a.e. $f \in \mathbb{H}$. Denoting by $g_0 \in \mathcal{G}$ a solution of the original optimization problem, we get for all $y_0 \in K$,*

$$1 = Q_{g^*}^{(1)} \le Q_{g^*}^{(2)}(y_0) \le Q_{g_0} \le Q_{g^*}.$$

Proof. First, the last part of the inequality, $Q_{g_0} \le Q_{g^*}$, holds automatically due to the optimality of g_0. Second, note that

$$\operatorname*{esssup}_{f\in\mathbb{H}} \frac{\sup_{y\in K} f(y)}{g(f)\sup_{\tilde{y}\in K} Z^{(m)}(\tilde{y})} \le \operatorname*{esssup}_{f\in\mathbb{H}} \frac{\sup_{y\in K} f(y)}{g(f) Z^{(m)}(y_0)}$$

and thus $Q_g^{(1)} \leq Q_g^{(2)}(y_0)$ for any g and any $y_0 \in K$. In particular, these inequalities hold for $g = g^*$. Therefore, we get that $Q_{g^*}^{(1)} \leq Q_{g^*}^{(2)}(y_0)$ for any $y_0 \in K$.
Third, we show that $Q_{g^*}^{(1)} = 1$. On the one hand, we have that

$$\sup_{y\in K} \frac{f(y)}{g^*(f)} = c \quad \text{for } g^*H\text{-a.e. } f \in \mathbb{H}. \tag{4.35}$$

On the other hand, with (4.8), we obtain

$$\sup_{y\in K} Z^{(m)}(y) = \sup_{y\in K} \max_{i\in\mathbb{N}} \frac{1}{\sum_{j=1}^{i} E_j} \frac{f_i(y)}{g^*(f_i)} = cE_1^{-1} \quad \text{for all } m \in \mathbb{N}. \tag{4.36}$$

Combining equations (4.35) and (4.36) yields

$$\begin{aligned} Q_{g^*}^* &= \mathbb{E} \min \left\{ m \in \mathbb{N} : \operatorname*{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g^*(f)} \leq \sup_{y\in K} Z^{(m)}(y) \sum_{j=1}^{m+1} E_j \right\} \\ &= \mathbb{E} \min \left\{ m \in \mathbb{N} : c \leq c \cdot \sum_{j=1}^{m+1} E_j/E_1 \right\} = 1. \end{aligned}$$

To complete the proof of the proposition, we show that $Q_g^{(2)}(y_0) \leq Q_g$ for all g. We first consider the case that $\operatorname{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)} = \infty$. Since $c < \infty$, by Proposition 4.9, we get that $\sup_{y\in K} Z^{(m)}(y) \leq \sup_{y\in K} Z(y) < \infty$ with probability one. Thus, by the definition of $Q_g^{(1)}$ in (4.21), we get that $Q_g^{(1)} = \infty$. Consequently, $Q_g \geq Q_g^{(1)} = \infty$. It is thus proved that $Q_g^{(2)} \leq Q_g$.
Next, consider the case $\operatorname{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)} < \infty$. Since there exists some countable set $K_0 \subset K$ such that $\sup_{y\in K} f(y) = \sup_{y\in K_0} f(y)$ for H-a.e. $f \in \mathbb{H}$, we have that

$$\operatorname*{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)} = \sup_{y\in K_0} \operatorname*{esssup}_{f\in\mathbb{H}} \frac{f(y)}{g(f)}.$$

Therefore, for every $\varepsilon > 0$, there exists some $y(\varepsilon) \in K_0$ such that

$$\frac{1}{1+\varepsilon} \operatorname*{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)Z^{(m)}(y(\varepsilon))} \leq \operatorname*{esssup}_{f\in\mathbb{H}} \frac{f(y(\varepsilon))}{g(f)Z^{(m)}(y(\varepsilon))} \leq \operatorname*{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)Z^{(m)}(y)}. \tag{4.37}$$

Analogously to the proof of Proposition 4.3, we have that

$$\begin{aligned} &\mathbb{E} \min \left\{ m \in \mathbb{N} : \frac{1}{1+\varepsilon} \operatorname*{esssup}_{f\in\mathbb{H}} \frac{\sup_{y\in K} f(y)}{g(f)Z^{(m)}(y(\varepsilon))} \leq \sum_{j=1}^{m+1} E_j \right\} \\ &= 1 + \sum_{m=1}^{\infty} \left[1 - \mathbb{E}\left(1 \wedge \left(\frac{(1+\varepsilon)\cdot f_1(y(\varepsilon))}{g(f_1)} \Big/ \operatorname*{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)} \right) \right) \right]^m \\ &= \left[\mathbb{E}\left(1 \wedge \left(\frac{(1+\varepsilon)\cdot f_1(y(\varepsilon))}{g(f_1)} \Big/ \operatorname*{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)} \right) \right) \right]^{-1} \geq \frac{\operatorname{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)}}{1+\varepsilon}, \end{aligned}$$

where the last step follows from $\int_{\mathbb{H}} f(y(\varepsilon))\, H(\mathrm{d}f) = 1$. From (4.25), it is straightforward to verify that, for any $y_0 \in K$, $Q_g^{(2)}(y_0) = \operatorname{esssup}_{f\in\mathbb{H}} \sup_{y\in K} \frac{f(y)}{g(f)}$, which is independent from the choice of $y_0 \in K$. Hence, by using (4.37) and taking $\varepsilon \to 0$, we obtain $Q_g^{(2)}(y_0) \leq Q_g$ for any g, which implies that $Q_{g^*}^{(2)}(y_0) \leq Q_{g_0}$, due to the optimality of g^*. □

As Proposition 4.13 shows, approximating the optimal number of steps in the original problem (4.12) by the solution for the two example problems (4.21) and (4.22) might be quite vague. In particular, the achieved minimum of $Q_g^{(1)}$ always equals to 1. In other words, some spectral functions that in fact contribute to Z are not taken into account in the calculation of $Q_g^{(1)}$ or $Q_g^{(2)}$.
To overcome the aforementioned pitfall, we proceed with a theoretical investigation to improve the lower bound of Q_g for a given density g. The idea is to evaluate Q_g by separating two types of spectral functions. First, we consider those functions that contribute to the max-stable process under the transformed spectral representation. Second, we deal with those functions that do not contribute, but are taken into consideration because of the stopping rule. To this end, we replace the processes $Z^{(m)}$ occurring in the construction of Z by the final process $Z^{(\infty)}$ given by (4.9). Theoretically, this will not affect the construction of the process because once we stop after m steps according to the stopping rule (4.10), we have $Z^{(m)} = Z^{(\infty)}$. However, the modified inequality is not a stopping rule. This is why we call the analysis below a "theoretical investigation".
We further assume that $\mathbb{H}$ corresponds to the space $C_+(K)$ of nonnegative continuous functions on K endowed with the Borel σ-algebra $\mathcal{H}$. This ensures the finiteness of c, and thus the solvability of all the optimization problems. Note that all the results of this section hold true for more general spaces $\mathbb{H}$, see Oesting et al. (2013).
In order to separately consider these two types of functions, we adopt the concepts of *K-extremal* and *K-subextremal points* introduced by Dombry and Éyi-Minko (2013) and Dombry and Éyi-Minko (2012) as follows.

Definition 4.14. Let Φ be some Poisson point process on $(0,\infty) \times C_+(K)$ with intensity measure $u^{-2}\,\mathrm{d}u \times \nu(\mathrm{d}h)$ where ν is a locally finite measure on $C_+(K)$. We call $(t^*, h^*) \in \Phi$ a K-extremal point and write $(t^*, h^*) \in \Phi_K^+$ if and only if

$$t^* h^*(y) = \max_{(t,h)\in\Phi} th(y) \quad \text{for some } y \in K.$$

Otherwise, i.e. if $t^*h^*(y) < \max_{(t,h)\in\Phi} th(y)$ for all $y \in K$, the pair $(t^*, h^*) \in \Phi$ is called a K-subextremal point and we write $(t^*, h^*) \in \Phi_K^-$.

In contrast to Dombry and Éyi-Minko (2013), we are interested in tuples (t, h) instead of the product th. Therefore, we generalize a result given in Dombry and Éyi-Minko (2013) and show that the random sets Φ_K^+ and Φ_K^- are point processes on $(0,\infty) \times C_+(K)$, i.e. $\Phi_K^+(S)$ and $\Phi_K^-(S)$ are random variables for any bounded set $S \in \mathcal{B} \times C_+(K)$; see the following proposition. The proof runs analogously using the fact that the mapping $\phi : (0,\infty) \times C_+(K) \to C_+(K),\ (t,h) \mapsto th(\cdot)$ is measurable and is therefore omitted.

Proposition 4.15. *Φ_K^+ and Φ_K^- are point processes on $(0,\infty) \times C_+(K)$.*

To apply the theory of extremal and subextremal points in the construction of the process $Z^{(\infty)}$, we define the Poisson point process

$$\Phi = \left\{ \left(t, \frac{f}{g(f)} \right) :\ (t,f) \in \tilde{\Pi} \right\}.$$

Similar to the proof of Lemma 3.2 in Dombry and Éyi-Minko (2012), the following lemma characterizes the points of the point process Φ_K^-.

Lemma 4.16. *Conditional on $Z^{(\infty)}$, the point process Φ_K^- is a Poisson point process on $(0,\infty) \times C_+(K)$ with intensity measure*

$$\frac{\mathrm{d}\tilde{\Lambda}^-}{\mathrm{d}t \times \mathrm{d}H}(t,h) = t^{-2} g(h) \cdot \mathbf{1}_{th(\cdot)/g(h) < Z^{(\infty)}(\cdot)}.$$

In addition, we calculate the expected number of points in the point process Φ_K^+ in the following lemma.

Lemma 4.17. *We have*

$$\mathbb{E}|\Phi_K^+| = \mathbb{E}_Z\left(\int_{C_+(K)} \sup_{y\in K} \frac{f(y)}{Z(y)}\, H(\mathrm{d}f)\right)$$

which does not depend on the choice of g.

Proof. Let $B = [t_0,\infty)\times C_+(K)$ with $t_0 > 0$. Then, $\Phi\cap B$ is finite and we have $\mathbb{E}\left|\Phi_K^+\cap B\right| = \mathbb{E}\left|\Phi\cap B\right| - \mathbb{E}\left|\Phi_K^-\cap B\right|$. Conditioning on $Z^{(\infty)}$, Lemma 4.16 yields

$$\begin{aligned}\mathbb{E}\left|\Phi_K^+\cap B\right| = &\int_{C_+(K)}\int_0^\infty t^{-2}\mathbf{1}_{\{t\ge t_0\}}\,\mathrm{d}t\, g(f)H(\mathrm{d}f)\\ &-\mathbb{E}_{Z^{(\infty)}}\left(\int_{C_+(K)}\int_0^\infty t^{-2}\mathbf{1}_{\{t\ge t_0\}}\mathbf{1}_{\left\{\frac{1}{t}>\sup_{y\in K}\frac{f(y)}{g(f)Z^{(\infty)}(y)}\right\}}\,\mathrm{d}t\, g(f)H(\mathrm{d}f)\right)\\ &=\mathbb{E}_Z\left(\int_{C_+(K)}\int_0^\infty t^{-2}\mathbf{1}_{\{t\ge t_0\}}\mathbf{1}_{\left\{\frac{1}{t}\le\sup_{y\in K}\frac{f(y)}{g(f)Z(y)}\right\}}\,\mathrm{d}t\, g(f)H(\mathrm{d}f)\right).\end{aligned}$$

Considering a monotone sequence $t_{0,n}\searrow 0$ as $n\to\infty$, the monotone convergence theorem yields

$$\begin{aligned}\mathbb{E}|\Phi_K^+| &= \mathbb{E}_Z\left(\int_{C_+(K)}\int_0^\infty t^{-2}\mathbf{1}_{\left\{t>1/\sup_{y\in K}\frac{f(y)}{g(f)Z(y)}\right\}}\,\mathrm{d}t\, g(f)\,H(\mathrm{d}f)\right)\\ &= \mathbb{E}_Z\left(\int_{C_+(K)}\sup_{y\in K}\frac{f(y)}{Z(y)}\,H(\mathrm{d}f)\right),\end{aligned}$$

which completes the proof. □

The first type of spectral functions that contribute to M_g correspond to the extremal points $(t^*,h^*)\in\Phi_K^+$. Thus, we can rewrite (4.12) as

$$Q_g = \mathbb{E}|\Phi_K^+| + \mathbb{E}\left(\left|\left\{(t,h)\in\Phi_K^-:\ \operatorname*{esssup}_{f\in C_+(K)}\sup_{y\in K}\frac{f(y)}{g(f)Z^{(\infty)}(y)}>\frac{1}{t}\right\}\right|\right). \tag{4.38}$$

The second term in the right hand side of (4.38) corresponds to the number of the second type of spectral functions: they do not contribute to the max-stable process but are counted due to the stopping criterion (4.10). Notice that the component $\mathbb{E}|\Phi_K^+|$ is independent of the choice of g. We thus modify the optimization problems by maintaining this component, while refining the second component in an analogous way as the modification in (4.20). This results in a refined version of the modified optimization problem (4.20) as

$$\tilde{Q}_g^* = \mathbb{E}|\Phi_K^+| + \mathbb{E}\left(\left|\left\{(t,h)\in\Phi_K^-:\ \operatorname*{esssup}_{f\in C_+(K)}\frac{\sup_{y\in K}f(y)}{g(f)T(Z^{(\infty)})}>\frac{1}{t}\right\}\right|\right). \tag{4.39}$$

The following proposition relates the minimizer of (4.39) to the solution of our previously modified optimization problem, $g^*\in\mathcal{G}^*$. In addition, it provides a formula for Q_g for any given g based on the results in Lemma 4.16 and 4.17.

Proposition 4.18. *1. For any g, we have*

$$Q_g = \mathbb{E}_Z\left(\operatorname*{esssup}_{f\in C_+(K)}\sup_{y\in K}\frac{f(y)}{g(f)Z(y)}\right). \tag{4.40}$$

2. For any max-linear function T, it holds

$$\arg\min_g \tilde{Q}_g^* \supset \arg\min_g \operatorname*{esssup}_{f\in C_+(K)} \frac{\sup_{y\in K} f(y)}{g(f)} = \mathcal{G}^*,$$

where $\tilde{Q}_g^*$ is as in (4.39) and $\mathcal{G}^* = \arg\min_g Q_g^*$.

Proof. Let $Z^{(\infty)}$ be given by (4.9). By Lemma 4.17, we obtain

$$\begin{aligned} Q_g = \mathbb{E}_Z &\int_{C_+(K)} \sup_{y\in K} \frac{f(y)}{Z(y)} H(\mathrm{d}f) \\ &+ \mathbb{E}\left(\left|\left\{(t,h)\in\Phi_K^- : \operatorname*{esssup}_{f\in C_+(K)} \sup_{y\in K} \frac{f(y)}{g(f)Z^{(\infty)}(y)} > t^{-1}\right\}\right|\right). \end{aligned}$$

Conditioning on $Z^{(\infty)}$, Lemma 4.16 yields

$$\begin{aligned} &\mathbb{E}\left(\left|\left\{(t,h)\in\Phi_K^- : \operatorname*{esssup}_{f\in C_+(K)} \sup_{y\in K} \frac{f(y)}{g(f)Z^{(\infty)}(y)} > t^{-1}\right\}\right|\right) \\ &= \mathbb{E}_Z\left(\int_{C_+(K)}\int_0^\infty t^{-2}\mathbf{1}_{t>1/\operatorname*{esssup}_{h\in C_+(K)}\sup_{y\in K}\frac{h(y)}{g(h)Z(y)}}\mathbf{1}_{t<1/\sup_{y\in K}\frac{f(y)}{g(f)Z(y)}}\,\mathrm{d}t\, g(f)H(\mathrm{d}f)\right) \\ &= \mathbb{E}_Z\left(\int_{C_+(K)}\left\{\operatorname*{esssup}_{h\in C_+(K)}\sup_{y\in K}\frac{h(y)g(f)}{g(h)Z(y)} - \sup_{y\in K}\frac{f(y)}{Z(y)}\right\}_+ H(\mathrm{d}f)\right) \\ &= \mathbb{E}_Z\left[\operatorname*{esssup}_{h\in C_+(K)}\sup_{y\in K}\frac{h(y)}{g(h)Z(y)}\right] - \mathbb{E}_Z\int_{C_+(K)}\sup_{y\in K}\frac{f(y)}{Z(y)}\,H(\mathrm{d}f). \end{aligned}$$

In the last step we used the fact that

$$\operatorname*{esssup}_{h\in C_+(K)}\sup_{y\in K}\frac{h(y)g(f)}{g(h)Z(y)} - \sup_{y\in K}\frac{f(y)}{Z(y)} \geq 0$$

for H-a.e. $f\in C_+(K)$. The first assertion follows.
Analogously to the first part, we get that

$$\begin{aligned} \tilde{Q}_g^* = \mathbb{E}_Z &\int_{C_+(K)} \sup_{y\in K} \frac{f(y)}{Z(y)} H(\mathrm{d}f) \\ &+ \mathbb{E}\left(\left|\left\{(t,h)\in\Phi_K^- : \operatorname*{esssup}_{f\in C_+(K)} \frac{\sup_{y\in K} f(y)}{g(f)T(Z^{(\infty)})} > t^{-1}\right\}\right|\right) \end{aligned}$$

and

$$\begin{aligned} &\mathbb{E}\left(\left|\left\{(t,h)\in\Phi_K^- : \operatorname*{esssup}_{f\in C_+(K)} \frac{\sup_{y\in K} f(y)}{g(f)T(Z^{(\infty)})} > t^{-1}\right\}\right|\right) \\ &= \mathbb{E}_Z\left(\int_{C_+(K)}\int_0^\infty t^{-2}\mathbf{1}_{t>1/\operatorname*{esssup}_{h\in C_+(K)}\sup_{y\in K}\frac{h(y)}{g(h)T(Z)}}\mathbf{1}_{t<1/\sup_{y\in K}\frac{f(y)}{g(f)Z(y)}}\,\mathrm{d}t\, g(f)H(\mathrm{d}f)\right) \\ &= \mathbb{E}_Z\left(\int_{C_+(K)}\left\{\operatorname*{esssup}_{h\in C_+(K)}\frac{\sup_{y\in K}h(y)}{g(h)T(Z)}g(f) - \sup_{y\in K}\frac{f(y)}{Z(y)}\right\}_+ H(\mathrm{d}f)\right) \\ &\geq \mathbb{E}_Z\left(\int_{C_+(K)}\left\{\frac{\sup_{y\in K}f(y)}{T(Z)} - \sup_{y\in K}\frac{f(y)}{Z(y)}\right\}_+ H(\mathrm{d}f)\right). \end{aligned} \tag{4.41}$$

Now, let $g\in\mathcal{G}^* = \arg\min_g \operatorname{esssup}_{f\in C_+(K)}(\sup_{y\in K} f(y)/g(f))$. Then, by Theorem 4.5, we have that $g(f) = c^{-1}\sup_{y\in K} f(y)$ for all $f\in C_+(K)$, H-a.e. Thus, we get equality in Equation (4.41) and hence $\mathcal{G}^*\subset\arg\min_g \tilde{Q}_g^*$. □

Proposition 4.18 leads to two implications in applications. Firstly, it facilitates the numerical calculation of $Q_g = \mathbb{E}(M_g)$ by simulation. While Equation (4.40) is difficult to be evaluated exactly in many cases, it may be used to obtain bounds for $\mathbb{E}M_g$ such as

$$\operatorname*{esssup}_{f\in C_+(K)} \sup_{y\in K} \frac{f(y)}{g(f)} \le Q_g = \mathbb{E}M_g \le \operatorname*{esssup}_{f\in C_+(K)} \sup_{y\in K} \frac{f(y)}{g(f)} \mathbb{E}\Big[\Big(\inf_{y\in K} Z(y)\Big)^{-1}\Big]. \tag{4.42}$$

This confirms the finding of Proposition 4.2 that $\operatorname{esssup}_{f\in C_+(K)} \sup_{y\in K} \frac{f(y)}{g(f)} < \infty$, is a necessary condition for $\mathbb{E}M_g$ to be finite.
In particular, the result can be applied to analyze Q_{g^*}. With $\frac{f(y)}{g^*(f)} \le c$, we obtain

$$Q_{g^*} \le c \cdot \mathbb{E}\big[\big(\inf_{y\in K} Z(y)\big)^{-1}\big]. \tag{4.43}$$

This yields that $Q_{g^*} < \infty$ if Z is sample-continuous, as in this case we have $c < \infty$ and $\mathbb{E}\big[\big(\inf_{y\in K} Z(y)\big)^{-1}\big] < \infty$ holds. In other words, the expectation of the stochastic number M_{g*} based on the normalized representation is finite. Note that the upper bound in (4.43) is reached under some mild conditions, see Oesting et al. (2013). We will see that all examples considered in the present chapter meet these conditions.
Secondly, Proposition 4.18 implies that the minimal value for $\tilde{Q}_g^*$ can be achieved by any $g^* \in \mathcal{G}^*$. As further $Q_g^* \le \tilde{Q}_g^*$ by construction, we obtain the following corollary.

Corollary 4.19. *The optimization problem given in* (4.39) *is solvable if and only if the optimization problem* (4.20) *is solvable (cf. Theorem 4.7). With solvability, the normalized spectral representation is an optimal solution to* (4.39) *and* (4.20).

This corollary further confirms that using the normalized spectral representation may lead to an efficient and exact simulation, because the refined optimization problem (4.39) is closer to the original optimization problem than the modified problem (4.20).
Lastly, we improve the lower bounds for $Q_{\tilde{g}}$. To this end, refined versions of the examples in (4.21) and (4.22) are considered:

$$\tilde{Q}_g^{(1)} = \mathbb{E}|\Phi_K^+| + \mathbb{E}\left(\left|\left\{(t,h) \in \Phi_K^- : \operatorname*{esssup}_{f\in C_+(K)} \frac{\sup_{y\in K} f(y)}{g(f)\sup_{\tilde{y}\in K} Z^{(\infty)}(\tilde{y})} > \frac{1}{t}\right\}\right|\right), \tag{4.44}$$

$$\tilde{Q}_g^{(2)}(y_0) = \mathbb{E}|\Phi_K^+| + \mathbb{E}\left(\left|\left\{(t,h) \in \Phi_K^- : \operatorname*{esssup}_{f\in C_+(K)} \frac{\sup_{y\in K} f(y)}{g(f) Z^{(\infty)}(y_0)} > \frac{1}{t}\right\}\right|\right). \tag{4.45}$$

Proposition 4.20. *For any $g_0 \in \mathcal{G}$, we have*

$$1 \le \tilde{Q}_{g^*}^{(1)} \le \inf_{y_0\in K} \tilde{Q}_{g^*}^{(2)}(y_0) \le Q_{g_0} \le Q_{g^*},$$

where $g^* \in \mathcal{G}^*$ is given by (4.28).

By definition, $\tilde{Q}_g^{(1)} \ge Q_g^{(1)}$ and $\tilde{Q}_g^{(2)} \ge Q_g^{(2)}$ for all g. Hence, the results in Proposition 4.20 give improved lower bounds of Q_{g_0}. The proof is analogous to Proposition 4.13 and is thus omitted.

4.5 Example: Moving Maxima Processes

In this section, we discuss the normalized spectral representation for the class of moving maxima processes which can be simulated via an algorithm of Schlather (2002). In this

case, the normalized spectral functions can be calculated explicitly and are convenient to handle, which allows a rather general implementation of the simulation procedure for processes on a grid, see the `R` package `RandomFields` (R Core Team, 2018; Schlather et al., 2019). The procedure is further compared to Schlather's (2002) algorithm both from a theoretical point of view and in a simulation study in Section 4.6.
Here, we focus on moving maxima processes on a compact set $K \subset \mathbb{R}^d$, i.e. processes of the form

$$Z(y) =_d \max_{(t,x)\in\Pi_{M2}} t h(y-x), \quad y \in K, \tag{4.46}$$

where Π_{M2} is a Poisson point process on $(0,\infty)\times\mathbb{R}^d$ with intensity $t^{-2}\,\mathrm{d}t\,\Lambda(\mathrm{d}x)$, Λ denotes the Lebesgue measure on $\mathbb{R}^d$ and $h : \mathbb{R}^d \to [0,\infty)$ is a so-called shape function satisfying $\int_{\mathbb{R}^d} h(x)\,\mathrm{d}x = 1$. Thus, Z has a spectral representation of form (4.1) with

$$H(A) = \lambda\left(\left\{x \in \mathbb{R}^d :\ h(\cdot - x) \in A\right\}\right), \quad A \in \mathbb{H}.$$

In the following, we will explicitly calculate the distribution of the normalized spectral functions for moving maxima processes. First, we note that, due to the specific structure of a moving maxima process Z defined in (4.46), its normalized spectral representation can be written as

$$Z(y) =_d \max_{t\in\Pi_0} ct\frac{h(y-X_t)}{\tilde{h}(X_t)}, \quad y \in K,$$

where Π_0 is a Poisson point process on $(0,\infty)$ with intensity measure $t^{-2}\,\mathrm{d}t$ and X_t, $t \in \Pi_0$, are independent random vectors with Lebesgue density $c^{-1}\tilde{h}(x)\,\mathrm{d}x$, $\tilde{h}(x) = \sup_{y\in K} h(y-x)$ and $c = \int_{\mathbb{R}^d} \tilde{h}(x)\,\mathrm{d}x$. Thus, both the function $\tilde{h}$ and the constant c are crucial for the simulation of the normalized spectral functions. Further, c also occurs in the stopping rule and thus influences the number of spectral functions considered in Algorithm 4.2. If h is continuous or K is discrete, the upper bound (4.43) is reached, i.e.

$$Q_{g^*} = \mathbb{E}M_{g^*} = c \cdot \mathbb{E}\left[\left(\inf_{y\in K} Z(y)\right)^{-1}\right].$$

In general, the term $\mathbb{E}[(\inf_{y\in K} Z(y))^{-1}]$ cannot be calculated analytically, but needs to be estimated via simulations. For the implementation of Algorithm 4.2, however, only the constant c and the function $\tilde{h}$ are needed, both of them depending on the shape function h and the geometry of the set K.
In the following, we will calculate c and $\tilde{h}$ under different assumptions on the domain K. We restrict ourselves to the case where the shape function h is radial symmetric and non-increasing, i.e. $h(x) = f_0(\|x\|)$ for a non-increasing function $f_0 : [0,\infty) \to [0,\infty)$. Then, in general,

$$\tilde{h}(x) := \sup_{y\in K} f_0(\|y-x\|) = f_0\left(\min_{y\in K}\|y-x\|\right).$$

First, consider the case that K is a d-dimensional ball $b(0,R)$ centered at the origin with radius R, i.e. $K = \{x \in \mathbb{R}^d : \|x\| \le R\}$. Then,

$$\begin{aligned} \tilde{h}(x) &= f_0(0)\mathbf{1}_{\{\|x\|\le R\}} + \mathbf{1}_{\{\|x\|>R\}} f_0(\|x\| - R), \quad x \in \mathbb{R}^d, \\ \text{and} \qquad c &= \mathrm{vol}(b(0,1)) \cdot \left[f_0(0)R^d + d\int_0^\infty (\tilde{r}+R)^{d-1} f_0(\tilde{r})\,\mathrm{d}\tilde{r}\right] < \infty \end{aligned} \tag{4.47}$$

where vol denotes the d-dimensional volume.

Second, consider the case that K is a d-dimensional cube, i.e. the case that $K = [-R, R]^d$ for some $R > 0$. Then, we get

$$\tilde{h}\left((x_1, \ldots, x_d)\right) = f_0\left(\|((|x_1| - R) \vee 0, \ldots, (|x_d| - R) \vee 0)\|\right). \tag{4.48}$$

We consider the subcases $d = 1$ and $d = 2$ to derive explicit formulae. If $d = 1$, then K satisfies $K = [-R, R] = b(0, R)$, and, according to the formulae above, we get that $\tilde{h}(x) = \mathbf{1}_{\{|x| \leq R\}} f_0(0) + \mathbf{1}_{\{|x| > R\}} f_0(|x| - R)$ and thus,

$$c = \int_{\mathbb{R}} \tilde{h}(x)\,\mathrm{d}x = 2Rf_0(0) + \int_{\{|x|>0\}} f_0(|x|)\,\mathrm{d}x = 2Rf_0(0) + 1.$$

If $d = 2$, we obtain

$$\begin{aligned}\tilde{h}(x) = \mathbf{1}_{\{|x_1| \vee |x_2| \leq R\}} f_0(0) + 2 \cdot \mathbf{1}_{\{|x_1| \wedge |x_2| \leq R, |x_1| \vee |x_2| > R\}} f_0((|x_1| \wedge |x_2|) - R) \\ + \mathbf{1}_{\{|x_1| \wedge |x_2| > R\}} f_0\left(\|(|x_1| - R, |x_2| - R)\|\right).\end{aligned}$$

Thus,

$$c = (2R)^2 \cdot f_0(0) + 4R \cdot \int_{\mathbb{R}} f_0(|x|)\,\mathrm{d}x + \int_{\mathbb{R}^2} f_0(\|x\|)\,\mathrm{d}x = 4R^2 f(0) + 4R \int_{\mathbb{R}} f_0(|x|)\,\mathrm{d}x + 1.$$

Next, we further specify explicit examples on the function f_0, under which the constant c can be further calculated.

Example 4.21. *1. Indicator function*
We consider the case that the shape function is the indicator function of a ball $b(0, r)$ with radius $r > 0$ centered at the origin, i.e. $f_0(\|x\|) = \mathbf{1}_{\{\|x\| \leq r\}}$. In this case we have $\tilde{h}(x) = \mathbf{1}_{\{K \oplus b(0,r)\}}(x)$ and $c = \mathrm{vol}(K \oplus b(0, r))$ where $\oplus$ denotes morphological dilation. Here, all the finite approximations derived from the normalized spectral representation coincide with the corresponding approximations resulting from the algorithm proposed by Schlather (2002). See Section 4.6 for details on this algorithm.

2. Smith model
As a second example, we consider the Gaussian extreme value process (Smith, 1990) where f_0 is a Gaussian density function. Here, for simplicity, we assume the shape function is the density of a multivariate normal random vector $Y \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathrm{Id})$ with $\sigma > 0$. Thus, it is a radial symmetric monotone function. Let $K = [-R, R]^d$ for some $R > 0$. Then, by the considerations above, we get that $\tilde{h}$ is of type (4.48) and for $d = 1, 2$, we obtain

$$c = \begin{cases} \sqrt{\frac{2}{\pi}} \frac{R}{\sigma} + 1, & d = 1 \\ \frac{2}{\pi} \left(\frac{R}{\sigma}\right)^2 + 2\sqrt{\frac{2}{\pi}} \frac{R}{\sigma} + 1, & d = 2. \end{cases}$$

Remark 4.22. Note that the results can easily be generalized to the case of mixed moving maxima processes (Schlather, 2002; Stoev and Taqqu, 2005), i.e. the case where the deterministic shape function h is replaced by independent copies of a random function. See Oesting et al. (2013) for details.

4.6 Simulation: Comparison to Other Algorithms

In this section, we investigate the number of spectral functions needed in simulating a max-stable process, considering the suggested normalized spectral representation as well as other algorithms. In Subsection 4.6.1, we compare it with the analogous number in Schlather's (2002) algorithm for moving maxima processes. Then, we compare simulations via the normalized spectral representation to the recent algorithms devised by Dieker and Mikosch (2015) and Dombry et al. (2016a) focusing on Brown–Resnick processes in Subsection 4.6.2. Besides the number of spectral functions, we further consider the actual computational costs accounting for the fact that simulation of the normalized spectral functions is not straightforward in this case.

4.6.1 Comparison to the Algorithm Proposed in Schlather (2002)

Let Z be a moving maxima process on $K \subset \mathbb{R}^d$ as defined in (4.46). Assume that the shape function h is bounded and has compact support, i.e. $h(x) < C$ for all $x \in \mathbb{R}^d$ for some $C > 0$ and $\operatorname{supp}(h) \subset b(0,r)\})$ for some $r > 0$. Schlather's (2002) algorithm considers the following equivalent representation

$$Z(y) =_d |K \oplus b(0,r)| \cdot \max_{1 \le k \le M} \frac{F_k(y - U_k)}{\sum_{i=1}^k E_i}, \quad y \in K,$$

where E_i are standard exponentially distributed random variables, F_i follow the law π, U_i are uniformly distributed on $K \oplus b(0,r)$ and all these random variables are independent. Following Schlather's (2002) algorithm, the number of simulated spectral functions is then a random number defined as

$$M = \min \left\{ m \in \mathbb{N} : \ \frac{C}{\sum_{i=1}^{m+1} E_i} \le \inf_{x \in K} \max_{1 \le k \le m} \frac{F_k(x - U_k)}{\sum_{i=1}^k E_i} \right\}.$$

Here, analogously to Proposition 4.18, the following result can be shown.

Proposition 4.23. *The expectation of M, defined as above, equals*

$$\mathbb{E}M = \mathbb{E}\left(\frac{|K \oplus b(0,r)| \cdot C}{\inf_{y \in K} Z(y)} \right).$$

Thus, the ratio between the expected numbers of spectral functions considered by the normalized spectral representation and by Schlather's (2002) algorithm is

$$\frac{Q^*_{g^*}}{\mathbb{E}M} = \frac{\mathbb{E}M_{g^*}}{\mathbb{E}M} = \frac{c}{|K \oplus b(0,r)| \cdot C}.$$

If the shape function h is bounded, but not compactly supported, the max-stable process Z can be approximated using a shape function which is cut off outside a compact set J, i.e. $h_{\text{cut}}(x) = h(x) \cdot \mathbf{1}_{\{x \in J\}}$. Let $\tilde{U}_k \sim_{i.i.d.} \text{Unif}(K \oplus \check{J})$ where $\check{J} = \{-x : \ x \in J\}$, and $Z_J(\cdot)$ be given by

$$Z_J(y) = |K \oplus \check{J}| \cdot \max_{n \in \mathbb{N}} \frac{h_{\text{cut}}(y - \tilde{U}_n)}{\sum_{k=1}^n E_k}, \quad y \in K,$$

Then, the number M_J of shape functions that need to be considered is finite a.s., and, by Proposition 4.23, its expectation equals $\mathbb{E}M_J = \mathbb{E}\left((\inf_{y \in K} Z_J(y))^{-1} \cdot |K \oplus \check{J}| \cdot C \right)$. Hence,

in the approximative case, the ratio of expected numbers of spectral functions considered by the two algorithms, can be written as the product

$$\frac{Q_{g^*}}{\mathbb{E}M_J} = \frac{\mathbb{E}M_{g^*}}{\mathbb{E}M_J} = A_{K,J} \cdot P_{K,J}, \tag{4.49}$$

$$\text{where } A_{K,J} = \frac{c}{|K \oplus \check{J}| \cdot C} \quad \text{and} \quad P_{K,J} = \frac{\mathbb{E}\left(\sup_{y\in K} Z(y)^{-1}\right)}{\mathbb{E}\left(\sup_{y\in K} Z_J(y)^{-1}\right)}.$$

The first factor $A_{K,J}$ refers to the domain of the Poisson point process, and the second factor $P_{K,J}$ refers to the precision of the approximation by Schlather's (2002) algorithm. As $h_{\text{cut}}(\cdot) \leq h(\cdot)$, we have that $P_{K,J} \leq 1$ with $\lim_{J \nearrow \mathbb{R}^d} P_{K,J} = 1$. Thus, we obtain the upper bound $Q_{g^*}/\mathbb{E}M_J \leq A_{K,J}$ which can be calculated via the formulae for c obtained in Section 4.5. The factor $P_{K,J}$, however, cannot be calculated explicitly in general, but needs to be accessed via simulation.

In view of these theoretical observations, we perform a simulation study for Smith's (1990) model described in Example 4.21 on a rectangle $[-R, R]^d$ for $d = 1, 2$. For the simulation algorithm of Schlather (2002), we need an approximation as described above. Here, a natural choice for cutting off the shape function is $J = [-k\sigma, k\sigma]^d$ for some $k \in \mathbb{N}$. Then, by Example 4.21, the first factor $A_{R,k} := A_{[-R,R]^d,[-k\sigma,k\sigma]^d}$ in Equation (4.49) equals $(R/\sigma + \sqrt{\pi/2})^d/(R/\sigma + k)^d$, i.e. $A_{R,k} < 1$ if and only if $k > \sqrt{\frac{\pi}{2}}$. In order to access $P_{R,k} := P_{[-R,R]^d,[-k\sigma,k\sigma]^d}$ and the exact values of Q_{g^*} and $\mathbb{E}M_{[-k\sigma,k\sigma]^d}$, we choose $\sigma = 1$ and simulate Z and $Z_{[-k,k]^d}$ for $k = 2, 3$ on a grid $K = \{-R, -R + h, \ldots, R - h, R\}^d$, $d = 1, 2$. For simplicity, the normalized spectral representation is chosen as if K was the rectangle $[-R, R]^d$.
In the case $d = 1$, for $h = 0.1$ and each $R \in \{1, 2, 5, 10, 50, 100\}$ we simulate each process $N = 5000$ times. The values of Q_{g^*} and $\mathbb{E}M_{[-k,k]^d}$ are estimated via the corresponding empirical means denoted by $\hat{Q}_{g^*}$ and $\widehat{\mathbb{E}M_k}$ (the corresponding emopirical standard deviations are denoted by $\widehat{s(M_{gj})}$ and $\widehat{s(M_k)}$, respectively). We use a plug-in estimator $\hat{P}_{R,k}$ for $P_{R,k}$ that is based on the empirical means of $\sup_{y\in K} Z(y)^{-1}$ and $\sup_{y\in K} Z_{[-k,k]^d}(y)^{-1}$. The results of the simulation study are shown in Table 4.1.
First, we note that – in accordance to Equation (4.49) – Q_{g^*} is always smaller than $\mathbb{E}M_{[-k,k]^d}$. For instance, for $R = 1$, the number of considered shape functions is decreased by 29% ($k = 2$) and 43% ($k = 3$), respectively. Furthermore, we observe that $P_{R,k}$ seems to be almost constant in R, namely $P_{R,2} \approx 0.95$ and $P_{R,3} \approx 1$ which shows that the approximation of Z by $Z_{[-3,3]}$ is sufficiently good for $h = 0.1$. Thus, the behavior of $Q_{g^*}/\mathbb{E}M_{[-k,k]}$ is basically driven by $A_{R,k}$ which tends to 1 as $R \to \infty$. For large R, $Q_{g^*}/\mathbb{E}M_{[-k,k]} \approx P_{R,k}$. This explains the surprising fact that $\mathbb{E}M_{[-2,2]} > \mathbb{E}M_{[-3,3]}$ even though the approximation of Z by $Z_{[-2,2]}$ is less accurate than that by $Z_{[-3,3]}$.
Next, we perform the simulation for $d = 2$, $R \in \{1, 2, 5, 10\}$ and $h = 0.25$. Each process is simulated $N = 2500$ times. The results are shown in Table 4.2. In general, the results are similar to our observations for $d = 1$. However, for $d = 2$ the improvements compared to Schlather's (2002) algorithm are even more distinct. In the case $R = 1$, the number of considered spectral functions is decreased by 45% ($k = 2$) and 69% ($k = 3$), respectively.

4.6.2 Comparison to the Algorithms Proposed in Dieker and Mikosch (2015) and Dombry et al. (2016a)

Recently, Dieker and Mikosch (2015) proposed an exact algorithm for the simulation of Brown–Resnick processes on a finite set. In Dombry et al. (2016a), a generalization of this

Table 4.1: Results for simulations of Z (via the normalized spectral representation) compared to $Z_{[-k,k]}$ with $k = 2$ (top) and $k = 3$ (bottom), on the 1-dimensional grid $\{-R, -R+0.1, \ldots, R-0.1, R\}$ for different R. For each case, $A_{R,k}$ and the estimates for Q_{g^*}, $\mathbb{E}M_{[-k,k]}$ and $P_{R,k}$ as well as the corresponding sample standard deviations are displayed, based on $N = 5000$ simulations of each process.

R	$\widehat{\mathbb{E}M_2}$	$\hat{Q}_{g^*}$	$A_{R,2}$	$\hat{P}_{R,2}$	$\widehat{s(M_2)}$	$\widehat{s(M_{g^*})}$
1	4.38	3.12 (-29 %)	0.75	0.94	2.84	1.85
2	7.57	5.73 (-24 %)	0.81	0.94	4.10	3.06
5	18.83	15.82 (-16 %)	0.89	0.95	7.93	6.68
10	40.57	35.63 (-12 %)	0.94	0.94	14.03	12.55
50	257.61	239.75 (- 7 %)	0.99	0.94	62.63	60.96
100	579.11	540.44 (- 7 %)	0.99	0.94	124.66	117.76

R	$\widehat{\mathbb{E}M_3}$	$\hat{Q}_{g^*}$	$A_{R,3}$	$\hat{P}_{R,3}$	$\widehat{s(M_3)}$	$\widehat{s(M_{g^*})}$
1	5.46	3.12 (-43 %)	0.56	1.00	3.78	1.85
2	8.93	5.73 (-36 %)	0.65	0.98	5.13	3.06
5	19.98	15.82 (-21 %)	0.78	1.02	8.71	6.68
10	41.16	35.63 (-13 %)	0.87	1.00	14.82	12.55
50	247.35	239.75 (- 3 %)	0.97	1.00	60.14	60.96
100	550.70	540.44 (- 2 %)	0.98	1.00	114.36	117.76

Table 4.2: Results for simulations of Z (via the normalized spectral representation) compared to $Z_{[-k,k]^2}$ with $k = 2$ (top) and $k = 3$ (bottom), on the 2-dimensional grid $\{-R, -R+0.25, \ldots, R-0.25, R\}^2$ for different R. For each case, $A_{R,k}$ and the estimates for Q_{g^*}, $\mathbb{E}M_{[-k,k]^2}$ and $P_{R,k}$ as well as the corresponding sample standard deviations are displayed, based on $N = 2500$ simulations of each process.

R	$\widehat{\mathbb{E}M_2}$	$\hat{Q}_{g^*}$	$A_{R,2}$	$\hat{P}_{R,2}$	$\widehat{s(M_2)}$	$\widehat{s(M_{g^*})}$
1	14.86	8.14 (-45 %)	0.56	0.96	9.26	4.65
2	40.17	26.32 (-34 %)	0.66	1.00	18.07	10.96
5	189.83	150.89 (-21 %)	0.80	0.99	49.32	40.96
10	727.33	636.03 (-13 %)	0.88	0.99	146.88	127.36

R	$\widehat{\mathbb{E}M_3}$	$\hat{Q}_{g^*}$	$A_{R,3}$	$\hat{P}_{R,3}$	$\widehat{s(M_3)}$	$\widehat{s(M_{g^*})}$
1	26.37	8.14 (-69 %)	0.32	0.96	16.63	4.65
2	61.07	26.32 (-57 %)	0.42	1.03	26.99	10.96
5	247.10	150.89 (-39 %)	0.61	1.00	65.75	40.96
10	839.44	636.03 (-24 %)	0.75	1.01	168.31	127.36

algorithm and a novel exact simulation procedure based on extremal functions have been presented. Denote the number of stochastic processes to be simulated for obtaining an exact simulation on K via the (generalized version of the) Dieker–Mikosch algorithm and via the extremal functions approach by $M^{(DM)}$ and $M^{(EF)}$, respectively. Then, it can be shown that

$$\mathbb{E}M^{(DM)} = |K| \cdot \mathbb{E}\left[\left(\inf_{y\in K} Z(y)\right)^{-1}\right] \quad \text{and} \quad \mathbb{E}M^{(EF)} = |K| \tag{4.50}$$

(cf. Dombry et al., 2016a). As $\mathbb{E}[(\inf_{y\in K} Z(y))^{-1}] \geq 1$ for any max-stable process with standard Fréchet margins, we have $\mathbb{E}M^{(DM)} \geq \mathbb{E}M^{(EF)}$. Further, the underlying stochastic processes follow mixtures of the same laws. Consequently, between these two approaches, the one via extremal functions is always preferred in terms of the average computational costs of simulation.

In this subsection, we will compare the computational costs of these algorithms to that of simulation via the normalized spectral representation for simulating Brown–Resnick processes. This is what the original Dieker–Mikosch algorithm was designed for. Let Z be a Brown–Resnick process on $K \subset \mathbb{R}^d$ associated to a variogram γ (Kabluchko et al., 2009), i.e. a max-stable process with representation (4.1) with H being the probability measure of the stochastic process

$$W(y) = \exp\left(G(y) - \frac{1}{2}\operatorname{Var}(G(y))\right), \quad y \in \mathbb{R}^d.$$

Here G is a centered Gaussian process with stationary increments and variogram γ defined as $\gamma(h) = \mathbb{E}(W(y+h) - W(y))^2$. In the following, we consider a Brown–Resnick process associated to the variogram $\gamma(h) = \|h\|$ on the rectangle $[0,1]^2$. We simulate $N = 500$ realizations of the process on the grids $\{0, 0.05, \ldots, 0.95\}^2$ (400 points) and $\{0, 0.01, \ldots, 0.99\}^2$ (10000 points), respectively, via each of the three algorithms on a 2.90 GHz processor.
We start with comparing the average numbers of spectral processes to be simulated, i.e. $Q_{g^*} = \mathbb{E}M_{g^*}$, $\mathbb{E}M^{(DM)}$ and $\mathbb{E}M^{(EF)}$. Since the terms c in (4.43) and $\mathbb{E}[(\inf_{y\in K} Z(y))^{-1}]$ in (4.43) and (4.50) often cannot be calculated explicitly, the above expectations need to be estimated via simulations. Furthermore, in the Dieker–Mikosch algorithm and the extremal functions approach, we can directly simulate the underlying spectral functions as a single log-Gaussian process, whereas we have to use the Algorithm 4.3 to simulate the normalized spectral functions. For each process $F^{(m)}$, we simulate n_{MCMC} log-Gaussian processes. Consequently, a fair comparison across the three algorithms should be based on the average computational costs accounting for the number of simulated log-Gaussian processes. The costs for the simulation via the normalized spectral representation, the Dieker–Mikosch algorithm and the extremal functions approach are thus

$$\begin{aligned} C^{(NSR)} &= n_{MCMC} \cdot \mathbb{E}M_{g^*} = n_{MCMC} \cdot c \cdot \mathbb{E}\left[\left(\inf_{y\in K} Z(y)\right)^{-1}\right], \\ C^{(DM)} &= \mathbb{E}M^{(DM)} = |K| \cdot \mathbb{E}\left[\left(\inf_{y\in K} Z(y)\right)^{-1}\right] \\ \text{and} \quad C^{(EF)} &= \mathbb{E}M^{(EF)} = |K|. \end{aligned}$$

We use $n_{MCMC} = 100$, 500 and 1000 in our simulation. The normalization constant c is finally estimated by the average of the maxima of the n_{MCMC} log-Gaussian processes simulated in each Markov chain.

Finally, we consider the exactness of the simulations. Notice that the Dieker–Mikosch algorithm and the extremal functions approach yield exact realizations of the Brown–Resnick process. This is true for the normalized spectral representation approach only if the Markov chains have converged. Therefore, we need to evaluate the quality of the simulations. As a measure of exactness, we first calculate the Kolmogorov–Smirnov distance between the standard Fréchet distribution Φ_1 and the empirical cumulative distribution function $F^y_{(500)}$ obtained from the $N = 500$ realizations $Z_1(y), \ldots, Z_N(y)$ at each location y. Then we calculate the average of these distances across all $y \in K$ i.e.

$$\overline{d_{KS}} = \frac{1}{|K|} \sum_{y \in K} \|F^y_{(500)} - \Phi_1\|_\infty.$$

Further, we consider the extremal coefficient (Smith, 1990; Schlather and Tawn, 2003), $\theta(y_1, y_2)$ defined by

$$\mathbb{P}(Z(y_1) \leq z, Z(y_2) \leq z) = \mathbb{P}(Z(y_1) \leq z)^{\theta(y_1,y_2)}, \quad z > 0,$$

as a measure of extremal dependence between $Z(y_1)$ and $Z(y_2)$. By definition, we have $\theta(y_1, y_2) \in [1, 2]$ with $\theta(y_1, y_2) = 1$ if $Z(y_1) = Z(y_2)$ a.s. and $\theta(y_1, y_2) = 2$ if $Z(y_1)$ and $Z(y_2)$ are independent. For a Brown-Resnick process Z, we have $\theta(y_1, y_2) = 2\Phi(\sqrt{\gamma(h)}/2)$ where Φ denotes the standard normal distribution function. We estimate the extremal coefficients making use of the relation

$$\theta(y_1, y_2) = \frac{1 + 2\nu^F(y_1, y_2)}{1 - 2\nu^F(y_1, y_2)}, \quad y_1, y_2 \in K, \tag{4.51}$$

between $\theta(y_1, y_2)$ and the F-madogram (Cooley et al., 2006)

$$\nu^F(y_1, y_2) = \frac{1}{2}\mathbb{E}\left|F^{y_1}(Z(y_1)) - F^{y_1}(Z(y_1))\right|, \qquad y_1, y_2 \in K,$$

with $F^{(y)}$ being the cumulative distribution function of $Z(y)$, $y \in K$. The F-madogram can be estimated non-parametrically by

$$\widehat{\nu^F}(y_1, y_2) = \frac{1}{2N(N+1)} \sum_{i=1}^{N} |R_i(y_1) - R_i(y_2)|, \quad y_1, y_2 \in K,$$

where $R_i(y)$ denotes the rank of $Z_i(y)$ (Ribatet, 2013). Plugging the estimator $\widehat{\nu^F}(y_1, y_2)$ into relation (4.51), we obtain an estimator $\hat{\theta}(y_1, y_2)$ and $\theta(y_1, y_2)$. As a measure of quality of the simulations, we calculate the root-mean-square error between the estimated and the theoretical extremal coefficients, i.e.

$$\overline{d_{EC}} = \left[\frac{1}{|\mathcal{Y}|} \sum_{(y_1,y_2)\in\mathcal{Y}} \left(\hat{\theta}(y_1, y_2) - 2\Phi(\sqrt{\|y_1 - y_2\|}/2)\right)^2\right]^{1/2}$$

for some finite set $\mathcal{Y} \subset K \times K \setminus \{(y, y) : y \in K\}$.

Table 4.3 reports the average number of simulated spectral functions $\overline{M}$, the average computational costs both in terms of the number of log-Gaussian processes simulated $\overline{C}$ and in terms of the CPU time for a single simulation (in seconds), the average Kolmogorov–Smirnov distance $\overline{d_{KS}}$ and the root-mean-square error $\overline{d_{EC}}$ based on 79800 pairs. We note that the average CPU time $\bar{t}$ is effectively proportional to the average number $\overline{C}$ of

Table 4.3: Results for simulations of Brown–Resnick processes on the 2-dimensional grids $\{0, 0.05, \ldots, 0.95\}^2$ (top) and $\{0, 0.01, \ldots, 0.99\}^2$ (bottom). For simulation via the normalized spectral representation with $n_{MCMC} = 100$, 500 and 1000 (NR_{100}, NR_{500} and NR_{1000}), the Dieker–Mikosch algorithm (DM) and simulation via extremal functions (EF), the average number $\overline{M}$ of simulated spectral functions, the average computational costs $\overline{C}$ (with sample standard deviation $s(C)$), the average CPU time $\bar{t}$ (with sample standard deviation $s(t)$), the average Kolmogorov–Smirnov distance $\overline{d_{KS}}$ and the root-mean-square error $\overline{d_{EC}}$ are displayed.

	$\overline{M}$	$\overline{C}$	$\bar{t}$	$\overline{d_{KS}}$	$\overline{d_{EC}}$	$s(C)$	$s(t)$
NSR_{100}	7	704	0.13	0.030	0.011	492	0.07
NSR_{500}	7	3677	0.56	0.050	0.014	2462	0.36
NSR_{1000}	7	7244	1.07	0.047	0.020	4971	0.71
DM	932	932	0.17	0.036	0.020	713	0.11
EF	390	390	0.09	0.036	0.013	355	0.05

	$\overline{M}$	$\overline{C}$	$\bar{t}$	$\overline{d_{KS}}$	$\overline{d_{EC}}$	$s(C)$	$s(t)$
NSR_{100}	9	896	108	0.041	0.012	666	81
NSR_{500}	9	4357	522	0.028	0.013	2888	346
NSR_{1000}	9	9194	1095	0.040	0.011	6131	737
DM	26270	26270	3126	0.027	0.012	21199	2542
EF	9361	9361	1116	0.054	0.015	8448	1014

simulated log-Gaussian processes. For the grid $\{0, 0.05, \ldots, 0.95\}^2$, we observe $\bar{t} \approx 2 \cdot 10^{-4} \overline{C}$ while we have $\bar{t} \approx 0.12\overline{C}$ for the grid $\{0, 0.01, \ldots, 0.99\}^2$. Smaller deviations from this proportion in case of the first grid are mainly due to some preparatory computations which have a larger effect in case of small $\overline{C}$. Thus, we can conclude that $\overline{C}$ is an appropriate measure of the computational costs of the simulations.

Comparing the results for the different algorithms in more detail, we first focus on the exactness of the simulations via the normalized spectral representation. As both the Kolmogorov–Smirnov distances and the root-mean-square errors for the extremal coefficients indicate, the Markov chains converge quite fast. In all the cases, both measures of the quality of the simulations via the normalized spectral representation are comparable to those for the exact simulations via the Dieker–Mikosch (2015) algorithm or via extremal functions. Therefore, the algorithms perform with equivalent exactness even though the algorithm based on the normalized spectral representation provides approximations only.

Secondly, the number of simulated normalized spectral functions in our algorithm is much lower than that in the other two algorithms. This difference becomes even more pronounced in the case of a dense grid: While the numbers of simulated processes in the Dieker–Mikosch (2015) algorithm and the algorithm via extremal functions grow (at least) linearly in $|K|$, that in our algorithm remains bounded and practically stable (increased from 7 to 9), as long as $K \subset [0, 1]^2$.

Lastly, even though the computational costs of a single normalized spectral function in our algorithm is n_{MCMC} times higher than that for the other two algorithms, on a dense grid, the total computational costs is much lower than that via the other two algorithms.

The aforementioned three features make the simulation via the normalized spectral representation very attractive, particularly when the process should be simulated at a large

number of locations on a dense grid. Simulation via extremal functions (Dombry et al., 2016a) is preferred only if the process should be simulated at a small or moderate number of locations.

4.7 Summary and Discussion

Whilst in the definition of a max-stable process an infinite number of spectral functions is involved, the minimal expected number of spectral functions that are actually needed for a simulation is an open problem. We consider two substitution problems, problems (4.20) and (4.39), and show that the unique normalized spectral representation is a solution in both cases. Consequently we propose a simulation algorithm based on the normalized spectral representation.

Our simulation result reveals two advantages of the proposed algorithm. First, it improves the algorithm of Schlather (2002) for (mixed) moving maxima processes, because in this case the normalized spectral functions can be simulated easily.

Secondly, it is competitive to other algorithms even if simulation of the underlying normalized spectral functions from the transformed measure g^*H is not straightforward, such as in the simulation of Brown–Resnick processes, even though simulations are no longer exact in this case. Similar results are also expected for other popular max-stable models, such as the extremal Gaussian (Schlather, 2002) and extremal t processes (Opitz, 2013).

Although the problem (4.39) is rather close to the original problem (4.12), it remains unclear whether the normalized spectral representation is also the solution to the original one. It is even not known whether different initial choices of the spectral representation in (4.1) may lead to the same solution via renormalizations g in (4.7) and whether the solution is unique.

Other representations of max-stable processes may also allow for exact simulations. For example, Wang and Stoev (2010) considered a representation based on a finite number of Fréchet variables when the spectral measure H is discrete. Our approach may be advantageous if the number of Fréchet variables gets large while the domain K is bounded. The assessment of the relative performance is left to future research.

Acknowledgements

The authors are grateful to an associate editor and two referees for valuable suggestions and comments. M. Oesting and M. Schlather have been financially supported by Volkswagen Stiftung within the project ‘Mesoscale Weather Extremes – Theory, Spatial Modeling and Prediction (WEX-MOP)’. The work of M. Oesting has also partly been funded by the ANR project ‘McSim’. Views expressed do not necessarily reflect official positions of De Nederlandsche Bank.

5 Sampling Sup-Normalized Spectral Functions for Brown–Resnick Processes

joint work with Martin Schlather and Claudia Schillings

Up to some minor modifications, this chapter is reprint of the research article Oesting et al. (2019) which has appeared in *Stat.*

Sup-normalized spectral functions form building blocks of max-stable and Pareto processes and therefore play an important role in modeling spatial extremes. For one of the most popular examples, the Brown–Resnick process, simulation is not straightforward. In this chapter, we generalize two approaches for simulation via Markov Chain Monte Carlo methods and rejection sampling by introducing new classes of proposal densities. In both cases, we provide an optimal choice of the proposal density with respect to sampling efficiency. The performance of the procedures is demonstrated in an example.

5.1 Introduction

Spatial and spatio-temporal extreme value analysis aims at investigating extremes of quantities described by stochastic processes. In the classical setting, the real-valued process of interest $X = \{X(t), t \in K\}$ is sample-continuous on a compact domain $K \subset \mathbb{R}^d$. Analysis of its extremes is often based on results of the limiting behavior of maxima of independent copies X_i, $i \in \mathbb{N}$. Provided that there exist continuous normalizing functions $a_n : K \to (0, \infty)$ and $b_n : K \to \mathbb{R}$ such that the process of normalized maxima $\{\max_{i=1}^n a_n^{-1}(t) \cdot (X_i(t) - b_n(t)), t \in K\}$ converges in distribution to some sample-continuous process Z with nondegenerate margins as $n \to \infty$, the limit process Z is necessarily max-stable and we say that X is in the max-domain of attraction of Z.
From univariate extreme value theory, it follows that the marginal distributions of Z are necessarily generalized extreme value (GEV) distributions (cf. de Haan and Ferreira, 2006, for instance). As max-stability is preserved under marginal transformations between different GEV distributions, without loss of generality, it can be assumed that Z has standard Fréchet margins, i.e. $\mathbb{P}(Z(t) \leq z) = \exp(-1/z)$, $z > 0$, for all $t \in K$. By de Haan (1984), any sample-continuous max-stable process with standard Fréchet margins can be represented as

$$Z(t) =_d \max_{i \in \mathbb{N}} \{U_i \cdot V_i(t)\}, \quad t \in K, \tag{5.1}$$

where the so-called spectral processes V_i, $i \in \mathbb{N}$, are independent copies of a nonnegative sample continuous stochastic process V on K satisfying $\mathbb{E}\{V_i(t)\} = 1$ for all $t \in K$, and $\sum_{i \in \mathbb{N}} \delta_{U_i}$ is a Poisson point process on $(0, \infty)$ which is independent of the V_i and has intensity measure Λ given by $\Lambda\{(u, \infty)\} = u^{-1}$ for all $u > 0$.
Due to its complex structure, many characteristics of the max-stable process Z in (5.1) cannot be calculated analytically, but need to be assessed via simulations. In order to

simulate Z efficiently, Oesting et al. (2018b) suggest to make use of the sup-normalized spectral representation

$$Z(t) =_d \max_{i\in\mathbb{N}} \left\{ U_i \cdot c_\infty \cdot \frac{V_i^{\max}(t)}{\|V_i^{\max}\|_\infty} \right\}, \quad t \in K, \tag{5.2}$$

where the U_i are the same as above, the processes $V_i^{\max}$ are independently and identically distributed, independently of the U_i, with distribution $\mathbb{P}(V^{\max} \in \cdot)$ given by

$$\mathbb{P}(V^{\max} \in B) = c_\infty^{-1} \cdot \int_B \|v\|_\infty \mathbb{P}(V \in \mathrm{d}v), \quad B \in \mathcal{C}(K), \tag{5.3}$$

and $\|f\|_\infty = \sup_{t\in K} f(t)$ for every $f \in C(K)$, where $C(K)$ denotes the set of all real-valued continuous functions on K equipped with the supremum norm $\|\cdot\|_\infty$ and corresponding σ-algebra $\mathcal{C}(K)$. Here, the normalizing constant $c_\infty = \mathbb{E}\{\|V\|_\infty\}$ is the so-called extremal coefficient of the max-stable process Z over the domain K. In a simulation study, Oesting et al. (2018b) demonstrate that simulation based on the sup-normalized spectral representation is competitive to other state-of-the-art algorithms such as simulation based on extremal functions (Dombry et al., 2016a) provided that the normalized spectral process $V^{\max}$ can be simulated efficiently.

The law of the processes $V^{\max}$ also occurs when analyzing the extremes of a stochastic process X in an alternative way focusing on exceedances over a high threshold: If X is in the max-domain of attraction of the max-stable process Z in (5.1), we have

$$\mathcal{L}\left(x^{-1}X(\cdot) \mid \|X\|_\infty > x\right) \longrightarrow_w \mathcal{L}\left(P\frac{V^{\max}(\cdot)}{\|V^{\max}\|_\infty}\right)$$

as $x \to \infty$, where P is a standard Pareto random variable and $V^{\max}$ is an independent process with law given in (5.3). The limit process $PV^{\max}(\cdot)/\|V^{\max}\|_\infty$ is called Pareto process (cf. Ferreira and de Haan, 2014; Dombry and Ribatet, 2015).

Arising as sup-normalized spectral process for both max-stable and Pareto processes, the process $V^{\max}$ plays an important role in modelling and analyzing spatial extremes. As a crucial building block of spatial and spatio-temporal models, this process needs to be simulated in an efficient way. Due to the measure transformation in (5.3), however, sampling of $V^{\max}$ is not straightforward even in cases where the underlying spectral process V can be simulated easily.
In the present paper, we focus on the simulation of $V^{\max}$ for the very popular class of log Gaussian spectral processes, i.e. $V(t) = \exp(W(t))$ for some Gaussian process W such that $\mathbb{E}\{\exp(W(t))\} = 1$ for all $t \in K$. The resulting subclass of max-stable processes Z in (5.1) comprises the only possible nontrivial limits of normalized maxima of rescaled Gaussian processes, the class of Brown–Resnick processes (Kabluchko et al., 2009; Kabluchko, 2011). In order to obtain Brown–Resnick processes that can be extended to stationary processes on $\mathbb{R}^d$, Kabluchko et al. (2009) consider $W(t) = G(t) - \mathrm{Var}\{G(t)\}/2$, $t \in K$, with G being a centered Gaussian process on $\mathbb{R}^d$ with stationary increments and variogram

$$\gamma(h) = \frac{1}{2}\mathbb{E}\left\{(G(t+h) - G(t))^2\right\}, \quad t, h \in \mathbb{R}^d.$$

It is important to note that the law of the resulting max-stable process Z does not depend on the variance of W, but only on γ. Therefore, Z is called the Brown–Resnick process associated with the variogram γ.

Recently, Ho and Dombry (2017) introduced a two-step procedure to simulate the corresponding sup-normalized process

$$\frac{V^{\max}(\cdot)}{\|V^{\max}\|_\infty} = \exp\left(W^{\max}(\cdot) - \|W^{\max}\|_\infty\right)$$

efficiently if the finite domain $K = \{t_1, \dots, t_N\}$ is of small or moderate size:

1. Sample the index i of the component where the vector $\mathbf{V}^{\max} = (V^{\max}(t_k))_{k=1,\dots,N}$ assumes its maximum, i.e. select one of the events $\mathbf{V}^{\max} \in S_i = \{\mathbf{s} \in (0,\infty)^N : \|\mathbf{s}\|_\infty = s_i\}$, $i = 1, \dots, N$. Provided that the covariance matrix $\mathbf{C}$ of the Gaussian vector $\mathbf{W} = (W(t_k))_{i=k}^N$ is nonsingular, we have that this index is a.s. unique and that the probabilities of the corresponding events can be calculated in terms of the matrix $\mathbf{Q} \in \mathbb{R}^{N\times N}$ and the vector $\mathbf{m} \in \mathbb{R}^N$ given by

$$\mathbf{Q} = \mathbf{C}^{-1} - \frac{\mathbf{C}^{-1}\mathbf{1}_N\mathbf{1}_N^\top\mathbf{C}^{-1}}{\mathbf{1}_N^\top\mathbf{C}^{-1}\mathbf{1}_N} \qquad \text{and} \qquad \mathbf{m} = -\Big(\frac{1}{2}\boldsymbol{\sigma} + \frac{1 - \frac{1}{2}\boldsymbol{\sigma}^\top\mathbf{C}^{-1}\mathbf{1}_N}{\mathbf{1}_N^\top\mathbf{C}^{-1}\mathbf{1}_N}\mathbf{1}_N^\top\Big)\mathbf{C}^{-1}$$

   where $\boldsymbol{\sigma} = (\mathrm{Var}(W(t_k)))_{k=1,\dots,N}$ denotes the variance vector of $\mathbf{W}$ and, furthermore, $\mathbf{1}_N = (1, \dots, 1)^\top \in \mathbb{R}^N$. More precisely, by Ho and Dombry (2017),

$$\begin{aligned}&\mathbb{P}\left(\mathbf{V}^{\max} \in S_i\right)\\ &= \frac{\det(\mathbf{Q}_{-i})^{-1/2}\exp\{\frac{1}{2}\mathbf{m}_{-i}^\top\mathbf{Q}_{-i}^{-1}\mathbf{m}_{-i}\}\Phi_{N-1}(\mathbf{0}_{N-1}; \mathbf{Q}_{-i}^{-1}\mathbf{m}_{-i}, \mathbf{Q}_{-i}^{-1})}{\sum_{j=1}^N \det(\mathbf{Q}_{-j})^{-1/2}\exp\{\frac{1}{2}\mathbf{m}_{-j}^\top\mathbf{Q}_{-j}^{-1}\mathbf{m}_{-j}\}\Phi_{N-1}(\mathbf{0}_{N-1}; \mathbf{Q}_{-j}^{-1}\mathbf{m}_{-j}, \mathbf{Q}_{-j}^{-1})},\end{aligned}$$

   where $\mathbf{m}_{-j}$ denotes the vector $\mathbf{m}$ after removing the jth component, $\mathbf{Q}_{-j}$ denotes the matrix $\mathbf{Q}$ after removing the jth row and jth column and $\Phi_{N-1}(\mathbf{0}_{N-1}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the distribution function of an $(N-1)$-dimensional Gaussian distribution with mean vector $\boldsymbol{\mu} \in \mathbb{R}^{N-1}$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{(N-1)\times(N-1)}$ evaluated at $\mathbf{0}_{N-1} = (0, \dots, 0) \in \mathbb{R}^{N-1}$.

2. Conditional on $\mathbf{V}^{\max} \in S_i$, we have $V^{\max}(t_i)/\|\mathbf{V}^{\max}\|_\infty = 1$ and the distribution of the vector $\mathbf{M} = (\log(V^{\max}(t_j)))_{j\neq i} - \log(\|\mathbf{V}^{\max}\|)$ is an $(N-1)$-dimensional Gaussian distribution with mean vector $\mathbf{Q}_{-i}^{-1}\mathbf{m}_{-i}$ and covariance matrix $\mathbf{Q}_{-i}^{-1}$ conditional on $\boldsymbol{M}$ being nonpositive.

However, the first step includes computationally expensive operations such as the evaluation of $(N-1)$-dimensional Gaussian distribution functions and the inversion of matrices of sizes $N \times N$ and $(N-1) \times (N-1)$. Furthermore, an efficient implementation of the second step is not straightforward. Thus, the procedure is feasible for small or moderate N only.

In this paper, we will introduce alternative procedures for the simulation of $V^{\max}$, or, equivalently, $W^{\max} = \log V^{\max}$, that are supposed to work for larger N, as well. To this end, we will modify a Markov Chain Monte Carlo (MCMC) algorithm proposed by Oesting et al. (2018b) and a rejection sampling approach based on ideas of de Fondeville and Davison (2018). Both procedures have originally been designed to sample sup-normalized spectral functions in general. Here, we will adapt them to the specific case of Brown–Resnick processes.

5.2 Simulating $W^{\max}$ via MCMC algorithms

Based on the Brown–Resnick process as our main example, we consider a max-stable process Z with spectral process $V = e^W$ for some sample-continuous process W. Henceforth,

we will always assume that the simulation domain $K = \{t_1, \dots, t_N\} \subset \mathbb{R}^d$ is finite and that the corresponding spectral vector $\mathbf{W}$ possesses a density f w.r.t. some measure μ on $\mathbb{R}^N$. Then, by (5.3), the transformed spectral vector $\mathbf{W}^{\max} = \log(\mathbf{V}^{\max})$, where the logarithm is applied componentwise, has the multivariate density

$$f_{\max}(\boldsymbol{w}) = c_\infty^{-1} \max_{i=1}^{N} \exp(w_i) f(\boldsymbol{w}), \qquad \boldsymbol{w} \in \mathbb{R}^N,$$

which obviously has the same support as f, i.e. $\operatorname{supp}(f_{\max}) = \operatorname{supp}(f)$.
As direct sampling from the density $f_{\max}$ is rather sophisticated and the normalizing constant c_∞ is not readily available, it is quite appealing to choose an MCMC approach for simulation. In the present paper, we focus on Metropolis-Hastings algorithms with independence sampler (cf. Tierney, 1994, for example). Denoting the strictly positive proposal density on $\operatorname{supp}(f)$ by f_{prop}, the algorithm is of the following form:

Algorithm 5.1: MCMC Approach (Metropolis–Hastings)

Input: proposal density f_{prop}
Simulate $\boldsymbol{w}^{(0)}$ according to the density f_{prop}.
for $k = 1, \dots, n_{MCMC}$ **do**
  Sample $\boldsymbol{w}$ from f_{prop} and set

$$\boldsymbol{w}^{(k)} = \begin{cases} \boldsymbol{w} & \text{with probability } \alpha(\boldsymbol{w}^{(k-1)}, \boldsymbol{w}), \\ \boldsymbol{w}^{(k-1)} & \text{with probability } 1 - \alpha(\boldsymbol{w}^{(k-1)}, \boldsymbol{w}), \end{cases}$$

  where the acceptance probability $\alpha(\cdot, \cdot)$ is given by (5.4).
Output: Markov chain $(\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(n_{MCMC})})$.

Here, the acceptance ratio $\alpha(\widetilde{\boldsymbol{w}}, \boldsymbol{w})$ for a new proposal $\boldsymbol{w} \in \operatorname{supp}(f)$ given a current state $\widetilde{\boldsymbol{w}} \in \operatorname{supp}(f)$ is

$$\alpha(\widetilde{\boldsymbol{w}}, \boldsymbol{w}) = \min\left\{ \frac{f_{\max}(\boldsymbol{w}) / f_{\text{prop}}(\boldsymbol{w})}{f_{\max}(\widetilde{\boldsymbol{w}}) / f_{\text{prop}}(\widetilde{\boldsymbol{w}})}, 1 \right\}, \tag{5.4}$$

using the convention that a ratio is interpreted as 0 if both the enumerator and the denominator are equal to 0. This choice of $\alpha(\widetilde{\boldsymbol{w}}, \boldsymbol{w})$ ensures reversibility of the resulting Markov chain $\{\boldsymbol{w}^{(k)}\}_{k \in \mathbb{N}}$ with respect to the distribution of $\mathbf{W}^{\max}$. Further, it allows for a direct transition from any state $\widetilde{\boldsymbol{w}} \in \operatorname{supp}(f)$ to any other state $\boldsymbol{w} \in \operatorname{supp}(f)$. Consequently, the chain is irreducible and aperiodic and, thus, its distribution converges to the desired stationary distribution, that is, for a.e. initial state $\boldsymbol{w}^{(0)} \in \operatorname{supp}(f)$, we have that

$$\| P^n(\boldsymbol{w}^{(0)}, \cdot) - \mathbb{P}(\mathbf{W}^{\max} \in \cdot) \|_{\text{TV}} \overset{n \to \infty}{\longrightarrow} 0, \tag{5.5}$$

where $P^n(\boldsymbol{w}^{(0)}, \cdot)$ denotes the distribution of the n-th state of a Markov chain with initial state $\boldsymbol{w}^{(0)}$ and $\| \cdot \|_{\text{TV}}$ is the total variation norm.

As a general approach for the simulation of sup-normalized spectral processes of arbitrary max-stable processes, Oesting et al. (2018b) propose to use Algorithm 5.1 with the density f of the original spectral vector $\mathbf{W}$ as proposal density (Algorithm 1A) and the Metropolis-Hastings acceptance ratio in (5.4) simplifies to

$$\alpha(\widetilde{\boldsymbol{w}}, \boldsymbol{w}) = \min\left\{ \frac{\max_{i=1}^N e^{w_i}}{\max_{i=1}^N e^{\widetilde{w}_i}}, 1 \right\}, \qquad \boldsymbol{w}, \widetilde{\boldsymbol{w}} \in \mathbb{R}^N. \tag{5.6}$$

As the proposal density $f_{\text{prop}} = f$ is strictly positive on $\operatorname{supp}(f)$, convergence of the distribution of the Markov chain to the distribution of $\mathbf{W}^{\max}$ as in (5.5) is ensured. If the support of f is unbounded, however, there is no uniform geometric rate of convergence of the chain in (5.5), as we have

$$\operatorname*{essinf}_{\boldsymbol{w}\in\mathbb{R}^N} \frac{f(\boldsymbol{w})}{f_{\max}(\boldsymbol{w})} = \operatorname*{essinf}_{\boldsymbol{w}\in\mathbb{R}^N} \left(c_\infty \cdot \min_{i=1}^{N} e^{-w_i} \right) = 0$$

(Mengersen and Tweedie, 1996). In particular, this holds true for the case of a Brown–Resnick process where $\mathbf{W}$ is a Gaussian vector.
Furthermore, due to the structure of the acceptance ratio in (5.6), the Markov chain may get stuck, once a state $\widetilde{\boldsymbol{w}}$ with a large maximum $\max_{i=1}^{N} \exp(\widetilde{w}_i)$ is reached. This might lead to rather poor mixing properties of the chain. Even though independent realizations could still be obtained by starting new independent Markov chains (cf. Oesting et al., 2018b), such a behavior is undesirable having chains in high dimension N with potentially long burn-in periods in mind.

While the algorithm in Oesting et al. (2018b) is designed to be applicable in a general framework, we will use a specific transformation to construct a Markov chain with stronger mixing and faster convergence to the target distribution. For many popular models such as Brown–Resnick processes, this transformation is easily applicable. More precisely, we consider the related densities f_i, $i = 1, \dots, N$, with $f_i(\boldsymbol{w}) = \exp(w_i) f(\boldsymbol{w})$. These densities are closely related to the distributions P_i that have been studied in Dombry et al. (2016a). Hence, we propose to approach the target distribution with density $f_{\max} = c_\infty^{-1} \max_{i=1}^{N} f_i$ by Algorithm 5.1 using a mixture

$$f_{\text{prop}} = \sum_{i=1}^{N} p_i f_i \tag{5.7}$$

as proposal density, where the weights $p_i \geq 0$, $i = 1, \dots, N$, are such that $\sum_{i=1}^{N} p_i = 1$. The corresponding acceptance probability in (5.4) is then

$$\widetilde{\alpha}(\widetilde{\boldsymbol{w}}, \boldsymbol{w}) = \min \left\{ \frac{\max_{i=1}^{N} e^{w_i} / \sum_{i=1}^{N} p_i e^{w_i}}{\max_{i=1}^{N} e^{\widetilde{w}_i} / \sum_{i=1}^{N} p_i e^{\widetilde{w}_i}}, 1 \right\}. \tag{5.8}$$

With the proposal density being strictly positive on $\operatorname{supp}(f)$, we can see that the distribution of the Markov chain again converges to its stationary distribution with density $f_{\max}$. As we further have

$$\inf_{\boldsymbol{w}\in\mathbb{R}^N} \frac{f_{\text{prop}}(\boldsymbol{w})}{f_{\max}(\boldsymbol{w})} = \inf_{\boldsymbol{w}\in\mathbb{R}^N} \frac{\sum_{i=1}^{N} p_i e^{w_i}}{c_\infty^{-1} \max_{i=1}^{N} e^{w_i}} = c_\infty \cdot \min_{i=1}^{N} p_i > 0, \tag{5.9}$$

provided that $p_i > 0$ for $i = 1, \dots, N$, the results found by Mengersen and Tweedie (1996) even ensure a uniform geometric rate of convergence for any starting value $\boldsymbol{w}^{(0)} \in \operatorname{supp}(f)$ in contrast to the case where $f_{\text{prop}} = f$.

In order to obtain a chain with good mixing properties, we choose p_i such that the acceptance rate in Algorithm 5.1 is high provided that the current state $\boldsymbol{w}^{(k)}$ is (approximately) distributed according to the stationary distribution. To this end, we minimize the relative deviation between f_{prop} and $f_{\max}$ under $f_{\max}$, i.e. we minimize

$$D(p_1, \dots, p_N) = \int_{\mathbb{R}^N} \left(\frac{f_{\text{prop}}(\boldsymbol{w})}{f_{\max}(\boldsymbol{w})} - 1 \right)^2 f_{\max}(\boldsymbol{w}) \, \mathrm{d}\boldsymbol{w},$$

under the constraint $\sum_{i=1}^N p_i = 1$. Introducing a Lagrange multiplier $\lambda \in \mathbb{R}$, minimizing

$$\begin{aligned} D(p_1,\dots,p_N) &= \mathbb{E}\left\{ \left(\frac{\sum_{i=1}^N p_i e^{W(t_i)}}{c_\infty^{-1} e^{\max_{j=1}^N W(t_j)}} - 1 \right)^2 c_\infty^{-1} e^{\max_{j=1}^N W(t_j)} \right\} \\ &= c_\infty \sum_{i=1}^N \sum_{k=1}^N p_i p_k \mathbb{E}\left\{ e^{W(t_i)+W(t_k)-\max_{j=1}^N W(t_j)} \right\} - 1 \end{aligned}$$

results in solving the linear system

$$\begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{1}_N \\ \mathbf{1}_N^\top & 0 \end{pmatrix} \begin{pmatrix} \mathbf{p} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{0}_N \\ 1 \end{pmatrix}, \tag{5.10}$$

where $\mathbf{p} = (p_1,\dots,p_N)^\top$ and $\boldsymbol{\Sigma} = (\sigma_{ik})_{1\le i,k\le N}$ with

$$\sigma_{ik} = \mathbb{E}\{ e^{W(t_i)+W(t_k)-\max_{j=1}^N W(t_j)} \}. \tag{5.11}$$

Provided that the matrix $\boldsymbol{\Sigma}$ is nonsingular, the solution of (5.10) is given by

$$\mathbf{p} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}_N}{\mathbf{1}_N^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}_N} \tag{5.12}$$

(cf. Cressie, 1993, for instance). This solution does not necessarily satisfy the additional restriction $p_i \ge 0$ for all $i = 1,\dots,N$. In case that Σ is singular or the vector $\mathbf{p}$ has at least one negative entry, the full optimization problem

$$\begin{aligned} &\min \mathbf{p}^\top \boldsymbol{\Sigma} \mathbf{p} \\ \text{s.t.}\quad &\mathbf{1}_N^\top \mathbf{p} = 1 \\ &p_i \ge 0 \qquad \forall i = 1,\dots,N, \end{aligned} \tag{QP}$$

has to be solved. Using the Karush–Kuhn–Tucker optimality conditions, the quadratic program (QP) can be transformed into a linear program with additional (nonlinear) complementary slackness conditions. It can be solved by modified simplex methods. Alternatively, the problem (QP) can be solved by the dual method by Goldfarb and Idnani (1983).

Remark 5.1. In order to ensure a geometric rate of convergence of the distribution of the Markov chain, we might replace the condition $p_i \ge 0$ for all $i = 1,\dots,N$, in (QP) by $p_i \ge \varepsilon$ for some given $\varepsilon > 0$. Then, a geometric rate of convergence follows from (5.9) as described above.

In the case of Brown–Resnick processes, for simplicity, we consider the case that the random vector $\mathbf{W}$ possesses a full Lebesgue density. In particular, the covariance matrix $\mathbf{C} = (\mathrm{Cov}(W(t_i), W(t_j)))_{1\le t_i,t_j\le N}$ of $\mathbf{W}$ is assumed to be nonsingular. Then, the target density is

$$\begin{aligned} f_{\max}(\boldsymbol{w}) &= c_\infty^{-1} \max_{i=1}^N \exp(w_i) f(\boldsymbol{w}) \\ &= \frac{c_\infty^{-1} \max_{i=1}^N \exp(w_i)}{(2\pi)^{N/2} \det(\mathbf{C})^{1/2}} \exp\left\{ -\frac{1}{2} \left(\boldsymbol{w} + \frac{\boldsymbol{\sigma}}{2} \right)^\top \mathbf{C}^{-1} \left(\boldsymbol{w} + \frac{\boldsymbol{\sigma}}{2} \right) \right\}, \quad \boldsymbol{w} \in \mathbb{R}^N, \end{aligned}$$

where $\boldsymbol{\sigma} = (\mathrm{Var}(W(t_k)))_{k=1,\dots,N}$ is again the variance vector of $\mathbf{W}$. Now, the densities $f_i(\boldsymbol{w})) = e^{w_i} f(\boldsymbol{w})$ which form the proposal density are just shifted Gaussian distributions:

$$\begin{aligned} f_i(\boldsymbol{w}) &= \frac{\exp(w_i)}{(2\pi)^{\frac{N}{2}} \det(\mathbf{C})^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{w}+\frac{\boldsymbol{\sigma}}{2}\right)^\top \mathbf{C}^{-1}\left(\boldsymbol{w}+\frac{\boldsymbol{\sigma}}{2}\right)\right\} \\ &= \frac{1}{(2\pi)^{\frac{N}{2}} \det(\mathbf{C})^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{w}-\mathbf{C}_{\cdot i}+\frac{\boldsymbol{\sigma}}{2}\right)^\top \mathbf{C}^{-1}\left(\boldsymbol{w}-\mathbf{C}_{\cdot i}+\frac{\boldsymbol{\sigma}}{2}\right)\right\}, \end{aligned}$$

cf. Lemma 1 in the Supplementary Material of Dombry et al. (2016a), i.e. we have

$$\mathcal{L}(\mathbf{W}^{(i)}) = \mathcal{L}(\mathbf{W} + C_{\cdot i}) \tag{5.13}$$

where the Gaussian vectors $\mathbf{W}^{(i)}$ and $\mathbf{W}$ possess densities f_i and f, respectively. The calculation of the optimal weights p_i is based on the expectation in (5.11) which typically cannot be calculated analytically, but needs to be assessed numerically via simulations. Such a numerical evaluation, however, is challenging as the random variable $\exp(W(t_i) + W(t_k) - \max_{j=1}^N W(t_j))$ is unbounded. To circumvent these computational difficulties, we make use of the identity

$$\begin{aligned} \sigma_{ik} &= \mathbb{E}\left\{e^{W(t_i)+W(t_k)-\max_{j=1}^N W(t_j)}\right\} \\ &= \int_{\mathbb{R}^N} \frac{e^{w_i+w_k-\max_{j=1}^N w_j}}{(2\pi)^{\frac{N}{2}} \det(\mathbf{C})^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{w}+\frac{\boldsymbol{\sigma}}{2}\right)^\top \mathbf{C}^{-1}\left(\boldsymbol{w}+\frac{\boldsymbol{\sigma}}{2}\right)\right\} \mathrm{d}\boldsymbol{w} \\ &= \int_{\mathbb{R}^N} \frac{e^{w_k-\max_{j=1}^N w_j}}{(2\pi)^{\frac{N}{2}} \det(\mathbf{C})^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{w}-\mathbf{C}_{\cdot i}+\frac{\boldsymbol{\sigma}}{2}\right)^\top \mathbf{C}^{-1}\left(\boldsymbol{w}-\mathbf{C}_{\cdot i}+\frac{\boldsymbol{\sigma}}{2}\right)\right\} \mathrm{d}\boldsymbol{w} \\ &= \int_{\mathbb{R}^N} e^{w_k-\max_{j=1}^N w_j} f_i(\boldsymbol{w})\,\mathrm{d}\boldsymbol{w} = \mathbb{E}\left\{\exp(W^{(i)}(t_k) - \max_{j=1}^N W^{(i)}(t_j))\right\}. \end{aligned} \tag{5.14}$$

The expression on the right-hand side of (5.14) can be conveniently assessed numerically as the random variable $\exp(W^{(i)}(t_k) - \max_{j=1}^N W^{(i)}(t_j))$ is bounded by 1.

Remark 5.2. Note that both (5.13) and the final result in (5.14) still hold true if $\mathbf{W}$ does not possess a full Lebesgue density, but exactly one component W_{i^*} is degenerate and the reduced covariance matrix $(C_{ij})_{i,j\neq i^*}$ is nonsingular. This situation appears in several examples such as W being a fractional Brownian motion where $W(0) = 0$ a.s.

In summary, we propose the procedure below to simulate the normalized spectral vector $\mathbf{W}^{\max}$ for the Brown–Resnick process (Algorithm 1B):

1. Calculate $\mathbf{p}$ by solving the quadratic program (QP) where the entries of the matrix $\boldsymbol{\Sigma}$ are given by (5.14). Provided that all its components are nonnegative, the solution $\mathbf{p}$ has the form (5.12).

2. Run Algorithm 5.1 with proposal density $f_{\mathrm{prop}} = \sum_{i=1}^N p_i f_i$ and acceptance probability given by (5.8). The output of the algorithm is a Markov chain whose stationary distribution is the distribution of $\mathbf{W}^{\max}$.

5.3 Exact Simulation via Rejection Sampling

In this section, we present an alternative procedure to generate samples from $\mathbf{W}^{\max}$ with probability density $f_{\max}$. In contrast to Section 5.2 where we generated a Markov chain

consisting of dependent samples with the desired distribution as stationary distribution, here, we aim to produce independent realizations from the exact target distribution. To this end, we make use of a rejection sampling approach (cf. Devroye, 1986, for instance) based on a proposal density $\tilde{f}_{\text{prop}}$ satisfying

$$f_{\max}(\boldsymbol{w}) \leq (c_\infty \cdot C)^{-1} \tilde{f}_{\text{prop}}(\boldsymbol{w}), \qquad \text{for all } \boldsymbol{w} \in \mathbb{R}^N, \tag{5.15}$$

for some $C > 0$.

Algorithm 5.2: Rejection Sampling Approach

Input: proposal density $\tilde{f}_{\text{prop}}$ and constant $C > 0$ satisfying (5.15)
repeat
  Simulate $\boldsymbol{w}^*$ according to the density $\tilde{f}_{\text{prop}}$.
  Generate a uniform random number u in $[0, 1]$.
until $u \cdot \tilde{f}_{\text{prop}}(\boldsymbol{w}^*) \leq C \cdot c_\infty \cdot f_{\max}(\boldsymbol{w}^*)$;
Output: exact sample $\boldsymbol{w}^*$ from distribution with density $f_{\max}$

Thus, on average, $(c_\infty \cdot C)^{-1}$ simulations from the proposal distribution are needed to obtain an exact sample from the target distribution. Of course, to minimize the computational burden, for a given proposal density $\tilde{f}_{\text{prop}}$, the constant C should be chosen maximal subject to (5.15), i.e.

$$C = \inf_{\boldsymbol{w} \in \mathbb{R}^N} \frac{\tilde{f}_{\text{prop}}(\boldsymbol{w})}{c_\infty f_{\max}(\boldsymbol{w})}.$$

Recently, de Fondeville and Davison (2018) followed a similar idea and suggested to base the simulation of a general sup-normalized spectral process $V(\cdot)^{\max}/\|V^{\max}\|_\infty$ on the relation

$$\mathbb{P}\left(\frac{V^{\max}}{\|V^{\max}\|_\infty} \in \mathrm{d}v\right) = \frac{\|\tilde{V}\|_\infty}{\mathbb{E}\|\tilde{V}\|_\infty} \mathbb{P}\left(\frac{\tilde{V}}{\|\tilde{V}\|_\infty} \in \mathrm{d}v\right) \tag{5.16}$$

where $\tilde{V}$ is a spectral process normalized with respect to another homogeneous functional r instead of the supremum norm, i.e. $r(\tilde{V}) = 1$ a.s. If $\|\tilde{V}\|_\infty$ is a.s. bounded from above by some constant, from the relation (5.16), we obtain an inequality of the same type as (5.15) for the densities of $V(\cdot)^{\max}/\|V^{\max}\|_\infty$ and $\tilde{V}(\cdot)/\|\tilde{V}\|_\infty$ instead of $f_{\max}$ and f_{prop}, respectively. Thus, samples of $\tilde{V}(\cdot)/\|\tilde{V}\|_\infty$ can be used as proposals for an exact rejection sampling procedure. For instance, the sum-normalized spectral vector $\tilde{\mathbf{V}}$, i.e. the vector which is normalized w.r.t. the functional $r(f) = \|f\|_1 = \sum_{k=1}^N |f(t_k)|$, can be chosen as it is easy to simulate in many cases (cf. Dombry et al., 2016a) and satisfies $\|\tilde{\mathbf{V}}\|_\infty \leq 1$ almost surely.

For a Brown–Resnick process, it is well-known that the sum-normalized process $\tilde{\mathbf{V}}$ has the same distribution as $\exp(\mathbf{W}^{\text{prop}})/\|\exp(\mathbf{W}^{\text{prop}})\|_1$ where $\mathbf{W}^{\text{prop}}$ has the density f_{prop} from (5.7) in Section 5.2 with equal weights $p_1 = \ldots = p_N = 1/N$ (see also Dieker and Mikosch, 2015). Thus, in this case, the procedure proposed by de Fondeville and Davison (2018) with $r(f) = \|f\|_1$ is equivalent to performing rejection sampling for $\mathbf{W}^{\max}$ with $\tilde{f}_{\text{prop}} = \sum_{i=1}^N \frac{1}{N} f_i$ as proposal distribution (Algorithm 2A). From Equation (5.9), it follows that rejection sampling can also be performed with $\tilde{f}_{\text{prop}} = \sum_{i=1}^N p_i f_i$ and arbitrary positive weights $p_1, \ldots, p_N$ summing up to 1, since we have (5.15) with $C = \min_{i=1}^N p_i$. Thus, accepting a proposal $\boldsymbol{w}^*$ in the rejection sampling procedure with probability

$$\min_{i=1}^N p_i \cdot \frac{c_\infty f_{\max}(\boldsymbol{w}^*)}{f_{\text{prop}}(\boldsymbol{w}^*)} = \frac{\min_{i=1}^N p_i \cdot \max_{i=1}^N e^{w_i^*}}{\sum_{i=1}^N p_i e^{w_i^*}},$$

we will obtain a sample of independent realizations from the exact target distribution $f_{\max}$. The rejection rate, however, is pretty high. In order to obtain one realization of $\mathbf{W}^{\max}$, on average $\left(c_\infty \cdot \min_{i=1}^N p_i\right)^{-1}$ simulations from f^*_{prop} are required. It can be easily seen that the computational costs are indeed minimal for the choice $p_1 = \ldots = p_N = 1/N$, i.e. the choice in the approach based on the sum-normalized representation. In this case, one realization of $\mathbf{W}^{\max}$ on average requires to sample $c_\infty^{-1} N$ times from f_{prop}. Therefore, this approach becomes rather inefficient if we have a large number N of points on a dense grid.

In order to reduce the large computational costs of rejection sampling which are mainly due to the fact that $\min_{i=1}^N f_i(\boldsymbol{w})/f_{\max}(\boldsymbol{w})$ gets small as $\|\boldsymbol{w}\| \to \infty$, we replace each density f_i by the modified multivariate Gaussian density $g_{i,\varepsilon}$ whose variance is increased by the factor $(1-\varepsilon)^{-1} \geq 1$ for some $\varepsilon \in [0,1)$:

$$\begin{aligned} g_{i,\varepsilon}(\boldsymbol{w}) &= \frac{(1-\varepsilon)^{N/2}}{(2\pi)^{N/2}\det(\mathbf{C})^{1/2}} \exp\left(-\frac{1}{2}(1-\varepsilon)\cdot\left(\boldsymbol{w}-\mathbf{C}_{\cdot i}+\frac{\boldsymbol{\sigma}}{2}\right)^\top \mathbf{C}^{-1}\left(\boldsymbol{w}-\mathbf{C}_{\cdot i}+\frac{\boldsymbol{\sigma}}{2}\right)\right) \\ &= \frac{(1-\varepsilon)^{N/2}}{(2\pi)^{N/2}\det(\mathbf{C})^{1/2}} \exp\left((1-\varepsilon)w_i\right)\cdot\exp\left(-\frac{1}{2}(1-\varepsilon)\cdot\left(\boldsymbol{w}+\frac{\boldsymbol{\sigma}}{2}\right)^\top \mathbf{C}^{-1}\left(\boldsymbol{w}+\frac{\boldsymbol{\sigma}}{2}\right)\right). \end{aligned}$$

Analogously to f_{prop} for the MCMC approach in Section 5.2, we propose a mixture

$$\tilde{f}_{\text{prop}} = \sum\nolimits_{i=1}^N p_i g_{i,\varepsilon}$$

with $p_i \geq 0$ and $\sum_{i=1}^N p_i = 1$ as proposal density for the rejection sampling algorithm. A proposal $\boldsymbol{w}^*$ is then accepted with probability

$$C(\mathbf{p},\varepsilon)\cdot\frac{c_\infty \cdot f_{\max}(\boldsymbol{w}^*)}{\sum_{i=1}^N p_i g_{i,\varepsilon}(\boldsymbol{w}^*)} = C(\mathbf{p},\varepsilon)\cdot\frac{\exp\left(-\frac{\varepsilon}{2}\left(\boldsymbol{w}+\frac{\boldsymbol{\sigma}}{2}\right)^\top \mathbf{C}^{-1}\left(\boldsymbol{w}+\frac{\boldsymbol{\sigma}}{2}\right)\right)}{(1-\varepsilon)^{\frac{N}{2}}\cdot\sum_{i=1}^N p_i \exp\left((1-\varepsilon)w_i - \max_{j=1}^N w_j^*\right)}, \tag{5.17}$$

where

$$\begin{aligned} C(\mathbf{p},\varepsilon) &= \inf_{\boldsymbol{w}\in\mathbb{R}^N} \frac{\sum_{i=1}^N p_i g_{i,\varepsilon}(\boldsymbol{w})}{c_\infty f_{\max}(\boldsymbol{w})} = \inf_{\boldsymbol{w}\in\mathbb{R}^N} \min_{j=1}^N \frac{\sum_{i=1}^N p_i g_{i,\varepsilon}(\boldsymbol{w})}{f_j(\boldsymbol{w})} \\ &= \inf_{\boldsymbol{w}\in\mathbb{R}^N} (1-\varepsilon)^{\frac{N}{2}} \sum_{i=1}^N p_i \exp\left((1-\varepsilon)w_i - \max_{j=1}^N w_j + \frac{\varepsilon}{2}\left(\boldsymbol{w}+\frac{\boldsymbol{\sigma}}{2}\right)^\top \mathbf{C}^{-1}\left(\boldsymbol{w}+\frac{\boldsymbol{\sigma}}{2}\right)\right). \end{aligned} \tag{5.18}$$

Thus, to summarize, for appropriately chosen $\varepsilon > 0$ and $\mathbf{p} \geq 0$ such that $\|\mathbf{p}\|_1 = 1$, we propose to run Algorithm 5.2 with proposal density $\tilde{f}_{\text{prop}} = \sum_{i=1}^N p_i g_{i,\varepsilon}$ and $C = C(\mathbf{p},\varepsilon)$ according to (5.18).

Remark 5.3. To further reduce the computational costs in the simulation, we might even choose a more flexible approach. For instance, instead of using a mixture of a finite number of functions $g_{1,\varepsilon},\ldots,g_{N,\varepsilon}$, one could consider arbitrary mixtures

$$\tilde{f}_{\text{prop}}(\boldsymbol{w}) = \int_{\mathbb{R}^d} g_{t,\varepsilon}(\boldsymbol{w})\nu(\mathrm{d}t), \qquad \boldsymbol{w}\in\mathbb{R}^N,$$

where $g_{t,\varepsilon}(\boldsymbol{w}) = \int_{\mathbb{R}} g_{N+1,\varepsilon}(\boldsymbol{w}, w_{N+1})\mathrm{d}w_{N+1}$ on the enlarged domain $\{t_1,\ldots,t_N,t\}$ and ν is a probability measure on $\mathbb{R}^d$. Furthermore, depending on $t\in\mathbb{R}^d$, different values for

$\varepsilon = \varepsilon(t) \in [0,1)$ might be chosen. However, due to the complexity of the optimization problems involved, we restrict ourselves to the situation above where ν is a probability measure on $K = \{t_1, \dots, t_N\}$ and ε is constant in space.

Using the procedure described above, on average, $(c_\infty \cdot C(\mathbf{p}, \varepsilon))^{-1}$ simulations from the proposal distribution are needed to obtain one exact sample from the target distribution, i.e. the computational complexity of the algorithm depends on the choices of $\mathbf{p}$ and ε. The remainder of this section will be devoted to this question.

Choice of $\mathbf{p}$ and ε For a given $\varepsilon \geq 0$, the computational costs of the algorithm can be minimized by choosing $\mathbf{p} = \mathbf{p}^*(\varepsilon)$ such that the constant $C(\mathbf{p}, \varepsilon)$ given in (5.18) is maximal, i.e. by choosing $\mathbf{p}$ as the solution of the nonlinear optimization problem

$$\begin{aligned} &\max_{\mathbf{p} \in \mathbb{R}^N} \quad C(\mathbf{p}, \varepsilon) \\ \text{s.t.} \quad &\|\mathbf{p}\|_1 = 1 \\ &p_i \geq 0 \qquad \forall i = 1, \dots, N \end{aligned} \tag{NP}$$

Optimizing further w.r.t. $\varepsilon \in [0,1)$, we obtain the optimal choice $(\mathbf{p}, \varepsilon) = (\mathbf{p}^*(\varepsilon^*), \varepsilon^*)$ where $\varepsilon^* = \operatorname{argmax}_{\varepsilon \in [0,1)} C(\mathbf{p}^*(\varepsilon), \varepsilon)$.
As the above optimization problem includes optimization steps with respect to $\boldsymbol{w} \in \mathbb{R}^N$, $\mathbf{p} \in \{\mathbf{x} \in [0,1]^N : \ \|\mathbf{x}\|_1 = 1\}$ and $\varepsilon \in [0,1]$, none of which can be solved analytically, the solution is quite involved. In order to reduce the computational burden, we simplify the problem by maximizing an analytically simpler lower bound. To this end, we decompose the convex combination $\sum_{i=1}^N p_i g_{i,\varepsilon} / f_j$ into sums over disjoint index subsets of the form $I = \{i_1, \dots, i_m\} \subset \{1, \dots, N\}$. For a convex combination of $(g_{i_k,\varepsilon})_{k=1,\dots,m}$ with weight vector $\boldsymbol{\lambda} = (\lambda_k)_{k=1,\dots,m} \in [0,1]^m$, we obtain the lower bound

$$\begin{aligned} &\inf_{\boldsymbol{w} \in \mathbb{R}^N} \frac{\sum_{k=1}^m \lambda_k g_{i_k,\varepsilon}(\boldsymbol{w})}{f_j(\boldsymbol{w})} \\ &= \inf_{\boldsymbol{w} \in \mathbb{R}^N} (1-\varepsilon)^{N/2} \cdot \left(\sum\nolimits_{k=1}^m \lambda_k e^{(1-\varepsilon) w_{i_k} - w_j} \right) \cdot \exp\left(\frac{\varepsilon}{2} \left(\boldsymbol{w} + \frac{\boldsymbol{\sigma}}{2} \right)^\top \mathbf{C}^{-1} \left(\boldsymbol{w} + \frac{\boldsymbol{\sigma}}{2} \right) \right) \\ &\geq \inf_{\boldsymbol{w} \in \mathbb{R}^N} (1-\varepsilon)^{N/2} \cdot \exp\left((1-\varepsilon) \sum_{k=1}^m \lambda_k w_{i_k} - w_j \right) \cdot \exp\left(\frac{\varepsilon}{2} \left(\boldsymbol{w} + \frac{\boldsymbol{\sigma}}{2} \right)^\top \mathbf{C}^{-1} \left(\boldsymbol{w} + \frac{\boldsymbol{\sigma}}{2} \right) \right) \\ &=: c_I^{(j)}(\varepsilon, \boldsymbol{\lambda}), \end{aligned}$$

where we made use of the convexity of the exponential function. Setting

$$\kappa_I^{(j)}(\varepsilon, \boldsymbol{\lambda}) = (1-\varepsilon) \sum_{k=1}^m \lambda_k \mathbf{C}_{\cdot i_k} - \mathbf{C}_{\cdot j},$$

this bound can be calculated explicitly:

$$\begin{aligned} &c_I^{(j)}(\varepsilon, \boldsymbol{\lambda}) \\ &= \inf_{\boldsymbol{w} \in \mathbb{R}^N} (1-\varepsilon)^{N/2} \cdot \exp\left\{ \frac{\varepsilon}{2} \left(\boldsymbol{w} + \frac{1}{\varepsilon} \kappa_I^{(j)}(\varepsilon, \boldsymbol{\lambda}) + \frac{\boldsymbol{\sigma}}{2} \right)^\top \mathbf{C}^{-1} \left(\boldsymbol{w} + \frac{1}{\varepsilon} \kappa_I^{(j)}(\varepsilon, \boldsymbol{\lambda}) + \frac{\boldsymbol{\sigma}}{2} \right) \right\} \\ &\qquad \cdot \exp\left(-(\kappa_I^{(j)}(\varepsilon, \boldsymbol{\lambda}))^\top \mathbf{C}^{-1} \frac{\boldsymbol{\sigma}}{2} - \frac{1}{2\varepsilon} (\kappa_I^{(j)}(\varepsilon, \boldsymbol{\lambda}))^\top \mathbf{C}^{-1} \kappa_I^{(j)}(\varepsilon, \boldsymbol{\lambda}) \right) \end{aligned}$$

$$= (1-\varepsilon)^{N/2} \cdot \exp\left(-\frac{1-\varepsilon}{2} \sum_{k=1}^{m} \lambda_k C_{i_k i_k} + \frac{1}{2} C_{jj} - \frac{(1-\varepsilon)^2}{2\varepsilon} \sum_{k=1}^{m} \lambda_k^2 C_{i_k i_k} - \frac{1}{2\varepsilon} C_{jj} \right)$$
$$\cdot \exp\left(-\frac{(1-\varepsilon)^2}{2\varepsilon} \sum_{k=1}^{m} \sum_{l \neq k} \lambda_k \lambda_l C_{i_k i_l} + \frac{1-\varepsilon}{\varepsilon} \sum_{k=1}^{m} \lambda_k C_{i_k j} \right)$$
$$= (1-\varepsilon)^{N/2} \exp\left(-\frac{1-\varepsilon}{\varepsilon} \sum_{k=1}^{m} \lambda_k \gamma(t_{i_k} - t_j) + \frac{(1-\varepsilon)^2}{2\varepsilon} \sum_{k=1}^{m} \sum_{l=1}^{m} \lambda_k \lambda_l \gamma(t_{i_k} - t_{i_l}) \right). \quad (5.19)$$

Hence, for $\boldsymbol{w} \in \mathbb{R}^N$,

$$\sum_{i \in I} p_i g_{i,\varepsilon}(\boldsymbol{w}) = \|\mathbf{p}_I\|_1 \cdot \sum_{i \in I} \frac{p_i}{\|\mathbf{p}_I\|_1} g_{i,\varepsilon}(\boldsymbol{w}) \geq \|\mathbf{p}_I\|_1 \cdot c_I^{(j)} \left(\varepsilon, \frac{\mathbf{p}_I}{\|\mathbf{p}_I\|_1} \right) \cdot f_j(\boldsymbol{w}),$$

where $\mathbf{p}_I = (p_i)_{i \in I}$ for every subset $I \subset \{1, \dots, N\}$. Now, for each $j \in \{1, \dots, N\}$, let $J^{(j)}$ be a partition of $\{1, \dots, N\}$, so that

$$\begin{aligned} c_\infty \cdot C(\mathbf{p}, \varepsilon) &= \inf_{\boldsymbol{w} \in \mathbb{R}^N} \frac{g_{\text{prop}}(\boldsymbol{w})}{f_{\max}(\boldsymbol{w})} = \inf_{\boldsymbol{w} \in \mathbb{R}^N} \min_{j=1}^{N} \frac{\sum_{I \in J^{(j)}} \sum_{i \in I} p_i g_{i,\varepsilon}(\boldsymbol{w})}{c_\infty^{-1} f_j(\boldsymbol{w})} \\ &\geq \inf_{\boldsymbol{w} \in \mathbb{R}^N} \min_{j=1}^{N} \frac{\sum_{I \in J^{(j)}} \left(\sum_{i \in I} p_i \right) \cdot c_I^{(j)} \left(\varepsilon, \frac{\mathbf{p}_I}{\|\mathbf{p}_I\|_1} \right) \cdot f_j(\boldsymbol{w})}{c_\infty^{-1} f_j(\boldsymbol{w})} \\ &= c_\infty \cdot \min_{j=1}^{N} \sum_{I \in J^{(j)}} \left(\sum_{i \in I} p_i \right) \cdot c_I^{(j)} \left(\varepsilon, \frac{\mathbf{p}_I}{\|\mathbf{p}_I\|_1} \right). \end{aligned} \quad (5.20)$$

Thus, the RHS of (5.20) provides an explicit lower bound for the average acceptance probability for any choice of the $J^{(j)}$.

Remark 5.4. Assume that, for some index $j \in \{1, \dots, N\}$, there is some index set $I = \{i_1, \dots, i_m\} \subset \{1, \dots, N\}$ such that $\gamma(t_i - t_j) = \Gamma$ for all $i \in I$. Then, Equation (5.19) provides the bound

$$\begin{aligned} \sum_{k=1}^{m} p_{i_k} g_{i_k,\varepsilon}(\boldsymbol{w}) &\geq \|\mathbf{p}_I\|_1 c_I^{(j)} \left(\varepsilon, \frac{\mathbf{p}_I}{\|\mathbf{p}_I\|_1} \right) f_j(\boldsymbol{w}) \\ &= \|\mathbf{p}_I\|_1 \exp\left(-\frac{1-\varepsilon}{\varepsilon} \Gamma + \frac{(1-\varepsilon)^2}{2\varepsilon} \sum_{k=1}^{m} \sum_{l=1}^{m} \frac{p_{i_k} p_{i_l}}{\|\mathbf{p}_I\|_1^2} \gamma(t_{i_k} - t_{i_l}) \right) f_j(\boldsymbol{w}) \end{aligned} \quad (5.21)$$

for all $\boldsymbol{w} \in \mathbb{R}^N$. Alternatively, for the same index set I, we could bound each summand separately, i.e.

$$\sum_{k=1}^{m} p_{i_k} g_{i_k,\varepsilon}(\boldsymbol{w}) \geq \sum_{k=1}^{m} p_{i_k} c_{\{i_k\}}^{(j)}(\varepsilon, 1) f_j(\boldsymbol{w}) = \|\mathbf{p}_I\|_1 \exp\left(-\frac{1-\varepsilon}{\varepsilon} \Gamma \right) f_j(\boldsymbol{w}). \quad (5.22)$$

Note that, for all $\boldsymbol{w} \in \mathbb{R}^N$, the RHS of (5.22) is less than the RHS of (5.21), i.e. the lower bound is less sharp. Therefore, we prefer pooling locations with the same distance to t_j rather than considering them separately in order to have the bound in (5.20) as sharp as possible.

In view of (5.20), instead of considering the exact value $C(\mathbf{p}, \varepsilon)$ which is needed to calculate the minimal rejection rate, but cannot be given explicitly, we might maximize the function

$$C_{\text{groups}}(\mathbf{p}, \varepsilon) = \min_{j=1}^{N} \sum_{I \in J^{(j)}} \sum_{i \in I} p_i c_I^{(j)} \left(\varepsilon, \frac{\mathbf{p}_I}{\|\mathbf{p}_I\|} \right)$$

for fixed partitions $J^{(1)}, \dots, J^{(N)}$. Due to the complex dependence of $c_I^{(j)}$ on $\mathbf{p}$, the resulting optimization problem is nonlinear in $\mathbf{p}$ even for fixed ε. To circumvent this difficulty, for each $I \in J^{(j)}$, we fix $|I|$-dimensional weight vectors $\boldsymbol{\lambda}(I)$ and consider the function

$$C_{\text{groups}}^{\text{fix}}(\mathbf{p}, \varepsilon; \boldsymbol{\lambda}) = \min_{j=1}^{N} \sum_{I \in J^{(j)}} \sum_{i \in I} p_i c_I^{(j)} \left(\varepsilon, \boldsymbol{\lambda}(I) \right) = \min_{j=1}^{N} \sum_{i=1}^{N} p_i c_{ij}(\varepsilon; \boldsymbol{\lambda}) = \min \mathbf{p}^\top \mathbf{c}(\varepsilon; \boldsymbol{\lambda})$$

with $\mathbf{c}(\varepsilon; \boldsymbol{\lambda}) = \{c_{ij}(\varepsilon; \boldsymbol{\lambda})\}_{1 \le i,j \le N}$ and $c_{ij}(\varepsilon; \boldsymbol{\lambda}) = c_I^{(j)}(\varepsilon; \boldsymbol{\lambda}(I))$ for the unique set $I \in J^{(j)}$ such that $i \in I$.

Analogously to the solution above, we first maximize $C_{\text{groups}}^{\text{fix}}(\cdot, \varepsilon; \boldsymbol{\lambda})$ for fixed $\varepsilon \in [0, 1)$ and $\boldsymbol{\lambda}$, i.e. we consider the optimization problem

$$\begin{aligned} &\max_{\mathbf{p} \in \mathbb{R}^N} \min_{j=1}^{N} c_{\cdot j}(\varepsilon; \boldsymbol{\lambda})^\top \mathbf{p} \\ &\text{s.t.} \quad \|\mathbf{p}\|_1 = 1 \\ &\qquad p_i \ge 0 \qquad \forall i = 1, \dots, N\,. \end{aligned} \tag{LP1}$$

To convert the linear program to standard form, we introduce an additional variable $z \in \mathbb{R}$, unconstrained in sign, leading to the equivalent program

$$\begin{aligned} &\max_{\mathbf{p} \in \mathbb{R}^N, z \in \mathbb{R}} z \\ &\text{s.t.} \quad z \le c_{\cdot j}(\varepsilon; \boldsymbol{\lambda})^\top \mathbf{p} \qquad \forall j = 1, \dots, N \\ &\qquad \mathbf{1}_N^\top \mathbf{p} = 1 \\ &\qquad p_i \ge 0 \qquad \forall i = 1, \dots, N\,. \end{aligned} \tag{LP2}$$

The standard form of (LP2) is then given by

$$\begin{aligned} &\max_{\substack{p \in \mathbb{R}^N \\ s, z^+, z^- \in \mathbb{R}}} z^+ - z^- \\ &\text{s.t.} \quad z^+ - z^- + s = c_{\cdot j}(\varepsilon; \boldsymbol{\lambda})^\top \mathbf{p} \qquad \forall j = 1, \dots, N \\ &\qquad \mathbf{1}_N^\top \mathbf{p} = 1 \\ &p_1, \dots, p_N, s, z^+, z^- \ge 0\,. \end{aligned} \tag{LP2S}$$

Such a linear program in standard form can be solved by standard techniques such as the simplex algorithm. Compared to the optimization of $C_{\text{groups}}(\cdot, \varepsilon)$, the complementary one-dimensional problem of maximizing $C_{\text{groups}}(\mathbf{p}, \cdot)$ for fixed $\mathbf{p}$ can be solved rather easily.

To summarize, starting from some $\varepsilon^* > 0$ and $\mathbf{p}^* = N^{-1}\mathbf{1}_N$, we propose to apply the following two steps repeatedly (Algorithm 2B):

1. Define

$$\boldsymbol{\lambda}_I = \frac{\mathbf{p}_I^*}{\|\mathbf{p}_I^*\|_1}, \quad I \in J^{(1)} \cup \ldots \cup J^{(N)}$$

and set

$$\mathbf{p}^* = \operatorname*{argmax}_{\mathbf{p}} C_{\text{groups}}^{\text{fix}}(\mathbf{p}, \varepsilon^*; \boldsymbol{\lambda}),$$

i.e. the solution of the optimization problem (LP1) (or (LP2) or (LP2S), equivalently).

2. Set $\varepsilon^* = \operatorname{argmax}_\varepsilon C_{\text{groups}}(\mathbf{p}^*, \varepsilon)$.

Even though $C_{\text{groups}}(\mathbf{p}, \varepsilon)$ might be significantly smaller than $C(\mathbf{p}, \varepsilon)$, in some cases, this bound is already sufficient to improve the results for $\varepsilon = 0$ that have been discussed in the beginning of this section, where we have already seen that the corresponding optimal weight vector equals $\mathbf{p}^* = N^{-1}\mathbf{1}_N$ and that $C(\mathbf{p}^*, 0) = 1/N$. We show an example to illustrate that this choice is not necessarily optimal, i.e. there is some $\varepsilon > 0$ and a vector $\mathbf{p}$ of weights such that $C(\mathbf{p}, \varepsilon) \geq C_{\text{groups}}(\mathbf{p}, \varepsilon) > 1/N$.

Example 5.5 (Fractional Brownian Motion)**.** *Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be N equidistant locations in $[0, 1]$ and $\gamma(h) = |h|^\alpha$ for some $\alpha > 1$. Choose $\mathbf{p} = N^{-1}\mathbf{1}_N$, $\varepsilon = \sqrt{2}N^{-1}$ and set $J^{(1)} = \ldots = J^{(N)} = \{\{1\}, \ldots, \{N\}\}$. Then, for every x_i there are at least $\lfloor N^{-1/\alpha} \cdot N \rfloor$ locations $\mathbf{x}_j$ such that $\gamma(\mathbf{x}_i - \mathbf{x}_j) \leq 1/N$. Thus, we obtain for large N that*

$$\begin{aligned} C(\mathbf{p}, \varepsilon) &\geq C_{\text{groups}}(\mathbf{p}, \varepsilon) \\ &\geq \lfloor N^{-1/\alpha} \cdot N \rfloor \cdot \frac{1}{N} \cdot \left(1 - \frac{\sqrt{2}}{N}\right)^{N/2} \exp\left(-\frac{N}{\sqrt{2}} \cdot \frac{1}{N}\right) \sim N^{-1/\alpha} \exp(-\sqrt{2}) \end{aligned}$$

which is eventually larger than $1/N$ as $\alpha > 1$.

5.4 Illustration

Finally, we illustrate the performance of Algorithm 5.1 and the rejection sampling algorithm in an example. Taking up Example 5.5 in higher dimensions, we consider the case that Z is a Brown–Resnick process associated with the variogram

$$\gamma(h) = \left\| \frac{h}{5} \right\|^{1.5}, \quad h \in \mathbb{R}^2,$$

on the grid $K = \{0, 0.2, \ldots, 5\} \times \{0, 0.2, \ldots, 5\}$ ($N = 676$ points). We run four different algorithms:

1A. Algorithm 5.1 with proposal density $f_{\text{prop}} = f$ as proposed by Oesting et al. (2018b)

1B. Algorithm 5.1 with proposal density $f_{\text{prop}} = \sum_{i=1}^N p_i f_i$ where $\mathbf{p}$ is given as the solution of (QP) (cf. Section 5.2).

2A. Algorithm 5.2 with proposal density $f_{\text{prop}} = \frac{1}{N}\sum_{i=1}^N f_i$ and $C = 1/N$ (equivalent to the procedure proposed in de Fondeville and Davison (2018) based on sum-normalized spectral functions)

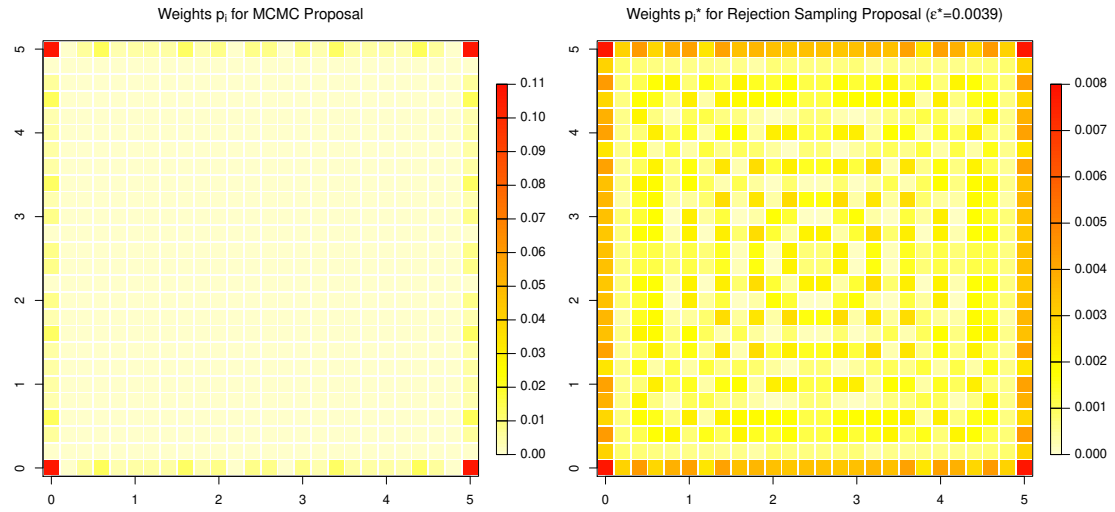

Figure 5.1: Vectors $\mathbf{p}$ (left) and $\mathbf{p}^*$ (right) of optimal weights used in Algorithms 1B and 2B, respectively, for a Brown–Resnick process associated with the variogram $\gamma(h) = \|h/5\|^{1.5}$ on the set $K = \{0, 0.2, \dots, 5\}^2$.

2B. Algorithm 5.2 with proposal density $\tilde{f}_{\text{prop}} = \sum_{i=1}^{N} p_i^*(\varepsilon^*) g_{i,\varepsilon^*}$ and $C = C(\mathbf{p}^*, \varepsilon^*)$ in (5.18) where $\mathbf{p}^*(\varepsilon^*) \in [0,1]^N$ and $\varepsilon^* \in (0,1)$ are obtained as described in Section 5.3

Even though the laws of the Brown–Resnick process Z and the normalized spectral process $V^{\max}$ do not depend on the variance, but only on the variogram of the underlying Gaussian process G, the choice of the Gaussian process may affect the performance of the algorithms. Here, we choose the Gaussian process G whose law is uniquely defined via the construction

$$G(t) = G_0(t) - \frac{1}{4}\left(G_0\left(\begin{pmatrix}0\\0\end{pmatrix}\right) + G_0\left(\begin{pmatrix}5\\0\end{pmatrix}\right) + G_0\left(\begin{pmatrix}0\\5\end{pmatrix}\right) + G_0\left(\begin{pmatrix}5\\5\end{pmatrix}\right)\right), \quad t \in K,$$

where G_0 is an arbitrary centered Gaussian process with variogram γ. Oesting and Strokorb (2018) show that this process has a smaller maximal variance and is thus preferable in the context of simulation.

We first calculate the optimal weights $\mathbf{p} = (p_1, \dots, p_{676})^\top$ as a solution of (QP) (used in Algorithm 1B) as well as the optimal weights $\mathbf{p}^*(\varepsilon^*)$ as a solution of (LP1) and the optimal variance modification ε^* (used in Algorithm 2B). The results for $\mathbf{p}$ and $\mathbf{p}^*(\varepsilon^*)$ are displayed in Figure 5.1. It can be seen that, in both cases, the weights are not spatially constant, but are larger on the boundary of the convex hull $\text{conv}(K) = [0,5] \times [0,5]$ with the maximum in the corners of the square. This observation is well in line with the fact that these points have the largest contribution to $\max_{t \in K} \exp(G(t) - \text{Var}(G(t))/2)$ since the variance of G attains its maximum there (see also Oesting and Strokorb, 2018).

We run Algorithm 5.1 with both proposal densities as specified above (Algorithms 1A and 1B, respectively) to obtain two different Markov chains $\{W_1^{(k)}\}_{k=1,\dots,1\,000\,000}$ and $\{W_2^{(k)}\}_{k=1,\dots,1\,000\,000}$ of length $n_{MCMC} = 1\,000\,000$. It can be seen that the empirical acceptance rate of the second chain (0.855) is remarkably higher than the one of the first chain (0.656). The consequences on the mixing properties of the chains can be discovered by analyzing the empirical autocorrelation functions of the time series $\{\|\exp(W_1^{(k)})\|_\infty\}_{k=1,\dots,1\,000\,000}$ and $\{\|\exp(W_2^{(k)})\|_\infty\}_{k=1,\dots,1\,000\,000}$ which are shown in Figure 5.2. Here, the empirical autocorrelation in the second chain is drastically reduced in comparison with the first chain, indicating that two states of the chain can be regarded as nearly uncorrelated after roughly five steps.

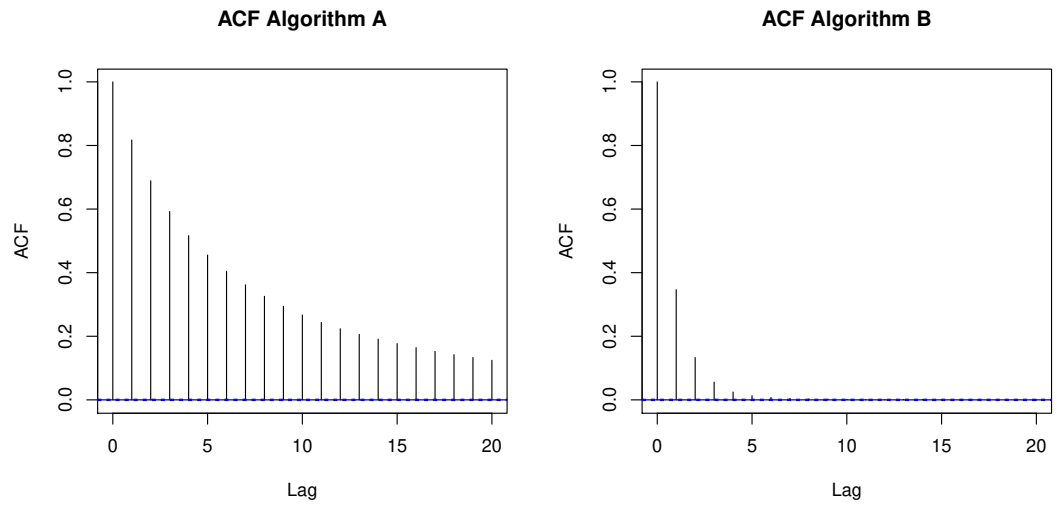

Figure 5.2: Empirical autocorrelation functions of the two different time series $\{\|\exp(W_1^{(k)})\|_\infty\}_{k=1,\ldots,1\,000\,000}$ (left) and $\{\|\exp(W_2^{(k)})\|_\infty\}_{k=1,\ldots,1\,000\,000}$ (right) obtained via Algorithms 1A and 1B, respectively. Here, by the optimal choice of $\mathbf{p}$, the autocorrelation is clearly reduced.

The rejection sampling algorithm (Algorithm 2A and Algorithm 2B) automatically generates independent realizations from the multivariate target density $f_{\max}$. Therefore, we will compare them with respect to their computational complexity. To this end, we run them to generate a sample of size 100 000 and count the average number of simulations of Gaussian vectors from the proposal density to generate one realization of $\mathbf{W}^{\max}$. In case of Algorithm 2A, this number is 203.1 which is close to the theoretical expression $c_\infty^{-1} \cdot N$. For Algorithm 2B, the number is improved by a factor of approximately 4.4, leading to an average number of 45.9 Gaussian vectors to be simulated to obtain one realization from the target distribution. This improvement is well in line with the corresponding value $C(\mathbf{p}^*, \varepsilon^*) \approx 0.0065 \approx 4.4 \cdot N^{-1}$ where N^{-1} corresponds to the constant C in (5.15) for Algorithm 2A.

As the example illustrates, the two modifications we suggested may lead to significant improvements of MCMC and rejection algorithms that have been proposed so far. Here, only the modified rejection sampling algorithm ensures independence of exact samples from the target distribution. However, as the example indicates, the MCMC algorithm might be particularly attractive in practice as a thinned chain results in nearly independent samples even if the thinning rate is rather small. Note that we also tried other examples such as a Brownian sheet ($\alpha = 1$). However, we found that significant improvements in the rejection sampling procedure become apparent only for $\alpha > 1$, see also Example 5.5.

Acknowledgements

The authors are grateful to Kirstin Strokorb, Dimitri Schwab and Jonas Brehmer for numerous valuable comments. M. Schlather has been financially supported by Volkswagen Stiftung within the project “Mesoscale Weather Extremes – Theory, Spatial Modeling and Prediction (WEX-MOP).”

6 Exact Simulation of Max-Stable Processes

joint work with Clément Dombry and Sebastian Engelke

This chapter is based on the research article Dombry et al. (2016a) that has been published in *Biometrika* and the corresponding supplementary material. Besides some minor modifications, also some changes in the structure of the chapter have been made: the different parts of the appendix and the supplementary material have been shifted to appropriate places in the main body of the chapter.

Max-stable processes play an important role as models for spatial extreme events. Their complex structure as the pointwise maximum over an infinite number of random functions makes their simulation difficult. Algorithms based on finite approximations are often inexact and computationally inefficient. In this chapter, we present a new algorithm for exact simulation of a max-stable process at a finite number of locations. It relies on the idea of simulating only the extremal functions, that is, those functions in the construction of a max-stable process that effectively contribute to the pointwise maximum. We further generalize the algorithm by Dieker and Mikosch (2015) for Brown–Resnick processes and use it for exact simulation via the spectral measure. We study the complexity of both algorithms, prove that our new approach via extremal functions is always more efficient, and provide closed-form expressions for their implementation that cover most popular models for max-stable processes and multivariate extreme value distributions. For simulation on dense grids, an adaptive design of the extremal function algorithm is proposed.

6.1 Introduction

Max-stable processes have become widely used tools to model spatial extreme events. Occurring naturally in the context of extremes as limits of maxima of independent copies of stochastic processes, they have found many applications in environmental science; see for instance Coles (1993), Buishand et al. (2008), Blanchet and Davison (2011) and Davison et al. (2012).
Any sample continuous max-stable process Z with unit Fréchet margins on some compact domain $\mathcal{X} \subset \mathbb{R}^d$ is characterized by a point process representation (de Haan, 1984)

$$Z(x) = \max_{i \geq 1} \zeta_i \psi_i(x), \quad x \in \mathcal{X}, \tag{6.1}$$

where $\{(\zeta_i, \psi_i), i = 1, 2, \dots\}$ is a Poisson point process on $(0, \infty) \times C_+(\mathcal{X})$ with intensity measure $\zeta^{-2}\, \mathrm{d}\zeta \times \nu(\mathrm{d}\psi)$ for some locally finite measure ν on the space $C_+(\mathcal{X})$ of continuous non-negative functions on $\mathcal{X}$ equipped with the Borel σ-algebra $\mathcal{C}_+(\mathcal{X})$ such that

$$\int_{C_+(\mathcal{X})} \psi(x) \nu(\mathrm{d}\psi) = 1, \quad x \in \mathcal{X}. \tag{6.2}$$

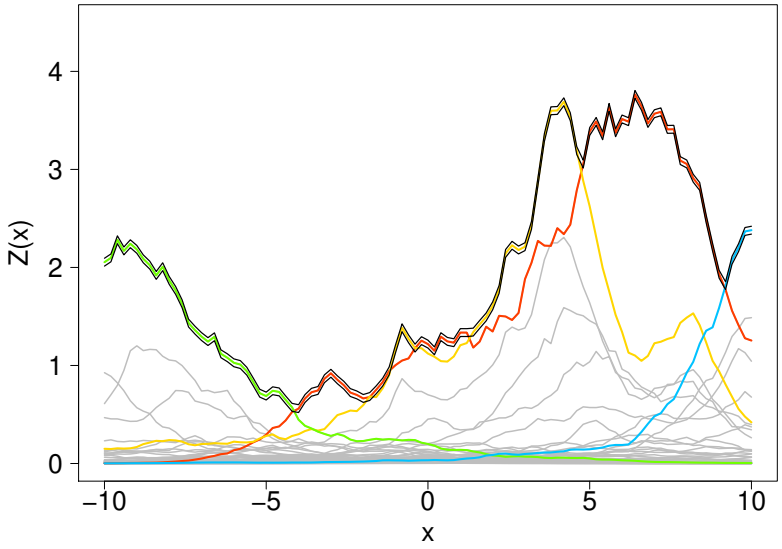

Figure 6.1: The Poisson point process $\{(\zeta_i, \psi_i), i \geq 1\}$ (grey). Only finitely many (colored) (ζ_i, ψ_i) contribute to the maximum process Z (bordered in black).

Figure 6.1 shows a realization of Z composed of those random functions of the above point process that are maximal at some location. Due to this complex structure of max-stable processes, in many cases, analytical expressions are only available for lower-dimensional distributions and related characteristics need to be assessed by simulations. Moreover, non-conditional simulation is an important part of conditional simulation procedures that can be used to predict extreme events given some additional information (e.g., Dombry et al., 2013; Oesting and Schlather, 2014). Thus, there is a need for fast and accurate simulation algorithms.

As the spectral representation (6.1) involves an infinite number of functions, exact simulation of Z is in general not straightforward and finite approximations are used in practice. For the widely used Brown–Resnick processes (Kabluchko et al., 2009), Engelke et al. (2011) and Oesting et al. (2012) exploit the fact that the representation (6.1) is not unique in order to propose simulation procedures based on equivalent representations. However, often these approximations do not provide satisfactory results in terms of accuracy or computational effort. The effect of the approximation can be illustrated in Fig. 6.1, where an approximate algorithm might miss one or several of the colored processes and the resulting maximum process would be strictly smaller than the exact realization Z.

Exact simulation procedures can so far be implemented only in special cases. Schlather (2002) proposes an algorithm that simulates the points $\{\zeta_i, i \geq 1\}$ in (6.1) in descending order until some stopping rule takes effect. If ν is the probability measure of a stochastic process whose supremum on $\mathcal{X}$ is almost surely bounded or if Z is a mixed moving maxima process with uniformly bounded and compactly supported shape function, this procedure allows exact simulation of Z. For extremal-t processes (Opitz, 2013), the elliptical structure of Gaussian processes can be exploited to obtain exact samples (Thibaud and Opitz, 2015). Oesting et al. (2018b) focus on a class of equivalent representations for general max-stable processes that, in principle, allow for optimally efficient exact simulation. They propose to simulate max-stable processes via the normalized spectral representation with all the spectral functions sharing the same supremum. Being efficient with respect to the number of spectral functions, the simulation of a single normalized function might be rather intricate in some cases including Brown–Resnick processes. For the latter, Dieker

and Mikosch (2015) derived a new representation that enables exact simulation at finitely many locations.
Several articles focus on the simulation of finite dimensional max-stable distributions or, equivalently, of their associated extreme value copula. Ghoudi et al. (1998) and Capéraà et al. (2000) propose simulation procedures for certain bivariate extreme value distributions. Stephenson (2003) considers extreme value distributions of logistic type. Boldi (2009) provides a method for exact simulation from the spectral measure of extremal Dirichlet and logistic distributions.
In this chapter, we consider the problem of exact simulation of a general max-stable process Z at a finite number of locations. We introduce a new procedure based on the idea to simulate only the extremal functions (cf. Dombry and Éyi-Minko, 2012, 2013) out of the infinite set $\{\zeta_i\psi_i, i \geq 1\}$, i.e., those functions that satisfy $\zeta_i\psi_i(x) = Z(x)$ for some $x \in \mathcal{X}$, the colored functions in Fig. 6.1. In contrast to all existing simulation procedures, the process Z is not simulated simultaneously, but successively at different locations, rejecting all those functions that are not compatible with the process at the locations simulated so far. We propose also a second procedure that relies on sampling from the spectral measure on the L_1-sphere of a multivariate extreme value distribution. Interestingly, in the case of Brown–Resnick processes, this second procedure turns out to be identical to the algorithm by Dieker and Mikosch (2015). We prove that the new procedure based on extremal functions is computationally more efficient than simulation via the spectral measure. Both procedures are based on random functions following the same type of distribution that can be easily simulated for most popular max-stable models. Both algorithms also apply very efficiently to exact simulation of finite-dimensional max-stable distributions or, equivalently, of the associated extreme value copulas.

6.2 Simulation via Extremal Functions

In Sections 6.2 and 6.3, we will propose two procedures for exact simulation of arbitrary max-stable processes and distributions. More precisely, for a fixed number $N \in \mathbb{N}$ of distinct locations $x = (x_1, \ldots, x_N) \in \mathcal{X}^N$, we aim at obtaining exact simulation of the max-stable random vector

$$Z(x) = \{Z(x_1), \ldots, Z(x_N)\}, \tag{6.3}$$

where Z is a sample-continuous process given by the spectral representation (6.1). Without loss of generality, we may restrict to processes with unit Fréchet margins as any sample-continuous max-stable process can be obtained from a process with unit Fréchet margins via marginal transformations. The first procedure, presented in this section, relies on conditional distributions of the Poisson point process underlying the max-stable process. This allows for exact simulation of (6.3) by simulating at each location only the unique function that actually attains the maximum, see Fig. 6.1. In the following, we will briefly present some results on the distribution of this function, the so-called extremal function. Throughout, we write $f(x) = \{f(x_1), \ldots, f(x_N)\}$ for the restriction of a generic, possibly random, function f to the locations $x \in \mathcal{X}^N$.

Starting from representation (6.1), we use a point process approach and recall that the $C_+(\mathcal{X})$-valued point process $\Phi = \{\phi_i\}_{i\geq 1}$ with $\phi_i = \zeta_i\psi_i$ is a Poisson point process with intensity

$$\mu(A) = \int_{C_+(\mathcal{X})} \int_0^\infty \mathbf{1}_{\{\zeta\psi \in A\}} \zeta^{-2} \,\mathrm{d}\zeta \,\nu(\mathrm{d}\psi), \qquad A \in \mathcal{C}_+(\mathcal{X}), \tag{6.4}$$

where $\mathbf{1}_{\{L\}}$ denotes the indicator function of a logical expression L, i.e. $\mathbf{1}_{\{L\}} = 1$ if L is true and $\mathbf{1}_{\{L\}} = 0$ otherwise.

Definition 6.1. Let $K \subset \mathcal{X}$ be a nonempty compact subset. A function $\phi \in \Phi$ is called K-extremal if there is some $x \in K$ such that $\phi(x) = Z(x)$, otherwise the function is called K-subextremal. We denote by Φ_K^+ the set of K-extremal functions and by Φ_K^- the set of K-subextremal functions.

It can be shown that Φ_K^+ and Φ_K^- are properly defined Poisson point processes. When $K = \{x_0\}$, $x_0 \in \mathcal{X}$, is reduced to a single point, it is easy to show that $\Phi_{\{x_0\}}^+$ is also almost surely reduced to a single point which we denote by $\phi_{x_0}^+$, termed the extremal function at x_0. The distribution of $\phi_{x_0}^+$ is given in the next proposition.

Proposition 6.2 (Dombry and Éyi-Minko (2013), Proposition 4.2)**.** *The random variables $Z(x_0)$ and $\phi_{x_0}^+/Z(x_0)$ are independent. Furthermore, $Z(x_0)$ has a unit Fréchet distribution and the distribution of $\phi_{x_0}^+/Z(x_0)$ is*

$$P_{x_0}(A) = \mathbb{P}\left\{\phi_{x_0}^+/Z(x_0) \in A\right\} = \int_{C_+(\mathcal{X})} \mathbf{1}_{\{f/f(x_0)\in A\}} f(x_0)\, \nu(\mathrm{d}f), \quad A \in \mathcal{C}_+(\mathcal{X}). \tag{6.5}$$

By definition, $\phi_{x_0}^+(x_0) = Z(x_0)$, so the distribution P_{x_0} is supported by the subset of functions $\{f \in C_+(\mathcal{X}), f(x_0) = 1\}$.

Proposition 6.3. *The restricted point process $\Phi \cap \{f \in C_+(\mathcal{X}), f(x_0) > 0\}$ is a Poisson point process with intensity*

$$\int_A \mathbf{1}_{\{f(x_0)>0\}} \mu(\mathrm{d}f) = \int_{C_+(\mathcal{X})} \int_0^\infty \mathbf{1}_{\{\zeta f\in A\}} \zeta^{-2}\, \mathrm{d}\zeta\, P_{x_0}(\mathrm{d}f), \quad A \in \mathcal{C}_+(\mathcal{X}). \tag{6.6}$$

Proof. The fact that the restricted point process $\Phi \cap \{f \in C_+(\mathcal{X}), f(x_0) > 0\}$ is a Poisson point process with intensity $\mathbf{1}_{\{f(x_0)>0\}}\mu(\mathrm{d}f)$ is standard. We prove Equation (6.6). For $A \in \mathcal{C}_+(\mathcal{X})$,

$$\begin{aligned}
&\int_{C_+(\mathcal{X})} \int_0^\infty \mathbf{1}_{\{\zeta f\in A\}} \zeta^{-2}\, \mathrm{d}\zeta\, P_{x_0}(\mathrm{d}f) = \int_{C_+(\mathcal{X})} \int_0^\infty \mathbf{1}_{\{\zeta f/f(x_0)\in A\}} \zeta^{-2}\, \mathrm{d}\zeta\, f(x_0)\, \nu(\mathrm{d}f) \\
&= \int_{C_+(\mathcal{X})} \int_0^\infty \mathbf{1}_{\{\tilde\zeta f\in A\}} \tilde\zeta^{-2}\, \mathrm{d}\tilde\zeta\, \mathbf{1}_{\{f(x_0)>0\}}\, \nu(\mathrm{d}f) = \int_{C_+(\mathcal{X})} \mathbf{1}_{\{f\in A\}} \mathbf{1}_{\{f(x_0)>0\}}\, \mu(\mathrm{d}f).
\end{aligned}$$

Here, we use successively Equation (6.5), the change of variable $\tilde\zeta = \zeta/f(x_0)$ with $f(x_0) > 0$ and Equation (6.4) for the last equality. □

Remark 6.4. As a consequence of (6.6), independent copies $Y_1, Y_2, \ldots$ of processes with distribution P_{x_0} result in a point process $\{\zeta_i Y_i\}_{i\geq 1}$ which has the same distribution as the restricted point process $\Phi \cap \{f \in C_+(\mathcal{X}), f(x_0) > 0\}$. If $\nu(\{f \in C_+(\mathcal{X}), f(x_0) = 0\}) = 0$, then Φ consists only of functions with positive value at x_0 and Φ has the same distribution as $\{\zeta_i Y_i\}_{i\geq 1}$. This provides an alternative point process representation of the max-stable process Z in terms of a random process Y such that $Y(x_0) = 1$ almost surely. Engelke et al. (2014, 2015) exploit this representation for statistical inference on Z.

These preliminary considerations on extremal functions enable us to introduce a procedure for exact simulation of the max-stable process Z at locations $x \in \mathcal{X}^N$. More precisely, for $n = 1, \ldots, N$, we consider the extremal and subextremal point processes $\Phi_n^+ = \Phi_{\{x_1,\ldots,x_n\}}^+$

and $\Phi_n^- = \Phi^-_{\{x_1,\dots,x_n\}}$. We have that Φ_n^+ equals $\{\phi^+_{x_i}\}_{1\le i\le n}$ where the cardinality of this set may be less than n as several locations may share the same extremal function. We define the nth-step maximum process

$$Z_n(x) = \max_{\phi\in\Phi_n^+} \phi(x) = \max_{1\le i\le n} \phi^+_{x_i}(x), \quad x \in \mathcal{X}. \tag{6.7}$$

By the definition of extremal functions we have $Z(x_i) = \phi^+_{x_i}(x_i)$ and clearly

$$Z(x_i) = Z_n(x_i), \quad i = 1,\dots,n. \tag{6.8}$$

Hence, in order to exactly simulate Z at locations x, it is enough to exactly simulate Φ_N^+. We will proceed inductively and simulate the sequence $(\phi^+_{x_n})_{1\le n\le N}$ according to the following theorem.

Theorem 6.5. *The distribution of $(\phi^+_{x_n})_{1\le n\le N}$ is given by the following sequential procedure. The initial distribution for the the extremal function $\phi^+_{x_1}$ has the same distribution as F_1Y_1 where F_1 is a unit Fréchet random variable and Y_1 an independent random process with distribution P_{x_1} given by* (6.5).
For $1 \le n \le N-1$, the conditional distribution of $\phi^+_{x_{n+1}}$ given $(\phi^+_{x_i})_{1\le i\le n}$ is equal to the distribution of

$$\tilde\phi^+_{x_{n+1}} = \begin{cases} \operatorname{argmax}_{\phi\in\tilde\Phi_{n+1}} \phi(x_{n+1}), & \tilde\Phi_{n+1} \neq \emptyset, \\ \operatorname{argmax}_{\phi\in\Phi_n^+} \phi(x_{n+1}), & \tilde\Phi_{n+1} = \emptyset, \end{cases}$$

where $\tilde\Phi_{n+1}$ is a Poisson point process with intensity

$$\mathbf{1}_{\{f(x_i)<Z_n(x_i),\ 1\le i\le n\}}\mathbf{1}_{\{f(x_{n+1})>Z_n(x_{n+1})\}}\mu(\mathrm{d}f) \tag{6.9}$$

and Z_n is defined by (6.7).

Proof. The distribution of $\phi^+_{x_1}$ is given in Proposition 6.2. We prove the result for the conditional distribution of $\phi^+_{x_{n+1}}$ given $(\phi^+_{x_i})_{1\le i\le n}$. Recall that $\Phi_n^+ = \{\phi^+_{x_1},\dots,\phi^+_{x_n}\}$. Then, by Lemma 3.2 in Dombry and Éyi-Minko (2012), the conditional distribution of Φ_n^- given Φ_n^+ is equal to the distribution of a Poisson point process with intensity

$$\mathbf{1}_{\{f(x_i)<Z(x_i),\ 1\le i\le n\}}\mu(\mathrm{d}f) = \mathbf{1}_{\{f(x_i)<Z_n(x_i),\ 1\le i\le n\}}\mu(\mathrm{d}f), \tag{6.10}$$

where the equality follows from Equation (6.8). In order to determine $\phi^+_{x_{n+1}}$ we focus on the functions $\phi \in \Phi_n^-$ satisfying $\phi(x_{n+1}) > Z_n(x_{n+1})$ and consider the restriction

$$\tilde\Phi_{n+1} = \Phi_n^- \cap \{f \in C_+(\mathcal{X}),\ f(x_{n+1}) > Z_n(x_{n+1})\}.$$

It follows from Equation (6.10) that conditionally on $(\phi^+_{x_i})_{1\le i\le n}$, $\tilde\Phi_{n+1}$ is a Poisson point process with intensity given by Equation (6.9). We distinguish two cases. If $\tilde\Phi_{n+1} = \emptyset$ then there is no function in Φ_n^- exceeding Z_n at point x_{n+1}, that is, $Z(x_{n+1}) = Z_n(x_{n+1})$ and $\phi^+_{x_{n+1}} = \operatorname{argmax}_{\phi\in\Phi_n^+} \phi(x_{n+1})$. If $\tilde\Phi_{n+1} \neq \emptyset$ then there is some function in Φ_n^- exceeding Z_n at point x_{n+1}, that is, $Z(x_{n+1}) > Z_n(x_{n+1})$ and $\phi^+_{x_{n+1}} = \operatorname{argmax}_{\phi\in\tilde\Phi_{n+1}} \phi(x_{n+1})$. This concludes the proof of Theorem 6.5. □

From Theorem 6.5 one can deduce Algorithm 6.1 for exact simulation of the max-stable process Z at locations $x = (x_1,\dots,x_N)$. According to Proposition 6.3 and Remark 6.4, the distribution $P_{x_{n+1}}$ can be used to simulate $\tilde\Phi_{n+1}$ with intensity (6.9). Hence, the algorithm requires only that one can simulate from the distributions $P_{x_1},\dots P_{x_N}$, which

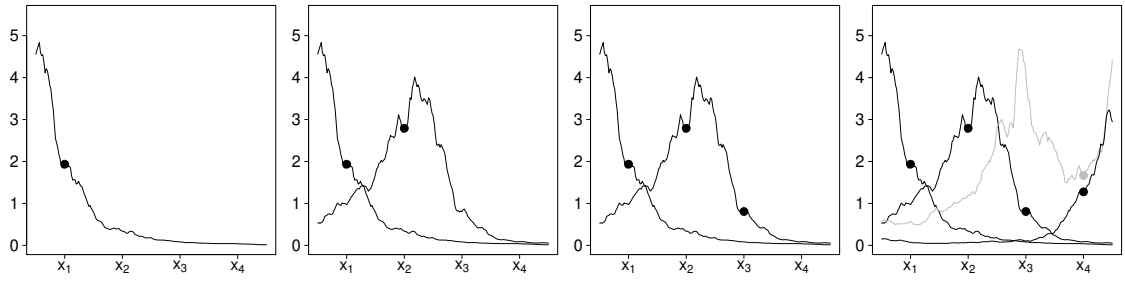

Figure 6.2: Simulation of Z via Algorithm 6.1 at locations (x_1, x_2, x_3, x_4). Initial process $\phi^+_{x_1}$ is always accepted (first panel). Second process $\phi^+_{x_2}$ is accepted as it exceeds $Z_1 = \phi^+_{x_1}$ at x_2 but not at x_1 (second panel). Third process $\phi^+_{x_3}$ is equal to $\phi^+_{x_2}$ since $\tilde{\Phi}_3 = \emptyset$ (third panel). First sample of P_{x_4} (grey line) is rejected since it exceeds Z_3 at x_3; second sample is valid and thus called $\phi^+_{x_4}$ (fourth panel).

can be easily done for the most popular max-stable models. See Section 6.4 for details. Figure 6.2 illustrates the procedure.

Algorithm 6.1: Exact simulation of a max-stable process Z at $x = (x_1, \ldots, x_N)$ via extremal functions

Simulate $\zeta^{-1} \sim \mathrm{Exp}(1)$ and $Y \sim P_{x_1}$.
Set $Z(x) = \zeta Y(x)$.
for $n = 2, \ldots, N$ **do**
  Simulate $\zeta^{-1} \sim \mathrm{Exp}(1)$.
  while $\zeta > Z(x_n)$ **do**
    Simulate $Y \sim P_{x_n}$.
    if $\zeta Y(x_i) < Z(x_i)$ *for all* $i = 1, \ldots, n-1$ **then**
      Update $Z(x)$ to the componentwise maximum $\max\{Z(x), \zeta Y(x)\}$.
    Simulate $e \sim \mathrm{Exp}(1)$ and update ζ^{-1} to $\zeta^{-1} + e$.
return Z

6.3 Simulation via the Spectral Measure

Dieker and Mikosch (2015) presented the first procedure for exact simulation of the finite-dimensional distributions of stationary Brown–Resnick processes. Applying change of measure arguments for Gaussian processes, they found an alternative representation of these processes that can be simulated easily. In this section, we introduce an approach relying on the spectral representation on the L_1-sphere that can be applied for general max-stable distributions. In the case of stationary Brown–Resnick processes, in Remark 6.11 we retrieve the algorithm of Dieker and Mikosch (2015).

Let us recall the spectral decomposition of the max-stable random vector $Z(x)$ with $x \in \mathcal{X}^N$; for details see (Resnick, 1987, Chapter 5). Following Equation (6.1), the vector $Z(x) = \max_{i \geq 1} \zeta_i \psi_i(x)$ is generated by the Poisson point process $\Phi_x = \{\zeta_i \psi_i(x), i \geq 1\}$ whose intensity measure on the cone $D = [0, \infty)^N \setminus \{0\}$ is denoted by μ_x. Due to its homogeneity, the exponent measure μ_x can be factorized into a radial part on $(0, \infty)$ and an angular part on the unit L_1-sphere $S_{N-1} = \{z \in D : \|z\| = 1\}$, where $\|z\| = z_1 + \cdots + z_N$, for $z = (z_1, \ldots, z_N) \in D$. More precisely, a change to polar coordinates under the map

$U : D \to (0,\infty) \times S_{N-1}$, $U(z) = (\|z\|, z/\|z\|)$ yields

$$\mu_x(F) = \int_{U(F)} \mu_x \circ U^{-1}(\mathrm{d}r, \mathrm{d}s) = N \int_{U(F)} r^{-2}\,\mathrm{d}r\, H(\mathrm{d}s), \tag{6.11}$$

for any Borel subset $F \subset D$. The probability measure H on S_{N-1} is called the spectral measure of $Z(x)$, and it satisfies

$$\int_{S_{N-1}} s_j H(ds) = N^{-1}, \quad j = 1, \dots, N.$$

Equation (6.11) shows that we can represent the process Φ_x as

$$\Phi_x = \{U^{-1}(R_i, Q_i) : i \geq 1\} = \{R_i Q_i : i \geq 1\}, \tag{6.12}$$

where $\{R_i : \ i = 1, 2, \dots\}$ is a Poisson point process on $(0,\infty)$ with intensity $Nr^{-2}\mathrm{d}r$ and $Q_1, Q_2, \dots$ are independently sampled from the spectral measure H on S_{N-1}. The advantage of this representation is that the components of Q_i are bounded by 1. This ensures that $Z(x) = \max_{i\geq 1} R_i Q_i$ can be simulated exactly by generating the largest R_i first until no more of the remaining points $R_i Q_i$ can contribute to the maximum.
The only difficulty is thus to generate the random variables Q_i from the probability measure H on the $(N-1)$-dimensional positive sphere S_{N-1}. The following theorem gives such an explicit representation for the Q_i for general max-stable distributions $Z(x)$ based on the distributions P_{x_k}, $k = 1, \dots, N$, in (6.5).

Theorem 6.6. *Let $T_1, T_2, \dots$ be independent copies of a random variable T with uniform distribution on the discrete set $\{1, \dots, N\}$. Further, for any $k = 1, \dots, N$, let $Y_1^{(k)}, Y_2^{(k)}, \dots$ be independent random processes with distribution P_{x_k} as in* (6.5)*. Then, the S_{N-1}-valued random variables*

$$Q_i = \frac{Y_i^{(T_i)}(x)}{\|Y_i^{(T_i)}(x)\|}, \quad i \geq 1,$$

are independent with distribution H. Consequently, with $\{R_i,\ i \geq 1\}$ as above,

$$Z(x) = \max_{i\geq 1} R_i \frac{Y_i^{(T_i)}(x)}{\|Y_i^{(T_i)}(x)\|}. \tag{6.13}$$

Proof. For any $k = 1, \dots, N$, Equation (6.5) implies

$$\int_{C_+(\mathcal{X})} f(x_k) \mathbf{1}_{\{f(x)/\|f(x)\| \in A\}}\, \nu(\mathrm{d}f) = \int_{C_+(\mathcal{X})} \mathbf{1}_{\{f(x)/\|f(x)\| \in A\}}\, P_{x_k}(\mathrm{d}f). \tag{6.14}$$

We compute the μ_x-measure of the set $U^{-1}\{(u,\infty) \times A\}$ for $u > 0$ and a Borel set $A \subset S_{N-1}$.

$$\begin{aligned}\mu_x[U^{-1}\{(u,\infty) \times A\}] &= \int_{C_+(\mathcal{X})} \int_0^\infty \mathbf{1}_{\{\zeta\|f(x)\| > u\}} \mathbf{1}_{\{f(x)/\|f(x)\| \in A\}} \zeta^{-2} \mathrm{d}\zeta\, \nu(\mathrm{d}f) \\ &= \frac{1}{u} \int_{C_+(\mathcal{X})} \|f(x)\| \mathbf{1}_{\{f(x)/\|f(x)\| \in A\}}\, \nu(\mathrm{d}f) = \frac{1}{u} \sum_{k=1}^N \int_{C_+(\mathcal{X})} f(x_k) \mathbf{1}_{\{f(x)/\|f(x)\| \in A\}}\, \nu(\mathrm{d}f)\end{aligned}$$

$$= \frac{1}{u}\sum_{k=1}^{N}\int_{C_+(\mathcal{X})} \mathbf{1}_{\{f(x)/\|f(x)\|\in A\}}\, P_{x_k}(\mathrm{d}f) = \frac{N}{u}\frac{1}{N}\sum_{k=1}^{N}\int_{C_+(\mathcal{X})} \mathbf{1}_{\{f(x)/\|f(x)\|\in A\}}\, P_{x_k}(\mathrm{d}f), \tag{6.15}$$

where the penultimate equality follows from (6.14). Let $Y^{(k)}$, $k = 1, \dots, N$, be independent random processes with distribution P_{x_k}, respectively, and let T be an independent uniform random variable on $\{1, \dots, N\}$, then Equation (6.15) can be restated as

$$\mu_x[U^{-1}\{(u,\infty)\times A\}] = \frac{N}{u}\mathbb{P}\left\{Y^{(T)}(x)/\|Y^{(T)}(x)\| \in A\right\}.$$

Comparing this with (6.11) yields the assertion of the theorem. □

Theorem 6.6 shows how to simulate from the spectral measure H. It requires only to be able to simulate from the distributions P_{x_k}, $k = 1, \dots, N$. Algorithm 6.2, an adaptation of Schlather's (2002) algorithm, provides an exact sample from the max-stable process Z at locations x.

Algorithm 6.2: Exact simulation of a max-stable process Z at $x = (x_1, \dots, x_N)$

Simulate $\zeta^{-1} \sim \mathrm{Exp}(N)$ and set $Z(x) = 0$.
while $\zeta > \min\{Z(x_1), \dots, Z(x_N)\}$ **do**
  Simulate T uniform on $\{1, \dots, N\}$ and Y according to the law P_{x_T}.
  Update $Z(x)$ by the componentwise $\max\{Z(x), \zeta Y(x)/\|Y(x)\|\}$.
  Simulate $e \sim \mathrm{Exp}(N)$ and update ζ^{-1} by $\zeta^{-1} + e$.
return Z

Both Algorithm 6.1 and Algorithm 6.2 include the simulation of random functions with distributions P_{x_0} in (6.5), $x_0 \in \mathcal{X}$. In Section 6.4 we provide closed-form expressions for various important examples of max-stable process and multivariate extreme value distributions.

6.4 Examples

6.4.1 Moving maximum process

The parameter space is $\mathcal{X} = \mathbb{Z}^d$ or $\mathbb{R}^d$ and λ denotes the counting measure or the Lebesgue measure, respectively. A moving maximum process on $\mathcal{X}$ is a max-stable process of the form

$$Z(x) = \max_{i\geq 1} \zeta_i h(x - \chi_i), \quad x \in \mathcal{X}, \tag{6.16}$$

where $\{(\zeta_i, \chi_i), i = 1, 2, \dots\}$ is a Poisson point process on $(0,\infty)\times\mathcal{X}$ with intensity measure $\zeta^{-2}\mathrm{d}\zeta \times \lambda(\mathrm{d}\chi)$ and $h : \mathcal{X} \to [0,\infty)$ is a continuous function satisfying $\int_{\mathcal{X}} h(x)\lambda(\mathrm{d}x) = 1$. A popular example is the Gaussian extreme value process proposed in Smith (1990) where h is a multivariate Gaussian density on $\mathbb{R}^d$.

Proposition 6.7. *Consider the moving maximum process* (6.16). *For all $x_0 \in \mathcal{X}$, the distribution P_{x_0} is equal to the distribution of the random function*

$$\frac{h(\cdot + \chi - x_0)}{h(\chi)}, \qquad \chi \sim h(u)\lambda(\mathrm{d}u).$$

Proof. In the case of the moving maximum process (6.16), the measure ν associated with the representation (6.1) is

$$\nu(A) = \int_{\mathcal{X}} \mathbf{1}_{\{h(\cdot-\chi)\in A\}} \lambda(\mathrm{d}\chi), \quad A \in \mathcal{C}_+(\mathcal{X}).$$

We deduce from Proposition 6.2,

$$\begin{aligned} P_{x_0}(A) &= \int_{C_+(\mathcal{X})} \mathbf{1}_{\{f/f(x_0)\in A\}} f(x_0)\, \nu(\mathrm{d}f) = \int_{\mathcal{X}} \mathbf{1}_{\{h(\cdot-\chi)/h(x_0-\chi)\in A\}} h(x_0-\chi)\, \lambda(\mathrm{d}\chi) \\ &= \int_{\mathcal{X}} \mathbf{1}_{\{h(\cdot+u-x_0)/h(u)\in A\}} h(u)\, \lambda(\mathrm{d}u) \end{aligned}$$

where the last line follows from the simple change of variable $x_0 - \chi = u$. This proves the result since $h(u)\lambda(\mathrm{d}u)$ is a density function on $\mathcal{X}$. □

6.4.2 Brown–Resnick process

We consider max-stable processes obtained by representation (6.1) where ν is a probability measure on $C_+(\mathcal{X})$ given by

$$\nu(A) = \mathbb{P}\left(\exp\left\{W(\cdot) - \frac{\sigma^2(\cdot)}{2}\right\} \in A\right), \quad A \in \mathcal{C}_+(\mathcal{X}), \tag{6.17}$$

with $\{W(x),\ x \in \mathcal{X}\}$ being a sample-continuous centered Gaussian process on $\mathcal{X}$ with variance $\sigma^2(x) = \mathbb{E}\{W(x)^2\}$. In other words, ν is the distribution of the log-normal process $Y(x) = \exp\big\{W(x) - \sigma^2(x)/2\big\}$, $x \in \mathcal{X}$.
An interesting phenomenon arises when $\mathcal{X} = \mathbb{Z}^d$ or $\mathbb{R}^d$ and W has stationary increments: Kabluchko et al. (2009) show that the associated max-stable process Z is then stationary with distribution depending only on the semi-variogram

$$\gamma(h) = \frac{1}{2}\mathbb{E}\left[\{W(h) - W(0)\}^2\right], \quad h \in \mathcal{X}.$$

The stationary max-stable process Z is called a Brown–Resnick process. However, our results apply both in the stationary and non-stationary case (cf., Kabluchko, 2011) and unless stated otherwise we do not assume that W has stationary increments.

Proposition 6.8. *Consider the Brown–Resnick type model* (6.17)*. For all $x_0 \in \mathcal{X}$, the distribution P_{x_0} is equal to the distribution of the log-normal process*

$$\widetilde{Y}(x) = \exp\Big[W(x) - W(x_0) - \frac{1}{2}\operatorname{Var}\{W(x) - W(x_0)\}\Big], \quad x \in \mathcal{X}.$$

The proof of Proposition 6.8 relies on the following lemma on exponential changes of measures for Gaussian processes. Note that the distribution of P_{x_0} is strongly connected to the notion of conditional intensity introduced in Dombry et al. (2013) and that the formula are similar.

Lemma 6.9. *The distribution of the random process $(W(x))_{x\in\mathcal{X}}$ under the transformed probability measure $\widehat{\mathbb{P}} = e^{W(x_0)-\sigma^2(x_0)/2}\mathrm{d}\mathbb{P}$ is equal to the distribution of the Gaussian random process*

$$W(x) + c(x_0, x), \quad x \in \mathcal{X},$$

where $c(x, y)$ denotes the covariance between $W(x)$ and $W(y)$.

Proof. We need to consider finite dimensional distributions only and we compute for some $x_1, \dots, x_k \in \mathcal{X}$ the Laplace transform of $(W(x_i))_{1 \le i \le k}$ under the transformed probability measure $\widehat{\mathbb{P}}$. For all $\theta = (\theta_1, \dots, \theta_k) \in \mathbb{R}^k$, we have

$$\widehat{\mathbb{E}}\left\{e^{\sum_{i=1}^k \theta_i W(x_i)}\right\} = \mathbb{E}\left\{e^{W(x_0)-\sigma^2(x_0)/2} e^{\sum_{i=1}^k \theta_i W(x_i)}\right\} = \exp\left(\frac{1}{2}\tilde{\theta}^\top \Sigma \tilde{\theta} - \frac{1}{2}\sigma^2(x_0)\right), \tag{6.18}$$

with $\tilde{\theta} = (1, \theta) \in \mathbb{R}^{k+1}$ and $\tilde{\Sigma} = (c(x_i, x_j))_{0 \le i,j \le k}$ the covariance matrix. We introduce the block decomposition

$$\tilde{\Sigma} = \begin{pmatrix} \sigma^2(x_0) & \Sigma_{0,k} \\ \Sigma_{k,0} & \Sigma \end{pmatrix}$$

with $\Sigma = (c(x_i, x_j))_{1 \le i,j \le k}$ and $\Sigma_{k,0} = \Sigma_{0,k}^\top = (c(x_0, x_i))_{1 \le i \le k}$. The exponent in Equation (6.18) can be rewritten as

$$\frac{1}{2}\tilde{\theta}^\top \tilde{\Sigma} \tilde{\theta} - \frac{1}{2}\sigma^2(x_0) = \frac{1}{2}\left\{\sigma^2(x_0) + \theta^\top \Sigma \theta + 2\theta^\top \Sigma_{k,0}\right\} - \frac{1}{2}\sigma^2(x_0) = \theta^\top \Sigma_{k,0} + \frac{1}{2}\theta^\top \Sigma \theta.$$

We recognize the Laplace transform of a Gaussian random vector with mean $\Sigma_{k,0}$ and covariance matrix Σ whence the lemma follows. □

Proof of Proposition 6.8. Equations (6.5) and (6.17) together with Lemma 6.9 yield, for all $A \in \mathcal{C}_+(\mathcal{X})$,

$$\begin{aligned}
P_{x_0}(A) &= \int_{C_+(\mathcal{X})} \mathbf{1}_{\{f/f(x) \in A\}} f(x)\, \nu(\mathrm{d}f) \\
&= \mathbb{E}\left[e^{W(x_0)-\frac{1}{2}\sigma^2(x_0)} \mathbf{1}_{\left\{\exp(W(\cdot)-\frac{1}{2}\sigma^2(\cdot))/\exp(W(x_0)+\frac{1}{2}\sigma^2(x_0)) \in A\right\}}\right] \\
&= \widehat{\mathbb{P}}\left(\exp\left[W(\cdot) - W(x_0) - \frac{1}{2}\left\{\sigma^2(\cdot) - \sigma^2(x_0)\right\}\right] \in A\right) \\
&= \mathbb{P}\left(\exp\left[W(\cdot) + c(x_0, \cdot) - W(x_0) - c(x_0, x_0) - \frac{1}{2}\left\{\sigma^2(\cdot) - \sigma^2(x_0)\right\}\right] \in A\right) \\
&= \mathbb{P}\left(\exp\left[W(\cdot) - W(x_0) - \frac{1}{2}\left\{\sigma^2(\cdot) + \sigma^2(x_0) - 2c(x_0, \cdot)\right\}\right] \in A\right).
\end{aligned}$$

Using the fact that for all $x \in \mathcal{X}$

$$\sigma^2(x) + \sigma^2(x_0) - 2c(x_0, x) = \mathrm{Var}[W(x) - W(x_0)]$$

we deduce that P_{x_0} is equal to the distribution of the log-normal process

$$\widetilde{Y}(x) = \exp\left[W(x) - W(x_0) - \frac{1}{2}\mathrm{Var}\left\{W(x) - W(x_0)\right\}\right], \quad x \in \mathcal{X}.$$

This proves Proposition 6.8. □

Remark 6.10. The finite dimensional margins of Brown–Resnick processes are Hüsler and Reiss (1989) distributions and the above therefore provides a method for their exact simulation.

Remark 6.11. For stationary Brown–Resnick processes, i.e., when W has stationary increments with $W(0) = 0$, it is easy to deduce from Proposition 6.8 that P_{x_0} is equal to the distribution of

$$\exp\left\{W(x - x_0) - \gamma(x - x_0)\right\}, \quad x \in \mathcal{X}.$$

Thus, Theorem 6.6 yields

$$Z(x) = \max_{i\geq 1} R_i \frac{\exp\{W_i(x-T_i) - \gamma(x-T_i)\}}{\sum_{\ell=1}^{N} \exp\{W_i(x_\ell - T_i) - \gamma(x_\ell - T_i)\}}$$

where $\{R_i : i = 1, 2, \dots\}$ is a Poisson point process on $(0, \infty)$ with intensity $Nr^{-2}\mathrm{d}r$, $T_1, T_2, \dots$ are independent with uniform distribution on $\{x_1, \dots, x_N\}$ and $W_1, W_2, \dots$ are independent copies of W. The same representation appears in Dieker and Mikosch (2015), so Algorithm 6.2 is identical to Dieker and Mikosch procedure in this case.

6.4.3 Extremal-t process

We consider the so called extremal-t max-stable process (cf. Opitz, 2013) defined by representation (6.1) with ν the distribution of the random process

$$Y(x) = c_\alpha \max\{0, W(x)\}^\alpha, \quad x \in \mathcal{X}, \tag{6.19}$$

where $\alpha > 0$, $c_\alpha = \pi^{1/2} 2^{-(\alpha-2)/2} / \Gamma\{(1+\alpha)/2\}$, and W a sample-continuous centered Gaussian process on $\mathcal{X}$ with unit variance and covariance function c. For $\alpha = 1$, the corresponding max-stable process in (6.1) coincides with the widely used extremal Gaussian process of Schlather (2002).

Proposition 6.12. *Consider the extremal-t model* (6.19). *For all $x_0 \in \mathcal{X}$, the distribution P_{x_0} is equal to the distribution of* $\max(T, 0)^\alpha$, *where $T = (T(x))_{x\in\mathcal{X}}$ is a Student process with $\alpha + 1$ degrees of freedom, location and scale functions given respectively by*

$$\mu(x) = c(x_0, x), \quad \hat{c}(x_1, x_2) = \frac{c(x_1, x_2) - c(x_0, x_1)c(x_0, x_2)}{(\alpha + 1)}.$$

It is worth noting that the formula for P_{x_0} provided in Proposition 6.12 is similar to the formula for the conditional intensity of the extremal-t process that was computed in Ribatet (2013).
The proof of Proposition 6.12 is based on the following lemma.

Lemma 6.13. *The distribution of the random process $(W(x)/W(x_0))_{x\in\mathcal{X}}$ under the transformed probability measure $\widehat{\mathbb{P}} = c_\alpha W(x_0)_+^\alpha \mathrm{d}\mathbb{P}$ is equal to the distribution of a Student process with $\alpha + 1$ degrees of freedom, location and scale functions given respectively by*

$$\mu_k = \Sigma_{k,0} \quad \textit{and} \quad \widehat{\Sigma}_k = \frac{\Sigma_k - \Sigma_{k,0}\Sigma_{0,k}}{\alpha + 1},$$

where $\Sigma_k = (c(x_i, x_j))_{1\leq i,j\leq k}$ and $\Sigma_{k,0} = \Sigma_{0,k}^\top = (c(x_0, x_i))_{1\leq i\leq k}$.

Proof. We consider finite dimensional distributions only. Let $k \geq 1$ and $x_1, \dots, x_k \in \mathcal{X}$. We first assume that the covariance matrix $\widetilde{\Sigma} = (c(x_i, x_j))_{0\leq i,j\leq k}$ is non singular so that $(W(x_i))_{0\leq i\leq k}$ has density

$$\tilde{g}(y) = (2\pi)^{-(k+1)/2} \det(\widetilde{\Sigma})^{-1/2} \exp\left(-\frac{1}{2} y^\top \widetilde{\Sigma}^{-1} y\right) \quad \text{with } y = (y_i)_{0\leq i\leq k}.$$

Setting $z = (y_i/y_0)_{1\leq i\leq k}$, we have for all Borel sets $A_1, \dots, A_k \subset \mathbb{R}$

$$\widehat{\mathbb{P}}\left\{\frac{W(x_i)}{W(x_0)} \in A_i,\ i = 1, \dots, k\right\} = \int_{\mathbb{R}^{k+1}} \mathbf{1}_{\{y_i/y_0 \in A_i,\ i=1,\dots,k\}} c_\alpha (y_0)_+^\alpha \tilde{g}(\boldsymbol{y})\, \mathrm{d}y$$

$$= \int_{\mathbb{R}^k} \mathbf{1}_{\{z_i \in A_i,\ i=1,\dots,k\}} \Big\{ \int_0^\infty c_\alpha (y_0)_+^\alpha \tilde{g}(y_0, y_0 z)\, y_0^k \,\mathrm{d}y_0 \Big\} \,\mathrm{d}z$$

We deduce that under $\widehat{\mathbb{P}}$, the random vector $(W(x_i)/W(x_0))_{1\le i\le k}$ has density

$$\begin{aligned} g(z) &= \int_0^\infty c_\alpha y_0^{k+\alpha} \tilde{g}(y_0, y_0 z)\,\mathrm{d}y_0 \\ &= c_\alpha (2\pi)^{-(k+1)/2} \det(\widetilde{\Sigma})^{-1/2} \int_0^\infty y_0^{k+\alpha} \exp\Big(-\frac{\tilde{z}^\top \widetilde{\Sigma}^{-1} \tilde{z}}{2} y_0^2 \Big)\,\mathrm{d}y_0 \end{aligned}$$

with $\tilde{z} = (1, z)$. Using the change of variable $u = \frac{1}{2}\tilde{z}^\top \widetilde{\Sigma}^{-1} \tilde{z} y_0^2$, we get

$$\begin{aligned} \int_0^\infty y_0^{k+\alpha} \exp\Big(-\frac{\tilde{z}^\top \widetilde{\Sigma}^{-1} \tilde{z}}{2} y_0^2 \Big)\,\mathrm{d}y_0 &= \frac{1}{2} \Big(\frac{\tilde{z}^\top \widetilde{\Sigma}^{-1} \tilde{z}}{2} \Big)^{-\frac{\alpha+k+1}{2}} \int_0^\infty u^{(k+\alpha-1)/2} \exp(-u)\,\mathrm{d}u \\ &= \frac{1}{2} \Big(\frac{\tilde{z}^\top \widetilde{\Sigma}^{-1} \tilde{z}}{2} \Big)^{-\frac{\alpha+k+1}{2}} \Gamma\Big(\frac{k+\alpha+1}{2} \Big) \end{aligned}$$

and we obtain after simplification

$$g(z) = \pi^{-k/2} \frac{\Gamma\left(\frac{k+\alpha+1}{2}\right)}{\Gamma\left(\frac{\alpha+1}{2}\right)} \det(\widetilde{\Sigma})^{-1/2} \left\{ \tilde{z}^\top \widetilde{\Sigma}^{-1} \tilde{z} \right\}^{-\frac{\alpha+k+1}{2}}.$$

Introducing the block decomposition $\widetilde{\Sigma} = \begin{pmatrix} 1 & \Sigma_{0,k} \\ \Sigma_{k,0} & \Sigma_k \end{pmatrix}$, the inverse matrix is

$$\widetilde{\Sigma}^{-1} = \begin{pmatrix} 1 + \Sigma_{0,k}(\Sigma_k - \Sigma_{k,0}\Sigma_{0,k})^{-1}\Sigma_{k,0} & -\Sigma_{0,k}(\Sigma_k - \Sigma_{k,0}\Sigma_{0,k})^{-1} \\ -(\Sigma_k - \Sigma_{k,0}\Sigma_{0,k})^{-1}\Sigma_{k,0} & (\Sigma_k - \Sigma_{k,0}\Sigma_{0,k})^{-1} \end{pmatrix}.$$

By the definition of μ_k and $\widehat{\Sigma}_k$, we have

$$\widetilde{\Sigma}^{-1} = \frac{1}{1+\alpha} \begin{pmatrix} 1 + \alpha + \mu_k^\top \widehat{\Sigma}_k^{-1} \mu_k & -\mu_k^\top \widehat{\Sigma}_k^{-1} \\ -\widehat{\Sigma}_k^{-1} \mu_k & \widehat{\Sigma}_k^{-1} \end{pmatrix}$$

and

$$\tilde{z}^\top \widetilde{\Sigma}^{-1} \tilde{z} = (1, z)^\top \widetilde{\Sigma}^{-1} (1, z) = 1 + \frac{(z-\mu_k)^\top \widehat{\Sigma}_k^{-1} (z - \mu_k)}{\alpha + 1}$$

Finally, we obtain after simplification

$$g(z) = \pi^{-k/2} (\alpha+1)^{-k/2} \frac{\Gamma\left(\frac{k+\alpha+1}{2}\right)}{\Gamma\left(\frac{\alpha+1}{2}\right)} \det(\widehat{\Sigma}_k)^{-1/2} \left\{ 1 + \frac{(z-\mu_k)^\top \widehat{\Sigma}_k^{-1} (z-\mu_k)}{\alpha+1} \right\}^{-\frac{\alpha+k+1}{2}}.$$

We recognize the k-variate Student density with $\alpha+1$ degrees of freedom, location parameter μ_k and scale matrix $\hat{\Sigma}_k$. □

Proof of Proposition 6.12. Consider the set

$$A = \{ f \in C_+(\mathcal{X}) :\ f(x_1) \in A_1, \dots, f(x_k) \in A_k \}.$$

Equations (6.5) and (6.19) together with Lemma 6.13 yield,

$$\begin{aligned} P_{x_0}(A) &= \int_{C_+(\mathcal{X})} \mathbf{1}_{\{f/f(x)\in A\}} f(x)\, \nu(\mathrm{d}f) = \mathbb{E}\left[c_\alpha W(x_0)_+^\alpha \mathbf{1}_{\{W(x_i)_+^\alpha / W(x_0)_+^\alpha \in A_i,\ i=1,\dots,k\}} \right] \\ &= \widehat{\mathbb{P}}\left\{ W(x_i)_+^\alpha / W(x_0)_+^\alpha \in A_i,\ i = 1, \dots, k \right\} = \mathbb{P}\left\{ (T_i)_+^\alpha \in A_i,\ i = 1, \dots, k \right\} \end{aligned}$$

where $T = (T_1, \dots, T_k)$ has a multivariate Student distribution with $\alpha+1$ degrees of freedom, location parameter μ_k and dispersion matrix $\widehat{\Sigma}_k$. This proves the result. □

6.4.4 Multivariate extreme value distributions

In this section, we review some popular models for multivariate extreme value distributions, i.e., the case when $\mathcal{X} = \{1, \dots, N\}$ in (6.1) is a finite set for some fixed $N \in \mathbb{N}$. For these models, we explicitly calculate the measure P_{j_0} for any $j_0 = 1, \dots, N$. Unless otherwise stated, all random vectors are N-dimensional in this section. For more details on the models, we refer to Gudendorf and Segers (2010).

Logistic model

The symmetric logistic model in dimension N with parameter $\theta \in (0, 1]$ corresponds to the max-stable random vector with cumulative distribution function

$$\mathbb{P}\{Z \le z\} = \exp\left\{-\left(\sum\nolimits_{j=1}^{N} z_j^{-1/\theta}\right)^{\theta}\right\}, \quad z = (z_1, \dots, z_N) \in (0, \infty)^N. \tag{6.20}$$

Proposition 6.14. *Let $\beta = 1/\theta$. In the logistic model* (6.20)*, the probability measure P_{j_0} for any $j_0 = 1, \dots, N$ is equal to the distribution of the random vector*

$$\left(\frac{F_1}{F_{j_0}}, \dots, \frac{F_N}{F_{j_0}}\right)$$

where $F_1, \dots, F_N$ are independent, F_j, $j \neq j_0$, follows a Fréchet(β, c_β) distribution with scale parameter $c_\beta = \Gamma(1 - 1/\beta)^{-1}$ and $(F_{j_0}/c_\beta)^{-\beta}$ follows a Gamma$(1 - 1/\beta, 1)$ *distribution.*

Proof. It is easily shown that the logistic model admits the representation

$$Z = \max_{i \ge 1} \zeta_i F_i$$

where the F_i are independent random vectors with independent Fréchet(β, c_β)-distributed components. To check this, we compute

$$\begin{aligned}
\mathbb{E}\left(\max_{j=1}^{N} \frac{F_j}{z_j}\right) &= \int_0^\infty \mathbb{P}\left(\max_{j=1}^{N} \frac{F_j}{z_j} > u\right) \mathrm{d}u = \int_0^\infty \left\{1 - \prod\nolimits_{j=1}^{N} \mathbb{P}(F_j < z_j u)\right\} \mathrm{d}u \\
&= \int_0^\infty \left\{1 - \prod\nolimits_{j=1}^{N} e^{-(z_j u / c_\beta)^{-\beta}}\right\} \mathrm{d}u = \int_0^\infty \left\{1 - e^{-u^{-\beta} \sum_{j=1}^{N} (z_j / c_\beta)^{-\beta}}\right\} \mathrm{d}u \\
&= \left(\sum\nolimits_{j=1}^{N} z_j^{-\beta}\right)^{1/\beta}.
\end{aligned}$$

For the computation of the last integral, we recognize the expectation of a Fréchet distribution. Next we use the fact that P_{j_0} is the distribution of F/F_{j_0} under the transformed density

$$y_{j_0} \prod_{k=1}^{N} \frac{\beta}{c_\beta} \left(\frac{y_k}{c_\beta}\right)^{-1-\beta} e^{-(y_k/c_\beta)^{-\beta}}.$$

We recognize a product measure where the jth margin, $j \neq j_0$, has a Fréchet(β, c_β) distribution. The j_0th marginal has density

$$y_{j_0} \frac{\beta}{c_\beta} \left(\frac{y_{j_0}}{c_\beta}\right)^{-1-\beta} e^{-(y_{j_0}/c_\beta)^{-\beta}}$$

and a simple change of variable reveals that this expression is the density of $c_\beta Z^{-1/\beta}$ with $Z \sim$ Gamma$(1 - 1/\beta, 1)$. □

Remark 6.15. The asymmetric logistic distribution can be represented as the mixture of symmetric logistic distributions; see Theorem 1 in Stephenson (2003), for instance. As a consequence, Proposition 6.14 also enables exact simulation of asymmetric logistic distributions.

Negative logistic model

The negative logistic model in dimension N with parameter $\theta > 0$ corresponds to the max-stable random vector Z with cumulative distribution function

$$\mathbb{P}\{Z \le z\} = \exp\left\{ \sum_{\emptyset \neq J \subset \{1,\dots,N\}} (-1)^{|J|} \left(\sum\nolimits_{j\in J} z_j^{\theta} \right)^{-1/\theta} \right\}, \quad z \in (0,\infty)^N. \tag{6.21}$$

Proposition 6.16. *In the negative logistic model* (6.21)*, the probability measure P_{j_0} for any $j_0 = 1,\dots,N$ is equal to the distribution of the random vector*

$$\left(\frac{W_1}{W_{j_0}}, \dots, \frac{W_N}{W_{j_0}} \right)$$

where $W_1,\dots,W_N$ are independent, W_j, $j \neq j_0$, follows a Weibull(θ, c_θ) *distribution with scale parameter $c_\theta = \Gamma(1+1/\theta)^{-1}$ and $(W_{j_0}/c_\theta)^\theta$ follows a $\Gamma(1+1/\theta, 1)$ distribution.*

Proof. Similarly to the logistic model, we have the spectral representation

$$Z = \max_{i\ge 1} \zeta_i W_i$$

where the W_i are independent random vectors with independent Weibull(θ, c_θ)-distributed components with scale parameter $c_\theta = \frac{1}{\Gamma(1+1/\theta)}$. To check this, we compute

$$\begin{aligned}
\mathbb{E}\left(\max_{j=1}^{N} \frac{W_j}{z_j} \right) &= \int_0^\infty \mathbb{P}\left(\max_{j=1}^{N} \frac{W_j}{z_j} > u \right) \mathrm{d}u = \int_0^\infty \left\{ 1 - \prod_{j=1}^{N} \mathbb{P}(W_j < z_j u) \right\} \mathrm{d}u \\
&= \int_0^\infty \left[1 - \prod_{j=1}^{N} \left\{ 1 - e^{-(z_j u/c_\theta)^\theta} \right\} \right] \mathrm{d}u = -\sum_J (-1)^{|J|} \int_0^\infty e^{-u^\theta \sum_{j\in J} (z_j/c_\theta)^\theta} \mathrm{d}u \\
&= -\sum_J (-1)^{|J|} \left\{ \sum_{j\in J} (z_j/c_\theta)^\theta \right\}^{-1/\theta} \Gamma(1+1/\theta) = -\sum_J (-1)^{|J|} \left(\sum_{j\in J} z_j^\theta \right)^{-1/\theta}.
\end{aligned}$$

For the computation of the last integral, we recognize the expectation of a Weibull distribution. As for the logistic model, P_{j_0} is the distribution of W/W_{j_0} under the transformed density

$$y_{j_0} \prod_{k=1}^{N} \frac{\theta}{c_\theta} \left(\frac{y_k}{c_\theta} \right)^{\theta-1} e^{-(y_k/c_\theta)^\theta}.$$

We recognize a product measure where the jth margin, $j \neq j_0$ has a Weibull(θ, c_θ) distribution. The j_0th marginal has density

$$y_{j_0} \frac{\theta}{c_\theta} \left(\frac{y_{j_0}}{c_\theta} \right)^{\theta-1} e^{-(y_{j_0}/c_\theta)^\theta}$$

and a simple change of variable reveals that this expression is the density of $c_\theta Z^{1/\theta}$ with $Z \sim \text{Gamma}(1+1/\theta, 1)$. □

Dirichlet mixture model

The Dirichlet mixture model was introduced by Boldi and Davison (2007). In dimension N, the model corresponds to the max-stable random vector given by

$$Z = \max_{i \geq 1} \zeta_i (N Y_i) \tag{6.22}$$

where the Y_i's are independent identically distributed random vectors on the simplex

$$S_{N-1} = \left\{ y \in [0,1]^n : \sum\nolimits_{j=1}^{N} y_j = 1 \right\}.$$

The distribution of each Y_i is a mixture of m Dirichlet models, i.e., its Lebesgue density is of the form

$$h(y) = \sum_{k=1}^{m} \pi_k \mathrm{diri}(y \mid \alpha_{1k}, \ldots, \alpha_{Nk}), \quad y = (y_1, \ldots, y_N) \in S_{N-1}, \tag{6.23}$$

where $\pi_k \geq 0$, $k = 1, \ldots, m$ such that $\sum_{k=1}^{m} \pi_k = 1$, $\alpha_{ik} > 0$, $i = 1, \ldots, N$, $k = 1, \ldots, m$, and

$$\mathrm{diri}(y \mid \alpha_1, \ldots, \alpha_N) = \frac{1}{B(\alpha)} \prod_{j=1}^{N} y_j^{\alpha_j - 1}, \quad B(\alpha) = \frac{\prod_{j=1}^{N} \Gamma(\alpha_j)}{\Gamma(\sum_{j=1}^{N} \alpha_j)}. \tag{6.24}$$

Here, the parameters π_k and α_{ik}, $i = 1, \ldots, N$, $k = 1, \ldots, m$, are such that

$$E(Y_j) = \sum_{k=1}^{m} \pi_k \frac{\alpha_{jk}}{\sum_{i=1}^{N} \alpha_{ik}} = \frac{1}{N}, \quad j = 1, \ldots, N.$$

Proposition 6.17. *In the Dirichlet model* (6.22)*, we have for any* $j_0 = 1, \ldots, N$ *that* $P_{j_0} = \sum_{k=1}^{m} \hat{\pi}_k P_{j_0}^{(k)}$ *where* $\hat{\pi}_k = \pi_k \alpha_{j_0 k} / (\sum_{i=1}^{N} \alpha_{ik})$ *and* $P_{j_0}^{(k)}$ *is equal to the distribution of the random vector*

$$\left(\frac{G_1^{(k)}}{G_{j_0}^{(k)}}, \ldots, \frac{G_N^{(k)}}{G_{j_0}^{(k)}} \right)$$

and $G_1^{(k)}, \ldots, G_N^{(k)}$ *are independent random variables with*

$$G_{j_0}^{(k)} \sim \mathrm{Gamma}(\alpha_{j_0 k} + 1, 1), \quad G_j \sim \mathrm{Gamma}(\alpha_{jk}, 1), \quad j \neq j_0.$$

Proof. By definition, P_{j_0} has the form

$$\begin{aligned} P_{j_0}(A) &= N \sum_{k=1}^{m} \pi_k \int_{S_{N-1}} y_{j_0} \mathbf{1}_{\{y/y_{j_0} \in A\}} \mathrm{diri}(y \mid \alpha_{1k}, \ldots, \alpha_{Nk}) \, \mathrm{d}y \\ &= N \sum_{k=1}^{m} \hat{\pi}_k \frac{\int_{S_{N-1}} y_{j_0} \mathbf{1}_{\{y/y_{j_0} \in A\}} \mathrm{diri}(y \mid \alpha_{1k}, \ldots, \alpha_{Nk}) \, \mathrm{d}y}{\int_{S_{N-1}} y_{j_0} \mathrm{diri}(y \mid \alpha_{1k}, \ldots, \alpha_{Nk}) \, \mathrm{d}y}, \quad A \subset (0, \infty)^N. \end{aligned}$$

Thus, P_{j_0} is given as the mixture $P_{j_0} = \sum_{k=1}^{m} \hat{\pi}_k P_{j_0}^{(k)}$, where for each $k = 1, \ldots, m$, the probability measure $P_{j_0}^{(k)}$ is equal to the distribution of the random vector $\tilde{Y}^{(k)} / \tilde{Y}_{j_0}^{(k)}$, and $\tilde{Y}^{(k)}$ has a transformed density proportional to $y_{j_0} \prod_{j=1}^{N} y_j^{\alpha_j - 1}$. We recognize the Dirichlet distribution with parameters $\tilde{\alpha}_{1k}, \ldots, \tilde{\alpha}_{Nk}$ given by

$$\tilde{\alpha}_{j_0 k} = \alpha_{j_0 k} + 1 \quad \text{and} \quad \tilde{\alpha}_{jk} = \alpha_{jk} \quad j \neq j_0.$$

It is well known that Dirichlet distributions can be expressed in terms of Gamma distributions. More precisely, we have the stochastic representation

$$\tilde{Y}^{(k)} = \left(G_1^{(k)} \Big/ \sum_{j=1}^{N} G_j^{(k)}, \dots, G_N^{(k)} \Big/ \sum_{j=1}^{N} G_j^{(k)} \right),$$

where $G_j^{(k)}$ are independent Gamma$(\tilde{\alpha}_{jk}, 1)$ random variables. The result follows since $P_{j_0}^{(k)}$ is the distribution of $\tilde{Y}^{(k)}/\tilde{Y}_{j_0}^{(k)}$. □

6.5 Complexity of the Algorithms

In this section, we assess the complexity of Algorithms 6.1 and 6.2 as a function of the number N of simulation sites. Both algorithms contain the simulation of exponential random variables e and the simulation of N-dimensional random vectors $Y(x)$ according to a mixture of the laws $P_{x_1}, \dots, P_{x_N}$. The simulation of e involves much less computational effort than the simulation of Y and can therefore be neglected in the analysis of the algorithmic complexity. We thus consider the number $C_1(N)$ and $C_2(N)$ of random vectors $Y(x)$ that must be simulated by Algorithm 6.1 and 6.2 respectively to obtain one exact simulation of $Z(x)$. The following proposition provides simple expressions for the expectations $E(C_1(N))$ and $E(C_2(N))$.

Proposition 6.18. *The expected number of random vectors $Y(x)$ that are needed for exact simulation of Z at $x = (x_1, \dots, x_N)$ are:*

Algorithm 6.1: $\mathbb{E}\{C_1(N)\} = N$

Algorithm 6.2: $\mathbb{E}\{C_2(N)\} = N\mathbb{E}\left\{ \max_{i=1,\dots,N} Z(x_i)^{-1} \right\}$

Furthermore, $\mathbb{E}\{C_1(N)\} \le \mathbb{E}\{C_2(N)\}$, with equality if and only if $Z(x_1) = \cdots = Z(x_N)$ almost surely.

Proof. In order to analyze the complexity of Algorithm 6.1, we consider each step of the algorithm separately. In the nth step, i.e. for sampling the process perfectly at site x_n, we simulate Poisson points ζ and stochastic processes Y, until one of the following two conditions is satisfied:

(a) $\zeta < Z_{n-1}(x_n)$. This condition is checked directly after the simulation of ζ and, in this case, no stochastic process Y needs to be simulated.

(b) $\zeta > Z_{n-1}(x_n)$ and $\zeta Y(x_i) \le Z(x_i)$ for all $1 \le i < n-1$. In this case, Z is updated and ζY is an extremal function as it contributes to Z at site x_n (and possibly also at some of the sites $x_{n+1}, \dots, x_N$).

Thus, any stochastic process that is simulated is either rejected, i.e. it is not considered as contribution to Z as it does not respect all the values $Z(x_1), \dots, Z(x_{n-1})$, or it leads to an extremal function. Denoting by $\Phi^{(n)} = \{(\xi_i^{(n)}, \psi_i^{(n)}), i \ge 1\}$ a Poisson point process on $(0,\infty) \times C_+(\mathcal{X})$ with intensity measure $\xi^{-2}\mathrm{d}\xi\, P_{x_n}(\mathrm{d}\psi)$, the random number $C_1(N)$ of processes simulated in Algorithm 6.1 satisfies

$$C_1(N) = |\Phi^+_{\{x_1,\dots,x_N\}}| + \sum\nolimits_{n=2}^{N} \left| \left\{ i \ge 1 :\ \xi_i^{(n)} > Z(x_n),\ \xi_i^{(n)} > \min_{j=1}^{n-1} \frac{Z(x_j)}{\psi_i^{(n)}(x_j)} \right\} \right|. \quad (6.25)$$

In this formula, the term $|\Phi^+_{\{x_1,\dots,x_N\}}|$ is the number of extremal functions that need to be simulated, and the term with index n in the sum is the number of functions that are simulated but rejected since $\xi_i^{(n)}\psi_i^{(n)}(x_j) > Z(x_j)$ for some $j \le n-1$. For the computation of the expectation of the second term, conditionally on $\Phi^+_{\{x_1,\dots,x_{n-1}\}}$, i.e. for fixed $Z(x_j)$, $1 \le j \le n-1$, the two sets

$$\begin{aligned}\Phi_1^{(n)} &= \{(\xi_i^{(n)}, \psi_i^{(n)}) : \xi_i^{(n)}\psi_i^{(n)}(x_j) > Z(x_j) \text{ for some } j = 1,\dots,n-1\}\\ \text{and } \Phi_2^{(n)} &= \{(\xi_i^{(n)}, \psi_i^{(n)}) : \xi_i^{(n)}\psi_i^{(n)}(x_j) \le Z(x_j) \text{ for all } j = 1,\dots,n-1\}\end{aligned}$$

are restrictions of the Poisson point process $\Phi^{(n)}$ to disjoint sets and, thus, are independent Poisson point processes with intensities $\xi^{-2}\mathbf{1}_{\{\xi>\min_{j=1}^{n-1}(Z(x_j)/\psi(x_j))\}}\mathrm{d}\xi\, P_{x_n}(\mathrm{d}\psi)$ and $\xi^{-2}\mathbf{1}_{\{\xi<\min_{j=1}^{n-1}(Z(x_j)/\psi(x_j))\}}\mathrm{d}\xi\, P_{x_n}(\mathrm{d}\psi)$, respectively. Conditioning further on $\Phi_2^{(n)}$, $Z(x_n)$ is also fixed and we obtain

$$\begin{aligned}&\mathbb{E}\left[\left|\left\{(\xi_i^{(n)}, \psi_i^{(n)}) : \xi_i^{(n)} > Z(x_n),\, \xi_i^{(n)} > \min_{j=1}^{n-1}\frac{Z(x_j)}{\psi_i^{(n)}(x_j)}\right\}\right|\right]\\ &= \mathbb{E}\left(\mathbb{E}\left[\left|\left\{(\xi,\psi)\in\Phi_1^{(n)} : \xi > Z(x_n)\right\}\right| \;\middle|\; \Phi^+_{\{x_1,\dots,x_{n-1}\}}, \Phi_2^{(n)}\right]\right)\\ &= \mathbb{E}\left[\int\int \xi^{-2}\mathbf{1}_{\{\xi>Z(x_n)\}}\mathbf{1}_{\left\{\xi>\min_{j=1}^{n-1}\frac{Z(x_j)}{\psi(x_j)}\right\}}\mathrm{d}\xi\, P_{x_n}(\mathrm{d}\psi)\right] = \mathbb{E}\left[\min\left\{\frac{1}{Z(x_n)}, \max_{j=1}^{n-1}\frac{Y_n(x_j)}{Z(x_j)}\right\}\right]\end{aligned}$$

where $Y_n \sim P_{x_n}$ and Z are independent. The relation $\min\{a,b\} = a+b-\max\{a,b\}$, $a,b\in\mathbb{R}$, and the fact that $Y_n(x_n) = 1$ almost surely yield

$$\begin{aligned}&\mathbb{E}\left[\left|\left\{(\xi_i^{(n)}, \psi_i^{(n)}) : \xi_i^{(n)} > Z(x_n),\, \xi_i^{(n)} > \min_{j=1}^{n-1}\frac{Z(x_j)}{\psi_i^{(n)}(x_j)}\right\}\right|\right]\\ &= \mathbb{E}\left\{\frac{1}{Z(x_n)}\right\} + \mathbb{E}\left\{\max_{j=1}^{n-1}\frac{Y_n(x_j)}{Z(x_j)}\right\} - \mathbb{E}\left\{\max_{j=1}^{n}\frac{Y_n(x_j)}{Z(x_j)}\right\}\\ &= 1 + \mathbb{E}\left(|\Phi^+_{\{x_1,\dots,x_{n-1}\}}|\right) - \mathbb{E}\left(|\Phi^+_{\{x_1,\dots,x_n\}}|\right),\end{aligned}$$

as $\mathbb{E}(|\Phi^+_{\{x_1,\dots,x_n\}}|) = \mathbb{E}\{\max_{j=1}^n Y_n(x_j)/Z(x_j)\}$ by Lemma 4.5 in Oesting et al. (2018b). Thus, by (6.25), we obtain

$$\begin{aligned}\mathbb{E}\{C_1(N)\} &= \mathbb{E}\left(|\Phi^+_{\{x_1,\dots,x_N\}}|\right) + \sum_{n=2}^{N}\left\{1 + \mathbb{E}\left(|\Phi^+_{\{x_1,\dots,x_{n-1}\}}|\right) - \mathbb{E}\left(|\Phi^+_{\{x_1,\dots,x_n\}}|\right)\right\}\\ &= N - 1 + \mathbb{E}\left(|\Phi^+_{\{x_1\}}|\right) = N.\end{aligned}$$

Moreover, by (6.2), we have that $\mathbb{E}Z(x_i)^{-1} = 1$ for $i = 1,\dots,N$, and, thus,

$$\mathbb{E}\left\{\max_{i=1}^{N} Z(x_i)^{-1}\right\} \ge 1,$$

with equality if only if $Z(x_1) = \cdots = Z(x_N)$ holds almost surely.
The expectation of $C_2(N)$ can be calculated similarly to Proposition 4.6 in Oesting et al. (2018b). □

Remark 6.19. The expectation of $C_1(N)$ does not depend on $Z(x)$. Further characteristics of its distribution such as its variance however depend strongly on the model

ε	$\bar{\mu}_1$	$\bar{\mu}_2$	$\bar{\sigma}_1$	$\bar{\sigma}_2$
0.25	291	1037	195	605
0.5	79	260	53	149
1	25	68	15	40
2	9	21	5	12

Table 6.1: Empirical means $\bar{\mu}_1$ and $\bar{\mu}_2$ and standard deviations $\bar{\sigma}_1$ and $\bar{\sigma}_2$ of the number of random vectors to be simulated to obtain an exact sample of a Brown–Resnick process on the grid $(\varepsilon\mathbb{Z}\cap[-2,2])\times(\varepsilon\mathbb{Z}\cap[-2,2])$ via Algorithm 6.1 and Algorithm 6.2, respectively.

and apparently cannot be readily expressed by an explicit formula in the general case. The following simple examples may provide some further insight into the distribution of $C_1(N)$. In the case of independent random variables $Z(x_i)$, $1 \leq i \leq N$, the extremal function at x_i is $Z(x_i)\mathbf{1}_{\{x_i\}}(\cdot)$ which is why, at each step of Algorithm 6.1, a new extremal function is simulated and accepted, whence $C_1(N) \equiv N$. In contrast, for completely dependent random variables $Z(x_i) \equiv Z(x_1)$, there is only one extremal function, namely the constant function $x \mapsto Z(x_1)$, the one simulated at the first location. At each further step of Algorithm 6.1, all proposed extremal functions are rejected. The number $C_1(N)$ follows a geometric distribution with success probability $1/N$ and, thus, $\mathbb{E}\{C_1(N)\} = N$ and $\mathrm{Var}\{C_1(N)\} = N(N-1)$. In this case, $C_1(N)$ and $C_2(N)$ share not only the same mean, cf. Proposition 6.18), but also the same distribution.

We conclude this section with some comments on the complexity of our algorithms and a comparison with other exact simulation procedures. Proposition 6.18 shows that, for any max-stable process, Algorithm 6.1 is more efficient than Algorithm 6.2 in terms of the expected number of simulated functions. As the spectral functions follow either of the laws $P_{x_1}, \ldots, P_{x_N}$ or a mixture of these, the simulation of a single spectral function is equally complex in both cases. Thus, the new Algorithm 6.1 based on extremal functions is always preferable to Algorithm 6.2.

The differences in complexity of the two algorithms are further illustrated in a simulation study. We consider exact simulations of the Brown–Resnick process associated to the variogram $\gamma(h) = \|h\|$ on a grid $(\varepsilon\mathbb{Z} \cap [-2,2]) \times (\varepsilon\mathbb{Z} \cap [-2,2])$. For $\varepsilon \in \{0.25, 0.5, 1, 2\}$, Algorithms 6.1 and 6.2 are run 10000 times. The empirical means $\bar{\mu}_1$ and $\bar{\mu}_2$ and standard deviations $\bar{\sigma}_1$ and $\bar{\sigma}_2$ of $C_1(N)$ and $C_2(N)$, respectively, are reported in Table 6.1. It can be seen that both the mean and the standard deviation are remarkably larger in case of Algorithm 6.2. The corresponding histograms for $C_1(N)$ and $C_2(N)$ are displayed in Figure 6.3 for the cases $\varepsilon = 0.5$ and $\varepsilon = 1$.

Finally, we briefly comment on exact simulation via the normalized spectral representation proposed by Oesting et al. (2018b). By Proposition 4.6 in Oesting et al. (2018b), the number $C_3(N)$ of simulated normalized spectral functions in this algorithm satisfies

$$\mathbb{E}\{C_3(N)\} = \Big\{ \int \max_{i=1,\ldots,N} \psi(x_i)\, \nu(\mathrm{d}\psi) \Big\} \mathbb{E}\Big\{ \max_{i=1,\ldots,N} Z(x_i)^{-1} \Big\} \tag{6.26}$$

and, thus, depends both on the geometry of the set $\{x_1, \ldots, x_N\}$ and on the law of the max-stable process Z. In general, the numbers $E\{C_3(N)\}$ and $E\{C_1(N)\}$ can however not directly be used to compare the complexity of simulation via the normalized spectral representation and simulation via extremal functions, because the distribution and the simulation complexity of a single random function are different for the two algorithms. As

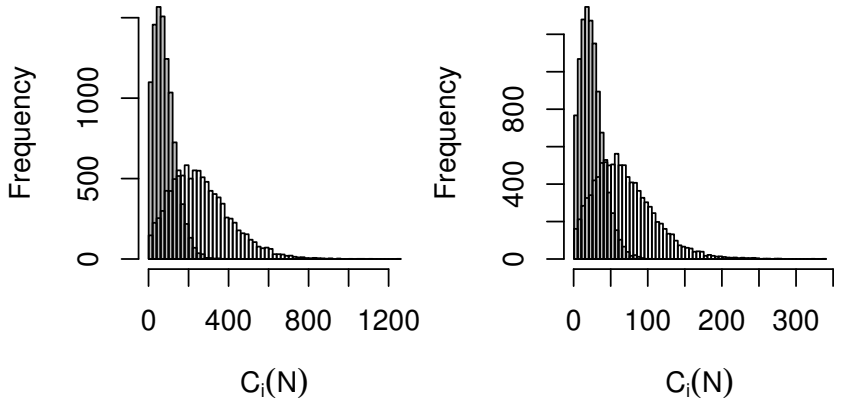

Figure 6.3: Histograms for $C_1(N)$ (grey) and $C_2(N)$ (white) based on 10000 exact simulations of an original Brown–Resnick process on the grid $(\varepsilon\mathbb{Z} \cap [-2,2]) \times (\varepsilon\mathbb{Z} \cap [-2,2])$ for $\varepsilon = 1$ (left) and $\varepsilon = 0.5$ (right), respectively.

we have seen in Section 6.4, the random functions in Algorithms 6.1 and 6.2 with distributions P_{x_0} in (6.5), $x_0 \in \mathcal{X}$, can be simulated efficiently for the most popular max-stable process and extreme value copula models. For the normalized spectral function, exact and efficient simulation procedures are only known for some cases such as mixed moving maxima processes, but are unavailable in other cases like Brown–Resnick or extremal-t processes. For this reason, simulation via Algoritm 6.1 is often preferable to simulation via the normalized spectral representation when $E\{C_1(N)\} > E\{C_3(N)\}$.

6.6 Simulation on Dense Grids

In many applications, one is interested in simulating a max-stable process Z on a dense grid, e.g., $x = \mathcal{X} \cap (\varepsilon\mathbb{Z})^d$. As discussed in Section 6.5, on average, this requires the simulation of $EC_1(N) = N$ random functions in Algorithm 6.1, that is, the simulation of N random vectors of size N. For small ε, N will be large and the procedure can become very time-consuming. Thus, one might be interested in aborting Algorithm 6.1 after $m < N$ steps, ensuring exactness of the simulation only at locations $x_1, \ldots, x_m$. In this case, an alternative design of the algorithm which efficiently chooses the subset of m locations might improve the probability of an exact sample at all N locations.
For comparison of two designs, we introduce the random number

$$N_0 = \min\{m \in \{1, \ldots, N\} : Z_m(x) = Z_N(x)\}.$$

For $n \geq N_0$, the algorithm does not provide any new extremal functions, but all the simulated functions are rejected. Hence, N_0 is the optimal number of iterations before aborting the algorithm. One design is preferable to another if its corresponding random number N_0 is stochastically smaller. An efficient design should thus simulate the extremal functions at an early stage of the algorithm. Based on the intuition that $\phi^+_{x_{n+1}}$ is likely not to be contained in Φ^+_n if $Z_n(x_{n+1})$ is small, we propose the following adaptive numbering $x^{(1)}, \ldots, x^{(N)}$ of points in Algorithm 6.1:

$$\begin{aligned} x^{(1)} &= x_1, \\ x^{(n+1)} &= \operatorname{argmin}\left\{Z_n(x) : x \in \{x_1, \ldots, x_N\} \setminus \{x^{(1)}, \ldots, x^{(n)}\}\right\}, \quad n = 1, \ldots, N-1. \end{aligned} \tag{6.27}$$

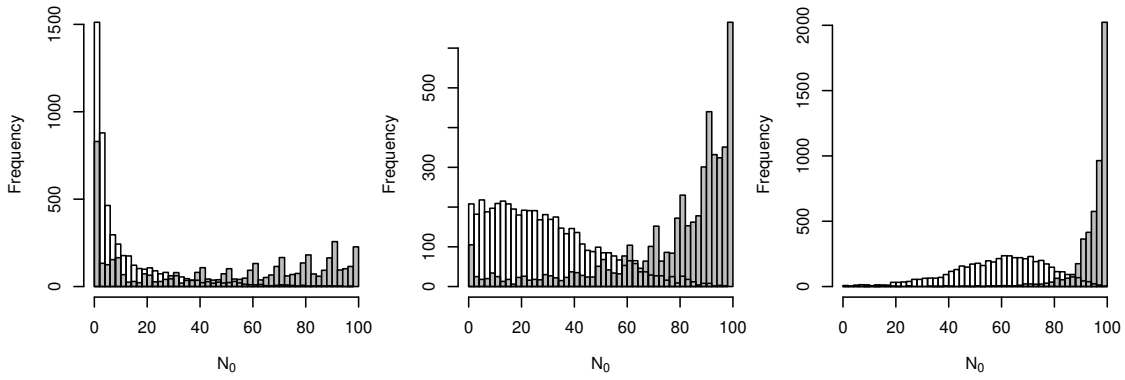

Figure 6.4: Histogram of N_0 based on 5000 realizations of a Brown–Resnick process associated to the semi-variogram $\gamma(h) = c\|h\|^\alpha$ with $c = 1$ and $\alpha = 1.5$ (left), $c = 2.5$ and $\alpha = 1$ (middle) and $c = 5$ and $\alpha = 0.5$ simulated via Algorithm 6.1 with the deterministic design (grey) and the adaptive design (6.27) (white), respectively.

We perform a simulation study to compare the adaptive version of Algorithm 6.1 introduced in (6.27) to a version, where the numbering of locations is deterministic. The simulation study is based on 5000 simulations of a Brown–Resnick process associated to a semivariogram of the type $\gamma(h) = c\|h\|^\alpha$ on the two-dimensional grid $\{0.05, 0.15, \ldots, 0.95\} \times \{0.05, 0.15, \ldots, 0.95\}$. We run Algorithm 6.1 with the deterministic design (the grid points are ordered by their coordinates in the lexicographical sense) and with the adaptive design (6.27). The simulation is repeated for different values of the parameter vector (c, α) representing strong dependence ($(c, \alpha) = (1, 1.5)$), moderate dependence ($(c, \alpha) = (2.5, 1)$) and weak dependence ($(c, \alpha) = (5, 0.5)$). The histograms of N_0 are shown in Figure 6.6. For each of the three parameter vectors, the number N_0 for the adaptive design is stochastically smaller than the corresponding number for the deterministic design.

Acknowledgement

We thank the editorial team and three referees for helpful comments, and the Swiss National Science Foundation and the Agence Nationale de la Recherche project McSim for financial support.

7 On the Distribution of a Max-Stable Process Conditional on Max-Linear Functionals

Up to minor modifications, this chapter is a reprint of the article Oesting (2015) that has been published in *Statistics & Probability Letters*.

Recently, Dombry and Éyi-Minko (2013) provided formulae for the distribution of a max-stable process conditional on its values at given sites and and proposed a methodology for sampling from this distribution. In this chapter, we generalize their results by allowing for conditions stemming from max-linear functionals of the process. Furthermore, we show that the conditional distribution of the extremal functions, i.e. the spectral functions attaining the imposed conditions, is closely related to the normalized spectral representation. The results are illustrated in several examples.

7.1 Introduction

During the last years, max-stable processes have become frequently used models for spatial extremes, in particular for applications in environmental sciences. In the context of the prediction of these processes given some data, the question of their conditional distribution arises. The conditions considered so far are restricted to values of the process at several sites. For this case, exact formulae in terms of the exponent measure of the max-stable process have been provided (Dombry and Éyi-Minko, 2013) and explicit computations have been implemented for several subclasses (cf. Dombry et al., 2013; Oesting and Schlather, 2014, for example).
In this chapter, we analyze the conditional distribution allowing for more general conditions given by max-linear functionals of the process. For instance, a condition on the maximum of the process may be considered. In this case, the analysis of the conditional distribution may provide further insight in characteristics of extreme events that exceed a certain value. More precisely, we consider a max-stable process $\{Z(x),\, x \in K\}$ on some compact set $K \subset \mathbb{R}^d$ which — without loss of generality — can be assumed to have unit Fréchet marginals, i.e. $\mathbb{P}(Z(x) \leq z) = \exp(-1/z)$, $z > 0$, for all $x \in K$. Further, we require Z to be sample-continuous, that is, all sample paths are in the space $C_+(K)$ of nonnegative continuous functions on K. Thus, Z possesses a spectral representation (see de Haan, 1984; Giné et al., 1990; Penrose, 1992, for example):

$$Z(t) = \max_{i \in \mathbb{N}} U_i W_i(t), \quad t \in K, \tag{7.1}$$

where $\{U_i,\, i \in \mathbb{N}\}$, is a Poisson point process on $(0, \infty)$ with intensity measure $u^{-2}\mathrm{d}u$ and W_i, $i \in \mathbb{N}$, are independent copies of a nonnegative sample-continuous stochastic process W with $\mathbb{E}W(t) = 1$ for all $t \in K$.
Assume that we observe values of continuous functionals $L_1, \ldots, L_n : C_+(K) \to [0, \infty)$ that are max-linear, i.e.

$$L_j(\max\{a_1 f_1, a_2 f_2\}) = \max\left\{a_1 L_j(f_1), a_2 L_j(f_2)\right\},$$

for all $a_1, a_2 \geq 0$, $f_1, f_2 \in C_+(K)$, $j = 1, \ldots, n$. An example for such a max-linear functional is $L(f) = \sup_{t \in K'} f(t)$ for some compact set $K' \subset K$, including the special cases $L(f) = \sup_{t \in K} f(t)$ and $L(f) = f(t_0)$ for some $t_0 \in K$. More generally, using continuity arguments and the compactness of K, it can be shown that every functional of this type is of the form

$$L(f) = \sup_{t \in K} h(t) f(t), \qquad f \in C_+(K), \tag{7.2}$$

for some bounded, but not necessarily continuous function $h : K \to [0, \infty)$.
In this paper, we analyze the conditional distribution of Z given $L_1(Z), \ldots, L_n(Z)$, i.e. the distribution of $Z \mid \boldsymbol{L}(Z)$ where $\boldsymbol{L}(Z) = (L_1(Z), \ldots, L_n(Z))^\top$. In Section 7.2, we provide formulae for the conditional distribution in terms of the exponent measure generalizing the results of Dombry and Éyi-Minko (2013). More explicit expressions for the case of one single condition are derived in Section 7.3 making use of connections to the normalized spectral representations to max-stable processes. Section 7.4 deals with the more general case of a finite number of conditions.

7.2 General Theory

In the following, we analyze the distribution of Z conditionally on $\boldsymbol{L}(Z) = \boldsymbol{z}$ for some $\boldsymbol{z} = (z_1, \ldots, z_n)^\top \in (0, \infty)^n$. Here, we note that, because of the max-linearity of L_j, $j = 1, \ldots, n$, the properties of $L_j(Z)$ are directly connected to those of $L_j(W)$. By Proposition 2.3 in Oesting et al. (2018b), the finiteness of $L(Z)$ implies that $\mathbb{E}L_j(W) < \infty$ and $\mathbb{P}(L_j(Z) \leq z) = \exp(-\mathbb{E}(L_j(W))/z)$, $z > 0$, i.e., $L_j(Z)$ follows a Fréchet distribution. To exclude the trivial case that the distribution of $L_j(Z)$ is degenerate, assume that $p_j = \mathbb{P}(L_j(W) > 0) > 0$ for every $j \in \{1, \ldots, n\}$. We consider the extended process $Z_L = (\{Z(t),\, t \in K\}, \boldsymbol{L}(Z))$ on $C_+(K) \times (0, \infty)^n$. By the max-linearity of L_j, we obtain that $L_j(Z) = \max_{i \in \mathbb{N}} U_i L_j(W_i)$. Thus,

$$Z_L = \max_{i \in \mathbb{N}} (\xi_i, \boldsymbol{L}(\xi_i)) \tag{7.3}$$

where the maximum is considered componentwise and $\Pi = \sum_{i \in \mathbb{N}} \delta_{(\xi_i, \boldsymbol{L}(\xi_i))}$ denotes a Poisson point process on $S = C_+(K) \times [0, \infty)^n$ with intensity measure

$$\Lambda(A \times B) = \int_0^\infty u^{-2} \mathbb{P}(uW \in A, u\boldsymbol{L}(W) \in B)\, \mathrm{d}u,$$

for Borel sets $A \subset C_+(K)$ and $B \subset [0, \infty)^n$ (cf. Kingman, 1993), that is, ξ_i corresponds to the product $U_i W_i$ in representation (7.1).
Perceiving the conditions $L_1(Z) = z_1, \ldots, L_n(Z) = z_n$ as conditions on the value of the process Z_L at specific "sites" according to representation (7.3) the results of Dombry and Éyi-Minko (2013) can be applied on Z_L, in order to derive the distribution of Z conditional on $\mathbf{L}(Z)$. To this end, for every non-empty index subset $J \subset \{1, \ldots, n\}$, we consider the J-extremal random point measure Π_J^+ and the J-subextremal random point process Π_J^-, defined by

$$\Pi_J^+ = \sum_{i \in \mathbb{N}} \delta_{\xi_i} \mathbf{1}_{\{L_j(\xi_i) = L_j(Z) \text{ for some } j \in J\}} \quad \text{and} \quad \Pi_J^- = \sum_{i \in \mathbb{N}} \delta_{\xi_i} \mathbf{1}_{\{L_j(\xi_i) < L_j(Z) \text{ for all } j \in J\}}.$$

It can be shown that Π_J^+ and Π_J^- are well-defined point processes on $C_+(K)$ (see Dombry and Éyi-Minko, 2013, Lemma A.3). Further, as $\Pi_{\{j\}}^+(C_K^+) = 1$ a.s. (cf. Dombry and Éyi-Minko, 2013, Proposition 2.5), $\Pi_{\{1,\ldots,n\}}^+$ is characterized via so-called hitting scenarios (cf.

Wang and Stoev, 2011; Dombry and Éyi-Minko, 2013), i.e. partitions $\tau = \{\tau_1, \dots, \tau_l\}$ of $\{1, \dots, n\}$ representing the situation that $\Pi^+_{\{1,\dots,n\}} = \{\xi_1^+, \dots, \xi_l^+\}$ s.t.

$$L_j(\xi_k^+) \begin{cases} = L_j(Z), & j \in \tau_k \\ < L_j(Z), & j \notin \tau_k \end{cases}, \qquad 1 \le j \le n,\ 1 \le k \le l.$$

Let $\Theta \in \mathcal{P}_{\{1,\dots,n\}}$ be the random partition realized by Z where $\mathcal{P}_{\{1,\dots,n\}}$ denotes the space of partitions of $\{1, \dots, n\}$. Based on the different hitting scenarios, conditional simulations of $Z \mid \boldsymbol{L}(Z) = \boldsymbol{z}$ can be performed via the following three-step procedure proposed by Dombry and Éyi-Minko (2013):

1. Draw a partition $\tau = \{\tau_k\}_{k=1}^l$ from the distribution of $\Theta \mid \boldsymbol{L}(Z) = \boldsymbol{z}$.
2. Simulate $(\xi_k^+\}_{k=1}^l$ as a realization of $\Pi^+_{\{1,\dots,n\}} \mid \boldsymbol{L}(Z) = \boldsymbol{z}, \Theta = \tau$.
3. Draw $\{\xi_i^-\}_{i\in\mathbb{N}}$ from the distribution of $\Pi^-_{\{1,\dots,n\}} \mid \boldsymbol{L}(Z) = \boldsymbol{z}$.

Then, $\max_{k=1}^l \xi_k^+ \vee \max_{i\in\mathbb{N}} \xi_i^-$ is a realization of $Z \mid \boldsymbol{L}(Z) = \boldsymbol{z}$.

The distributions involved in the algorithm are given in the following theorem which summarizes Theorem 3.1 and Theorem 3.2 in Dombry and Éyi-Minko (2013). First, we need some more notation. For any non-empty index subset $J \subset \{1, \dots, n\}$, define the mapping $R_J : [0,\infty)^n \to [0,\infty)^{|J|}$ as the projection on the components belonging to J, i.e. $R_J(\boldsymbol{z}) = \boldsymbol{z}_J$ where $\boldsymbol{z}_J = (z_j)_{j\in J}$ for $\boldsymbol{z} = (z_j)_{j=1}^n$. Further, let μ be the exponent measure of Z, i.e. $\mu(A) = \Lambda(A \times [0,\infty)^n)$, $A \subset C_+(K)$, and μ_J the exponent measure of $(L_j(Z))_{j\in J}$, i.e.

$$\mu_J(B) = \Lambda(C_+(K) \times R_J^{-1}(B)), \quad B \subset [0,\infty)^{|J|} \text{ Borel}.$$

The tail function of μ_J is denoted by $\bar\mu_J(\boldsymbol{z}_J) = \mu_J([0, \boldsymbol{z}_J)^c)$, $\boldsymbol{z}_J \in [0,\infty)^{|J|}$. Furthermore, let $\{P_J(\boldsymbol{z}_J; \cdot),\ \boldsymbol{z}_J \in [0,\infty)^{|J|}\}$ be a regular version of the measure $\mu(\mathrm{d}f)$ conditional on $(L_j(f))_{j\in J} = \boldsymbol{z}_J$. That is, $P_J(\boldsymbol{z}_J; \cdot)$ satisfies $\int_A \mathbf{1}_B(\boldsymbol{L}(f))\, \mu(\mathrm{d}f) = \int_B \int_A P_J(\boldsymbol{z}; \mathrm{d}f)\, \mu_J(\mathrm{d}\boldsymbol{z}_J)$ for all Borels sets $A \subset C_+(K)$ and $B \subset [0,\infty)^{|J|}$.

Theorem 7.1 (adapting Dombry and Éyi-Minko, 2013)**.** *With the above notation, the following holds true:*

1. $\mathbb{P}(\Theta = \tau \mid \boldsymbol{L}(Z) = \boldsymbol{z}) = \dfrac{\mathbb{P}(\Theta = \tau,\ \boldsymbol{L}(Z) \in \mathrm{d}\boldsymbol{z})}{\sum_{\tau' \in \mathcal{P}_{\{1,\dots,n\}}} \mathbb{P}(\Theta = \tau',\ \boldsymbol{L}(Z) \in \mathrm{d}\boldsymbol{z})}$,
   where the fraction is understood as a Radon-Nikodym derivative and

$$\begin{aligned} &\mathbb{P}(\Theta = \tau,\ \boldsymbol{L}(Z) \in \mathrm{d}\boldsymbol{z}) \\ &= \exp\left(-\bar\mu_{\{1,\dots,n\}}(\boldsymbol{z})\right) \prod_{k=1}^l \left\{P_{\tau_k}(\boldsymbol{z}_{\tau_k}; \{(L_j(f))_{j\in\tau_k^c} < \boldsymbol{z}_{\tau_k^c}\}) \mu_{\tau_k}(\mathrm{d}\boldsymbol{z}_{\tau_k})\right\}. \end{aligned}$$

2. $\mathbb{P}(\xi_k^+ \in \mathrm{d}f_k \mid \boldsymbol{L}(Z) = \boldsymbol{z},\ \Theta = \tau) = \displaystyle\prod_{k=1}^l \frac{\mathbf{1}_{\{\boldsymbol{L}_{\tau_k^c}(f_k) < \boldsymbol{z}_{\tau_k^c}\}} P_{\tau_k}(\boldsymbol{z}_{\tau_k}; \mathrm{d}f_k)}{P_{\tau_k}(\boldsymbol{z}_{\tau_k}; \{f :\ \boldsymbol{L}_{\tau_k^c}(f) < \boldsymbol{z}_{\tau_k^c}\})}$.

3. *Conditionally on $\boldsymbol{L}(Z) = \boldsymbol{z}$, $\Pi^-_{\{1,\dots,n\}}$ is a Poisson point process with intensity measure $\mathbf{1}_{\{\boldsymbol{L}(f) < \boldsymbol{z}\}}\, \mu(\mathrm{d}f)$, independently of Θ and $\Pi^+_{\{1,\dots,n\}}$.*

Remind that, for the conditional simulation of Z, we do not need the whole Poisson point process $\Pi^-_{\{1,\dots,n\}}$, but only the pointwise maximum $Z^-(t) = \max_{i\in\mathbb{N}} \xi_i^-(t)$, $t \in K$. By the

third part of Theorem 7.1, simulation of Z^- is closely related to unconditional simulation of Z. The only difference is that those points that violate the condition $\boldsymbol{L}(Z) = \boldsymbol{z}$ are neglected.

A specific spectral representation of Z proves beneficial for an exact simulation: By de Haan and Ferreira (2006), Cor. 9.4.5, any sample-continuous max-stable process with unit Fréchet margins can be written as

$$Z(t) = \max_{i\in\mathbb{N}} U_i V_i(t), \quad t \in K,$$

where V_i, $i \in \mathbb{N}$, are independent copies of some stochastic process V with $\sup_{t\in K} V(t) = c$ a.s. for some uniquely determined $c > 0$. Recently, Oesting et al. (2018b) revisited this representation – which they call the normalized spectral representation – and showed that the law of the process V can be expressed in terms of the law of the process W in representation (7.1) via the equality

$$V =_d c \cdot \left(\sup_{t\in K} \tilde{W}(t)\right)^{-1} \cdot \tilde{W},$$

where the law of $\tilde{W}$ is given by

$$\mathbb{P}(\tilde{W} \in A) = c^{-1} \int_A \sup_{t\in K} w(t)\, \mathbb{P}(W \in \mathrm{d}w), \quad A \subset C_+(K),$$

with $c = \mathbb{E}(\sup_{t\in K} W(t))$. For this representation, only a finite number of tupels (U_i, V_i) — those tupels that satisfy $cU_i > \inf_{t\in K} Z(t)$ — can contribute to the maximum Z. This fact allows for an exact simulation of Z in finite time and can also be used for an exact simulation in the third step of the procedure for conditional simulation described above. Further, as we will see in what follows, the distribution of the normalized spectral functions is closely connected to the conditional distribution of the extremal functions.

7.3 Conditioning on One Max-Linear Functional

We first consider to the special case that we have one condition $L(Z) = z$ for some $z > 0$, only. In this case, there is one extremal function ξ^+ satisfying $L(\xi^+) = z$ and there is no need to consider different hitting scenarios. The second part of Theorem 7.1 allows us to calculate the distribution of ξ^+:

$$\mathbb{P}(\xi^+ \in A \mid L(Z) = z) = \mathcal{P}_{\{1\}}(z; A) = \frac{\mathrm{d}}{\mathrm{d}z}\Lambda(A \times [0, z]) = \frac{\mathbb{E}\left(L(W)\mathbf{1}_{\{zW/L(W)\in A\}}\right)}{\mathbb{E}L(W)}, \quad (7.4)$$

for $A \subset C_+(K)$. Thus, we can write $\xi^+ = z\tilde{W}_L/L(\tilde{W}_L)$, where the law of $\tilde{W}_L$ is given by the Radon-Nikodym derivative $\mathbb{P}(\tilde{W}_L \in \mathrm{d}w)/\mathbb{P}(W \in \mathrm{d}w) = c_L^{-1}L(w)$ with $c_L = \mathbb{E}L(W)$.

Remark 7.2. Note that there is a connection between the law of the extremal function ξ^+ and a representation similar to the normalized spectral representation. Generalizing the functional $f \mapsto \sup_{t\in K} f(t)$ by a max-linear functional $L : C_+(K) \to [0, \infty)$ as above, we obtain the L-normalized spectral representation

$$Z(t) = \max_{i\in\mathbb{N}} U_i V_i^L(t), \quad t \in K, \quad (7.5)$$

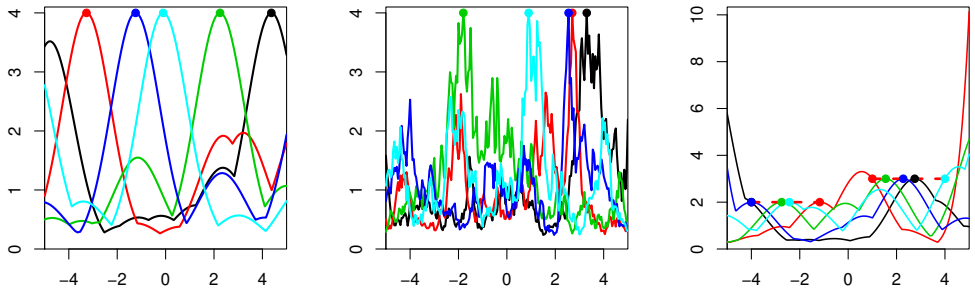

Figure 7.1: Examples of samples from the distribution of max-stable processes conditional on max-linear functionals containing five realizations each. The dots indicate the hitting of the condition. Left: A Smith process Z with $\sigma = 1$ conditional on $\max_{t\in[-5,5]} Z(t) = 4$. Middle: A Brown-Resnick process Z associated to the variogram $\gamma(h) = 2|h|^{1.3}$ conditional on $\max_{t\in[-5,5]} Z(t) = 4$. Right: A Smith process Z with $\sigma = 1$ conditional on $L_1(Z) = \max_{t\in[-4,-1]} Z(t) = 2$ and $L_2(Z) = \max_{t\in[1,4]} Z(t) = 3$.

where V_i^L, $i \in \mathbb{N}$, are independent copies of a stochastic process V_L satisfying

$$V_L = B\frac{c_L}{\mathbb{P}(L(W) > 0)L(\tilde{W}_L)}\tilde{W}_L + (1 - B)W_0$$

where B is some Bernoulli random variable with parameter $p = \mathbb{P}(L(W) > 0)$, $\tilde{W}_L$ has the same law as in (7.4) and W_0 has the distribution of $W \mid L(W) = 0$. Thus, the spectral functions V_i^L, $i \in \mathbb{N}$, of the L-normalized spectral representation are characterized by the fact that $L(V_i^L) = p^{-1}c_L$ with probability p and $L(V_i^L) = 0$ with probability $1 - p$. From (7.5), it can be seen that the law of extremal function ξ^+ conditional on $L(Z) = z$ equals the law of $c_L^{-1}pzV_L \mid L(V_L) > 0$, confirming Equation (7.4).

In practice, the simulation of the process $\tilde{W}_L$, which is needed for the simulation of the extremal function, and the process $\tilde{W}$, which can be used for the exact simulation of the subextremal functions, is not straightforward. Here, remind that $\tilde{W}$ is a special case of $\tilde{W}_L$ for the choice $L(f) = \max_{t\in K} f(t)$. One possible way of simulation is simulation via MCMC methods. For instance, one could use the Metropolis-Hastings algorithm with the distribution of W as proposal distribution.

We now present two examples for sampling of max-stable processes conditionally on the max-linear functional $L(f) = \max_{t\in K} f(t)$.

Example 7.3 (Smith process)**.** *We consider Smith's (1990) process of moving maxima type*

$$Z(t) = \max_{i\in\mathbb{N}} \tilde{U}_i f_\sigma(t - S_i), \quad t \in K,$$

on some finite interval $K = [-r, r]$, where $\sum_{i\in\mathbb{N}} \delta_{(\tilde{U}_i,S_i)}$ is a Poisson point process on $(0,\infty)\times\mathbb{R}$ with intensity u^{-2}duds and $f_\sigma(x) = (\sqrt{2\pi}\sigma)^{-1}\exp\left(-x^2/(2\sigma^2)\right)$, $x \in \mathbb{R}$, denotes the normal density with standard deviation $\sigma > 0$. As we consider a condition on the overall maximum of Z, we can make use of the normalized spectral representation of the Smith process which was calculated explicitly in Oesting et al. (2018b):

$$Z(t) =_d \max_{i\in\mathbb{N}} U_i \frac{cf(t - \tilde{S}_i)}{f(\max\{0, |\tilde{S}_i| - r\})}, \quad t \in K, \tag{7.6}$$

where $\tilde{S}_i$, $i \in \mathbb{N}$, are independent copies of some random variable $\tilde{S}$ whose law is given by $\mathbb{P}(\tilde{S} \le s) = c^{-1} \int_{-\infty}^{s} f(\max\{0, |x| - r\}) \mathrm{d}x$ and $c = 2rf(0) + 1$. Thus, the extremal function ξ^+ can be written as

$$\xi^+(t) = z \frac{f(t - \tilde{S})}{f(\max\{0, |\tilde{S}| - r\})}, \quad t \in K.$$

As the spectral functions in representation (7.6) can be simulated easily, we also use them to simulate the subextremal functions contributing to $Z \mid L(Z) = z$ exactly. An example is shown in Figure 7.1.

Example 7.4 (Brown-Resnick process)**.** *The second example we consider are Brown-Resnick processes (Brown and Resnick, 1977; Kabluchko et al., 2009) on some compact set $K \subset \mathbb{R}^d$:*

$$Z(t) = \max_{i \in \mathbb{N}} U_i \exp\left(B_i(t) - \frac{\sigma^2(t)}{2}\right), \quad t \in K,$$

where B_i, $i \in \mathbb{N}$, are independent copies of a centered Gaussian process $\{B(t), t \in \mathbb{R}^d\}$ with stationary increments, variance $\sigma^2(\cdot)$ and variogram $\gamma(h) = \mathrm{Var}\,(B(h) - B(0))$. Kabluchko et al. (2009) showed that the process Z, extended to $\mathbb{R}^d$, is stationary and its law depends on the variogram γ only. Figure 7.1 shows five realizations of a Brown-Resnick process Z on $[-5, 5]$ conditional on $\max_{t \in [-5,5]} Z(t) = 4$. Here, for the simulation as well of the extremal as of the subextremal functions, we use a Metropolis-Hastings algorithm. However, for the subextremal functions, also other (approximative) methods for the simulation of Brown-Resnick processes can be used, see Oesting et al. (2012).

7.4 Conditioning on a Finite Number of Max-Linear Functionals

In this section, we use the formulae given in Theorem 7.1 to explicitly calculate the distribution of the random partition Θ and the extremal functions ξ_k^+ in the general case of n conditions $L_1(Z) = z_1, \ldots, L_n(Z) = z_n$. Here, we restrict to the case that the distribution of the random vector $\boldsymbol{L}(W)$ is absolutely continuous and has density f_L with respect to the Lebesgue measure. Then, applying Theorem 7.1, yields to the following proposition.

Proposition 7.5. *Under the above assumptions, we obtain the following results.*

1. *For any partition $\tau = \{\tau_1, \ldots, \tau_l\} \in \mathcal{P}_{\{1,\ldots,n\}}$, we have*

$$\mathbb{P}\left(\Theta = \tau \mid \boldsymbol{L}(Z) = \boldsymbol{z}\right) \sim \prod_{k=1}^{l} \int_0^\infty v^{|\tau_k|} \int_{\left[0, v\boldsymbol{z}_{\tau_k^c}\right]} f_{\boldsymbol{L}}\left(v\boldsymbol{z}_{\tau_k}, \boldsymbol{y}_{\tau_k^c}\right) \mathrm{d}\boldsymbol{y}_{\tau_k^c} \mathrm{d}v.$$

2. *Conditional on $\Theta = \tau$ for a partition τ as in the first part, the distribution of the extremal functions ξ_k^+, $k = 1, \ldots, l$, is given by*

$$\begin{aligned} &\mathbb{P}(\xi_k^+ \in A \mid \boldsymbol{L}(Z) = \boldsymbol{z}, \Theta = \tau) \\ &\sim \int_0^\infty v^{|\tau_k|} \int_{\left[0, v\boldsymbol{z}_{\tau_k^c}\right]} \mathbb{P}\left(z_j W / L_j(W) \in A \mid \boldsymbol{L}(W) = \left(v\boldsymbol{z}_{\tau_k}, \boldsymbol{y}_{\tau_k^c}\right)\right) \\ &\qquad\qquad f_{\boldsymbol{L}}\left(v\boldsymbol{z}_{\tau_k}, \boldsymbol{y}_{\tau_k^c}\right) \mathrm{d}\boldsymbol{y}_{\tau_k^c} \mathrm{d}v, \end{aligned}$$

   for any Borel set $A \subset C_+(K)$ and an arbitrary index $j \in \tau_k$.

In the case $\tau = \{\{1, \ldots, n\}\}$, the inner integrals are read as a single evaluation of the integrand.

Alternatively to representation (7.1) and the corresponding spectral process W, we can also consider the L-normalized spectral representation (7.5). Let $V^{(j)}$ be the spectral process of the L_j-normalized spectral representation conditional on $L_j > 0$, i.e. $L_j(V^{(j)}) = c^{(j)}$ a.s. for some $c^{(j)} > 0$. Assuming that all the $(n-1)$-dimensional vectors $(L_1(V^{(j)}), \dots, L_{j-1}(V^{(j)}), L_{j+1}(V^{(j)}), \dots, L_n(V^{(j)}))^\top$, $j = 1, \dots, n$, possess densities $f_{\boldsymbol{L},j}$ w.r.t. the Lebesgue measure on $(0,\infty)^{n-1}$, Proposition 7.5 can be reformulated in terms of these lower-dimensional densities.

Proposition 7.6. *For any partition $\tau = \{\tau_1, \dots, \tau_l\}$, let $j_{k,\tau}$ be an element of τ_k with $k = 1, \dots, l$. Then, the following statements hold true.*

1. *The conditional distribution of the random partition Θ has the form*

$$\begin{aligned}&\mathbb{P}(\Theta = \tau \mid \boldsymbol{L}(Z) = \boldsymbol{z})\\ &\propto \prod_{k=1}^{l} \frac{(c^{(j_{k,\tau})})^{|\tau_k|}}{z_{j_{k,\tau}}^{|\tau_k|+1}} \int_{\left[0, c^{(j_{k,\tau})} \boldsymbol{z}_{\tau_k^c}/z_{j_{k,\tau}}\right]} f_{\boldsymbol{L},j}\left(c^{(j_{k,\tau})} \frac{\boldsymbol{z}_{\tau_k \setminus \{j_{k,\tau}\}}}{z_{j_{k,\tau}}}, \boldsymbol{y}_{\tau_k^c}\right) \mathrm{d}\boldsymbol{y}_{\tau_k^c}.\end{aligned}$$

2. *The distribution of the extremal functions ξ_k^+, $k = 1, \dots, l$, conditional on $\Theta = \tau$ for $\tau = \{\tau_1, \dots, \tau_l\}$, is given by*

$$\begin{aligned}&\mathbb{P}(\xi_k^+ \in A \mid \boldsymbol{L}(Z) = z, \Theta = \tau)\\ &\propto \int_{\left[0, c^{(j_{k,\tau})} \boldsymbol{z}_{\tau_k^c}/z_{j_{k,\tau}}\right]} f_{\boldsymbol{L},j}\left(c^{(j_{k,\tau})} \frac{\boldsymbol{z}_{\tau_k \setminus \{j_{k,\tau}\}}}{z_{j_{k,\tau}}}, \boldsymbol{y}_{\tau_k^c}\right)\\ &\qquad \mathbb{P}\left(\frac{z_{j_{k,\tau}} V^{(j_{k,\tau})}}{c^{(j_{k,\tau})}} \in A \,\middle|\, \boldsymbol{L}(V^{(j_{k,\tau})}) = \left(c^{(j_{k,\tau})} \frac{\boldsymbol{z}_{\tau_k}}{z_{j_{k,\tau}}}, \boldsymbol{y}_{\tau_k^c}\right)\right) \mathrm{d}\boldsymbol{y}_{\tau_k^c}.\end{aligned}$$

In general, sampling from the distribution of the extremal functions may be quite sophisticated even in the case of few conditions. We finally consider an example that allows for conditional simulation in the case of two conditions.

Example 7.7 (Smith process, cf. Example 7.3)**.** *We consider Smith's (1990) process $\{Z(t), t \in K\}$ on some real interval K conditional on $L_1(Z) = z_1$ and $L_2(Z) = z_2$ for some $z_1, z_2 > 0$. Here, we consider conditions of the type $L_1(f) = \max_{t \in A_1} f(t)$ and $L_2(f) = \max_{t \in A_2} f(t)$ for disjoint closed intervals $A_1, A_2 \subset K$. Using the notation $\tilde{f}_{\sigma,A}(x) = \max_{t \in A} f_\sigma(t - x)$ and $c_A = \int_{\mathbb{R}} \tilde{f}_{\sigma,A}(x)\mathrm{d}x$, the spectral function $V^{(j)}$ of the L_j-normalized spectral representation can be written as $V^{(j)}(t) = c_{A_j} f_\sigma(t - X_j)/\tilde{f}_{\sigma,A_j}(X_j)$, $t \in K$, $j = 1, 2$, where X_j is a random variable with Lebesgue density $c_{A_j}^{-1} \tilde{f}_{\sigma,A_j}$ (cf. Oesting et al., 2018b). Thus, the distribution of $L_k(V^{(j)})$, $j \neq k$, is given by*

$$\mathbb{P}(L_k(V^{(j)}) \in \cdot) = \int_{\mathbb{R}} \mathbf{1}_{\{c_{A_j} \tilde{f}_{\sigma,A_k}(x)/\tilde{f}_{\sigma,A_j}(x) \in \cdot\}} c_{A_j}^{-1} \tilde{f}_{\sigma,A_j}(x)\mathrm{d}x. \tag{7.7}$$

By Proposition 7.6, we obtain

$$\begin{aligned}\mathbb{P}(\Theta = \{\{1\},\{2\}\}) &= \frac{c_{A_1} c_{A_2}}{z_1^2 z_2^2} \mathbb{P}\left(L_2(V^{(1)}) \leq \frac{c_{A_1} z_2}{z_1}\right) \mathbb{P}\left(L_1(V^{(2)}) \leq \frac{c_{A_2} z_1}{z_2}\right),\\ \mathbb{P}(\Theta = \{\{1,2\}\}) &= \frac{c_{A_1}^2}{z_1^3} \mathbb{P}\left(L_2(V^{(1)}) \in \mathrm{d}\frac{c_{A_1} z_2}{z_1}\right) = \frac{c_{A_2}^2}{z_2^3} \mathbb{P}\left(L_1(V^{(2)}) \in \mathrm{d}\frac{c_{A_2} z_1}{z_2}\right),\end{aligned}$$

where the right-hand sides can be calculated by solving the integral in (7.7) *numerically. The extremal functions belonging to the partition* $\{\{1\},\{2\}\}$ *can be simulated easily by simulating* $V^{(1)}$ *and* $V^{(2)}$ *conditional on* $L_2(V^{(1)}) \leq c_{A_1} z_2/z_1$ *and* $L_1(V^{(2)}) \leq c_{A_2} z_1/z_2$*, respectively. Sampling the extremal function associated to* $\{\{1,2\}\}$ *is more involved as it includes the law of* $V^{(j)} \mid L_k(V^{(j)})$*. However, in many situations,* $V^{(j)}$ *is uniquely determined by the condition* $L_k(V^{(j)})/L_j(V^{(j)}) = z_k/z_j$*. Figure 7.1 shows five realizations of a Smith process* Z *conditional on* $L_1(Z) = 2$ *and* $L_2(Z) = 3$*.*

Acknowledgements

This work has been supported by the ANR project McSim. The author thanks Liliane Bel and Christian Lantuéjoul for fruitful discussions on the simulation of normalized spectral functions and an anonymous referee for valuable comments and remarks.

8 Sampling from a Max-Stable Process Conditional on a Homogeneous Functional

joint work with Liliane Bel and Christian Lantuéjoul

This chapter is based on the article Oesting et al. (2018a) that has been published in the *Scandinavian Journal of Statistics*. Here, the section on the application to climate data is omitted as it has mainly been implemented by Liliane Bel. Apart from this, only minor modifications have been made. The appendix of the paper has been put to Section 8.6 and Section 8.7 at the end of the chapter.

Conditional simulation of max-stable processes allows for the analysis of spatial extremes taking into account additional information provided by the conditions. Instead of observations at given sites as usually done, in this chapter, we consider a single condition given by a more general functional of the process as may occur in the context of climate models. As the problem turns out to be intractable analytically, we make use of Markov chain Monte Carlo methods to sample from the conditional distribution. Simulation studies indicate fast convergence of the Markov chains involved.

8.1 Introduction

Naturally occurring as limits of appropriately rescaled pointwise maxima of stochastic processes, max-stable processes have become popular models in spatial and spatio-temporal extremes. For instance, they have found application in modelling of extreme rainfall (Huser and Davison, 2014), temperatures (Davison and Gholamrezaee, 2012), snowfall (Gaume et al., 2013) or wind speeds (Engelke et al., 2015). A detailed review can be found in Davison et al. (2012). In this chapter, we will deal with a sample-continuous max-stable process $\{X(t),\ t \in K\}$ on some compact metric space K with α-Fréchet marginals, i.e. $\mathbb{P}(X(t) \leq x) = \exp\left(-\sigma(t)x^{-\alpha}\right)$, $x > 0$, for some $\sigma(t) > 0$, $\alpha > 0$. In the following, we will write $X(t) \sim \Phi_{\alpha,\sigma(t)}$, for short, using the convention that $\Phi_\alpha = \Phi_{\alpha,1}$. By de Haan (1984), the process X possesses a spectral representation which, without loss of generality, may be assumed to be of Penrose (1992) type:

$$X(t) = \max_{i \geq 1} U_i W_i(t), \quad t \in K, \tag{8.1}$$

where $\{U_i,\ i \geq 1\}$ are the points of a Poisson point process on $(0, \infty)$ with intensity measure $\alpha u^{-(\alpha+1)} \mathrm{d}u$ and W_i, $i \geq 1$, are independent copies of some nonnegative sample-continuous process $\{W(t),\ t \in K\}$ such that $\sigma(t) = \mathbb{E}W(t)^\alpha \in (0, \infty)$ for all $t \in K$.

In the framework of spatial and spatio-temporal modelling, simulations have proven to be a useful tool for gaining more insight in the structure of the problem at hand. In the situation that additional information on the process is available, simulations conditional on this information are required. For max-stable processes, expressions for the conditional

distribution and algorithms for conditional simulation have been developed recently. Wang and Stoev (2011) provide exact formulae and an exact and efficient algorithm for conditional simulation for the subclass of spectrally discrete, also called max-linear, models. In the general case of max-stable processes, Dombry and Éyi-Minko (2013) derive formulae for the conditional distribution in terms of the exponent measure. Further works on the problem of conditional simulation deal with algorithmic aspects and more explicit expressions for popular subclasses like Brown-Resnick and extremal Gaussian processes (Dombry et al., 2013), (mixed) moving maxima processes (Oesting and Schlather, 2014) and extremal t processes (Ribatet, 2013).
In all the cited works, the conditions for the simulation are given by observations at some sites, i.e. conditions of the type $X(t_1) = x_1, \dots, X(t_k) = x_k$ for some $t_1, \dots, t_k \in K$ and $x_1, \dots, x_k > 0$. Oesting (2015) showed that the same methodology can be used if the conditions stem from max-linear functionals such as the maximum of the process over some subregion. However, in many applications, some aggregated data may be given that are not of max-linear type. For instance, numerical models for meteorological and climate variables often provide average values on large grid cells rather than values at specific sites. Here, conditional sampling given these average values can be used for downscaling. For this purpose, Bechler et al. (2015) proposed to assign aggregated data to specific locations via transfer functions and to perform conditional simulations in the classical framework where the conditions are given by values at several sites. In this chapter, we will deal with a procedure to directly simulate a max-stable process conditional on a more general, single observation functional. More precisely, we consider a positively homogeneous functional $\ell : C_+(K) \to [0, \infty)$, that is, a functional satisfying $\ell(\lambda f) = \lambda \ell(f)$ for all $\lambda \geq 0$ and $f \in C_+(K)$ where $C_+(K)$ denotes the space of nonnegative continuous functions on K equipped with the topology of uniform convergence. Examples for such a functional are $\ell(f) = \sup_{t \in K} f(t)$, $\ell(f) = \inf_{t \in K} f(t)$ or $\ell(f) = \int_K f(t)\mu(\mathrm{d}t)$ for any positive and finite measure μ on K. There is already some related work in the literature such as in the papers by Dombry and Ribatet (2015) and Thibaud and Opitz (2015) who consider a process Y in the domain of attraction of a max-stable process X and analyze the conditional distribution $Y|\ell(Y) > y$, as $y \to \infty$. Here, we will focus on a max-stable process X and condition on $\ell(X) = x$, instead. Thus, we consider the exact value of the functional applied to the limiting max-stable process, while existing work typically considers exceedances of high thresholds by the functional applied to the underlying process as condition.

In general, the conditional distribution cannot be expressed in terms of the exponent measure of the max-stable process. Thus, the approach of Dombry and Éyi-Minko (2013) cannot be applied here. Instead, we will develop a simulation algorithm based on Markov chain Monte Carlo techniques. The problem of conditional simulation of max-stable problem will be approached in three steps: In Section 8.2, we will first restrict to max-linear models. Then, in Section 8.3, we will extend our focus to conditionally max-linear models, before covering the general max-stable case (Section 8.4). Finally, we conclude the chapter and provide some perspectives for future research (Section 8.5).

8.2 Max-linear Models

8.2.1 Background

We first consider a max-linear, also called spectrally discrete, model (cf. Wang and Stoev, 2011):

$$X(t) = \max_{j=1,\dots,n} a_j(t) Z_j, \qquad t \in K, \tag{8.2}$$

where $Z_j \sim \Phi_\alpha$, $j = 1, \ldots, n$, are independent random variables and $a_j \in C_+(K)$ for $j = 1, \ldots, n$. Following Wang and Stoev (2011), henceforth, we will write $X = \boldsymbol{a} \odot \boldsymbol{Z}$ with $\boldsymbol{a} = (a_j(t))_{t\in K, j=1,\ldots,n}$ and $\boldsymbol{Z} = (Z_j)_{j=1,\ldots,n}$ instead of (8.2), for short. Further, for any vector $\boldsymbol{z} \in (0,\infty)^n$ and an index $i \in \{1, \ldots, n\}$, we use the notation $\boldsymbol{z}_{-i} = (z_j)_{j\in\{1,\ldots,n\}\setminus\{i\}}$. The definition (8.2) allows for the calculation of the finite-dimensional distributions yielding

$$\mathbb{P}(X(t_i) \le x_i,\ 1 \le i \le m) = \exp\left(-\sum_{j=1}^{n} \max_{i=1,\ldots,m} \left(\frac{a_j(t_i)}{x_i}\right)^\alpha\right), \qquad t_i \in K,\ x_i > 0.$$

In particular, X has marginal distributions of α-Fréchet type. Further, all the finite-dimensional marginal distributions of a max-stable process can be approximated arbitrarily well by a max-linear model using a sufficiently large number n of spectral functions a_j (cf. Wang and Stoev, 2011, among others).

For such max-linear models, Wang and Stoev (2011) calculated the conditional distribution of $\boldsymbol{Z}$ conditional on a finite number of observations $X(t_1) = x_1, \ldots, X(t_k) = x_k$ by considering so-called hitting scenarios describing which components of Z hit their upper bound given by the conditions. However, for a general condition of the type $\ell(X) = x$, such upper bounds for Z may not necessarily be hit. Consider for instance the process $X(t) = \max\{Z_1, tZ_2\}$, $t \in [0,1]$. The condition $\int_0^1 X(t)\,\mathrm{d}t = x$ implies that $Z_1 \le x$ and $Z_2 \le 2x$. Although Z_1 or Z_2 may be arbitrarily close to these bounds, they are hit with probability zero. Thus, an alternative approach has to be designed for sampling from the distribution of Z conditional on $\ell(X) = x$. First, we observe that this distribution has all its mass in the set

$$M_{+,a} = \{\boldsymbol{z} \in (0,\infty)^n : \ell(\boldsymbol{a} \odot \boldsymbol{z}) > 0\}.$$

Due to the homogeneity of ℓ, for each $(1, \boldsymbol{y}_{-1}) \in M_{+,\boldsymbol{a}}$, there is one and only one vector $\boldsymbol{z} \in (0,\infty)^n$ with $\boldsymbol{z}_{-1}/z_1 = \boldsymbol{y}_{-1}$ satisfying $\ell(\boldsymbol{a} \odot \boldsymbol{z}) = x$, namely

$$\boldsymbol{z} = \frac{x}{\ell(\boldsymbol{a} \odot (1, \boldsymbol{y}_{-1}))} \begin{pmatrix} 1 \\ \boldsymbol{y}_{-1} \end{pmatrix}.$$

Thus, the mapping $\boldsymbol{z} \mapsto (\ell(\boldsymbol{a} \odot \boldsymbol{z}), \boldsymbol{z}_{-1}/z_1)$ maps the set $M_{+,\boldsymbol{a}}$ bijectively onto the set $(0,\infty) \times M_{+,\boldsymbol{a}}^{(-1)}$ where

$$M_{+,\boldsymbol{a}}^{(-1)} = \left\{\boldsymbol{y}_{-1} \in (0,\infty)^{n-1} :\ \ell\left(\boldsymbol{a} \odot \begin{pmatrix} 1 \\ \boldsymbol{y}_{-1} \end{pmatrix}\right) > 0\right\}.$$

Thus, instead of sampling from the vector $\boldsymbol{Z}$ conditional on $\ell(X) = x$ directly, we can focus on the equivalent problem of sampling from the distribution of $\boldsymbol{Y}_{-1} = (Z_2/Z_1, \ldots, Z_n/Z_1)$ conditional on $\ell(X) = x$. In contrast to the conditional distribution of Z given $\ell(X) = x$ which is supported on some Lebesgue null subset of $(0,\infty)^n$, the conditional distribution of $\boldsymbol{Y}_{-1}$ is supported on $M_{+,\boldsymbol{a}}^{(-1)} \subset (0,\infty)^{n-1}$ and possesses a Lebesgue density $f_{\boldsymbol{Y}_{-1}|\ell(X)=x}(\cdot)$. Thus, the homogeneity of ℓ allows us to sample from a distribution on a lower-dimensional space. The sampling algorithm presented in the following subsection will make use of the fact that this Lebesgue density is proportional to the joint density of $(\ell(X), \boldsymbol{Y}_{-1})$, i.e. the product of the densities given in the following proposition.

Proposition 8.1. *The Lebesgue density of* $\boldsymbol{Y}_{-1}$ *is*

$$f_{\boldsymbol{Y}_{-1}}(\boldsymbol{y}_{-1}) = \frac{\alpha^{n-1}(n-1)!}{(\sum_{i=1}^n y_i^{-\alpha})^n} \prod_{i=1}^n y_i^{-(\alpha+1)}, \qquad \boldsymbol{y}_{-1} \in (0,\infty)^{n-1}. \tag{8.3}$$

Further, for $x > 0$*, the conditional density of* $\ell(X)$ *given* $\boldsymbol{Y}_{-1} = \boldsymbol{y}_{-1}$ *for* $\boldsymbol{y}_{-1} \in M_{+,\boldsymbol{a}}^{(-1)}$ *equals*

$$f_{\ell(X)|\boldsymbol{Y}_{-1}=\boldsymbol{y}_{-1}}(x) = \frac{\alpha x^{-\alpha n-1}}{(n-1)!} \left(\sum\nolimits_{i=1}^n \left(\frac{\ell(\boldsymbol{a} \odot \boldsymbol{y})}{y_i} \right)^\alpha \right)^n \exp\left(-\frac{1}{x^\alpha} \sum\nolimits_{i=1}^n \left(\frac{\ell(\boldsymbol{a} \odot \boldsymbol{y})}{y_i} \right)^\alpha \right), \tag{8.4}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ *with* $y_1 = 1$*.*

Proof. First, we use the fact that there is a diffeomorphic one-to-one mapping between the random vectors $\boldsymbol{Z}$ and $(Z_1, \boldsymbol{Y}_{-1})$ given by $(z_1, \ldots, z_n)^\top \mapsto (z_1, z_2/z_1, \ldots, z_n/z_1)^\top$. Thus, the joint density $f_{(Z_1,\boldsymbol{Y}_{-1})}(\cdot,\cdot)$ of Z_1 and $\boldsymbol{Y}_{-1}$ is obtained by the density transformation formula

$$f_{(Z_1,\boldsymbol{Y}_{-1})}(z, \boldsymbol{y}_{-1}) = z^{n-1} \prod_{i=1}^n \Phi_\alpha'(z y_i) = \alpha^n z^{-\alpha n-1} \exp\left(-\frac{1}{z^\alpha} \sum\nolimits_{i=1}^n y_i^{-\alpha} \right) \prod_{i=1}^n y_i^{-(\alpha+1)}, \tag{8.5}$$

where $z > 0$, $y_1 = 1$, $\boldsymbol{y}_{-1} = (y_2, \ldots, y_n)^\top \in (0,\infty)^{n-1}$ and Φ_α' denotes the α-Fréchet density function. Integrating (8.5) with respect to the first component yields

$$\begin{aligned} f_{\boldsymbol{Y}_{-1}}(\boldsymbol{y}_{-1}) &= \alpha^n \int_0^\infty z^{-\alpha n-1} \exp\left(-\frac{1}{z^\alpha} \sum_{i=1}^n y_i^{-\alpha} \right) \prod_{i=1}^n y_i^{-(\alpha+1)} \mathrm{d}z \\ &= \alpha^{n-1} \int_0^\infty u^{n-1} \exp\left(-u \sum\nolimits_{i=1}^n y_i^{-\alpha} \right) \prod_{i=1}^n y_i^{-(\alpha+1)} \mathrm{d}u \\ &= \frac{\alpha^{n-1}(n-1)!}{(\sum_{i=1}^n y_i^{-\alpha})^n} \prod_{i=1}^n y_i^{-(\alpha+1)}, \end{aligned}$$

which is (8.3). Thus, by (8.3) and (8.5), the conditional density $f_{Z_1|\boldsymbol{Y}_{-1}=\boldsymbol{y}_{-1}}(\cdot)$ is

$$f_{Z_1|\boldsymbol{Y}_{-1}=\boldsymbol{y}_{-1}}(z) = \frac{f_{(Z_1,\boldsymbol{Y}_{-1})}(z, \boldsymbol{y}_{-1})}{f_{\boldsymbol{Y}_{-1}}(\boldsymbol{y}_{-1})} = \frac{\alpha z^{-n\alpha-1}}{(n-1)!} \left(\sum_{i=1}^n y_i^{-\alpha} \right)^n \exp\left(-\frac{1}{z^\alpha} \sum_{i=1}^n y_i^{-\alpha} \right).$$

Conditional on $\boldsymbol{Y}_{-1} = \boldsymbol{y}_{-1}$ for some $\boldsymbol{y}_{-1} \in M_{+,\boldsymbol{a}}^{(-1)}$, the homogeneity of ℓ implies the linear relationship $\ell(\boldsymbol{a} \odot \boldsymbol{Z}) = \ell(\boldsymbol{a} \odot \boldsymbol{y}) Z_1$. Thus, by standard density transformation, $f_{\ell(X)|\boldsymbol{Y}_{-1}=\boldsymbol{y}_{-1}}(x) = \ell(\boldsymbol{a} \odot \boldsymbol{y})^{-1} f_{Z_1|\boldsymbol{Y}_{-1}=\boldsymbol{y}_{-1}}\left(\ell(\boldsymbol{a} \odot \boldsymbol{y})^{-1} x\right)$, which yields (8.4). □

8.2.2 The Algorithm

Due to the fact that the distribution of $\ell(X)$ is unknown, the target distribution $f_{\boldsymbol{Y}_{-1}|\ell(X)=x}$ is only known up to a constant. Thus, we will use a Metropolis-Hastings algorithm with independent samplers (cf. Tierney, 1994, for example) for sampling from the conditional distribution of $\boldsymbol{Y}_{-1}$ given $\ell(\boldsymbol{a} \odot \boldsymbol{Z}) = x$. Here, we will propose the transition from a current

state $\boldsymbol{y}_{-1}$ to a new state $\tilde{\boldsymbol{y}}_{-1}$ according to the unconditional density $f_{\boldsymbol{Y}_{-1}}(\tilde{\boldsymbol{y}}_{-1})$ which does not depend on the current state. The proposal will be accepted with probability

$$\begin{aligned} p(y, \tilde{y}) &= \min\left\{1, \frac{f_{\boldsymbol{Y}_{-1}|\ell(X)=x}(\tilde{\boldsymbol{y}}_{-1})/f_{\boldsymbol{Y}_{-1}}(\tilde{\boldsymbol{y}}_{-1})}{f_{\boldsymbol{Y}_{-1}|\ell(X)=x}(\boldsymbol{y}_{-1})/f_{\boldsymbol{Y}_{-1}}(\boldsymbol{y}_{-1})}\right\} = \min\left\{1, \frac{f_{\ell(X)|\boldsymbol{Y}_{-1}=\tilde{\boldsymbol{y}}_{-1}}(x)}{f_{\ell(X)|\boldsymbol{Y}_{-1}=\boldsymbol{y}_{-1}}(x)}\right\}, \\ &= \min\left\{1, \left(\frac{\ell(\boldsymbol{a}\odot\tilde{\boldsymbol{y}})^\alpha \sum_{i=1}^n \tilde{\boldsymbol{y}}_i^{-\alpha}}{\ell(\boldsymbol{a}\odot\boldsymbol{y})^\alpha \sum_{i=1}^n \boldsymbol{y}_i^{-\alpha}}\right)^n \frac{\exp\left(-x^{-\alpha}\sum_{i=1}^n(\tilde{\boldsymbol{y}}_i/\ell(\boldsymbol{a}\odot\tilde{\boldsymbol{y}}))^{-\alpha}\right)}{\exp\left(-x^{-\alpha}\sum_{i=1}^n(\boldsymbol{y}_i/\ell(\boldsymbol{a}\odot\boldsymbol{y}))^{-\alpha}\right)}\right\} \end{aligned} \tag{8.6}$$

by Proposition 8.1. Then, by construction, $f_{\boldsymbol{Y}_{-1}|\ell(X)=x}(\cdot)$ is the density of the stationary distribution of the resulting Markov chain. Further, for any state $\boldsymbol{y}_{-1} \in M_{+,\boldsymbol{a}}^{(-1)}$, the probability of proposing a new state that is not accepted is positive and, consequently, the Markov chain is aperiodic. As the proposal density allows for a direct transition from any state $\boldsymbol{y}_{-1} \in M_{+,\boldsymbol{a}}^{(-1)}$ to any other state $\tilde{\boldsymbol{y}}_{-1} \in M_{+,\boldsymbol{a}}^{(-1)}$, the chain is also irreducible. This implies that the distribution of the Markov chain converges to the desired conditional distribution in total variation norm. Using that

$$\begin{aligned} &\inf_{\boldsymbol{y}_{-1}\in M_{+,\boldsymbol{a}}^{(-1)}} \frac{f_{\boldsymbol{Y}_{-1}}(\boldsymbol{y}_{-1})}{f_{\boldsymbol{Y}_{-1}|\ell(X)=x}(\boldsymbol{y}_{-1})} = \inf_{\boldsymbol{y}_{-1}\in M_{+,\boldsymbol{a}}^{(-1)}} \kappa(x)\frac{1}{f_{\ell(X)|\boldsymbol{Y}_{-1}=\tilde{\boldsymbol{y}}_{-1}}(x)} \\ &= \inf_{\boldsymbol{y}_{-1}\in M_{+,\boldsymbol{a}}^{(-1)}} \kappa(x)\cdot\frac{(n-1)!}{\alpha}\cdot x\cdot\left(\frac{1}{x^\alpha}\sum_{i=1}^n\left(\frac{\ell(\boldsymbol{a}\odot\boldsymbol{y})}{y_i}\right)^\alpha\right)^{-n}\exp\left(\frac{1}{x^\alpha}\sum_{i=1}^n\left(\frac{\ell(\boldsymbol{a}\odot\boldsymbol{y})}{y_i}\right)^\alpha\right) \\ &\geq \kappa(x)\cdot\frac{(n-1)!}{\alpha}\cdot x\cdot\left(\frac{e}{n}\right)^n > 0 \end{aligned}$$

for some $\kappa(x) > 0$, we even get a geometric rate of convergence by Theorem 2.1 in Mengersen and Tweedie (1996). Hence, Algorithm 8.1 below generates a Markov chain $\{X^{(k)}(t), t \in K\}_{k=1,\dots,M}$ whose distribution converges to the desired conditional distribution of $\boldsymbol{Y}_{-1}$.

Algorithm 8.1: Conditional sampling from max-linear model with M iterations

Initial Step: Simulate independent random numbers $z_1, \dots, z_n \sim \Phi_\alpha$.
Set $\boldsymbol{y}^{(0)} = \boldsymbol{z}/z_1$. Repeat the simulation if necessary until $\ell(\boldsymbol{a}\odot\boldsymbol{y}^{(0)}) > 0$.
Update Step: **for** $k = 1, \dots, M$ **do**
Sample independent random numbers $\tilde{z}_1, \dots, \tilde{z}_n \sim \Phi_\alpha$.
Set $\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{z}}/\tilde{z}_1$ and

$$\boldsymbol{y}^{(k)} = \begin{cases} \tilde{\boldsymbol{y}} & \text{with probability } p(\boldsymbol{y}^{(k-1)}, \tilde{\boldsymbol{y}}) \\ \boldsymbol{y}^{(k-1)} & \text{else} \end{cases}$$

where $p(\cdot,\cdot)$ is defined as in Equation (8.6).
return $\{X^{(k)}(\cdot)\}_{k=1,\dots,M}$ *where* $X^{(k)}(\cdot) = (x\boldsymbol{a}\odot\boldsymbol{y}^{(k)})/\ell(\boldsymbol{a}\odot\boldsymbol{y}^{(k)})$

Remark 8.2. For specific choices of ℓ and $\boldsymbol{a}$, it may happen that $\ell(X) = 0$ with some positive probability. In this case, repeating the simulation in the first step of Algorithm 8.1 may be necessary to ensure that the initial state $\boldsymbol{y}$ is in the support of the density $f_{\boldsymbol{Y}|\ell(X)=x}$ and, consequently, the acceptance probability $p(\boldsymbol{y}, \tilde{\boldsymbol{y}})$ is almost surely well-defined in each iteration step.

8.2.3 Simulation Study

We aim to analyze the performance of Algorithm 8.1 empirically in a simulation study. To this end, we compare the results of the algorithm with the exact conditional distribution which can be calculated analytically or at least be simulated in some special cases. Here, we consider the distribution of a max-linear model at m locations, i.e.

$$X_i = \max_{j=1,\dots,n} a_{ij} Z_j, \qquad i = 1, \dots, m,$$

conditional on its maximum

$$\check{X} = \max_{i=1,\dots,m} X_i = \max_{j=1,\dots,n} b_j Z_j,$$

where $b_j = \max_{i=1,\dots,m} a_{ij}$. Then, there exists a well-defined random variable J with values in $\{1, \dots, n\}$ such that $\check{X} = b_J Z_J$. The conditional distribution of this random variable given $\check{X} = x$ is

$$\mathbb{P}(J = j \mid \check{X} = x) = \frac{b_j^\alpha}{\sum_{k=1}^n b_k^\alpha}, \qquad j = 1, \dots, n,$$

and, thus, does not depend on the value of $\check{X}$ (cf. Resnick and Roy, 1990). Conditional on $J = j$ and $\check{X} = x$, we have $Z_j = x/b_j$. The other $Z_{j'}$, $j' \neq j$, follow the conditional law of independent Φ_α-variables given $Z_{j'} < x/b_{j'}$. Consequently, after sampling from the conditional distribution of the random index J, we can easily simulate from the exact conditional distribution of the vector $(Z_j)_{j=1,\dots,n}$, and, thus, of the vector $(X_i)_{i=1,\dots,m}$, given $\check{X} = x$.

In our simulation study, we choose coefficients a_{ij} which are independently drawn from a uniform distribution on $[0, 1]$ for $m = 10$ and $n = 100$. Given $\boldsymbol{a} = (a_{ij})_{i=1,\dots,m, j=1,\dots,n}$, a realization $(x_i)_{i=1,\dots,m}$ of $X = \boldsymbol{a} \odot \boldsymbol{Z}$ is simulated with independent unit Fréchet distributed random numbers $Z_1, \dots, Z_n$. Then, a sample of size 5 000 is drawn from the conditional distribution of $\min_{i=1,\dots,m} X_i$ given $\check{X} = \max_{i=1,\dots,m} x_i$ via Algorithm 8.1. Here, as a more detailed analysis of convergence and mixing properties of the algorithm suggests (cf. Subsection 8.6), we use a burn-in period of 1000 steps and thin the chain by using every 250th step only. Thus, we consider the states of the Markov chain after 1 000, 1 250, 1 500, ..., 1 250 750 iterations as an independent sample of size 5 000 from the conditional distribution.

In order to compare the output of Algorithm 8.1 to the exact conditional distribution, we take an independent sample of size 5000 from the exact conditional distribution and we consider the p-value of the two-sample Kolmogorov-Smirnov test. This procedure is repeated 500 times yielding different coefficient functions $\boldsymbol{a}$ and different values of $\max_{i=1,\dots,m} x_i$. Under the null hypothesis that Algorithm 8.1 produces independent samples from the desired stationary distribution, i.e. the samples from the algorithm and the exact conditional distribution follow the same law, the p-value are uniformly distributed on the interval $[0, 1]$, i.e. the histogram of the p-values shown in Figure 8.1 is flat. The fact that the histogram fully lies in the region of acceptance of the hypothesis test with a significance level of 5 % (indicated by the dashed red lines) obtained by a Monte Carlo experiment, confirms the performance of the algorithm.

8.3 Extension to Conditionally Max-Linear Models

In this section, we deal with a generalization of Algorithm 8.1 which can be applied to models of type (8.2) where the coefficient functions $a_j(\cdot)$, $j = 1, \dots, n$, are random.

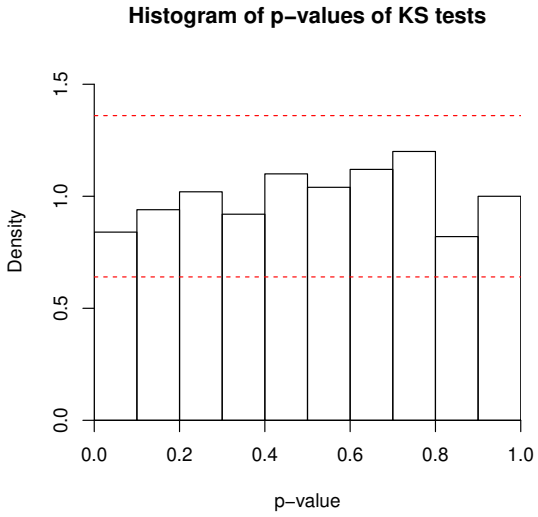

Figure 8.1: Histogram of 500 p-values of Kolmogorov-Smirnov tests comparing samples of size 5000 from the exact conditional distribution for a max-linear model of $n = 100$ coefficients with the output of Algorithm 8.1.

More precisely, let $Z_1, \dots, Z_n$ be independent α-Fréchet distributed random variables and $\{A_j(t),\, t \in K\}$, $j = 1, \dots, n$, be independent sample-continuous processes. Then, the process

$$X(t) = \max_{j=1,\dots,n} A_j(t) Z_j, \quad t \in K, \tag{8.7}$$

is max-linear conditionally on $A_1(\cdot) = a_1(\cdot), \dots, A_n(\cdot) = a_n(\cdot)$. However, in general, it is not max-stable and does not have marginal distributions of α-Fréchet type any more. Instead, the finite-dimensional distributions are given by

$$\mathbb{P}(X(t_i) \le x_i,\, 1 \le i \le m) = \mathbb{E}\left(\exp\left(-\sum\nolimits_{j=1}^{n} \max_{i=1,\dots,m} \left(\frac{A_j(t_i)}{x_i}\right)^{\alpha}\right)\right), \tag{8.8}$$

for all $t_1, \dots, t_m \in K$, $x_1, \dots, x_m > 0$, $m \ge 1$. Although not being max-stable themselves, conditionally max-linear models (8.7) allow for a straightforward approximation of general max-stable processes as we will see in Section 8.4.

A first approach to perform conditional sampling from model (8.7) might be to reduce the problem to the problem of conditional sampling from a max-linear model by drawing the vector of coefficient functions $\boldsymbol{A} = (A_1(\cdot), \dots, A_n(\cdot))^\top$ in a first step before applying Algorithm 8.1. However, in this case, $\boldsymbol{A}$ has to be drawn from the conditional law of A given $\ell(X) = x$. For this law, we obtain

$$\mathbb{P}(\boldsymbol{A} \in \mathrm{d}\boldsymbol{a} \mid \ell(X) = x) \propto f_{\ell(\boldsymbol{a} \odot \boldsymbol{Z})}(x) \mathbb{P}(\boldsymbol{A} \in \mathrm{d}\boldsymbol{a}) = \mathbb{E}\left(f_{\ell(\boldsymbol{a} \odot \boldsymbol{Z}) \mid Y_{-1}}(x)\right) \mathbb{P}(\boldsymbol{A} \in \mathrm{d}\boldsymbol{a}). \tag{8.9}$$

In general, the expectation in (8.9) cannot be calculated explicitly.

As an alternative to such a two-step procedure, we could also directly sample from the joint conditional distribution of $\boldsymbol{A}$ and $\boldsymbol{Z}$ given $\ell(\boldsymbol{A} \odot \boldsymbol{Z}) = x$. Analogously to Section 8.2, we note this distribution has all its mass in the set

$$\bigcup_{\boldsymbol{a} \in \mathrm{supp}(\mathbb{P}(\boldsymbol{A} \in \cdot))} \{\boldsymbol{a}\} \times M_{+,\boldsymbol{a}}.$$

Again, for every fixed vector of coefficient functions $\boldsymbol{A} = \boldsymbol{a}$, the vector Z is uniquely determined by $\boldsymbol{Y} = \boldsymbol{Z}/Z_1$. Consequently, we will sample from the joint conditional distribution of $\boldsymbol{A}$ and $\boldsymbol{Y}_{-1}$. Up to the normalizing constant $\mathbb{P}(\ell(\boldsymbol{A} \odot \boldsymbol{Z}) \in \mathrm{d}x)$, this distribution can be calculated by the means of Proposition 8.1 yielding

$$\begin{aligned}&\mathbb{P}(\boldsymbol{Y}_{-1} \in \mathrm{d}\boldsymbol{y}_{-1},\, \boldsymbol{A} \in \mathrm{d}\boldsymbol{a} \mid \ell(\boldsymbol{A} \odot \boldsymbol{Z}) = x) \propto \mathbb{P}(\boldsymbol{Y}_{-1} \in \mathrm{d}\boldsymbol{y}_{-1},\, \boldsymbol{A} \in \mathrm{d}\boldsymbol{a},\, \ell(\boldsymbol{A} \odot \boldsymbol{Z}) \in \mathrm{d}x)\\ &= f_{\ell(\boldsymbol{a}\odot\boldsymbol{Z})|\boldsymbol{Y}_{-1}=\boldsymbol{y}_{-1}}(x) f_{\boldsymbol{Y}_{-1}}(\boldsymbol{y}_{-1}) \mathbb{P}(\boldsymbol{A} \in \mathrm{d}\boldsymbol{a}) \mathrm{d}\boldsymbol{y}_{-1} \mathrm{d}x \qquad (8.10)\end{aligned}$$

and the first and the second factor are given by formulae (8.4) and (8.3), respectively. Analogously to Subsection 8.2.2, it can be shown that the distribution of the Markov chain generated in Algorithm 8.2 converges to the desired conditional distribution.

Algorithm 8.2: Conditional sampling from a conditionally max-linear model with M iterations

Initial Step: Simulate $\boldsymbol{a}^{(0)}$ according to the distribution of $\boldsymbol{A}$ and independent random numbers $z_1, \ldots, z_n \sim \Phi_\alpha$. Set $\boldsymbol{y}^{(0)} = \boldsymbol{z}/z_1$. Repeat the simulation, if necessary, until $\ell(\boldsymbol{a} \odot \boldsymbol{z}) > 0$.

Update Step: **for** $k = 1, \ldots, M$ **do**

Draw $\tilde{\boldsymbol{a}}$ from the distribution of $\boldsymbol{A}$ and, independently, $\tilde{z}_1, \ldots, \tilde{z}_n \sim \Phi_\alpha$. Set $\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{z}}/\tilde{z}_1$ and

$$(\boldsymbol{a}^{(k)}, \boldsymbol{y}^{(k)}) = \begin{cases} (\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{y}}) & \text{with probability } p((\boldsymbol{a}^{(k-1)}, \boldsymbol{y}^{(k-1)})), (\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{y}})) \\ (\boldsymbol{a}^{(k-1)}, \boldsymbol{y}^{(k-1)}) & \text{else} \end{cases}$$

where

$$p((\boldsymbol{a}, \boldsymbol{y}), (\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{y}})) = \min\left\{1, \frac{\left(\sum_{i=1}^n \left(\frac{\ell(\tilde{\boldsymbol{a}}\odot\tilde{\boldsymbol{y}})}{\tilde{y}_i}\right)^\alpha\right)^n \exp\left(-\frac{1}{x^\alpha}\sum_{i=1}^n \left(\frac{\ell(\tilde{\boldsymbol{a}}\odot\tilde{\boldsymbol{y}})}{\tilde{y}_i}\right)^\alpha\right)}{\left(\sum_{i=1}^n \left(\frac{\ell(\boldsymbol{a}\odot\boldsymbol{y})}{y_i}\right)^\alpha\right)^n \exp\left(-\frac{1}{x^\alpha}\sum_{i=1}^n \left(\frac{\ell(\boldsymbol{a}\odot\boldsymbol{y})}{y_i}\right)^\alpha\right)}\right\}$$

return $\{X^{(k)}(\cdot)\}_{k=1,\ldots,M}$ *where* $X^{(k)}(\cdot) = (x\boldsymbol{a}^{(k)} \odot \boldsymbol{y}^{(k)})/\ell(\boldsymbol{a}^{(k)} \odot \boldsymbol{y}^{(k)})$

8.4 General Max-Stable Processes

8.4.1 An Exact Algorithm for Max-Stable Processes

We now consider a not necessarily spectrally discrete max-stable process $\{X(t),\, t \in K\}$ with α-Fréchet marginals and sample paths in $C_+(K)$, i.e. a process X given by a spectral representation (8.1). One of the main difficulties of such a general max-stable process compared to a (conditionally) max-linear model is the fact that it is represented as a maximum over an infinite number of functions. However, the following two well-known observations allow us to restrict to a finite (but random) number of functions and, thus, form the basis for the algorithm:

First, let $\boldsymbol{E} = (E_i)_{i\geq 1}$ be a sequence of independent standard exponential random variables and $\Psi_k(\boldsymbol{E}) = \left(\sum_{i=1}^k E_i\right)^{-1/\alpha}$ for $k = 1, 2, \ldots$. Then, $\{\Psi_k(\boldsymbol{E}),\, k \geq 1\}$ are the points of a Poisson point process on $(0, \infty)$ with intensity measure $\alpha u^{-(\alpha+1)}\mathrm{d}u$. Moreover, by construction, the sequence $(\Psi_k(\boldsymbol{E}))_{k\geq 1}$ is monotonically decreasing. Thus, the points $\{U_i,\, i \geq 1\}$ from the construction (8.1) of X can be simulated in a descending order.

Second, for a given max-stable process X, the spectral representation (8.1) is not unique, but allows for several choices of the spectral process W. By de Haan and Ferreira (2006, Cor. 9.4.5), for any sample-continuous max-stable process, there exists a corresponding spectral process W such that

$$\sup_{t\in K} W(t) \le C \tag{8.11}$$

almost surely for some $C > 0$. Moreover, any spectral representation can be transformed such that equality in (8.11) holds almost surely for $C = (\mathbb{E}\sup_{t\in K} W(t)^\alpha)^{1/\alpha}$ (Oesting et al., 2018b). We have that

$$X(t) = \max_{k\in\mathbb{N}} \Psi_k(\mathbf{E}) W_k(t),\ t \in K,$$

and by (8.11), those $(\Psi_k(\mathbf{E}), W_k)$ that satisfy $C\Psi_k(\mathbf{E}) \le \inf_{t\in K} X(t)$ do not contribute to the above maximum X. Thus, a finite (but possibly random) number of points need to be taken into account when simulating $X(t)$ over the set K. More precisely, for any positive sequence $\boldsymbol{E} = \{E_k\}_{k\ge1}$ and any collection $\boldsymbol{W} = \{W_k\}_{k\ge1}$ of continuous functions bounded by C, we consider the operation $\boldsymbol{W} \odot \Psi(\boldsymbol{E})$ which – in analogy to the max-linear case with finite vectors – is defined by

$$(\boldsymbol{W} \odot \Psi(\boldsymbol{E}))(t) = \max_{k\ge1} W_k(t)\Psi_k(\boldsymbol{E}), \quad t \in K.$$

By definition, $\boldsymbol{W} \odot \Psi(\boldsymbol{E})$ depends only on a potentially random, but finite number of components, namely the first $N(\boldsymbol{W}, \boldsymbol{E})$ elements of $\boldsymbol{W}$ and $\boldsymbol{E}$ where

$$N(\boldsymbol{W}, \boldsymbol{E}) = \min\left\{n \ge 1 :\ \Psi_n(\boldsymbol{E})C \le \inf_{t\in K}(\boldsymbol{W} \odot \Psi(\boldsymbol{E}))(t)\right\}. \tag{8.12}$$

Assuming from now on that C and $\boldsymbol{W}$ are chosen such that (8.11) is satisfied, for E_k, $k \ge 1$, being independently standard exponentially distributed and W_k, $k \ge 1$, being independent processes with the same distribution as $\boldsymbol{W}$, the definitions above imply that

$$\{X(t)\}_{t\in K} =_d \{(\boldsymbol{W} \odot \Psi(\boldsymbol{E}))(t)\}_{t\in K} = \left\{\max_{i=1,\dots,N(\boldsymbol{W},\boldsymbol{E})} \Psi_i(\boldsymbol{E})W_i(t)\right\}_{t\in K}.$$

As $N(\boldsymbol{W}, \boldsymbol{E})$ is a stopping time, i.e. the event $N(\boldsymbol{E}, \boldsymbol{W}) \le n$ depends only on $E_1, \dots, E_n$ and $W_1, \dots, W_n$, this finite representation can be used for an exact simulation of X (cf. Oesting et al., 2018b). Further, it allows the conditional distribution of X given $\ell(X) = x$ to be simulated as the limit distribution of a Markov chain similarly to the case of a (conditionally) max-linear model. Here, the Poisson points $\Psi_k(\boldsymbol{E})$, $k \ge 1$, take the role of the random variables Z_k in the max-linear model, while the spectral functions W_k correspond to the random coefficient functions A_k. Analogously to the max-linear case, the conditional distribution of $\boldsymbol{E} \mid \boldsymbol{W} = \boldsymbol{w}$, $N(\boldsymbol{W}, \boldsymbol{E}) = n$, $\ell(\boldsymbol{W} \odot \Psi(\boldsymbol{E})) = x$ is supported in the set

$$M_{+,\boldsymbol{w},n} = \{\boldsymbol{e} \in (0,\infty)^n :\ N(\boldsymbol{w}, \boldsymbol{e}) = n,\ \ell(\boldsymbol{w} \odot \Psi(\boldsymbol{e})) > 0\}.$$

Moreover, for each $(1, \boldsymbol{d}_{-1}) \in M_{+,\boldsymbol{w},n}$, there is exactly one vector $\boldsymbol{e} \in (0,\infty)^n$ such that $\boldsymbol{e}/e_1 = \boldsymbol{d}$ and $\ell(\boldsymbol{w} \odot \Psi(\boldsymbol{e})) = x$, namely $\boldsymbol{e} = (\ell(\boldsymbol{w} \odot \Psi(\boldsymbol{d}))/x)^\alpha(1, \boldsymbol{d}_{-1})$. Thus, the mapping $\boldsymbol{e} \mapsto (\ell(\boldsymbol{w} \odot \Psi(\boldsymbol{e})), e_2/e_1, \dots, e_n/e_1)$ maps $M_{+,\boldsymbol{w},n}$ bijectively onto $(0,\infty) \times M^{(-1)}_{+,\boldsymbol{w},n}$ where

$$M^{(-1)}_{+,\boldsymbol{w},n} = \left\{\boldsymbol{d}_{-1} \in (0,\infty)^{n-1} :\ \begin{pmatrix} 1 \\ \boldsymbol{d}_{-1} \end{pmatrix} \in M_{+,\boldsymbol{w},n}\right\}.$$

Hence, similarly to the case of a max-linear model, we will sample from the vector $\boldsymbol{D}_{-1} = (E_2/E_1, \dots, E_n/E_1)^\top$ given that $\boldsymbol{W} = \boldsymbol{w}$, $N(\boldsymbol{W}, \boldsymbol{E}) = n$ and $\ell(\boldsymbol{W} \odot \Psi(\boldsymbol{E})) = x$.

The corresponding density $f_{\boldsymbol{D}_{-1}|\boldsymbol{w},n,x}(\cdot)$ is proportional to the product of the conditional density of $\boldsymbol{D}_{-1}$ given $\boldsymbol{W}=\boldsymbol{w}$ and $N(\boldsymbol{E},\boldsymbol{W})=n$ and the conditional density of $\ell(\boldsymbol{W}\odot\Psi(\boldsymbol{E}))$ given $\boldsymbol{D}_{-1}=\boldsymbol{d}_{-1}$, $\boldsymbol{W}=\boldsymbol{w}$ and $N(\boldsymbol{E},\boldsymbol{W})=n$.
These densities are given in the following proposition which can be proved analogously to the proof of Proposition 8.1.

Proposition 8.3. *For $n\geq 1$ and $\boldsymbol{w}\in(C_+(K))^n$, the conditional density of $\boldsymbol{D}_{-1}$ given $\boldsymbol{W}=\boldsymbol{w}$ and $N(\boldsymbol{E},\boldsymbol{W})=n$ is*

$$f_{\boldsymbol{D}_{-1}|\boldsymbol{w},n}(\boldsymbol{d}_{-1})=\frac{(n-1)!}{(\sum_{i=1}^n d_i)^n}\frac{\mathbf{1}_{\{N(\boldsymbol{d},\boldsymbol{w})=n\}}}{\mathbb{P}(N(\boldsymbol{D},\boldsymbol{w})=n)},\quad \boldsymbol{d}_{-1}\in(0,\infty)^{n-1}, \tag{8.13}$$

where $d_1=1$ and $\boldsymbol{d}_{-1}=(d_2,\dots,d_n)^\top$. The conditional density of $\ell(X)$ given $\boldsymbol{D}_{-1}=\boldsymbol{d}_{-1}$, $\boldsymbol{W}=\boldsymbol{w}$ and $N(\boldsymbol{E},\boldsymbol{W})=n$ equals

$$\begin{aligned}&f_{\ell(X)|\boldsymbol{d}_{-1},\boldsymbol{w},n}(x) \\ &=\frac{\alpha/x}{(n-1)!}\left(\sum_{i=1}^n\left(\frac{\ell(\boldsymbol{w}\odot\Psi(\boldsymbol{d}))}{x}\right)^\alpha d_i\right)^n\exp\left(-\sum_{i=1}^n\left(\frac{\ell(\boldsymbol{w}\odot\Psi(\boldsymbol{d}))}{x}\right)^\alpha d_i\right),\quad x>0.\end{aligned} \tag{8.14}$$

Similarly to the first approach in the case of conditionally max-linear models, this result suggests to sample via the following two-step procedure which reduces the problem to a problem similar to the max-linear case:

1. Draw some realization of the number $N(\boldsymbol{W},\boldsymbol{E})$ and the vector $(W_k)_{k=1}^{N(\boldsymbol{W},\boldsymbol{E})}$ of stochastic processes conditional on $\ell(\boldsymbol{W}\odot\Psi(\boldsymbol{E}))=x$.

2. Conditional on $(W_k)_{k=1}^{N(\boldsymbol{W},\boldsymbol{E})}$, $N(\boldsymbol{W},\boldsymbol{E})$ and $\ell(\boldsymbol{W}\odot\Psi(\boldsymbol{E}))=x$, simulate the vector $\boldsymbol{E}=(E_1,\dots,E_{N(\boldsymbol{W},\boldsymbol{E})})^\top$.

However, the first step would require to sample from the distribution

$$\begin{aligned}&\mathbb{P}(\boldsymbol{W}\in\mathrm{d}\boldsymbol{w},\,N(\boldsymbol{W},\boldsymbol{E})=n\mid\ell(\boldsymbol{W}\odot\Psi(\boldsymbol{E}))=x)\\ &\propto f_{\ell(X)|w,n}(x)\mathbb{P}(\boldsymbol{W}\in\mathrm{d}\boldsymbol{w},\,N(\boldsymbol{W},\boldsymbol{E})=n)\\ &=\int_{M_{+,\boldsymbol{w},n}^{(-1)}}f_{\ell(X)|\boldsymbol{d}_{-1},\boldsymbol{w},n}(x)f_{\boldsymbol{D}_{-1}|\boldsymbol{w},n}(\boldsymbol{d}_{-1})\mathrm{d}\boldsymbol{d}_{-1}\mathbb{P}(\boldsymbol{W}\in\mathrm{d}\boldsymbol{w},\,N(\boldsymbol{W},\boldsymbol{E})=n)\\ &=\mathbb{E}\Bigg\{\frac{\alpha/x}{(n-1)!}\left[\left(\frac{\ell(\boldsymbol{w}\odot\Psi(\boldsymbol{D}))}{x}\right)^\alpha\sum_{i=1}^n D_i\right]^n\\ &\quad\cdot\exp\left[-\left(\frac{\ell(\boldsymbol{w}\odot\Psi(\boldsymbol{D}))}{x}\right)^\alpha\sum_{i=1}^n D_i\right]\,\Bigg|\,N(\boldsymbol{w},\boldsymbol{D})=n\Bigg\}\mathbb{P}(\boldsymbol{W}\in\mathrm{d}\boldsymbol{w},\,N(\boldsymbol{W},\boldsymbol{E})=n),\end{aligned} \tag{8.15}$$

where we used the results from Proposition 8.3. In general, the expectation in (8.15) cannot be calculated explicitly.
Thus, we directly sample from the joint conditional distribution of $\boldsymbol{E}$, $\boldsymbol{W}$ and $N(\boldsymbol{W},\boldsymbol{E})$ conditional on $\ell(\boldsymbol{W}\odot\Psi(\boldsymbol{E}))=x$, which is supported in the set

$$\bigcup_{n\in\mathbb{N}}\bigcup_{\boldsymbol{w}\in(C_+(K))^n}M_{+,\boldsymbol{w},n}\times\{\boldsymbol{w}\}\times\{n\}.$$

Sampling from $\boldsymbol{D}$ again instead of $\boldsymbol{E}$ requires the density

$$\begin{aligned}&\mathbb{P}(\boldsymbol{D}_{-1}\in\mathrm{d}\boldsymbol{d}_{-1},\,\boldsymbol{W}\in\mathrm{d}\boldsymbol{w},\,N(\boldsymbol{W},\boldsymbol{E})=n\mid\ell(\boldsymbol{W}\odot\Psi(\boldsymbol{E}))=x)\\ &\propto\mathbb{P}(\boldsymbol{D}_{-1}\in\mathrm{d}\boldsymbol{d}_{-1},\,\boldsymbol{W}\in\mathrm{d}\boldsymbol{w},\;N(\boldsymbol{W},\boldsymbol{E})=n,\,\ell(\boldsymbol{W}\odot\Psi(\boldsymbol{E}))\in\mathrm{d}x)\end{aligned}$$

$$= f_{\ell(X)|\boldsymbol{d}_{-1},\boldsymbol{w},n}(x)\mathbb{P}(\boldsymbol{D}_{-1} \in \mathrm{d}\boldsymbol{d}_{-1},\, \boldsymbol{W} \in \mathrm{d}\boldsymbol{w},\, N(\boldsymbol{W},\boldsymbol{E}) = n)\mathrm{d}x, \tag{8.16}$$

where the first factor is given by Equation (8.14).
Noting that we can sample from the unconditional distribution of $(\boldsymbol{D}_{-1}, \boldsymbol{W}, N(\boldsymbol{W},\boldsymbol{E}))$ by simulating the max-stable process X and extracting these quantities, we can use this as the proposal distribution. This leads to the following algorithm generating a Markov chain whose distribution converges to the conditional distribution of $X \mid \ell(X) = x$.

Algorithm 8.3: Conditional sampling from a max-stable process with M iterations

Initial Step: Simulate $(\boldsymbol{e}^{(0)}, \boldsymbol{w}^{(0)}, n^{(0)})$ according to the law of $(\boldsymbol{E}, \boldsymbol{W}, N(\boldsymbol{W},\boldsymbol{E}))$ by simulating X and set $\boldsymbol{d}^{(0)} = \boldsymbol{e}^{(0)}/e_1^{(0)}$.
Update Step: **for** $k = 1, \dots, M$ **do**
Simulate $(\tilde{\boldsymbol{e}}, \tilde{\boldsymbol{w}}, \tilde{n})$ according to the law of $(\boldsymbol{E}, \boldsymbol{W}, N(\boldsymbol{W},\boldsymbol{E}))$ by simulating X and set $\tilde{\boldsymbol{d}} = \tilde{\boldsymbol{e}}/\tilde{e}_1$. Update

$$(\boldsymbol{d}^{(k)}, \boldsymbol{w}^{(k)}, n^{(k)}) = \begin{cases} (\tilde{\boldsymbol{d}}, \tilde{\boldsymbol{w}}, \tilde{n}) & \text{with probability} \\ & p((\boldsymbol{d}^{(k-1)}, \boldsymbol{w}^{(k-1)}, n^{(k-1)}), (\tilde{\boldsymbol{d}}, \tilde{\boldsymbol{w}}, \tilde{n})) \\ (\boldsymbol{d}^{(k-1)}, \boldsymbol{w}^{(k-1)}, n^{(k-1)}) & \text{else} \end{cases}$$

where

$$p((\boldsymbol{d}, \boldsymbol{w}, n), (\tilde{\boldsymbol{d}}, \tilde{\boldsymbol{w}}, \tilde{n})) = \min\left\{ \frac{\frac{1}{(\tilde{n}-1)!}\left(\left(\frac{\ell(\tilde{\boldsymbol{w}}\odot\Psi(\tilde{\boldsymbol{d}}))}{x}\right)^{\alpha}\sum_{i=1}^{\tilde{n}}\tilde{d}_i\right)^{\tilde{n}}\exp\left(-\left(\frac{\ell(\tilde{\boldsymbol{w}}\odot\Psi(\tilde{\boldsymbol{d}}))}{x}\right)^{\alpha}\sum_{i=1}^{\tilde{n}}\tilde{d}_i\right)}{\frac{1}{(n-1)!}\left(\left(\frac{\ell(\boldsymbol{w}\odot\Psi(\boldsymbol{d}))}{x}\right)^{\alpha}\sum_{i=1}^{n}d_i\right)^{n}\exp\left(-\left(\frac{\ell(\boldsymbol{w}\odot\Psi(\boldsymbol{d}))}{x}\right)^{\alpha}\sum_{i=1}^{n}d_i\right)}, 1\right\}.$$

return $\{X^{(k)}(\cdot)\}_{k=1,\dots,M}$ *where* $X^{(k)}(\cdot) = \boldsymbol{w} \odot \Psi((\ell(\boldsymbol{w}\odot\Psi(\boldsymbol{d}))/x)^{\alpha}\boldsymbol{d})$

8.4.2 Simulation Study

In order to study the performance of Algorithm 8.3, we consider an extremal-t process (Opitz, 2013) given by

$$X(t) = \max_{i\geq 1} c_\nu \cdot U_i \cdot \max\{0, G_i(t)\}^\nu, \quad t \in K \subset \mathbb{R}^2, \tag{8.17}$$

where $\{U_i,\ i \geq 1\}$ are the points of a Poisson point process on $(0,\infty)$ with intensity $u^{-2}\mathrm{d}u$, $\nu > 0$, $c_\nu = \sqrt{\pi}2^{-(\nu-2)/2}\Gamma\left(\frac{\nu+1}{2}\right)^{-1}$ and G_i, $i \geq 1$, are independent copies of a centered Gaussian process G with unit variance and correlation function ρ. Conditioning on a max-linear functional such as $\ell(X) = \sup_{t\in K} X(t)$, exact conditional simulation can be performed by the algorithm introduced in Oesting (2015) and the results can be compared to those of Algorithm 8.3.
To apply Algorithm 8.3, we need to find a constant C such that the corresponding spectral processes W_i are bounded. Such a constant does not exist for the spectral processes

$$W_i(t) = c_\nu \cdot \max\{0, G_i(t)\}^\nu, \quad t \in K$$

given in (8.17). Applying an appropriate transform of the probability measure, an equivalent representation of the same max-stable process with bounded spectral processes can be found (cf. Oesting et al., 2018b). However, the resulting transformed spectral processes are difficult to simulate. Thus, for simplicity, we will assume that the original spectral processes are bounded by a constant C of the form $C = c_\nu \cdot C_G^\nu$ where C_G is an appropriate bound for the standard Gaussian process G, e.g. $C_G = 4$.
In our simulation study, we consider an extremal-t process X with unit Fréchet margins, parameter $\nu = 2$ and correlation function $\rho(h) = \exp(-\|h\|/1.5)$ on the set $K = \{0, 0.2, \dots, 1\} \times \{0, 0.2, \dots, 1\}$. A realization $\{x(t)\}_{t\in K}$ is simulated from the unconditional distribution of X. Then, both Algorithm 8.3 and the exact algorithm in Oesting (2015) are performed to obtain two samples of size 5 000 from the conditional distribution of $\inf_{t\in K} X(t)$ given $\sup_{t\in K} X(t) = \sup_{t\in K} x(t)$. For Algorithm 8.3, we choose $C_G = 4$, and, based on some further analysis (cf. Subsection 8.7), a burn-in period of length 1 000. We further thin the chain by using every 200th step only, i.e. we select the states of the Markov chain after $1\,000, 1\,200, \dots, 1\,000\,800$ iterations as a sample.
The procedure is repeated 250 times and the p-values two-sample Kolmogorov-Smirnov test is calculated for each pair of samples. The histogram of the obtained p-values is shown in Figure 8.2. It fully lies in the region of acceptance of the test of the hypothesis that both distributions coincide with a significance level of 5 % which indicates a good performance of the algorithm.

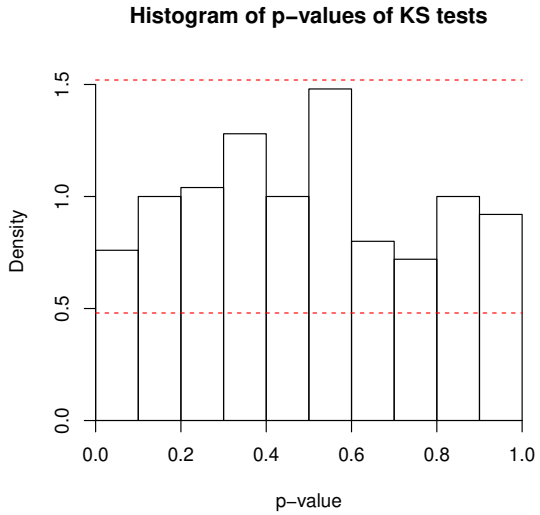

Figure 8.2: Histogram of 250 p-values of Kolmogorov-Smirnov tests comparing samples of size 5000 from the exact conditional distribution for an extremal-t process with the output of Algorithm 8.3.

8.5 Conclusion and Perspectives

We investigate the simulation of max-stable processes conditionally on the value of a single positively homogeneous functional of the process. An iterative algorithm based on Markov chain Monte Carlo techniques is used to sample from the conditional distribution. In simulation studies, we verify that simulations exhibit convergence to the right distribution

in a reasonable way. Further, the algorithm can be successfully applied to climate data for downscaling (see Oesting et al., 2018a).

Resorting to conditional simulation for downscaling has several advantages: both the spatial dependence structure and the extremal behaviour of the max-stable process are used to full capacity by the algorithm. Moreover, we can profit from standard downscaling techniques for post-processing the large scale conditions. Thus, the overall procedure takes benefit from various sources of information.

The methodology proposed in this work provides a mathematically rigorous way to sample from the conditional distribution if the condition is expressed as a single functional of the process. Nevertheless, in most applications, there are several conditions which need to be respected and performing conditional simulations for each condition separately is not satisfactory. Lifting this restriction remains a challenging problem from a mathematical point of view. In the application we choose the simplest way to deal with this restriction, that is, we separately sample in each region covered by a large scale grid cell. This approach is feasible because the range of precipitation is small compared to the cell size. In order to rectify the resulting discontinuities at the interfaces of different cells, smoothing techniques could be investigated.

Acknowledgements

This work has been financially supported by the ANR project McSim. The authors are grateful to two anonymous referees for their helpful comments and suggestions improving this article.

8.6 Diagnostics of Markov Chain in Algorithm 8.1

We further analyze convergence and mixing properties of the Markov chain in Algorithm 8.1. To this end, we consider three random matrices with $m = 10$ and $n = 100$ as in Subsection 8.2.3 and perform Algorithm 8.1 conditioning on $\check{X} = \max_{i=1,\dots,m} X_i$. More precisely, for the first, second and third matrix A, we condition on $\check{X}$ being equal to the 5 %-, 50 %- and 95 %-quantile of the unconditional distribution of $\check{X}$, respectively, which is given by $\mathbb{P}(\check{X} \leq x) = \exp(-\sum_{j=1}^{n}(b_j/x)^{\alpha})$ where $b_j = \max_{i=1,\dots,m} a_{ij}$. Then, each iteration provides a realization of the vector $(X_i)_{i=1}^m$. For each of the three Markov chains, Figure 8.5 shows the values of $\min_{i=1,\dots,m} X_i$ for the first 10 000 iteration steps of the algorithm. It can be seen that the behaviour of the chains appears stationary already after a small number of iterations. Besides this common effect, however, there are some obvious differences in the behaviour of the Markov chains. In the third case, where we condition on $\check{X}$ being large, we encounter much longer sequences of iterations with $\min_{i=1,\dots,m} X_i$ being constant than in the first case where $\check{X}$ is small. This difference becomes even more pronounced if we compare the third and the second Markov chain where $\check{X}$ is moderate.

To further investigate this phenomenon indicating different strengths of mixing, we calculate the empirical autocorrelation functions of the Markov chains based on the values for $\min_{i=1,\dots,m} X_i$ in iteration steps 5 000-500 000. The results displayed in Figure 8.3 confirm the observation made in Figure 8.5. While the autocorrelation drops below 0.05 after 70 and 30 steps for the first and second chain, respectively, there is much stronger dependence in the third chain where autocorrelation is higher than 0.05 for almost 200 steps.

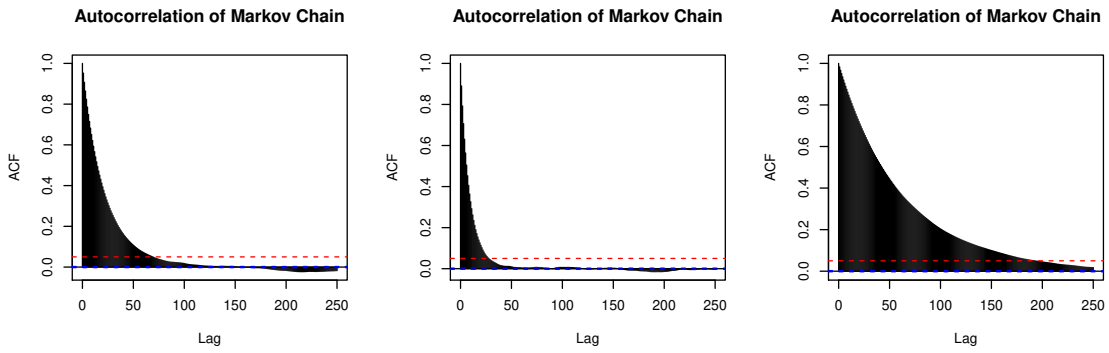

Figure 8.3: The empirical autocorrelation function for the values of $\min_{i=1,\dots,m} X_i$ provided by Algorithm 8.1 conditioning on $\check{X}$ being equal to the 5 %-, 50 %- and 95 %-quantile of its distribution (from left to right) for three random matrices A of size 10×100.

8.7 Diagnostics of Markov Chain in Algorithm 8.3

Similarly to Section 8.6, we analyze the Markov chains occurring in Algorithm 8.3. To this end, we consider the same extremal-t process model as in the simulation study in Subsection 8.4.2 and perform Algorithm 8.3 conditioning on $\check{X} = \sup_{t \in K} X(t)$. More precisely, we condition on $\check{X}$ being equal to the 5 %-, 50 %- and 95 %-quantile of the unconditional distribution of $\check{X}$, respectively. For each of the three scenarios, we run a Markov chain according to Algorithm 8.3 with $C_G = 4$.
For each of the three Markov chains, Figure 8.6 shows the values of $\min_{i=1,\dots,m} X_i$ for the first 10 000 iteration steps of the three chains. Further, the corresponding autocorrelation functions based on iteration steps 5 000-100 000 are displayed in Figure 8.4. The results are very similar to those obtained in Section 8.6.

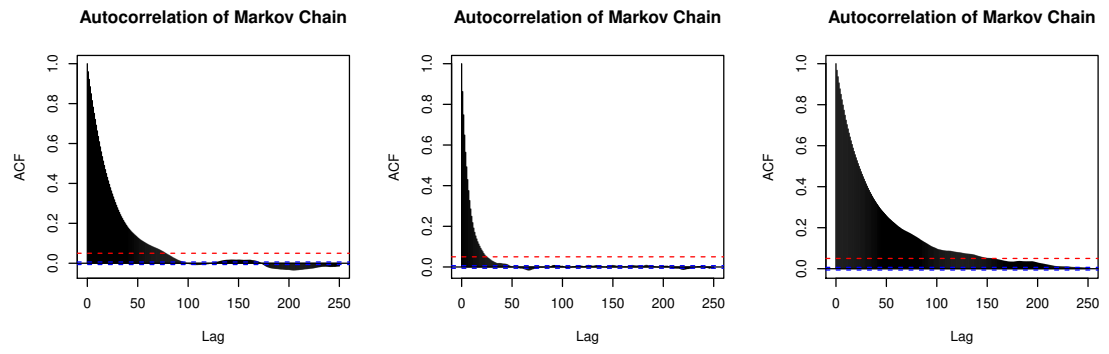

Figure 8.4: The empirical autocorrelation function for the values of $\inf_{t \in K} X(t)$ provided by Algorithm 8.3 conditioning on $\check{X}$ being equal to the 5 %-, 50 %- and 95 %-quantile of its distribution (from left to right) where X is an extremal-t process.

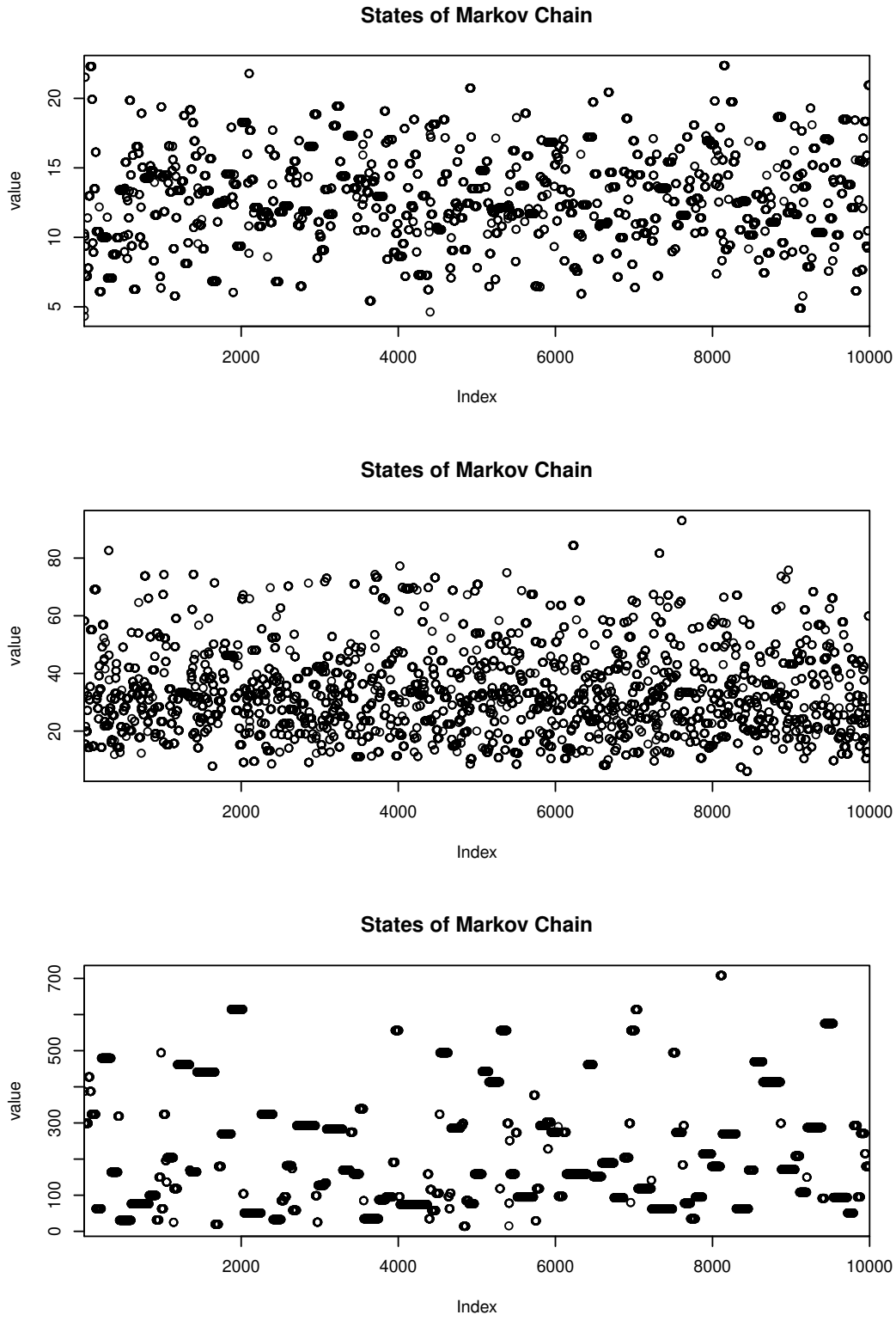

Figure 8.5: The values of $\min_{i=1}^{m} X_i$ in the first 10 000 iterations of Algorithm 8.1 conditioning on $\check{X}$ being equal to the 5 %-, 50 %- and 95 %-quantile of its distribution (from top to bottom) for three random matrices A of size 10×100.

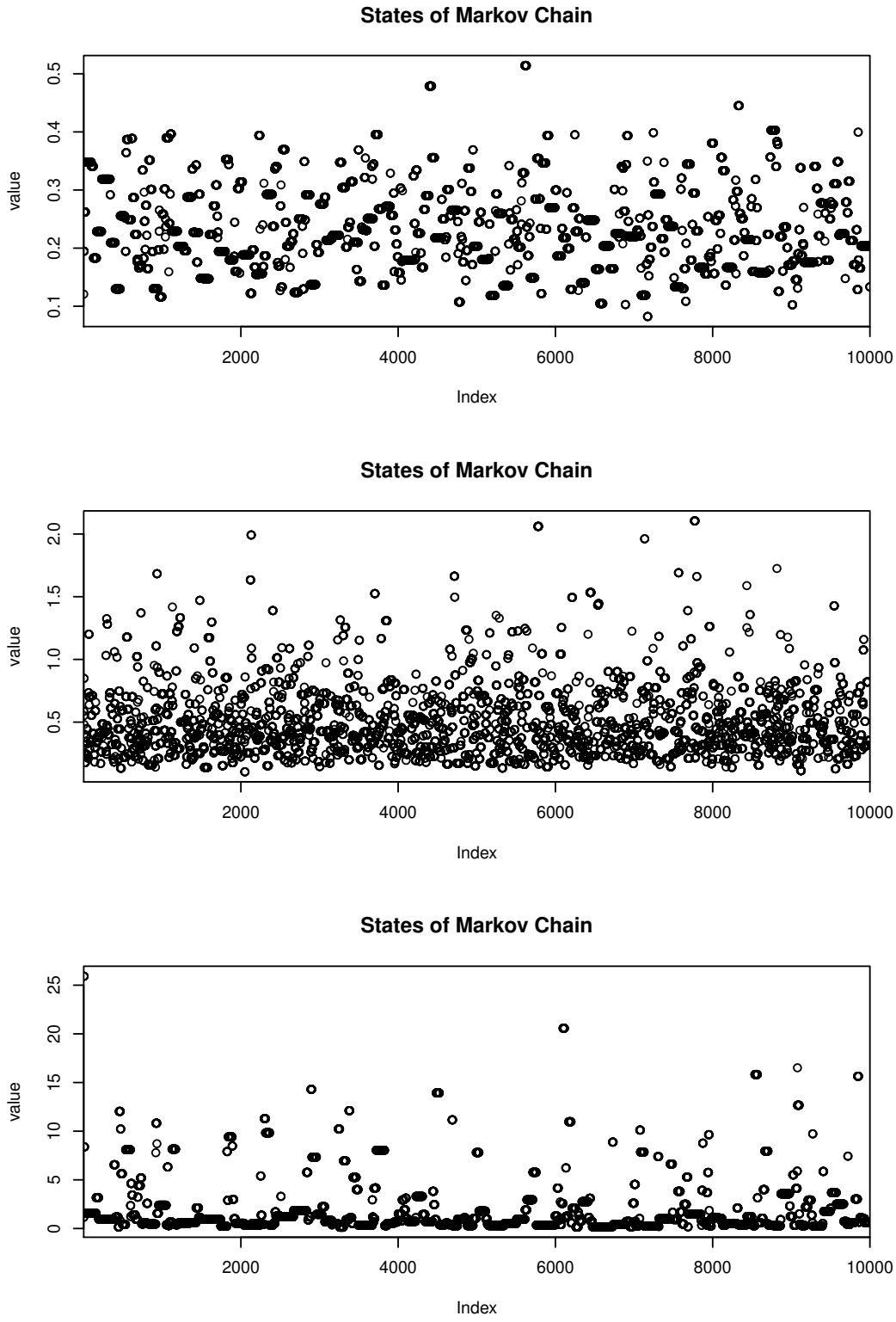

Figure 8.6: The values of $\inf_{t\in K} X(t)$ in the first 10 000 iterations of Algorithm 8.3 conditioning on $\check{X}$ being equal to the 5 %-, 50 %- and 95 %-quantile of its distribution (from top to bottom) where X is an extremal-t process.

9 Statistical Post-Processing of Forecasts for Extremes Using Bivariate Brown–Resnick Processes with an Application to Wind Gusts

joint work with Martin Schlather and Petra Friederichs

Up to some minor modifications, this chapter is a reprint of the article Oesting et al. (2017) that has been published in *Extremes*.

To improve the forecasts of weather extremes, in this chapter, we propose a joint spatial model for the observations and the forecasts, based on a bivariate Brown-Resnick process. As the class of stationary bivariate Brown-Resnick processes is fully characterized by the class of pseudo cross-variograms, we contribute to the theorical understanding of pseudo cross-variograms refining the knowledge of the asymptotic behaviour of all their components and introducing a parsimonious, but flexible parametric model. Both findings are of interest in classical geostatistics on their own. The proposed model is applied to real observation and forecast data for extreme wind gusts at 119 stations in Northern Germany.

9.1 Introduction

Spatial extremes may occur in various forms such us heavy rainfall, floods, heat waves or wind gusts. In view of their severe consequences, an adequate and precise forecast of these events is of great importance. However, the rareness of extreme events impedes any such task and, consequently, existing forecasts often lack accuracy. In meteorology, for example, forecasting extreme wind gusts, which are defined as peak wind speeds over a few seconds, is exacerbated by the short temporal and spatial ranges. Furthermore, numerical weather prediction (NWP) models provide estimates or diagnoses of wind gusts based on empirical knowledge only (cf. Brasseur, 2001). Although wind is a prognostic variable in NWP models, its values represents an average wind speed over a few minutes or longer depending on the grid spacing of the NWP model. Hence, post-processing procedures are needed that allow for an enhanced probabilistic forecast.

Occurring as limits of normalized pointwise maxima of stochastic processes, max-stable processes provide a suitable framework for the description of spatial extreme events, commonly used in environmental sciences (Coles, 1993; Coles and Tawn, 1996; Huser and Davison, 2014). Of particular interest is the subclass formed by Brown-Resnick processes which arise as limits of rescaled maxima of Gaussian processes (Brown and Resnick, 1977; Kabluchko et al., 2009; Kabluchko, 2011).
During the last years, max-stable processes have been frequently applied as models for spatial extremes in environmental sciences. For instance, Engelke et al. (2015) and Genton et al. (2015) recently used max-stable processes to model extreme wind speed observations.

The model we propose will go one step further, also taking into account the forecasts in two different aspects: First and in contrast to Engelke et al. (2015) and Genton et al. (2015), we consider the mean forecast to get a normalized version of the extreme observations. Second, besides the observable variable of interest itself, the corresponding forecast is included as second variable yielding a bivariate max-stable process. Here, we will focus on the class of bivariate Brown-Resnick processes (cf. Molchanov and Stucki, 2013; Genton et al., 2015) to exploit the statistical relation between observable data and the corresponding forecast. Modeling the behavior of observational data, a sample from the distribution of the observations conditional on the forecast is supposed to provide more realistic results than the original forecast and thus will appear as an appropriate probabilistic post-processed forecast.

The chapter is structured as follows: In Section 9.2, we present a univariate model for extreme observations, which may, in general, provide a first alternative to the original forecast. We introduce a model for the marginal distribution, i.e. the distribution of the observable variable of interest at a single location, motivating the normalization of its extremes by the mean forecast. The spatial dependence structure is incorporated into the model by the use of univariate Brown-Resnick processes. Section 9.3 is dedicated to the bivariate Brown-Resnick process which serves as a joint model for both the maximally observed and forecasted quantities. We deduce a necessary condition on the asymptotic behavior of the pseudo cross-variogram and provide a flexible cross-variogram model which leads to a stationary bivariate Brown-Resnick process. In Section 9.4, we describe how the model can be fitted to data. Based on this model, we propose a post-processing procedure which is presented in Section 9.5. Further, we provide tools to verify the procedure and the underlying models. Finally, the methods presented in Sections 9.4 and 9.5 are applied to real observation and forecast data for extreme wind gusts provided by the German's National Meteorological Service, Deutscher Wetterdienst (DWD) (Section 9.6).

9.2 Modeling by a Univariate Random Field

In this section, we present a spatial model for the observed pointwise maximum $V_{\max}^{\text{obs}}$ within a specific time period. To this end, we assume that, for each location and time period, the maximum $V_{\max}^{\text{obs}}$ is based on observations at N equidistant instants of times per period, that is, we have $V_{\max}^{\text{obs}} = \max_{t=1,\dots,N} V_t^{\text{obs}}$ for $V_1^{\text{obs}}, \dots, V_N^{\text{obs}} \sim F_\vartheta$ for some parameter ϑ. Here, the probability distributions F_ϑ are supposed to form a location-scale family with finite second moments, i.e. $\vartheta = (m, s) \in \mathbb{R} \times (0, \infty)$ with $F_{(m,s)}(x) = F_{(0,1)}\left(\frac{x-m}{s}\right)$, $x \in \mathbb{R}$, and $F_{(0,1)}$ is standardized to mean zero and unit variance. We assume $\vartheta = (m, s)$ to be temporally constant at each location within the same time period, but allow the values to vary among different locations and different time periods. The values of m and s will essentially be estimated from the bulk of the distribution, not the tail, and thus, they can often be extracted accurately from forecasts. Within the same time period and at the same location, the observable variables $V_1^{\text{obs}}, \dots, V_N^{\text{obs}}$ are assumed to be subsequent N elements of a stationary time series $(V_t^{\text{obs}})_{t\in\mathbb{Z}}$. Furthermore, we assume that the standardized distribution $F_{(0,1)}$ belongs to the max-domain of attraction of some univariate extreme value distribution G_ξ, $\xi \in \mathbb{R}$, i.e. there are sequences $(a_n)_{n\in\mathbb{N}}$, $a_n > 0$, and $(b_n)_{n\in\mathbb{N}}$, $b_n \in \mathbb{R}$, such that

$$F_{(0,1)}^n(a_n x + b_n) \overset{n\to\infty}{\longrightarrow} G_\xi(x), \quad 1 + \xi x > 0,$$

where

$$G_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}), & \xi \neq 0, \\ \exp(-\exp(-x)), & \xi = 0, \end{cases}$$

for $1 + \xi x > 0$. As the second moment of $F_{(0,1)}$ is assumed to be finite, we have $\xi < 0.5$. Under some conditions on the regularity and the dependence of the stationary sequence $V_1^{\text{obs}}, V_2^{\text{obs}}, \ldots$, we obtain that

$$\mathbb{P}\left(\frac{\max_{i=1,\ldots,n} V_i^{\text{obs}} - m - \widetilde{b}_n s}{\widetilde{a}_n s} \leq x \right) \overset{n\to\infty}{\longrightarrow} G_\xi(x), \quad 1 + \xi x > 0, \tag{9.1}$$

where $\widetilde{a}_n = a_n \theta^{-\xi}$ and $\widetilde{b}_n = b_n - \xi^{-1}(1 - \theta^{-\xi})$ for some $\theta \in (0, 1]$ called extremal index (cf. Coles, 2001; Leadbetter et al., 1983).
Let $m = m(l, p)$ and $s = s(l, p)$ be the mean of the variable at location l and period p and its standard deviation, respectively. Let

$$G_{\xi,\mu,\sigma}(x) = G_\xi\left(\frac{x-\mu}{\sigma}\right), \qquad 1 + \xi\frac{x-\mu}{\sigma} > 0$$

be the generalized extreme value distribution (GEV). Then, considering the maximum $V_{\max}^{\text{obs}} = V_{\max}^{\text{obs}}(l, p)$ for large N, we have approximately that

$$\frac{V_{\max}^{\text{obs}}(l,p) - m(l,p)}{s(l,p)} \sim G_{\xi^{\text{obs}},\mu^{\text{obs}}(l),\sigma^{\text{obs}}(l)}. \tag{9.2}$$

Here, the GEV parameters are assumed to be the same for every time period, which, in general, enables us to estimate the parameters for current and future time periods from past data. As common in many applications, the extreme value index ξ is also assumed to be constant in space. Under the ideal assumption that $V_i^{\text{obs}} \sim F_{(m,s)}$ and that m and s can be determined exactly, the GEV parameters ξ^{obs}, μ^{obs} and σ^{obs} are constant in space, as well. In practice, however, the observed variable of interest is subject to measurement errors whose distribution is spatially varying. Further, m and s often need to be extracted from forecasts with limited spatial resolution. To account for these difficulties, we allow $\mu^{\text{obs}}(l)$ and $\sigma^{\text{obs}}(l)$ to depend on the location l, while the extreme value index ξ^{obs} is assumed to be constant in space, as common in many applications. In contrast to μ^{obs} and σ^{obs}, $m(l, p)$ and $s(l, p)$ vary in space and time and may be interpreted as normalizing constants that will be the same for observation and forecasts. As $m(l, p)$ and $s(l, p)$ are defined as mean and standard deviation of the variable of interest, the parameters $\mu^{\text{obs}}(l)$ and $\sigma^{\text{obs}}(l)$ are uniquely determined. Marginal transformation yields that

$$X^{\text{obs}}(l,p) = \frac{1}{\xi^{\text{obs}}} \log\left(1 + \xi^{\text{obs}} \frac{V_{\max}^{\text{obs}}(l,p) - m(l,p) - s(l,p)\mu^{\text{obs}}(l)}{s(l,p)\sigma^{\text{obs}}(l)}\right) \tag{9.3}$$

is standard Gumbel distributed for every location l and time period p.

Perceiving the set of locations as a subset of $\mathbb{R}^2$ and the set of periods as a subset of $\mathbb{Z}$, the transformed observations can be regarded as realizations of a spatio-temporal random field $\{X^{\text{obs}}(l, p),\ l \in \mathbb{R}^2,\ p \in \mathbb{Z}\}$. While we assume that the spatial random fields $\{X^{\text{obs}}(l, p),\ l \in \mathbb{R}^2\}$, $p \in \mathbb{Z}$, are independent and identically distributed, we allow for a non-trivial spatial dependence structure. Here, we use the class of Brown-Resnick processes that can be defined for arbitrary dimensions D (Brown and Resnick, 1977; Kabluchko et al., 2009): Let $\Pi = \sum_{i\in\mathbb{N}} \delta_{U_i}$ be a Poisson point process on $\mathbb{R}$ with intensity $e^{-u}\,\mathrm{d}u$ and, independently of

Π, let W_i, $i \in \mathbb{N}$, be independent copies of a zero-mean Gaussian random field $\{W(s),\, s \in \mathbb{R}^D\}$ with stationary increments and semi-variogram $\gamma(\cdot)$ defined by

$$2\gamma(s) = \mathrm{Var}(W(s) - W(0)), \quad s \in \mathbb{R}^D.$$

Then, the random field Z defined by

$$Z(s) = \max_{i\in\mathbb{N}} \big(U_i + W_i(s) - \mathrm{Var}(W(s))/2\big), \quad s \in \mathbb{R}^D,$$

and called Brown-Resnick process associated to the semi-variogram γ, is stationary and max-stable with standard Gumbel margins and its law only depends on the semi-variogram γ (Kabluchko et al., 2009). For the application of the Brown-Resnick model to observed data with locations in $\mathbb{R}^2$, we propose to restrict to semi-variograms from a flexible parametric subclass, such as semi-variograms of the type

$$\gamma_\vartheta(h) = \|sA(b,\zeta)h\|^\alpha, \quad h \in \mathbb{R}^2, \tag{9.4}$$

with $\vartheta = (s, b, \zeta, \alpha)$ for $s, b > 0$, $\zeta \in (-\pi/4, \pi/4]$ and $\alpha \in (0, 2]$. Here, the matrix $A(b,\zeta) \in \mathbb{R}^{2\times 2}$ allows for geometric (elliptical) anisotropy, i.e.

$$A(b,\zeta) = \begin{pmatrix} \cos\zeta & \sin\zeta \\ -b\sin\zeta & b\cos\zeta \end{pmatrix} \tag{9.5}$$

(cf. Chilès and Delfiner, 2012, Subsection 2.5.2), and s is an overall scale factor.

9.3 Modeling by a Bivariate Random Field

In this section, we also take into account the dependence between the observed maximum $V_{\max}^{\mathrm{obs}}$ and its forecast $V_{\max}^{\mathrm{pred}}$. As $V_{\max}^{\mathrm{pred}}$ is a forecast for $V_{\max}^{\mathrm{obs}}$, it seems reasonable to use a GEV model similar to the one described in Section 9.2 with possibly different parameters ξ^{pred}, $\mu^{\mathrm{pred}}(\cdot)$ and $\sigma^{\mathrm{pred}}(\cdot)$, i.e.

$$\frac{V_{\max}^{\mathrm{pred}}(l,p) - m(l,p)}{s(l,p)} \sim G_{\xi^{\mathrm{pred}},\mu^{\mathrm{pred}}(l),\sigma^{\mathrm{pred}}(l)} \tag{9.6}$$

(cf. Equation (9.2)). Marginally transforming $V_{\max}^{\mathrm{pred}}$ analogously to (9.3) yields a random field $\{X^{\mathrm{pred}}(l,p),\, l \in \mathbb{R}^2,\, p \in \mathbb{Z}\}$ with standard Gumbel margins. Thus, we end up with bivariate spatial random fields $\{(X^{\mathrm{obs}}(l,p), X^{\mathrm{pred}}(l,p)),\, l \in \mathbb{R}^2\}$ which are assumed to be independent and identically distributed for $p \in \mathbb{Z}$.

A bivariate Brown-Resnick process can be constructed in the following way (cf. Molchanov and Stucki, 2013; Genton et al., 2015): Let $\sum_{i\in\mathbb{N}} \delta_{U_i}$ be a Poisson point process on $\mathbb{R}$ with intensity measure $e^{-u}\,\mathrm{d}u$. Further, let W_i, $i \in \mathbb{N}$, be independent copies of a bivariate centered Gaussian process $W = (W^{(1)}, W^{(2)})^\top = \{(W^{(1)}(s), W^{(2)}(s))^\top : s \in \mathbb{R}^D\}$ such that the pseudo cross-variogram (Clark et al., 1989; Papritz et al., 1993) $\gamma(h) = (\gamma_{ij}(h))_{i,j\in\{1,2\}}$ defined by

$$2\gamma_{ij}(h) = \mathrm{Var}(W^{(i)}(s+h) - W^{(j)}(s)), \quad h \in \mathbb{R}^D, \tag{9.7}$$

does not depend on $s \in \mathbb{R}^D$. Analogously to the univariate Brown-Resnick process, it can be shown that the bivariate Brown-Resnick process $Z = (Z^{(1)}, Z^{(2)})^\top$ defined by

$$Z^{(j)}(s) = \max_{i\in\mathbb{N}} \big(U_i + W_i^{(j)}(s) - \mathrm{Var}(W^{(j)}(s))/2\big), \quad s \in \mathbb{R}^D, \quad j = 1, 2, \tag{9.8}$$

is max-stable and stationary. Its law only depends on the pseudo cross-variogram γ.

Remark 9.1. The fact that $(\gamma_{ij}(h))_{i,j=1,2}$ can be defined independently of $s \in \mathbb{R}^D$ implies that W is intrinsically stationary, i.e. the process $\{W(s+h)-W(s) : s \in \mathbb{R}^D\}$ is stationary for every $h \in \mathbb{R}^D$. Both conditions, however, are not equivalent as the definition of $\gamma_{12}(h)$ might depend on $s \in \mathbb{R}^d$ even if W is intrinsically stationary. For instance, if both components are independent, we have $\gamma_{12}(h) = 2\gamma_{11}(s+h) + 2\gamma_{22}(s)$ for the off-diagonal element of pseudo cross-variogram. By way of contrast, intrinsic stationarity is equivalent to the cross variogram $h \mapsto (\mathbb{E}(W_i(s+h) - W_i(s))(W_j(s+h) - W_j(s)))_{i,j=1,2}$ being independent of s.

Indeed, Molchanov and Stucki (2013) already gave necessary and sufficient conditions for a multivariate process of Brown-Resnick type to be stationary. For a fixed intensity $e^{-u}\,\mathrm{d}u$ of the Poisson point process, the conditions on Gaussian processes given in Theorem 5.3 in Molchanov and Stucki (2013) can be shown to be equivalent to the conditions on the process W stated above (if we additionally require Z to have standard Gumbel margins) by a straightforward computation. Thus, the Gaussian processes in the above definition of bivariate Brown-Resnick processes are essentially the only ones that yield a stationary max-stable process.

In the following, we investigate the structure and the asymptotic behavior of bivariate variograms that are translation invariant, refining the result by Papritz et al. (1993) that $\lim_{h\to\infty} \gamma_{12}(h)/\gamma_{11}(h) = 1$ if γ_{11} is unbounded. This allows us to find valid models for bivariate Brown-Resnick processes. The following theorem, as well as the statements above, immediately extend to the general multivariate case.

Theorem 9.2. *Let $W = (W^{(1)}, W^{(2)})^\top$ be a bivariate second-order process on $\mathbb{R}^D$ with pseudo cross-variogram $(\gamma_{ij}(h))_{i,j\in\{1,2\}}$ which does not depend on $s \in \mathbb{R}^D$. Then,*

$$\sqrt{\gamma(h)} = \sqrt{(\gamma_{ij}(h))_{i,j\in\{1,2\}}} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \sqrt{\gamma_0(h)} + \begin{pmatrix} f_{11}(h) & f_{12}(h) \\ f_{21}(h) & f_{22}(h) \end{pmatrix}$$

for some univariate variogram γ_0 and bounded functions $f_{11}, f_{12}, f_{21}, f_{22} : \mathbb{R}^D \to \mathbb{R}$.

Proof. For $i, j \in \{1, 2\}$, and $h \in \mathbb{R}^D$, we obtain

$$\begin{aligned}
&\left(\sqrt{\gamma_{ii}(h)} - \sqrt{\gamma_{jj}(h)}\right)^2 = \gamma_{ii}(h) - 2\sqrt{\gamma_{ii}(h)\gamma_{jj}(h)} + \gamma_{jj}(h) \\
&\le \gamma_{ii}(h) - \operatorname{Cov}\left(W^{(i)}(h) - W^{(i)}(0), W^{(j)}(h) - W^{(j)}(0)\right) + \gamma_{jj}(h) \\
&= \frac{1}{2}\operatorname{Var}\left(W^{(i)}(h) - W^{(i)}(0) - W^{(j)}(h) + W^{(j)}(0)\right) \\
&= \gamma_{ij}(0) - \operatorname{Cov}\left(W^{(i)}(h) - W^{(j)}(h), W^{(i)}(0) - W^{(j)}(0)\right) + \gamma_{ij}(0) \le 4\gamma_{ij}(0),
\end{aligned}$$

where we used the Cauchy-Schwarz inequality for both inequalities. Analogously, we get the assessment

$$\begin{aligned}
&\left(\sqrt{\gamma_{ii}(h)} - \sqrt{\gamma_{ji}(h)}\right)^2 = \gamma_{ii}(h) - 2\sqrt{\gamma_{ii}(h)\gamma_{ji}(h)} + \gamma_{ji}(h) \\
&\le \gamma_{ii}(h) - \operatorname{Cov}(W^{(i)}(h) - W^{(i)}(0), W^{(j)}(h) - W^{(i)}(0)) + \gamma_{ji}(h) \\
&= \frac{1}{2}\operatorname{Var}\left(W^{(i)}(h) - W^{(j)}(h)\right) = \gamma_{ij}(0).
\end{aligned}$$

Thus, the assertion of the theorem follows with $\gamma_0 = \gamma_{11}$. □

As the components of a translation invariant bivariate pseudo cross-variogram only differ by a function that may increase only with a rate of order $O(\sqrt{\gamma_0(h)})$ (Theorem 9.2), a reasonable and not too restrictive model for the corresponding bivariate Gaussian random field $W = (W^{(1)}, W^{(2)})^\top$ is given by

$$W(s) = (1,1)^\top V_1(s) + V_2(s), \quad s \in \mathbb{R}^D,$$

where V_1 is a univariate Gaussian random field with stationary increments and semi-variogram γ_0 and V_2 is a bivariate stationary Gaussian random field with bivariate cross-covariance function $C(h) = (C_{ij}(h))_{i,j\in\{1,2\}}$, independent from V_1. Then, the pseudo cross-variogram γ of W has the form

$$\gamma_{ij}(h) = \gamma_0(h) + \frac{1}{2}C_{ii}(0) + \frac{1}{2}C_{jj}(0) - C_{ij}(h), \quad i,j \in \{1,2\},\ h \in \mathbb{R}^D.$$

Analogously to the univariate case, we propose to restrict to a parametric subclass of semi-variograms for γ_0 such as

$$\gamma_0(h) = \sigma^2 \frac{(\kappa^{-1}\|h\|)^2}{((\kappa^{-1}\|h\|)^2 + 1)^\beta}$$

where $\sigma, \kappa > 0$ and $\beta \in (0,1)$. Here, γ_0 is a valid univariate variogram as $h \to \|h\|^2$ is a variogram and $\lambda \mapsto \lambda/(\lambda+1)^\beta$ is a Bernstein function (cf. Berg et al., 1984; Schilling et al., 2010). Note that γ_0 is a variogram of power law type modified to be smooth at the origin.

For the bivariate cross-covariance C, we propose to use a parsimonious version of the bivariate Matérn model (cf. Gneiting et al., 2010), which is a bivariate generalization of one of the most widely used models in geostatistics, the Matérn model (cf. Guttorp and Gneiting, 2006; Stein, 1999, for example). In the bivariate Matérn model, each component of C is a Matérn covariance function which we parametrize in the way suggested by Handcock and Wallis (1994), i.e.

$$C_{ii}(h) = \sigma_i^2 \frac{2^{1-\nu_i}}{\Gamma(\nu_i)} \left(\frac{2\sqrt{\nu_i}}{a_i}\|h\|\right)^{\nu_i} K_{\nu_i}\left(\frac{2\sqrt{\nu_i}}{a_i}\|h\|\right), \quad i = 1,2,$$
$$C_{12}(h) = C_{21}(h) = \rho\sigma_1\sigma_2 \frac{2^{1-\nu_{12}}}{\Gamma(\nu_{12})} \left(\frac{2\sqrt{\nu_{12}}}{a_{12}}\|h\|\right)^{\nu_{12}} K_{\nu_{12}}\left(\frac{2\sqrt{\nu_{12}}}{a_{12}}\|h\|\right),$$

for $a_1, a_2, a_{12}, \sigma_1, \sigma_2, \nu_1, \nu_2, \nu_{12} > 0$ and suitable $\rho \in [-1,1]$.

Here, analogously to the parsimonious version of the bivariate Matérn model which is based on a different parametrization (Gneiting et al., 2010), we set $a_1 = a_{12} = a_2 = a > 0$ and $\nu_{12} = \frac{1}{2}(\nu_1 + \nu_2)$. Then, by Theorem 3 in (Gneiting et al., 2010), C is a valid bivariate cross-covariance model if and only if

$$\rho^2 \le \frac{(1+\nu_{12}^{-1})^{2\nu_{12}+2}}{(1+\nu_1^{-1})^{\nu_1+1}(1+\nu_2^{-1})^{\nu_2+1}}.$$

To increase the flexibility of the model, we further add a spatially constant effect with variance c^2 in the second component. Thus, C has the form

$$C_{11}(h) = \sigma_1^2 \frac{2^{1-\nu_1}}{\Gamma(\nu_1)} \left(\frac{2\sqrt{\nu_1}}{a}\|h\|\right)^{\nu_1} K_{\nu_1}\left(\frac{2\sqrt{\nu_1}}{a}\|h\|\right),$$
$$C_{12}(h) = C_{21}(h) = \rho\sigma_1\sigma_2 \frac{2^{1-\nu_{12}}}{\Gamma(\nu_{12})} \left(\frac{2\sqrt{\nu_{12}}}{a}\|h\|\right)^{\nu_{12}} K_{\nu_{12}}\left(\frac{2\sqrt{\nu_{12}}}{a}\|h\|\right),$$

$$C_{22}(h) = c^2 + \sigma_2^2 \frac{2^{1-\nu_2}}{\Gamma(\nu_2)} \left(\frac{2\sqrt{\nu_2}}{a} \|h\| \right)^{\nu_2} K_{\nu_2} \left(\frac{2\sqrt{\nu_2}}{a} \|h\| \right).$$

Note that as the common summand γ_0 is smooth at the origin, the behavior of γ_{ii} near the origin, i.e. the differentiability of $W^{(i)}$, depends only on the behavior of C which can be modeled flexibly by the smoothness parameters ν_1 and ν_2 of the bivariate Matérn model. In particular, as $\|h\| \to 0$ and for some $k(a, \nu) > 0$, we have

$$\gamma_{ii}(h) = \begin{cases} k(a, \nu_i)\|h\|^{2\nu_i} + O(\|h\|^2), & \nu_i < 1, \\ k(a, 1)\|h\|^2 \log \|h\| + O(\|h\|^2), & \nu_i = 1, \\ k(a, \nu_i)\|h\|^2 + o(\|h\|^2), & \nu_i > 1 \end{cases}$$

(cf. Stein, 1999). Furthermore, the sample paths are m times differentiable if and only if $\nu > m$ (Gelfand et al., 2010). The behavior of the γ_{ii} as $\|h\| \to \infty$, which has to be the same for all components by Theorem 9.2, is parameterized by β as we have $\gamma_{ii}(h)\|h\|^{-2(1-\beta)} \to 1$ as $\|h\| \to \infty$. To increase the applicability of our model to real data located in $\mathbb{R}^2$, we further allow for geometric anisotropy, replacing $\|h\|$ by $\|h^*\|$ where $h^* = A(b, \zeta)h$ and $A(b, \zeta)$ is the anisotropy matrix defined in (9.5). Thus, we obtain the variogram model $\gamma(\vartheta; \cdot)$ given by

$$\begin{aligned} \gamma_{ii}(\vartheta; h) &= \sigma^2 \frac{(\kappa^{-1}\|h^*\|)^2}{((\kappa\|h^*\|)^2 + 1)^\beta} + \sigma_i^2 \left(1 - \left(\frac{2\sqrt{\nu_i}}{a} \|h^*\| \right)^{\nu_1} K_{\nu_1} \left(\frac{2\sqrt{\nu_i}}{a} \|h^*\| \right) \right), \\ \gamma_{12}(\vartheta; h) &= \sigma^2 \frac{(\kappa^{-1}\|h^*\|)^2}{((\kappa\|h^*\|)^2 + 1)^\beta} + \frac{\sigma_1^2 + c^2 + \sigma_2^2}{2} \\ &\quad - \rho\sigma_1\sigma_2 \frac{2^{1-\nu_{12}}}{\Gamma(\nu_{12})} \left(\frac{2\sqrt{\nu_{12}}}{a} \|h^*\| \right)^{\nu_{12}} K_{\nu_{12}} \left(\frac{2\sqrt{\nu_{12}}}{a} \|h^*\| \right), \end{aligned} \tag{9.9}$$

for $i = 1, 2$ and $h \in \mathbb{R}^2$ where $\vartheta = (\sigma, \kappa, b, \zeta, \beta, c, \sigma_1, \nu_1, \sigma_2, \nu_2, a, \rho)$.

9.4 Model Fitting

In the following, we will assume that data $v_{\max}^{\text{obs}}(l_i, p)$ and $v_{\max}^{\text{pred}}(l_i, p)$ for the maximal observed and forecasted variable of interest at stations l_i, $i = 1, \ldots, n_l$ and time period $p = 1, \ldots, n_p$ are available.

9.4.1 Fitting of the Univariate Model

Let henceforth be $k \in \{\text{“obs”}, \text{“pred”}\}$. We concentrate here on the estimation of the GEV and max-stable parameters assuming that the unknown mean $m(l_i, d)$ and standard deviation $s(l_i, p)$ of the underlying distribution F have already been estimated by $\hat{m}(l_i, p)$ and $\hat{s}(l_i, p)$, respectively. An example for the later estimates can be found in Section 9.6. Given the estimates $\hat{m}(l_i, p)$ and $\hat{s}(l_i, p)$, we obtain the standardized data

$$y^k(l_i, p) = \frac{v_{\max}^k(l_i, p) - \hat{m}(l_i, p)}{\hat{s}(l_i, p)}, \qquad i = 1, \ldots, n_l,\ p = 1, \ldots, n_p, \tag{9.10}$$

which are assumed to be GEV distributed with parameters ξ^k, $\mu^k(l_i)$ and $\sigma^k(l_i)$. We assume that the parameters are independent between the stations. This can be justified by measurement errors and the fact, that the forecasts used to estimate $\hat{m}(l_i, d)$ and $\hat{s}(l_i, d)$ are not directly available for the station l_i but only for the closest grid point. Furthermore,

we face model errors, e.g. misrepresentation of orographic effects. Effects stemming from the environment of the measurement stations might even be the major cause for the variations. As genuine variation and measurement errors cannot be separated in our set up, we estimate the parameters separately for each station, via maximum likelihood. As the standardized data y^k are assumed to be temporally independent, by Smith (1985), the maximum likelihood estimators $(\hat{\xi}^k(l_i), 1 \le i \le n_l$, are asymptotically normally distributed if $\xi^k > -0.5$. Thus, under the hypothesis that $\hat{\xi}^k = \frac{1}{n_l}\sum_{i=1}^{n_l}\hat{\xi}^k(l_i)$ is the true shape parameter of the GEV at each station, the standardized residuals

$$\frac{\hat{\xi}^k(l_1) - \hat{\xi}^k}{(\widehat{\mathrm{Var}}(\hat{\xi}^k(l_1)))^{1/2}}, \ldots, \frac{\hat{\xi}^k(l_{n_l}) - \hat{\xi}^k}{(\widehat{\mathrm{Var}}(\hat{\xi}^k(l_{n_l})))^{1/2}}$$

are approximately standard normally distributed, where $\widehat{\mathrm{Var}}(\hat{\xi}^k(l_i))$ is the variance of $\hat{\xi}^k(l_i)$ estimated via the Hesse matrix of the log-likelihood function. Thus, the three hypotheses that the shape parameter, the location and the scale parameter are spatially constant can be checked indirectly via one-sample Kolmogorov-Smirnov tests of the corresponding residuals for the standard normal distribution. Here, although the data for different locations may be dependent, we assume that the normalized estimated parameters are independent.

By transformation (9.3), the estimates $\hat{\xi}^k$, $\hat{\mu}^k(l_i)$ and $\hat{\sigma}^k(l_i)$ yield normalized data

$$x^k(l_i,p) = \frac{1}{\hat{\xi}^k}\log\left(1 + \hat{\xi}^k\frac{y^k(l_i,p) - \hat{\mu}^k(l_i)}{\hat{\sigma}^k(l_i)}\right), \quad 1 \le i \le n_l,\ 1 \le p \le n_p. \tag{9.11}$$

These can be compared to a standard Gumbel distribution via Kolmogorov-Smirnov tests separately for each station as a goodness-of-fit test for the marginal model

$$V^k_{\max}(l_i,p) \sim G_{\hat{\xi}^k,\hat{\mu}^k_v(l_i,p),\hat{\sigma}^k_v(l_i,p)}, \tag{9.12}$$

$$\text{where} \quad \hat{\mu}^k_v(l,p) = \hat{m}(l,p) + \hat{s}(l,p)\hat{\mu}^k(l) \quad \text{and} \quad \hat{\sigma}^k_v(l,p) = \hat{s}(l,p)\hat{\sigma}^k(l). \tag{9.13}$$

In order to capture the spatial dependence structure, a univariate Brown-Resnick process associated to a variogram γ^k as defined in (9.4) is fitted to the transformed data $(x^k(l_i,p))_{1\le i\le n_l, 1\le p\le n_p}$. Note that there exist numerous methods of inference for Brown-Resnick processes, see, for example, Engelke et al. (2015) for a comparison of different estimators. The method we will use is based on the extremal coefficient function (Schlather and Tawn, 2003). For a stationary Brown-Resnick process associated to the semi-variogram γ^k, the pairwise extremal coefficients are given by

$$\theta^k(s_1,s_2) = \frac{\log\mathbb{P}(X^k(s_1)\le x, X^k(s_2)\le x)}{\log\mathbb{P}(X^k(s_1)\le x)} = 2\Phi\left(\sqrt{\frac{\gamma^k(s_1-s_2)}{2}}\right), \quad s_1,s_2\in\mathbb{R}^2, \tag{9.14}$$

where Φ denotes the standard normal distribution function (cf. Kabluchko et al., 2009). This relation can be used for fitting Brown-Resnick processes to real data as the extremal coefficients $\theta^k(s_1,s_2)$ can be estimated well via the relation

$$\theta^k(s_1,s_2) = \frac{1+2\nu^{F,k}(s_1,s_2)}{1-2\nu^{F,k}(s_1,s_2)}, \quad s_1,s_2\in\mathbb{R}^2, \tag{9.15}$$

where the F-madogram $\nu^{F,k}(s_1,s_2)$ is defined by

$$\nu^{F,k}(s_1,s_2) = \frac{1}{2}\mathbb{E}\left|F(X^k(s_1)) - F(X^k(s_2))\right|, \quad s_1,s_2\in\mathbb{R}^2, \tag{9.16}$$

and F is the marginal distribution function of $X^k(s)$ (Cooley et al., 2006). Thus, we obtain a plug-in estimator $\hat{\theta}^k(l_i, l_j)$ for the extremal coefficients $\theta^k(l_i, l_j)$, by replacing $\nu^{F,k}$ in (9.15) by an estimator $\hat{\nu}^{F,k}(l_i, l_j)$, $1 \leq i, j \leq n_l$. In order to avoid propagation of errors in marginal modeling, we choose the non-parametric estimator

$$\hat{\nu}^{F,k}(l_i, l_j) = \frac{1}{2 \cdot n_p \cdot (n_p - 1)} \sum_{p=1}^{n_p} \left| R_p(x^k(l_i, \cdot)) - R_p(x^k(l_j, \cdot)) \right| \tag{9.17}$$

where $R_p(x)$ denotes the rank of the p-th component of some vector x (cf. Ribatet, 2013). Then, the corresponding variogram parameter vector $\hat{\vartheta}^k$ can be estimated by a weighted least squares fit of $\hat{\theta}^k(l_i, l_j)$ to $\theta^k(l_i, l_j)$ as given in (9.14). As proposed by Smith (1990), we choose weights that depend on the (estimated) variance $\widehat{\mathrm{Var}}(\theta^k(l_i, l_j))$ of the estimator $\theta^k(l_i, l_j)$. Thus, we obtain the estimator

$$\hat{\vartheta}^k = \arg\min_{\vartheta} \sum_{1 \leq i < j \leq n_l} \left(\frac{\hat{\theta}^k(l_i, l_j) - 2\Phi\left(\sqrt{\gamma^k(l_i - l_j)/2}\right)}{\sqrt{\widehat{\mathrm{Var}}(\theta^k(l_i, l_j))}} \right)^2 . \tag{9.18}$$

We will further discuss the estimation of the variance of $\theta^k(l_i, l_j)$ in Section 9.6.

9.4.2 Fitting of the Bivariate Model

For fitting the bivariate Brown-Resnick process $\{(X^{\mathrm{obs}}(l), X^{\mathrm{pred}}(l))^\top : l \in \mathbb{R}^2\}$ we consider the extremal coefficients $\theta^{k_1,k_2}(s,t)$ of max-stable vectors $(X^{k_1}(s), X^{k_2}(t))^\top$ for $k_1, k_2 \in \{\text{“obs”}, \text{“pred”}\}$. The extremal coefficients can be estimated from the transformed data $x^{\mathrm{obs}}(l_i, p)$ and $x^{\mathrm{pred}}(l_i, p)$, $1 \leq i \leq n_l$, $1 \leq p \leq n_p$, in the same way as in the univariate case. The resulting estimates $\hat{\theta}^{k_1,k_2}(l_i, l_j)$, $1 \leq i, j \leq n_l$, $k_1, k_2 \in \{\text{“obs”}, \text{“pred”}\}$ are compared to the corresponding extremal coefficients of a bivariate Brown-Resnick process associated to the variogram $\gamma(\vartheta; \cdot)$ yielding the weighted least squares fit

$$\hat{\vartheta} = \arg\min_{\vartheta} \sum_{1 \leq i,j \leq n_l} \sum_{k_1,k_2 \in \{\text{“obs”}, \text{“pred”}\}} \left(\frac{\hat{\theta}^{k_1,k_2}(l_i, l_j) - 2\Phi(\sqrt{\gamma_{k_1,k_2}(\vartheta; l_i - l_j)/2})}{\widehat{\mathrm{Var}}(\theta^{k_1,k_2}(l_i, l_j))} \right)^2 .$$

9.5 The Post-Processing Procedure

As the bivariate Brown-Resnick process model developed in this chapter describes the joint distribution of the observed and forecasted maxima of the variable of interest, it allows for some spatial post-processing of the original forecast. In this section, we will describe the resulting post-processing procedure in more detail and provide some tools to verify the procedure and the underlying model.

9.5.1 Post-Processing via Conditional Simulation

Let $\hat{\xi}^{\mathrm{obs}}$, $\hat{\mu}^{\mathrm{obs}}(\cdot)$, $\hat{\sigma}^{\mathrm{obs}}(\cdot)$, $\hat{\xi}^{\mathrm{pred}}$, $\hat{\mu}^{\mathrm{pred}}(\cdot)$, $\hat{\sigma}^{\mathrm{pred}}(\cdot)$ and $\hat{\vartheta}$ be estimates for the GEV and variogram parameters derived from past training data. Further, assume that we have $v^{\mathrm{pred}}_{\max}(l_i, p)$, $\hat{m}(l_i, p)$ and $\hat{s}(l_i, p)$, $i = 1, \ldots, n_l$, based on forecasts for n_l locations $l_1, \ldots, l_{n_l}$ and a time period p in near future. Then, we obtain an arbitrary number K of realizations $(v_j(l_i))_{1 \leq i \leq n_l}$, $j = 1, \ldots, K$, of the modeled distribution of the maximal observation conditional on the forecast by the following three-step procedure:

1. Transform $v_{\max}^{\text{pred}}(\cdot, p)$ to standard Gumbel margins:

$$x^{\text{pred}}(\cdot) = \frac{1}{\hat{\xi}^{\text{pred}}} \log\left(1 + \hat{\xi}^{\text{pred}} \frac{v_{\max}^{\text{pred}}(\cdot, p) - \hat{\mu}_v^{\text{pred}}(\cdot, p)}{\hat{\sigma}_v^{\text{pred}}(\cdot, p)}\right),$$

   where $\hat{\mu}_v^{\text{pred}}$ and $\hat{\sigma}_v^{\text{pred}}$ are given by Equation (9.13) for $k = \text{pred}$.

2. Conditional simulation of a bivariate Brown-Resnick process given its second component: Simulate K independent realizations $(x_j^{\text{obs}}(\cdot), x_j^{\text{pred}}(\cdot))$, $j = 1, \ldots, K$, of a bivariate Brown-Resnick process associated to the pseudo cross-variogram $\gamma(\hat{\vartheta}^{\text{obs}}; \cdot)$ with standard Gumbel margins conditional on $x_j^{\text{pred}}(\cdot) = x^{\text{pred}}(\cdot)$.

3. Transform $x_j^{\text{obs}}(\cdot)$ to GEV margins: For $j = 1, \ldots, K$, set

$$v_j(\cdot, p) = \hat{\sigma}_v^{\text{obs}}(\cdot, p) \frac{\exp(\hat{\xi}^{\text{obs}} x_j^{\text{obs}}(\cdot)) - 1}{\hat{\xi}^{\text{obs}}} + \hat{\mu}_v^{\text{obs}}(\cdot, p),$$

   where $\hat{\mu}_v^{\text{obs}}$ and $\hat{\sigma}_v^{\text{obs}}$ are given by Equation (9.13) for $k = \text{pred}$.

The random fields obtained by this three-step procedure can be interpreted as post-processed probabilistic forecasts for the maxima of the variable of interest at time period p. While the first and the third steps only consist of marginal transformations, the conditional simulation in the second step is the challenging part of the procedure. For this step, the algorithm by Dombry et al. (2013) can be used. Note that the algorithm, which has originally been designed for conditional simulation of univariate Brown-Resnick processes, can directly be transferred to the multivariate case by perceiving the multivariate processes as univariate processes on a larger index set. However, the computations will be computationally expensive, in particular if the number of conditioning locations gets large.

9.5.2 Verification

In practical applications, the proposed post-processing procedure and the underlying model need to be verified. Here, we do not only consider the full bivariate Brown-Resnick model which forms the base of the post-processing procedure, but also intermediate models such as the marginal GEV model and the univariate model. This allows us to evaluate the effect of incorporating the spatial dependence structure and the forecasted maxima, respectively.
For the evaluation and verification of the different models, we choose a standard verification score in probabilistic prediction, the (negatively oriented) continuous ranked probability score (CRPS) (cf. Gneiting and Raftery, 2007):

$$CRPS(F, x) = \int_{-\infty}^{\infty} |y - x| \, F(\mathrm{d}y) - \frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |y_1 - y_2| \, F(\mathrm{d}y_1) \, F(\mathrm{d}y_2),$$

where F is a real-valued distribution and $x \in \mathbb{R}^m$ is an observation. The continuous ranked probability score is a strictly proper scoring rule, i.e.

$$\int CRPS(F, x) F(\mathrm{d}x) \leq \int CRPS(G, x) F(\mathrm{d}x)$$

for all distribution functions F and G with finite first moments and equality if and only if $F = G$. This indicates that the mean CRPS for different observations is the smaller,

the better the predicted distribution F fits to the true distribution of the observation data. The usefulness of the CRPS for evaluating extremes was shown by Friederichs and Thorarinsdottir (2012).

First, we evaluate the improvement in predictive quality by fitting the parameters of the GEV model given in (9.12) and (9.13) to the observations instead of the forecast, i.e. we calculate $\mathrm{CRPS}^{\mathrm{obs}}(l_i)$ and $\mathrm{CRPS}^{\mathrm{pred}}(l_i)$ where

$$\mathrm{CRPS}^{k}(l_i) = n_d^{-1} \sum_{d=1}^{n_d} \mathrm{CRPS}(G_{\hat{\xi}^k, \hat{\mu}_v^k(l_i,d), \hat{\sigma}_v^k(l_i,d)}, v_{\max}^{\mathrm{obs}}(l_i, d)).$$

for every station l_i, $1 \leq i \leq n_l$, and $k \in \{\text{“obs”}, \text{“pred”}\}$. For the calculation, we employ the closed formula for the CRPS of a GEV provided by Friederichs and Thorarinsdottir (2012). For $\xi \neq 0$, they obtain

$$\begin{aligned}\mathrm{CRPS}(G_{\xi,\mu,\sigma}, x) = &\left(x - \mu + \frac{\sigma}{\xi}\right)(2F(x) - 1) \\ &- \frac{\sigma}{\xi}\left(2^{\xi}\Gamma(1-\xi) - 2\Gamma_l(1-\xi, -\log F(x)\right)\end{aligned}$$

where Γ_l is the lower incomplete gamma function. Furthermore, the CRPS for the GEV fitted to the observations can be compared to the CRPS of the original forecast

$$\mathrm{CRPS}^{\mathrm{orig}}(l_i) = {n_p}^{-1} \sum_{p=1}^{n_p} \mathrm{CRPS}(F_{l_i,p}^{\mathrm{orig}}, v_{\max}^{\mathrm{obs}}(l_i, p)) \tag{9.19}$$

where $F_{l_i,p}^{\mathrm{orig}}$ denotes the distribution of the original (probabilistic) forecast for the maximum of the variable of interest at location l_i within time period p. If this forecast is given by an ensemble of values, such as the output of a numerical weather prediction model, for example, $F_{l_i,p}^{\mathrm{orig}}$ corresponds to the empirical distribution function of this sample. If the forecast corresponds to a single value, $\mathrm{CRPS}^{\mathrm{orig}}(l_i)$ reduces to the mean absolute error.

Finally, the full bivariate model and, thus, the proposed post-processing procedure can be evaluated and verified by comparing the CRPS

$$\mathrm{CRPS}^{\mathrm{biv}}(l_i) = {n_p}^{-1} \sum\nolimits_{p=1}^{n_p} \mathrm{CRPS}\big(F_{l_i,p|v_{\max}^{\mathrm{pred}}}, v_{\max}^{\mathrm{obs}}(l_i, p)\big)$$

where $F_{l_i,p|v_{\max}^{\mathrm{pred}}}$ is the distribution of the observed maximum at location l_i, $1 \leq i \leq n_l$ within time period p conditional on $v_{\max}^{\mathrm{pred}}$, that is, the distribution of the post-processed forecast, with the CRPS of the original forecast, $\mathrm{CRPS}^{\mathrm{orig}}(l_i)$.

9.6 Application to Real Data

In this section, we will apply the fitting and verification procedure described in Section 9.4 to real wind gust data consisting both of observation and forecast data. We will see that, even though the marginal distributions are fitted quite well, a forecast based on the single GEV for the observations is not able to outperform the forecast by the numerical weather prediction model. However, the results for the bivariate model indicate that the post-processing procedure proposed in Subsection 9.5.1 improves the predictive quality. We also discuss the uncertainty of the obtained estimates.

9.6.1 The Data

We consider observed as well as forecasted wind speed data provided by Germany's National Meteorological Service, the Deutscher Wetterdienst (DWD). We use observations from 218 DWD weather stations over Germany at 360 days from March 2011 to February 2012. The weather stations register mean and maximum wind speed on an hourly basis. Due to the inertia of the measuring instruments, the maximum wind speed approximately corresponds to the highest 3-second average wind speed. Here, we use the maximum wind speed $v_{\max}^{\text{obs}}(l,d)$ between 08 UTC and 18 UTC for each station l and each day d.

Furthermore, for each day, forecasts for the wind speed maxima and for the hourly mean wind speed both in 10m height above ground and for the 10-hour-period from 08 UTC to 18 UTC are available. The forecasts are provided by the COSMO-DE ensemble prediction system (EPS) operated by DWD. COSMO-DE (Baldauf et al., 2011) is a non-hydrostatic limited-area numerical weather prediction model that gives forecasts for the next 21 hours on a horizontal grid with a width of 2.8km covering Germany and neighboring countries. For each variable of interest, the COSMO-DE EPS yields forecasts consisting of 20 ensemble members stemming from COSMO-DE runs with five different physical parameterizations and four different lateral boundary conditions provided by global model forecasts. For more details on the Consortium for Small-scale Modeling see `http://www.cosmo-model.org/`, and Gebhardt et al. (2011) and Peralta et al. (2012), for COSMO-DE EPS.

The COSMO-DE EPS is initialized every 3 hours. Here, we take the forecasts that are initialized at 00 UTC. Using the forecasts for the nearest grid point of a station, we obtain forecasts $v_{\text{mean}}^{(1)}(l,d,\tau)$, ..., $v_{\text{mean}}^{(20)}(l,d,\tau)$, $\tau \in \{9,10,\ldots,18\}$, and $v_{\max}^{(1)}(l,d)$, ..., $v_{\max}^{(20)}(l,d)$ for every weather station l and every day d. Here, $v_{\text{mean}}^{(j)}(l,d,\tau)$ and $v_{\max}^{(j)}(l,d)$ denote the forecast for the mean wind speed between $(\tau-1)$ UTC and τ UTC and the maximal wind speed, respectively, at station l and day d, forecasted by the jth COSMO-DE ensemble member.

For the application of our model with a stationary spatial dependence structure, in the following, we will restrict ourselves to forecasted and observed data for 119 DWD stations north of 51°N, denoted by $l_1,\ldots,l_{119}$, as the northern part of Germany has a much more homogeneous topography than the southern part.

9.6.2 Applying the Univariate Model

As the wind speed observations correspond to 3-second averages, the daily maximal wind gusts $v_{\max}^{\text{obs}}$ can be perceived as the maximum of a long time series. Further, the distribution of a single wind speed is frequently modeled by a Weibull or a Gamma distribution (e.g., Conradsen et al., 1984; Pavia and O'Brien, 1986; Sloughter et al., 2007), that is, given a fixed shape parameter of the Weibull or Gamma distribution, respectively, which is spatially and temporally constant, the single observations may be assumed to come from a location-scale family of distributions. These considerations give support to the usage of the GEV model presented in Section 9.2 as a model for the maximal wind speed $V_{\max}^{k}(l_i,d)$, $i \in \{1,\ldots,119\}$, $d \in \{1,\ldots,360\}$. Fitting a GEV distribution to the standardized wind speeds $y^k(l_i,d)$ as defined in (9.10) needs the estimates $\hat{m}(l_i,d)$ and $\hat{s}(l_i,d)$ for the mean and the standard deviation of the underlying wind speed distribution. We aim to extract these characteristics from the forecast. Here, instead of direct estimates for the mean and

the standard deviation, we use

$$\hat{m}(l_i, d) = \max_{j=1}^{20} \frac{1}{10} \sum_{\tau=9}^{18} v_{\text{mean}}^{(j)}(l_i, d, \tau) \tag{9.20}$$

$$\text{and } \hat{s}(l_i, d) = \left(\frac{1}{199} \sum_{j=1}^{20} \sum_{\tau=9}^{18} \left[v_{\text{mean}}^{(j)}(l_i, d, \tau) - \frac{1}{200} \sum_{j'=1}^{20} \sum_{\tau'=9}^{18} v_{\text{mean}}^{(j')}(l_i, d, \tau') \right]^2 \right)^{1/2}. \tag{9.21}$$

Even though not providing consistent estimates for mean and standard deviation of the underlying distribution, $\hat{m}(l_i, d)$ and $\hat{s}(l_i, d)$ lead to a consistent normalization in the following sense: If the data $v_{\text{max}}^{\text{obs}}$ and the forecasts $v_{\text{mean}}^{(j)}$ are affinely transformed (i.e. the parameters m and s are modified) in the same way, the normalized data $y^{\text{obs}}(l_i, d)$ remain unchanged. This choice of $\hat{m}(l_i, d)$ and $\hat{s}(l_i, d)$ also ensures the identifiability of the GEV parameters $\mu^k(l_i)$ and $\sigma^k(l_i)$.

As described in Section 9.4, the GEV parameters for the standardized observations can be estimated via maximum likelihood and the hypotheses that these are spatially constant can be checked via Kolmogorov-Smirnov tests. For ξ^{obs}, we obtain a p-value of 0.194. The analogous tests for μ^{obs} and σ^{obs} both yield p-values smaller than $2.2 \cdot 10^{-16}$. Thus, the hypotheses that the residuals of the estimates of μ^{obs} and σ^{obs} follow a normal distribution both can be rejected and, consequently, we stick to the assumption that the GEV parameters, location and scale, differ among the stations.
In contrast to the location and the scale parameter, the shape parameter of the GEV will be assumed to be spatially constant in northern Germany with the value

$$\xi^{\text{obs}} = \hat{\xi}^{\text{obs}} = \frac{1}{119} \sum_{i=1}^{119} \hat{\xi}^{\text{obs}}(l_i) = 0.043$$

(The empirical standard deviation of the sample $\{\hat{\xi}^{\text{obs}}(l_i) : \ i = 1, \ldots, 119\}$ is 0.049). Note, however, that the estimated shape parameter differs significantly (to a 5%-level) from the mean value in case of 20 stations. For six of these stations, it even differs highly significantly (to a 1%-level), and four of them even to a 0.1%-level. The parameter estimates $\hat{\mu}(l_i)$ and $\hat{\sigma}(l_i)$, $1 \leq i \leq 119$ for the location and scale parameters, respectively, obtained by maximum likelihood estimation with fixed shape parameter $\xi^{\text{obs}} = \hat{\xi}^{\text{obs}}$ are depicted in Figure 9.1a. Note that the estimated vectors of location and scale parameters show a strong empirical correlation of 0.97. By (9.3), the data can be transformed to standard Gumbel margins. Kolmogorov-Smirnov tests performed separately for each station yield p-values of at least 0.098 with a mean value of 0.718 which indicates that the GEV model fits quite well for all the stations.

As a fit of the GEV distribution to the forecast is needed for both verification of the marginal model and the bivariate Brown-Resnick model, we repeat our analysis replacing the observed maximal wind speed $v_{\text{max}}^{\text{obs}}(l_i, d)$ by $v_{\text{max}}^{\text{pred}}(l_i, d)$, i.e. a forecast for the maximal wind speed at station l_i and day d. Here, we use the maximum over the 20 corresponding COSMO-DE ensemble members

$$v_{\text{max}}^{\text{pred}}(l_i, d) = \max_{j=1,\ldots,20} v_{\text{max}}^{(j)}(l_i, d), \quad 1 \leq i \leq 119, \ 1 \leq d \leq 360,$$

which suggests that the distribution of $v_{\text{max}}^{\text{pred}}$ should be close to a GEV distribution. Note that this choice of $v_{\text{max}}^{\text{pred}}$ is in complete accordance to the choice of $\hat{m}(l_i, d)$ as maximal mean of all the ensemble members in Equation (9.20).

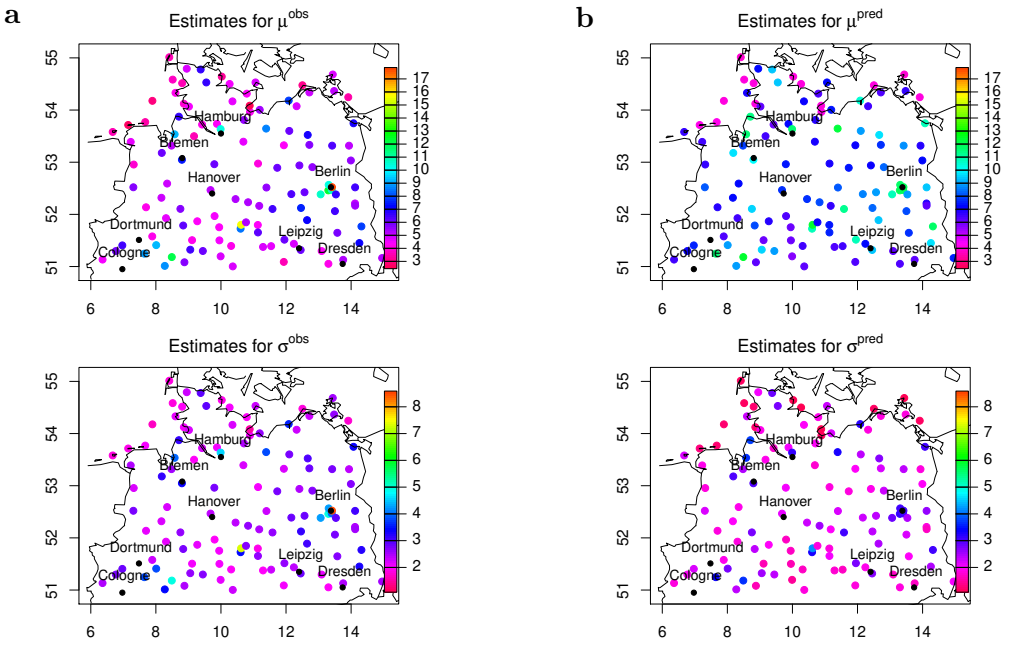

Figure 9.1: **a** Estimates $\hat{\mu}^{\text{obs}}(l_i)$ and $\hat{\sigma}^{\text{obs}}(l_i)$ for the location and scale parameters corresponding to the observed maximal wind speed at the stations in the northern part of Germany. **b** Estimates $\hat{\mu}^{\text{pred}}(l_i)$ and $\hat{\sigma}^{\text{pred}}(l_i)$ for the location and scale parameters corresponding to the forecasted maximal wind speed at the stations in the northern part of Germany.

As the Kolmogorov-Smirnov test of the normalized estimates for ξ^{pred} yields a p-value of 0.53 and the estimates differ significantly from the mean for seven stations (for three of them very significantly), we may assume a shape parameter of

$$\xi^{\text{pred}} = \hat{\xi}^{\text{pred}} = \frac{1}{119} \sum_{i=1}^{119} \xi^{\text{pred}}(l_i) = 0.028$$

(The empirical standard deviation of the sample $\{\hat{\xi}^{\text{pred}}(l_i) : \ i = 1, \ldots, 119\}$ is 0.044) at every station in northern Germany. However, the hypotheses that the estimates for the location and the scale parameter follow a normal distribution have been both rejected. The maximum likelihood estimates $\hat{\mu}^{\text{pred}}(l_i)$ and $\hat{\sigma}^{\text{pred}}(l_i)$, $1 \leq i \leq 119$, with fixed shape parameter are shown in Figure 9.1b. Here, the empirical correlation of the vectors of estimated location and scale parameters is just as strong as in case of the observations. Kolmogorov-Smirnov tests of the transformed data $x^{\text{pred}}(l_i, d)$ for every station yield p-values of at least 0.142 with and equal 0.748 in average which also indicates an appropriate fit.

The spatial dependence is modeled by a univariate Brown-Resnick process which is obtained by a weighted least squares fit of the extremal coefficient function. Here, the weights depend on the variance of the estimators $\hat{\vartheta}^{\text{obs}}(l_i, l_j)$ (see Section 9.4) estimated by a jackknife procedure where the extremal coefficients are reestimated leaving out one month of data. The estimated extremal coefficients $\hat{\theta}^{\text{obs}}$ and the fitted extremal coefficient function

$$\tilde{\theta}^{\text{obs}}(s,t) = 2\Phi\left(\sqrt{\frac{\gamma_{\hat{\vartheta}^{\text{obs}}}(s-t)}{2}}\right), \quad s,t \in \mathbb{R}^2.$$

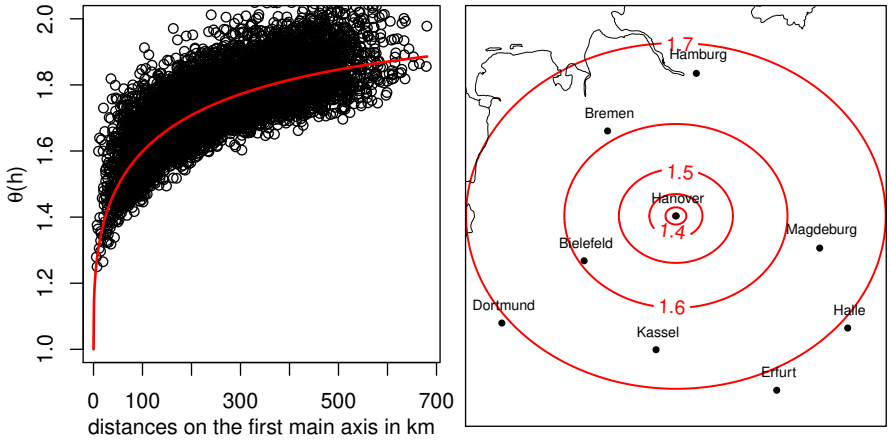

Figure 9.2: Left: The estimated extremal coefficients $\hat{\theta}^{\text{obs}}$ (black circles) and the fitted extremal coefficient function $\tilde{\theta}^{\text{obs}}$ (red line) of the normalized random field $X^{\text{obs}}(\cdot, d)$ of observed wind gusts. Right: Contour level plot of the fitted extremal coefficient function $\tilde{\theta}^{\text{obs}}(l_0, \cdot)$ where l_0 is located at Hanover.

are displayed in Figure 9.2. Here, the estimated coefficients seem to be fitted quite well.

For verification, we first calculate the mean CRPS for each of the two models given by (9.12), $\text{CRPS}^{\text{obs}}(l_i)$ and $\text{CRPS}^{\text{pred}}(l_i)$, for every station l_i, $1 \leq i \leq 119$. Then, the improvement or deterioration by using the GEV distributions of the observations instead of the forecasts is expressed in terms of the skill score (e.g., Gneiting and Raftery, 2007)

$$S(l_i) = 1 - \frac{\text{CRPS}^{\text{obs}}(l_i)}{\text{CRPS}^{\text{pred}}(l_i)}$$

which has the value 1 in case of an "optimal" model which equals $v_{\max}^{\text{obs}}$ a.s. and the value 0 if both models yield the same result. Here, $S_{l_i} > 0$ for 115 of 119 stations. For the skill score corresponding to the mean CRPS averaged over all the stations, we obtain

$$S = 1 - \frac{\sum_{i=1}^{119} \text{CRPS}_{l_i}^{\text{obs}}}{\sum_{i=1}^{119} \text{CRPS}_{l_i}^{\text{pred}}} \approx 0.293.$$

Note that, for simplicity, the reference model (9.12) for the predictions is based on the maximal ensemble members $v_{\max}^{\text{pred}}(l_i, d)$ only and further information given by the maximal wind speed forecasted by the other ensemble members are neglected. Thus, we further compare the CRPS of the GEV model for the observations, $\text{CRPS}^{\text{obs}}(l_i)$, with the CRPS of the original COSMO-DE ensemble, $\text{CRPS}^{\text{orig}}(l_i)$, taking the ensemble forecast as a probabilistic forecast with equal probability for each ensemble member. Here, the skill $\tilde{S}(l_i) = 1 - \text{CRPS}^{\text{obs}}(l_i)/\text{CRPS}^{\text{orig}}(l_i)$ is positive for 37 of 119 only, with the skill of the averaged CRPS being approximately -0.032. As the skill score is slightly negative, the COSMO-DE ensemble forecast seems to contain more information than our marginal model.

Note that, for a fair comparison, we should avoid the validation of our model on the same data that have been used for the model fit. Hence, we perform cross validation: Separately for every month, the GEV parameters are reestimated leaving out the data for this month and using only the data for the other eleven months for the model fit.

The GEV parameters estimated for different months in this way show very little variation corroborating the assumption that they are constant in time. Further, the verification results above are confirmed: We obtain skill scores of 0.285 for the CRPS compared with the GEV model for the forecast and -0.048 compared to the COSMO-DE ensemble.

9.6.3 Applying the Bivariate Model

A bivariate Brown-Resnick process is fitted to the transformed data according to Section 9.4. Here, as a preliminary analysis suggests, the parameter ρ is set to the maximal value yielding a valid valid variogram, i.e.

$$\rho = \frac{(1+\nu_{12}^{-1})^{\nu_{12}+1}}{(1+\nu_1^{-1})^{\frac{1}{2}\nu_1+\frac{1}{2}}(1+\nu_2^{-1})^{\frac{1}{2}\nu_2+\frac{1}{2}}}$$

The estimate $\hat{\vartheta}$ of the remaining eleven pseudo cross-variogram parameters leads to the fitted extremal coefficient function

$$\begin{aligned}\tilde{\theta}(l_i,l_j) &= \left(\tilde{\theta}^{k_1,k_2}(l_i,l_j)\right)_{k_1,k_2\in\{\text{“obs”},\text{“pred”}\}} \\ &= 2\left(\Phi\left(\sqrt{\frac{\gamma_{k_1k_2}(\hat{\vartheta};l_i-l_j)}{2}}\right)\right)_{k_1,k_2\in\{\text{“obs”},\text{“pred”}\}}.\end{aligned}$$

Figure 9.3 presents the estimated extremal coefficients $\hat{\theta}^{k_1,k_2}(l_i,l_j)$, and the fitted extremal coefficient functions $\tilde{\theta}^{k_1,k_2}(\cdot,\cdot)$ for $k_1,k_2 \in \{\text{“obs”}, \text{“pred”}\}$. As illustrated, the fitted model seems to be appropriate with respect to the behavior of the extremal coefficient function. Figure 9.4 depicts a simulated realization of the corresponding Brown-Resnick process associated to the variogram $\gamma(\hat{\vartheta};\cdot)$ with standard Gumbel margins. The realization indicates a remarkable amount of positive correlation between x^{obs} and x^{pred} which emphasizes the gain of information by taking x^{pred} into account.
In order to verify the bivariate model, we apply the post-processing procedure proposed in Subsection 9.5.1. Due to the computational complexity of the conditional simulation, we do not simulate the observations at all stations simultaneously conditional on the forecast at all locations, but perform post-processing with sample size $K = 20$, i.e. the size of the original COSMO-DE ensemble, at each location separately conditioning on the forecast at the same location and two neighboring grid cells only. We calculate the CRPS of the post-processed distribution, $\text{CRPS}^{\text{biv}}(l_i)$, and compare it with $\text{CRPS}^{(NWP)}(l_i)$, i.e. the CRPS belonging to the empirical distribution of the COSMO-DE ensemble, yielding a positive skill score for 82 of 119 stations where the skill score related to the mean CRPS equals 0.128 (0.111 cross-validated). If we increase the sample size $K = 100$, we obtain an improved skill score of 0.164 (0.147 cross-validated), being positive for 104 of 119 stations. Thus, we conclude that the post-processing procedure based on the bivariate Brown-Resnick model is able to improve the forecast given by COSMO-DE ensemble.

9.6.4 Uncertainty Assessment

We assess the uncertainty in the estimation of the model parameters via parametric bootstrap, i.e. we simulate data sets from the fitted model and repeat the estimation procedure. As a detailed analysis of the numerical model producing the forecasts is beyond the scope of this chapter, we do not account for the uncertainty in the estimates $\hat{m}(l_i,d)$ and $\hat{s}(l_i,d)$,

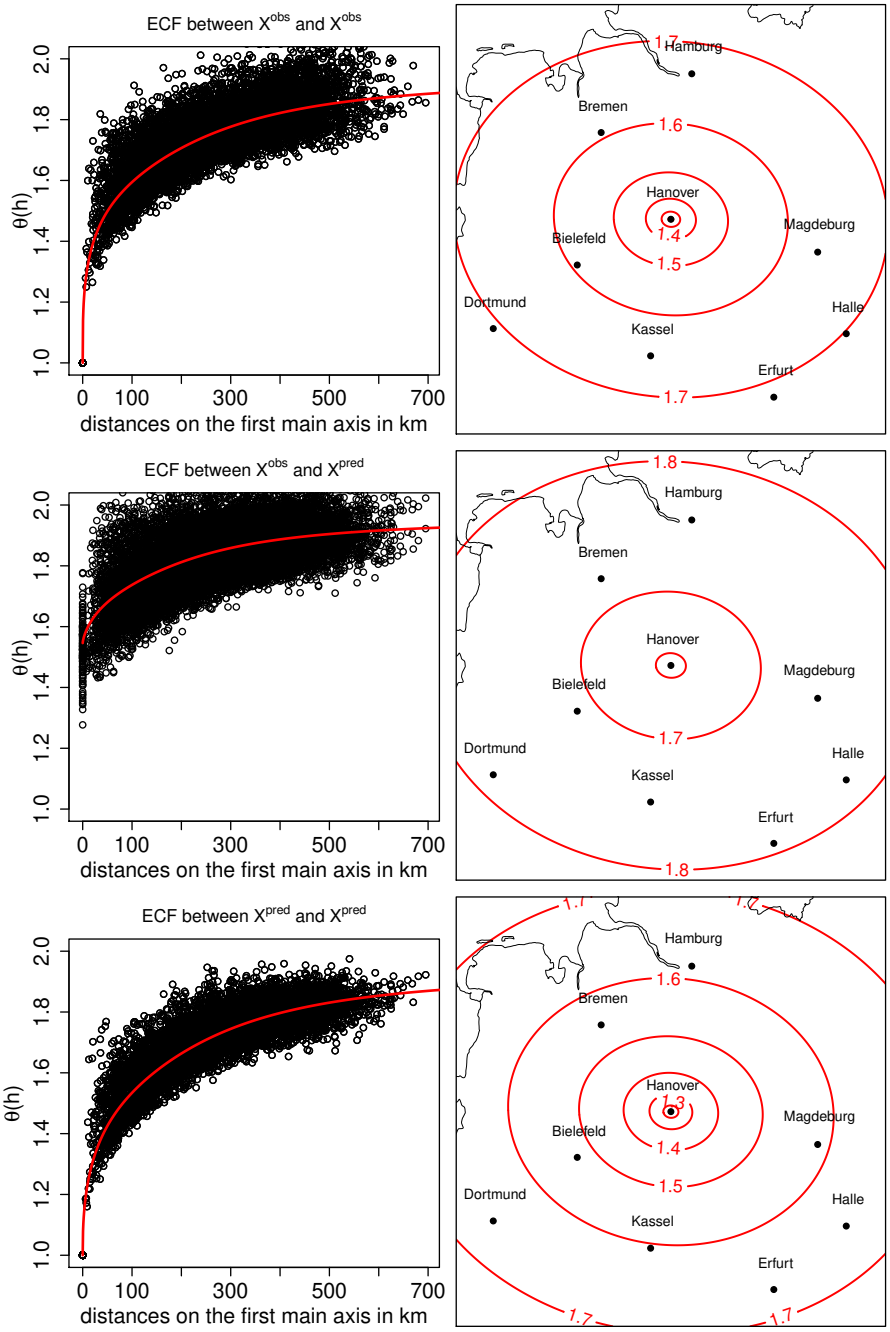

Figure 9.3: Left: The estimated extremal coefficients (black circles) and the fitted extremal coefficient function (red line) of the normalized bivariate random field $(X^{\text{obs}}, X^{\text{pred}})$ of observed and forecasted wind gusts. Right: Contour level plots of the fitted extremal coefficient function $\tilde{\theta}(l_0, \cdot)$ where l_0 is located at Hanover.

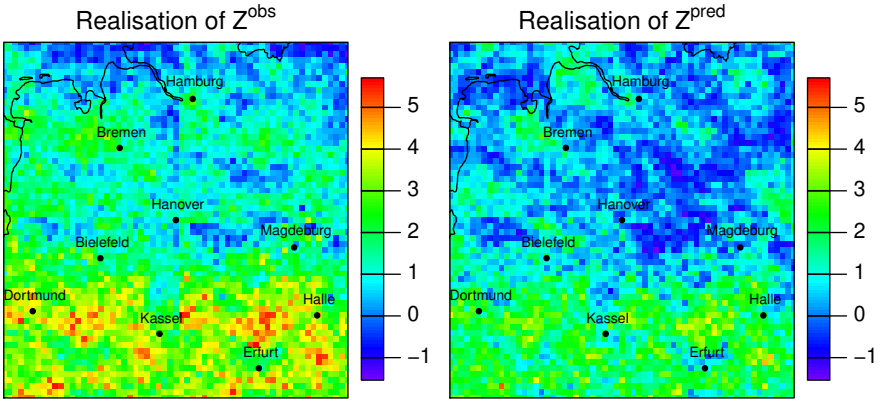

Figure 9.4: Simulated realization of a Brown-Resnick process associated to the variogram $\gamma(\vartheta;\cdot)$ with standard Gumbel margins.

	b	θ	β	σ	κ	c	σ_1	ν_1	σ_2	ν_2	a
orig.	1.2	-0.10	0.91	2.0	425	1.5	1.3	0.3	1.1	0.4	124
mean	1.2	0.05	0.72	1.3	307	1.5	1.4	0.3	1.2	0.5	306
std. dev.	0.2	0.36	0.23	0.8	214	0.1	0.4	0.5	0.4	0.7	283

Table 9.1: Results for parameters of the bivariate Brown–Resnick model obtained via parametric bootstrap

but focus on the normalized values

$$\left\{ \frac{v_{\max}^{\text{obs}}(l_i,d) - \hat{m}(l_i,d)}{\hat{s}(l_i,d)}, \frac{v_{\max}^{\text{obs}}(l_i,d) - \hat{m}(l_i,d)}{\hat{s}(l_i,d)} : \; i = 1,\dots,119, \; d = 1,\dots,360 \right\}.$$

To this end, we draw 360 independent realizations from the bivariate Brown-Resnick process fitted in Subsections 9.6.2 and 9.6.3 using the simulation algorithm by Dombry et al. (2016a). The procedure is repeated 100 times yielding 100 independent data sets of the same size as the original one. Following the steps described in Subsections 9.6.2 and 9.6.3, we thus obtain 100 independent estimates.
The sample of estimates $\widehat{\xi}^{\text{obs}}$ has mean 0.043 and a standard deviation of 0.02. In order to validate the p-value of the Kolmogorov-Smirnov tests for ξ^{obs}, we repeat these tests on the simulated data sets and obtain a p-value smaller than the original one (0.194) in 22 of 100 cases which supports the non-rejection of the hypothesis that ξ^{obs} is spatially constant. The p-values of the tests for μ^{obs} and σ^{obs} are confirmed, as well.
Analogously, the results for the marginal parameters for the forecast are verified: The sample of estimates $\widehat{\xi}^{\text{pred}}$ has mean 0.03 and a standard deviation of 0.02 while the original p-value (0.53) of the test for ξ^{obs} is undercut in 62 of 100 cases.

Further, we assess the estimation of the Brown-Resnick model parameters. The original parameter values and the sample means and standard deviations obtained from the parametric bootstrap are presented in Table 9.1. It can be seen that most parameters are recovered well by the estimation procedure. For some parameters of the bivariate Whittle-Matérn model, however, rather large variances are observed, a phenomenon which is often encountered in practice in accordance to the fact that not all parameters of the Whittle-Matérn model can be consistently estimated from observations in a fixed domain (cf.

Zhang, 2004). Even though some parameters show considerable variation, there is little variation in the bivariate extremal coefficient function as displayed in Figure 9.5.
Besides the uncertainty of the parameter estimates, we also assess the deviation of the non-parametrically estimated extremal coefficients from the parametrically estimated extremal coefficient function. To this end, for $k_1, k_2 \in \{\text{“obs”}, \text{“pred”}\}$, we calculate the root-mean-square error

$$\mathrm{RMSE}^{k_1,k_2} = \left[\frac{1}{119^2} \sum_{1 \le i,j \le 119} \left(\hat{\theta}^{k_1,k_2}(l_i, l_j) - 2\Phi(\sqrt{\gamma_{k_1,k_2}(\hat{\vartheta}; l_i, l_j)/2}) \right)^2 \right]^{1/2}.$$

For the original data, the corresponding root-mean-square errors are $\mathrm{RMSE}^{\mathrm{obs,obs}} = 0.074$, $\mathrm{RMSE}^{\mathrm{obs,pred}} = \mathrm{RMSE}^{\mathrm{pred,obs}} = 0.078$ and $\mathrm{RMSE}^{\mathrm{pred,pred}} = 0.061$. Repeating the calculation for the simulated data sets, we obtain sample mean root-mean-square errors 0.039 (for $\mathrm{RMSE}^{\mathrm{obs,obs}}$), 0.041 (for $\mathrm{RMSE}^{\mathrm{obs,pred}}$ and $\mathrm{RMSE}^{\mathrm{pred,obs}}$) and 0.037 (for $\mathrm{RMSE}^{\mathrm{pred,pred}}$) with standard deviations 0.0018, 0.0019 and 0.0018, respectively. Thus, the deviations for the real data are roughly twice as large as expected in case of the model being correct. One may conclude, that the extremal coefficient for real data is not only a function of the distance, but may be modulated by the topography or the climate at the stations.

Finally, note that the same methodology could be used to assess the uncertainty of the post-processed forecast. To this end, different sets of estimated parameters could be used as input for the post-processing procedure described in Subsection 9.5.1 to determine the uncertainty of its output. However, a full analysis would also require an assessment of the uncertainty in the estimates $\hat{m}(l_i, d)$ and $\hat{s}(l_i, d)$ which, as already mentioned above, is beyond the scope of this work.

Acknowledgments

This research has been financially supported by VolkswagenStiftung within the project “Mesoscale Weather Extremes – Theory, Spatial Modeling and Prediction (WEX-MOP)”. The research of M. Oesting has also partially been funded by the ANR project ‘McSim’. Observational data and COSMO-DE-EPS forecasts have been kindly provided by Deutscher Wetterdienst in Offenbach, Germany. The authors are grateful to the Associate Editor and an anonymous referee for their valuable suggestions improving this article.

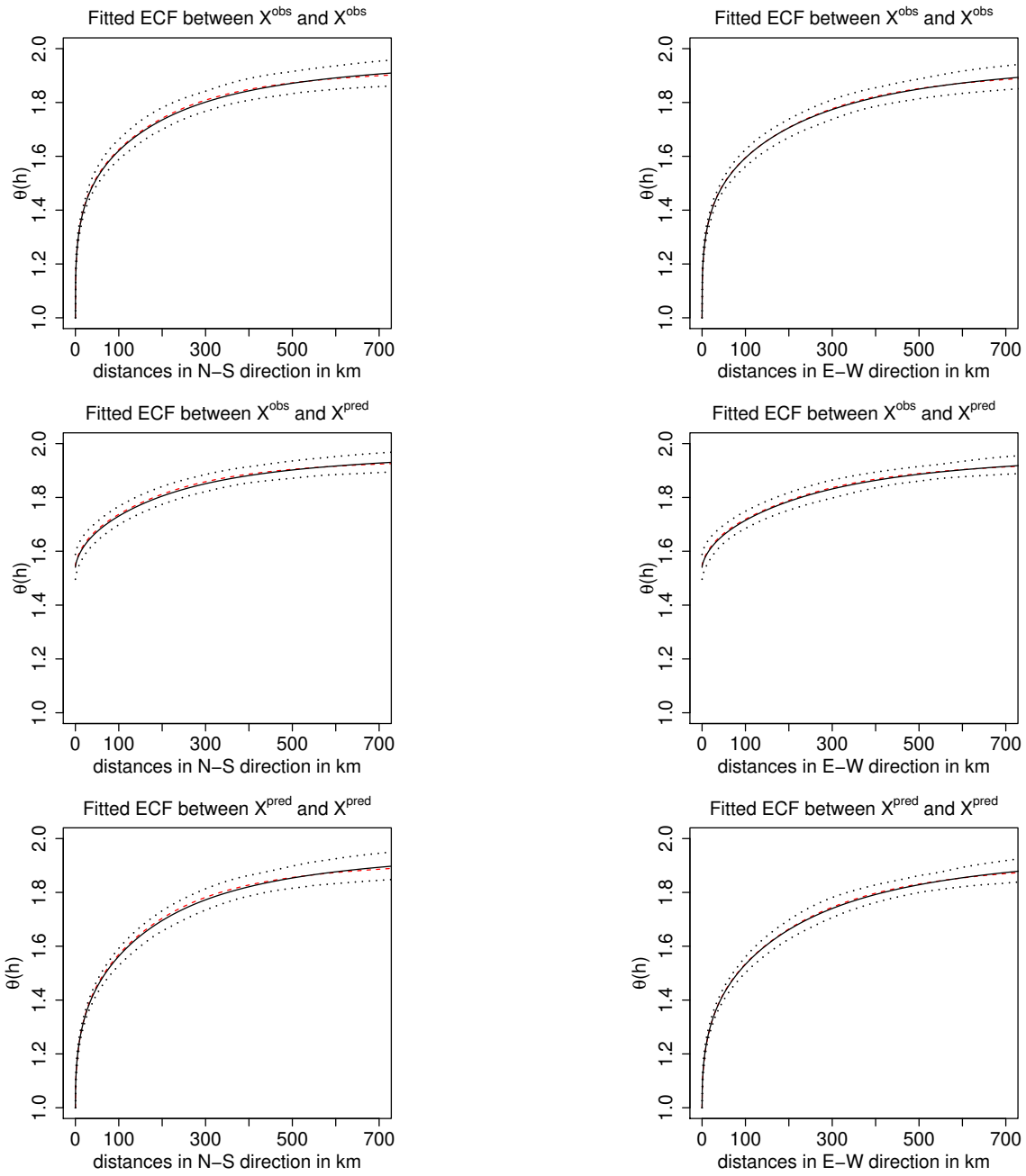

Figure 9.5: Results for the bivariate extremal coefficient functions obtained via parametric bootstrap as functions of the distance in north-south (left) and east-west (right) direction. The dashed lines are the true extremal coefficient functions while the solid lines gives the pointwise bootstrap sample mean of the fitted extremal coefficients. The dotted lines indicate the pointwise 0.025 and 0.975-quantiles, respectively.

Bibliography

Asadi, P., Davison, A. C., and Engelke, S. (2015). Extremes on river networks. *Ann. Appl. Stat.*, 9(4), 2023–2050.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T. (2011). Operational convective-scale numerical weather prediction with the COSMO model. *Mon. Wea. Rev.*, 139(12), 3887–3905.

Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *Ann. Probab.*, 2(5), 792–804.

Bechler, A., Vrac, M., and Bel, L. (2015). A spatial hybrid approach for downscaling of extreme precipitation fields. *J. Geophys. Res.*, 120(10), 4534–4550.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Chichester: John Wiley & Sons.

Berg, C., Christensen, J. P. R., and Ressel, P. (1984). *Harmonic Analysis on Semigroups*. New York: Springer-Verlag.

Blanchet, J. and Davison, A. C. (2011). Spatial Modeling of Extreme Snow Depth. *Ann. Appl. Stat.*, 5(3), 1699–1725.

Boldi, M.-O. (2009). A note on the representation of parametric models for multivariate extremes. *Extremes*, 12(3), 211–218.

Boldi, M.-O. and Davison, A. C. (2007). A mixture model for multivariate extremes. *J. R. Stat. Soc. B.*, 69(2), 217–229.

Brasseur, O. (2001). Development and application of a physical approach to estimating wind gusts. *Mon. Wea. Rev.*, 129(1), 5–25.

Brown, B. M. and Resnick, S. I. (1977). Extreme Values of Independent Stochastic Processes. *J. Appl. Probab.*, 14(4), 732–739.

Buishand, T., de Haan, L., and Zhou, C. (2008). On spatial extremes: With application to a rainfall problem. *Ann. Appl. Stat.*, 2(2), 624–642.

Capéraà, P., Fougères, A.-L., and Genest, C. (2000). Bivariate distributions with given extreme value attractor. *J. Multivariate Anal.*, 72, 30–49.

Castruccio, S., Huser, R., and Genton, M. G. (2016). High-order composite likelihood inference for max-stable distributions and processes. *J. Comput. Graph. Stat.*, 25(4), 1212–1229.

Chandler, R. E. and Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, 94(1), 167–183.

Chilès, J.-P. and Delfiner, P. (2012). *Geostatistics. Modeling Spatial Uncertainty.* Hoboken: John Wiley & Sons, second edition.

Clark, I., Basinger, K. L., and Harper, W. V. (1989). MUCK – a novel approach to cokriging. In *Geostatistical, sensitivity, and uncertainty methods for ground-water flow and radionuclide transport modeling.* San Francisco: Battelle Press.

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

Coles, S. G. (1993). Regional Modelling of Extreme Storms via max-stable processes. *J. R. Stat. Soc. B.*, 55(4), 797–816.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *J. R. Stat. Soc. B.*, 53(2), 377–392.

— (1996). Modelling extremes of the areal rainfall process. *J. R. Stat. Soc. B.*, 58(2), 329–347.

Conradsen, K., Nielsen, L. B., and Prahm, L. P. (1984). Review of Weibull statistics for estimation of wind speed distributions. *J. Clim. Appl. Meteorol.*, 23, 1173–1183.

Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. In Bertail, P., Doukhan, P., and Soulier, P., editors, *Dependence in Probability and Statistics*, pp. 373–390. New York: Springer-Verlag.

Cressie, N. A. C. (1993). *Statistics for Spatial Data.* New York: John Wiley & Sons.

Davison, A. C. and Gholamrezaee, M. M. (2012). Geostatistics of extremes. *Proc. R. Soc. Lond. A*, 468(2138), 581–608.

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical Modeling of Spatial Extremes (with discussion). *Stat. Sci.*, 27(2), 161–186.

de Fondeville, R. and Davison, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika*, 105(3), 575–592.

de Haan, L. (1984). A spectral representation for max-stable processes. *Ann. Probab.*, 12(4), 1194–1204.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction.* Berlin: Springer.

de Haan, L. and Lin, T. (2001). On convergence toward an extreme value distribution in C[0,1]. *Ann. Probab.*, 29(1), 467–483.

Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models.* Hoboken: John Wiley & Sons.

Devroye, L. (1986). *Non-Uniform Random Variate Generation.* New York: Springer-Verlag.

Dieker, A. B. and Mikosch, T. (2015). Exact simulation of Brown-Resnick random fields at a finite number of locations. *Extremes*, 18(2), 301–314.

Dombry, C., Engelke, S., and Oesting, M. (2016a). Exact Simulation of Max-Stable Processes. *Biometrika*, 103(2), 303–317.

— (2017a). Asymptotic properties of the maximum likelihood estimator for multivariate extreme value distributions. Available from `https://arxiv.org/abs/1612.05178v2`.

— (2017b). Bayesian inference for multivariate extreme value distributions. *Electron. J. Stat.*, 11(2), 4813–4844.

Dombry, C. and Éyi-Minko, F. (2012). Strong mixing properties of max-infinitely divisible random fields. *Stoch. Proc. Appl.*, 122(11), 3790–3811.

— (2013). Regular conditional distributions of continuous max-infinitely divisible random fields. *Electron. J. Probab.*, 18(7), 1–21.

Dombry, C., Éyi-Minko, F., and Ribatet, M. (2013). Conditional simulation of max-stable processes. *Biometrika*, 100(1), 111–124.

Dombry, C., Oesting, M., and Ribatet, M. (2016b). Conditional Simulation of Max-Stable Processes. In Dey, D. K. and Yan, J., editors, *Extreme Value Modeling and Risk Analysis: Methods and Applications*, pp. 215–238. Boca Raton: CRC Press.

Dombry, C. and Ribatet, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Stat. Interface*, 8(1), 9–17.

Einmahl, J. H. J., Kiriliouk, A., Krajina, A., and Segers, J. (2016). An M–estimator of spatial tail dependence. *J. R. Stat. Soc. B.*, 78(1), 275–298.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin Heidelberg: Springer-Verlag.

Engelke, S., De Fondeville, R., and Oesting, M. (2018). Extremal behaviour of aggregated data with an application to downscaling. *Biometrika*, 106(1), 127–144.

Engelke, S. and Ivanovs, J. (2017). Robust bounds in multivariate extremes. *Ann. Appl. Probab.*, 27(6), 3706–3734.

Engelke, S., Kabluchko, Z., and Schlather, M. (2011). An equivalent representation of the Brown–Resnick process. *Stat. Probabil. Lett.*, 81(8), 1150–1154.

Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *J. R. Stat. Soc. B.*, 77, 239–265.

Engelke, S., Malinowski, A., Oesting, M., and Schlather, M. (2014). Statistical inference for max-stable processes by conditioning on extreme events. *Adv. Appl. Probab.*, 46(2), 478–495.

Ferreira, A. and de Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4), 1717–1737.

Fisher, R. A. and 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(2), 180–190.

Fougères, A.-L., Mercadier, C., and Nolan, J. P. (2013). Dense classes of multivariate extreme value distributions. *J. Multivar. Anal.*, 116, 109–129.

Fougères, A.-L., Nolan, J. P., and Rootzén, H. (2009). Models for dependent extremes using stable mixtures. *Scand. J. Stat.*, 36(1), 42–59.

Friederichs, P. and Thorarinsdottir, T. L. (2012). Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction. *Environmetrics*, 23(7), 579–594.

Gaume, J., Eckert, N., Chambon, G., Naaim, M., and Bel, L. (2013). Mapping extreme snowfalls in the French Alps using max-stable processes. *Water Resour. Res.*, 49(2), 1079–1098.

Gebhardt, C., Theis, S., Paulat, M., and Ben Bouallègue, Z. (2011). Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmos. Res.*, 100(2), 168–177.

Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M. (2010). *Handbook of Spatial Statistics*. Boca Raton: CRC Press.

Genton, M. G., Ma, Y., and Sang, H. (2011). On the likelihood function of Gaussian max-stable processes. *Biometrika*, 98(2), 481–488.

Genton, M. G., Padoan, S. A., and Sang, H. (2015). Multivariate max-stable spatial processes. *Biometrika*, 102(1), 215–230.

Ghoudi, K., Khoudraji, A., and Rivest, L.-P. (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *Can. J. Stat.*, 26, 187–197.

Giné, E., Hahn, M., and Vatan, P. (1990). Max-infinitely divisible and max-stable sample continuous processes. *Probab. Th. Rel. Fields*, 87, 139–165.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Ann. Math.*, 44(3), 423–453.

Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Am. Stat. Assoc.*, 105(491), 1167–1177.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.*, 102(477), 359–378.

Goldfarb, D. and Idnani, A. (1983). A numerically stable dual method for solving strictly convex quadratic programs. *Math. Program.*, 27(1), 1–33.

Gradshteyn, I. S. and Ryzhik, I. (2007). *Table of Integrals, Series, and Products*. Amsterdam: Elsevier/Academic Press, 7th edition.

Gudendorf, G. and Segers, J. (2010). Extreme-value copulas. In Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., editors, *Copula Theory and its Applications*, pp. 127–145. Berlin: Springer-Verlag.

Gumbel, E. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Stat. Univ. Paris*, 9, 171–173.

Guttorp, P. and Gneiting, T. (2006). Studies in the history of probability and statistics XLIX. On the Matérn correlation family. *Biometrika*, 93(4), 989–995.

Handcock, M. S. and Wallis, J. R. (1994). An Approach to Statistical Spatial-Temporal Modeling of Meteorological Fields. *J. Am. Stat. Assoc.*, 89(426), 368–378.

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1952). *Inequalities*. Cambridge: Cambridge University Press, second edition.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109.

Ho, Z. W. O. and Dombry, C. (2017). Simple models for multivariate regular variations and the Hüsler–Reiss Pareto distribution. ArXiv preprint arXiv:1712.09225.

Hofert, M. and Mächler, M. (2011). Nested Archimedean copulas meet R: The nacopula package. *J. Stat. Softw.*, 39(9), 1–20.

Huser, R. and Davison, A. (2013). Composite likelihood estimation for the Brown-Resnick process. *Biometrika*, 100, 511–518.

Huser, R. and Davison, A. C. (2014). Space–time modelling of extreme events. *J. R. Stat. Soc. B.*, 76(2), 439–461.

Huser, R., Davison, A. C., and Genton, M. G. (2016). Likelihood estimators for multivariate extremes. *Extremes*, 19(1), 79–103.

Huser, R., Dombry, C., Ribatet, M., and Genton, M. G. (2019). Full likelihood inference for max-stable data. *Stat*, 8(1), e218.

Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: between independence and complete dependence. *Stat. Probabil. Lett.*, 7, 283–286.

Kabluchko, Z. (2009). Spectral representations of sum-and max-stable processes. *Extremes*, 12(4), 401–424.

— (2011). Extremes of independent Gaussian processes. *Extremes*, 14(3), 285–310.

Kabluchko, Z. and Schlather, M. (2010). Ergodic properties of max-infinitely divisible processes. *Stoch. Proc. Appl.*, 120(3), 281–295.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary Max-Stable Fields Associated to Negative Definite Functions. *Ann. Probab.*, 37(5), 2042–2065.

Kabluchko, Z. and Stoev, S. (2016). Stochastic integral representations and classification of sum-and max-infinitely divisible processes. *Bernoulli*, 22(1), 107–142.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *J. Am. Stat. Assoc.*, 90, 773–795.

Kingman, J. F. C. (1993). *Poisson Processes*. Oxford: Oxford University Press.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer-Verlag.

Liu, Z., Blanchet, J. H., Dieker, A. B., and Mikosch, T. (2019+). On logarithmically optimal exact simulation of max-stable and related random fields on a compact set. *Bernoulli*. To appear.

Mengersen, K. L. and Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Stat.*, 24(1), 101–121.

Molchanov, I. (2008). Convex geometry of max-stable distributions. *Extremes*, 11(3), 235–259.

Molchanov, I. and Stucki, K. (2013). Stationarity of multivariate particle systems. *Stochastic Process. Appl.*, 123(6), 2272–2285.

Nikoloulopoulos, A. K., Joe, H., and Li, H. (2009). Extreme value properties of multivariate t copulas. *Extremes*, 12(2), 129–148.

Northrop, P. J. and Attalides, N. (2016). Posterior propriety in Bayesian extreme value analyses using reference priors. *Stat. Sin.*, 26(2), 721–743.

Oesting, M. (2015). On the distribution of a max-stable process conditional on max-linear functionals. *Stat. Probabil. Lett.*, 100, 158–163.

— (2018). Equivalent representations of max-stable processes via ℓ^p-norms. *J. Appl. Probab.*, 55(1), 54–68.

Oesting, M., Bel, L., and Lantuéjoul, C. (2018a). Sampling from a Max-Stable Process Conditional on a Homogeneous Functional with an Application for Downscaling Climate Data. *Scand. J. Stat.*, 45(2), 382–404.

Oesting, M., Kabluchko, Z., and Schlather, M. (2012). Simulation of Brown-Resnick Processes. *Extremes*, 15(1), 89–107.

Oesting, M. and Schlather, M. (2014). Conditional Sampling for max-stable processes with a mixed moving maxima representation. *Extremes*, 17(1), 157–192.

Oesting, M., Schlather, M., and Friederichs, P. (2017). Statistical post-processing of forecasts for extremes using bivariate Brown–Resnick processes with an application to wind gusts. *Extremes*, 20(2), 309–332.

Oesting, M., Schlather, M., and Schillings, C. (2019). Sampling Sup-Normalized Spectral Functions for Brown–Resnick Processes. *Stat*, 8(1), e228.

Oesting, M., Schlather, M., and Zhou, C. (2013). On the Normalized Spectral Representation of Max-Stable Processes on a Compact Set. Technical Report.
Available from `http://arxiv.org/abs/1310.1813v1`; basis for Oesting et al. (2018b).

— (2018b). Exact and Fast Simulation of Max-Stable Processes on a Compact Set Using the Normalized Spectral Representation. *Bernoulli*, 24(2), 1497–1530.

Oesting, M. and Strokorb, K. (2018). Efficient simulation of Brown–Resnick processes based on variance reduction of Gaussian processes. *Adv. in Appl. Probab.*, 50(4), 1155–1175.

— (2019). A comparative tour through the simulation algorithms for max-stable processes. Available from `https://arxiv.org/abs/1809.09042`.

Opitz, T. (2013). Extremal t processes: Elliptical domain of attraction and a spectral representation. *J. Multivar. Anal.*, 122, 409–413.

Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010). Likelihood-Based Inference for Max-Stable Processes. *J. Am. Stat. Assoc.*, 105(489), 263–277.

Papritz, A., Künsch, H., and Webster, R. (1993). On the pseudo cross-variogram. *Math. Geol.*, 25(8), 1015–1026.

Pavia, E. G. and O'Brien, J. J. (1986). Weibull statistics of wind speed over the ocean. *J. Clim. Appl. Meteorol.*, 25, 1324–1332.

Penrose, M. D. (1992). Semi-min-stable processes. *Ann. Probab.*, 20(3), 1450–1463.

Peralta, C., Ben Bouallégue, Z., Theis, S. E., Gebhardt, C., and Buchhold, M. (2012). Accounting for initial condition uncertainties in COSMO-DE-EPS. *J. Geophys. Res.*, 117, D07108.

Pickands, J. I. (1975). Statistical inference using extreme order statistics. *Ann. Stat.*, 3(1), 119–131.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`.

Ramos, H. M., Ollero, J., and Sordo, M. A. (2000). A sufficient condition for generalized Lorenz order. *J. Econ. Theory*, 90(2), 286–292.

Reich, B. J. and Shaby, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. *Ann. Appl. Stat.*, 6(4), 1430–1451.

Reich, B. J., Shaby, B. A., and Cooley, D. (2014). A hierarchical model for serially-dependent extremes: A study of heat waves in the western US. *J. Agric. Biol. Envir. S.*, 19(1), 119–135.

Resnick, S. and Roy, R. (1990). Multivariate extremal processes, leader processes and dynamic choice models. *Adv. in Appl. Probab.*, 22(2), 309–331.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. New York: Springer-Verlag.

Resnick, S. I. and Roy, R. (1991). Random usc functions, max-stable processes and continuous choice. *Ann. Appl. Probab.*, 1(2), 267–292.

Ressel, P. (2013). Homogeneous distributions – And a spectral representation of classical mean values and stable tail dependence functions. *J. Multivar. Anal.*, 117, 246–256.

Ribatet, M. (2013). Spatial extremes: Max-stable processes at work. *J. Soc. Fr. Stat.*, 154(2), 156–177.

Ribatet, M., Cooley, D., and Davison, A. C. (2012). Bayesian Inference from Composite Likelihoods, with an Application to Spatial Extremes. *Stat. Sin.*, 22(2), 813–845.

Rootzén, H., Segers, J., and Wadsworth, J. L. (2018). Multivariate peaks over thresholds models. *Extremes*, 21(1), 115–145.

Rootzén, H. and Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, 12(5), 917–930.

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. New York: Chapman & Hall.

Schilling, R. L., Song, R., and Vondraček, Z. (2010). *Bernstein Functions: Theory and Applications*. Berlin: Gruyter.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1), 33–44.

Schlather, M., Malinowski, A., Oesting, M., Boecker, D., Strokorb, K., Engelke, S., Martini, J., Ballani, F., Moreva, O., Menck, P., Gross, S., Ober, U., Ribeiro, P., Singleton, R., Pfaff, B., and R Core Team (2019). *RandomFields: Simulation and Analysis of Random Fields*. R package version 3.3.1, URL `http://CRAN.R-project.org/package=RandomFields`.

Schlather, M. and Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90(1), 139–156.

Sebille, Q., Fougères, A.-L., and Mercadier, C. (2017). Modeling extreme rainfall: A comparative study of spatial extreme value models. *Spat. Stat.*, 21(Part A), 187–208.

Shaby, B. A. and Reich, B. J. (2012). Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. *Environmetrics*, 23(8), 638–648.

Shi, D. (1995). Fisher Information for a Multivariate Extreme Value Distribution. *Biometrika*, 82(3), 644–649.

Sloughter, J. M. L., Raftery, A. E., Gneiting, T., and Fraley, C. (2007). Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Wea. Rev.*, 135, 3209–3220.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1), 67–90.

— (1990). Max–Stable Processes and Spatial Extremes. Unpublished manuscript.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer-Verlag.

Stephenson, A. (2003). Simulating Multivariate Extreme Value Distributions of Logistic Type. *Extremes*, 6(1), 49–59.

Stephenson, A. and Tawn, J. (2004). Bayesian Inference for Extremes: Accounting for the Three Extremal Types. *Extremes*, 7(4), 291–307.

— (2005). Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika*, 92(1), 213–227.

Stephenson, A. G. (2009). High-dimensional parametric modelling of multivariate extreme events. *Aust. N.Z. J. Stat.*, 51(1), 77–88.

Stephenson, A. G., Shaby, B. A., Reich, B. J., and Sullivan, A. L. (2015). Estimating spatially varying severity thresholds of a forest fire danger rating system using max-stable extreme-event modeling. *J. Appl. Meteor. Climatol.*, 54(2), 395–407.

Stoev, S. A. (2008). On the ergodicity and mixing of max-stable processes. *Stoch. Proc. Appl.*, 118(9), 1679–1705.

Stoev, S. A. and Taqqu, M. S. (2005). Extremal stochastic integrals: a parallel between max-stable processes and α-stable processes. *Extremes*, 8(4), 237–266.

Strokorb, K., Ballani, F., and Schlather, M. (2015). Tail correlation functions of max-stable processes. *Extremes*, 18(2), 241–271.

Thibaud, E., Aalto, J., Cooley, D. S., Davison, A. C., and Heikkinen, J. (2016). Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures. *Ann. Appl. Stat.*, 10(4), 2303–2324.

Thibaud, E. and Opitz, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102(4), 855–870.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Stat.*, 22(4), 1701–1728.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Wadsworth, J. L. (2015). On the occurrence times of componentwise maxima and bias in likelihood inference for multivariate max-stable distributions. *Biometrika*, 102(3), 705–711.

Wadsworth, J. L. and Tawn, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, 101(1), 1–15.

Wang, Y. and Stoev, S. (2010). On the structure and representations of max-stable processes. *Adv. Appl. Probab.*, 42(3), 855–877.

Wang, Y. and Stoev, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Adv. Appl. Probab.*, 43(2), 461–483.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Am. Stat. Assoc.*, 99(465), 250–261.