Max-stable processes for spatial extremes

Kirstin Strokorb and Marco Oesting

Max-stable processes are an analogue of max-stable distributions in a spatial setting. They are a mathematically motivated limiting process used to approximate block-maxima in time over space. In this capacity, they have become a first benchmark when answering questions on extremes that have a geostatistical nature. We summarize major developments of the last decades, covering theoretical properties, models and tools for inference. An indicative application demonstrates practical uses alongside hints where caution is needed. We close with some concluding remarks.

1 Introduction

Questions about the frequency and severity of adverse rare events come up naturally within the design of safety-critical infrastructure or contingency planning. And often they will not be limited to one or several variables of interest, but may have a geostatistical nature. Consider for instance the following two hypothetical situations:

Joint temperature exceedance Health officials are concerned about the probability that three regions with a particularly vulnerable population experience a temperature above 40°C within the same fortnight in summer. Current emergency resource plans are only fit for purpose if at most two of the three regions were affected at the same time. It is advised that plans need to be adapted if there is a higher than 2.5% chance of such an event during one summer.

Wind speeds across a bridge The construction manager of a large bridge with design life of 150 years needs advice about the strength of wind gusts which the bridge will be exposed to. Assuming that wind speeds are relatively heterogeneous across time, what would be the daily maximum wind speed across the entire bridge that one would expect to occur once in 1000 years? The regulatory framework demands that the bridge should be able to withstand such an event.

Both situations have in common that variables of interest (wind speeds or temperatures) can only be measured for a few scattered sites within the region of interest or closeby at a certain frequency. It is likely that in the past data have only been collected systematically at even fewer sites than exist today or even less frequently than today or are available only in aggregated form across time or space. On the other hand the practical questions involve information across the entire structure or region. They can be expressed via distributions of spatial maxima within a given time frame (daily maximal windspeed across the bridge, joint distribution of summer fortnightly temperature maxima across the regions).

In addition, the extreme events of interest (a joint exceedance of 40°C, a "one in 1000 years" event) are rare and likely not to be found within the range of the available data. This renders classical geostatistics built on classical theory of covariance models and Gaussian processes an inadequate companion for modelling of such situations. Instead, these problems fall into the remit of an area that we nowadays call *spatial extremes*. Essential reviews for this active area of research include [2, 24, 23, 18, 66]. Naturally, the "maxima-versus-threshold exceedances" paradigm from univariate and multivariate extreme value analysis (cf. Chapters 2 and 7) has its analog in spatial extremes. This chapter is concerned with the spatial maxima approach, cf. Chapter 16 for spatial threshold approaches.

There is one more important commonality of the two situations above, namely that a high degree of dependence among extreme outcomes is to be expected across different locations. For instance, large wind speeds at both ends of the bridge within a day will likely originate from the same storm or weather pattern. Likewise, a heatwave is likely to affect a large part of a country or continent during the same fortnight. This means, we anticipate encountering a phenomenon called *asymptotic dependence* between different sites of interest, cf. also Chapter 8. To be precise, let us denote the region of interest by $\mathcal{S}$ in what follows. In mathematical terms, we will consider $\mathcal{S}$ a compact subset of $\mathbb{R}^d$. In geostatistics $d = 2$ or $d = 3$ are common.

Definition 1.1. *Two variables $X(\boldsymbol{s}_1)$ and $X(\boldsymbol{s}_2)$ at sites $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{S}$, which have distribution functions $F_{\boldsymbol{s}_1}$ and $F_{\boldsymbol{s}_2}$ that are eventually continuous and strictly monotone, are called* asymptotically dependent *if the tail dependence measure*

$$\chi(\boldsymbol{s}_1, \boldsymbol{s}_2) = \lim_{q \uparrow 1} \mathrm{P}\big\{X(\boldsymbol{s}_1) > F_{\boldsymbol{s}_1}^{-1}(q) \,|\, X(\boldsymbol{s}_2) > F_{\boldsymbol{s}_2}^{-1}(q)\big\}$$

is larger than zero.

This chapter deals with modelling of extremal phenomena arising from a spatial process $X = \{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ that exhibits spatial asymptotic dependence using max-stable processes. It is a first important benchmark in spatial extremes. We will explain how such an approach can be used to address practical questions such as the ones stated above. What is essential for the block maxima approach to be eligible is that the practical problem of interest needs to have a formulation in terms of the distribution of the maxima process

$$M_n^X(\boldsymbol{s}) = \max\big\{X_1(\boldsymbol{s}), X_2(\boldsymbol{s}), \ldots, X_n(\boldsymbol{s})\big\}, \qquad \boldsymbol{s}\in\mathcal{S}$$

of repeated observations X_i of X. Many textbooks advise (for good reasons) to look at *annual* maxima, which means that n is chosen so that $\{1,\ldots,n\}$ represents an entire year (or season within a year). We would like to point out here that this does not necessarily need to be the case and may in fact lead to inefficient use of data in many situations. In Section 4 we will reflect on the choice of block size and what to consider for a sensible decision.

Naturally, many environmental processes will require both, spatial and temporal modelling on the scales of interest. In this chapter we deliberately focus on the spatial component with rather simplistic assumptions on time aspects. However, especially when climate is shifting, taking this into account in the modelling requires expert knowledge and should not be ignored. After all, we involve such techniques in order to draw conclusions for extrapolations on extremes in time. We need to be aware that the notion of a "one in 1000 years" event is only meaningful in a stationary climate and the notion of design life level has been proposed as an easier interpretable way of risk communication [109]. Chapter 18 is devoted to the ongoing research topic of space-time modelling of extremes. Nevertheless, techniques will often be based on a combination of purely temporal or purely spatial processes as a building block. This chapter is, as a starting point, concerned with the generalization of the block maxima paradigm for spatial processes.

2 Theory

The fundamental observation that typically justifies the approximation of a maxima process M_n^X by a max-stable process for large enough $n\in\mathbb{N}$ is the following. Let X_i, $i\geq 1$, be independent copies of a spatial stochastic process $X = \{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$. If there exist suitable normalizing functions $\mu_n : \mathcal{S}\to\mathbb{R}$ and $\sigma_n : \mathcal{S}\to[0,\infty)$, such that the sequence of stochastic processes

$$\left\{\frac{M_n^X(\boldsymbol{s}) - \mu_n(\boldsymbol{s})}{\sigma_n(\boldsymbol{s})}\right\}_{\boldsymbol{s}\in\mathcal{S}} \tag{1}$$

converges in distribution to a stochastic process $Z = \{Z(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ as $n\to\infty$, then the process Z will be *max-stable*, cf. [28, 32].

Definition 2.1. *A stochastic process $Z = \{Z(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ is* max-stable *if for every $n\in\mathbb{N}$ and $Z^{(1)},\ldots,Z^{(n)}$ independent copies of Z, there exist normalizing functions $a_n : \mathcal{S}\to\mathbb{R}$ and $b_n : \mathcal{S}\to[0,\infty)$, such that the maxima process*

$$M_n^Z(\boldsymbol{s}) = \max\big\{Z^{(1)}(\boldsymbol{s}), Z^{(2)}(\boldsymbol{s}), \ldots, Z^{(n)}(\boldsymbol{s})\big\}, \qquad \boldsymbol{s}\in\mathcal{S},$$

has the same distribution as $\{a_n(\boldsymbol{s}) + b_n(\boldsymbol{s})Z(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$.

2.1 Marginal standardization

In the most general situation, a max-stable process $Z = \{Z(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ is a collection of random variables such that all multivariate marginal distributions are max-stable. That is, the distribution of each random vector $(Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2), \ldots, Z(\boldsymbol{s}_m))^{\mathrm{T}}$, where Z is evaluated at m sites $\boldsymbol{s}_1,\boldsymbol{s}_2,\ldots,\boldsymbol{s}_m\in\mathcal{S}$, is max-stable in the multivariate sense, cf. Chapter 7. In particular, the univariate marginal distributions of Z are either degenerate to a point mass or GEV distributions, cf. Chapter 2. We will assume henceforth that our limiting distributions are non-degenerate and so only GEV distributions arise as univariate marginals.

Let us denote the extreme value index of $X(\boldsymbol{s})$ by $\xi(\boldsymbol{s})$. Then the normalizing functions in (1) can be chosen to ensure that the marginal distributions of Z are standardized to

$$Z(\boldsymbol{s}) \sim \mathrm{GEV}(0,1,\xi(\boldsymbol{s})).$$

In practice, the convergence of (1) amounts to the approximation of the process M_n^X by the max-stable process Z_n, where

$$Z_n(\boldsymbol{s}) = \mu_n(\boldsymbol{s}) + \sigma_n(\boldsymbol{s})Z(\boldsymbol{s}), \quad \boldsymbol{s}\in\mathcal{S}.$$

By our choice of standardization

$$Z_n(\boldsymbol{s}) \sim \mathrm{GEV}(\mu_n(\boldsymbol{s}), \sigma_n(\boldsymbol{s}), \xi(\boldsymbol{s})),$$

and the location, scale and shape parameter functions (μ_n,σ_n,ξ) are typically estimated from realizations of M_n^X, cf. Section 4.

Besides taking care of marginal distributions of Z_n, we need to understand its dependence structure if we want to understand the joint distribution of the entire process Z_n. This will give us insight, how likely joint extreme events within the region of interest are. In the spirit of Sklar's theorem and copula modelling, it is convenient to separate marginal and dependence modelling among multivariate and spatial extreme value distributions, at least when developing the theory, cf. also Chapter 7. Structural results about max-stable processes are typically easier to formulate when they have identical univariate marginal distributions. The most common choices in the literature are standard Gumbel margins ($\mathrm{GEV}(0,1,0)$)

or standard Fréchet margins (GEV$(1,1,1)$); we will adopt the latter convention here.

The marginally transformed process

$$Z^*(\boldsymbol{s}) = \{1+\xi(\boldsymbol{s})Z(\boldsymbol{s})\}^{1/\xi(\boldsymbol{s})} = \left\{1+\xi(\boldsymbol{s})\frac{Z_n(\boldsymbol{s})-\mu_n(\boldsymbol{s})}{\sigma_n(\boldsymbol{s})}\right\}^{1/\xi(\boldsymbol{s})}, \quad \boldsymbol{s}\in\mathcal{S},$$

has standard Fréchet margins and is therefore called a *simple max-stable process*. Conversely, the original process Z_n can be retrieved from the simple max-stable process Z^* via the order-preserving transformation

$$Z_n(\boldsymbol{s}) = \mu_n(\boldsymbol{s}) + \sigma_n(\boldsymbol{s})\frac{Z^*(\boldsymbol{s})^{\xi(\boldsymbol{s})}-1}{\xi(\boldsymbol{s})}, \quad \boldsymbol{s}\in\mathcal{S}. \quad (2)$$

2.2 Spectral representation

The following fundamental theorem characterizes *continuous* simple max-stable processes through a Poisson point process representation and can be traced back to the seminal contributions of [28, 136, 31, 84, 52]. In order to distinguish it from other representations, it is also referred to as *de Haan's* spectral representation, as it appears in [28] for the first time.

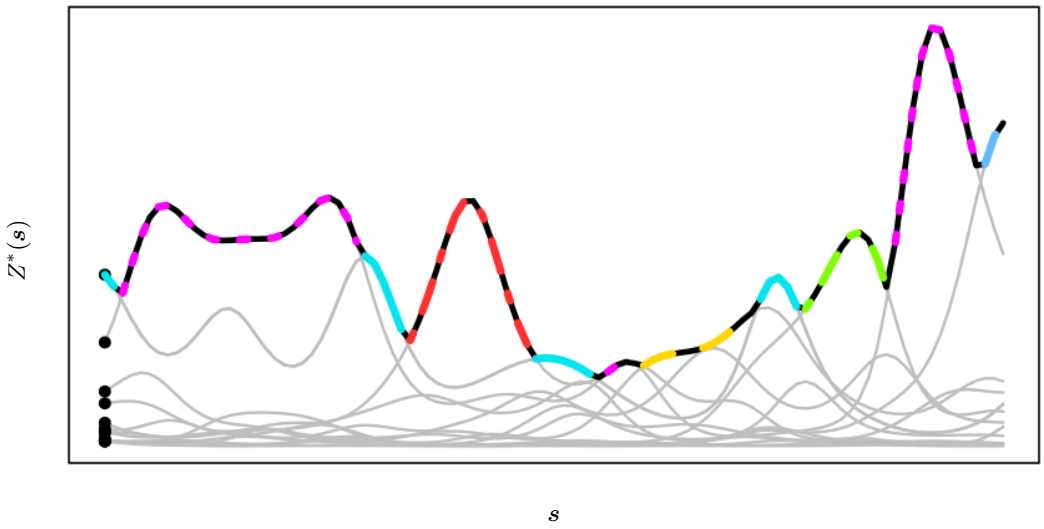

Figure 1: Conceptual illustration of de Haan's spectral representation (Theorem 2.1). Black dots on the left are inverted arrival times $(\Gamma_k^{-1})_{k\geq 1}$; grey lines represent the spatial profiles of the products $(\Gamma_k^{-1}V_k)_{k\geq 1}$; coloured lines of different type show how different profiles contribute to the resulting maximum process Z^*.

Theorem 2.1 (Spectral representation/LePage representation of continuous simple max-stable processes)**.** *A continuous stochastic process $Z^* = \{Z^*(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ is simple max-stable if and only if it can be represented in distribution as*

$$Z^*(\boldsymbol{s}) = \max_{k\geq 1}\frac{V_k(\boldsymbol{s})}{\Gamma_k}, \quad \boldsymbol{s}\in\mathcal{S},$$

where $(\Gamma_k)_{k\geq 1}$ are the arrival times of a unit rate Poisson process on $(0,\infty)$ and $(V_k)_{k\geq 1}$ are independent markings following the law of a continuous stochastic process $V = \{V(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$, which is almost surely non-negative and satisfies $\mathrm{E}\{V(\boldsymbol{s})\} = 1$ for all $\boldsymbol{s}\in\mathcal{S}$.

A memorable interpretation of the spectral representation of Z^* goes as follows [123, 130], cf. also Figure 1: The collection of $(\Gamma_k^{-1}V_k)_{k\geq 1}$ are potential storms affecting the region $\mathcal{S}$, where the multiplicator $1/\Gamma_k \in (0,\infty)$ determines the severity (or magnitude) of a storm and $V_k = \{V_k(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ determines its spatial profile across the region $\mathcal{S}$. Each V_k could affect, for instance, the entire region $\mathcal{S}$ or just a small proportion of it. Eventually, when all storms have passed, only the maximum storm at each site $\boldsymbol{s}\in\mathcal{S}$ counts towards the process Z^*. For each site $\boldsymbol{s}\in\mathcal{S}$ only the most severe storm is relevant in the aggregation.

The finite-dimensional distribution of the simple max-stable process Z^* with spectral process V are given by

$$\begin{aligned} &-\log \mathrm{P}\{Z^*(\boldsymbol{s}_i)\leq u_i \text{ for } i\in\{1,\dots,m\}\} \\ &\quad = -\log \mathrm{P}\left\{\nexists\, k\geq 1 : \max_{i=1,\dots,m}\frac{V_k(\boldsymbol{s}_i)}{u_i} > \Gamma_k\right\} \\ &\quad = \int_0^\infty \mathrm{P}\left\{\max_{i=1,\dots,m}\frac{V(\boldsymbol{s}_i)}{u_i} > r\right\}\mathrm{d}r \\ &\quad = \mathrm{E}\left\{\max_{i=1,\dots,m}\frac{V(\boldsymbol{s}_i)}{u_i}\right\} \end{aligned} \quad (3)$$

for $\boldsymbol{s}_1,\dots,\boldsymbol{s}_m\in\mathcal{S}$ and $u_1,\dots,u_m>0$.

While such a spectral representation as in Theorem 2.1 always *exists*, one needs to be careful, as it is far from being *unique*. Stochastic processes V and V' with different laws may lead to the same law of Z^*. It is only the law of the Poisson point process of the products $(\Gamma_k^{-1}V_k)_{k\geq 1}$, which is uniquely determined by the law of Z^*.

Unique spectral process by a homogeneous functional Instead, uniqueness of the spectral process V can only be artificially achieved by imposing a normalizing condition, cf. [41]. Consider a non-negative measurable functional r, which satisfies

$$r(\lambda f) = \lambda r(f) \quad \text{for} \quad f \geq 0 \text{ continuous}, \lambda > 0,$$

i.e. r is 1-homogeneous. For instance, $r(f)$ can be $\sup_{\boldsymbol{s}\in\mathcal{S}}(f(\boldsymbol{s})g(\boldsymbol{s}))$ or $\int_{\mathcal{S}} f(\boldsymbol{s})g(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$ for a bounded measurable weight function g on $\mathcal{S}$. Given such a functional r, and assuming that $r(V) > 0$ almost surely for *some* spectral process V, there exists always a (distributionally) unique spectral process V^r that satisfies

$$r(V^r) = c_r \quad \text{for a constant} \quad c_r > 0.$$

The constant c_r is then necessarily given by

$$c_r = \mathrm{E}\{r(V)\} = \lim_{u\to\infty} u\,\mathrm{P}\{r(Z^*) > u\}, \tag{4}$$

where V is *any* spectral process of Z^*. In fact, V^r can be obtained from V by the measure transform

$$V^r = c_r \frac{\widetilde{V}^r}{r(\widetilde{V}^r)}, \tag{5}$$

where

$$\mathrm{P}\{\widetilde{V}^r \in \mathrm{d}v\} = \frac{r(v)}{c_r}\,\mathrm{P}\{V \in \mathrm{d}v\}. \tag{6}$$

Some of these representations turn out to be convenient for the simulation of max-stable processes, cf. Section 5. The condition $r(V) > 0$ almost surely is always satisfied if r is a norm. The functional r is also considered as *risk functional* for the definition of Pareto processes (see Chapter 16).

Generalizations In Theorem 2.1 it is also possible to consider more general spaces than compact subsets of $\mathbb{R}^d$ and more general regularity conditions instead of continuity. For instance, [99] considers continuous paths on $\mathbb{R}$ instead of continuous paths on a compact set $\mathcal{S}$; continuity/separability in probability are considered in [128, 127, 142] from an alternative but essentially equivalent perspective of extremal integral representations, whereas [106, 80, 110] draw attention to upper-semicontinuous paths, cf. also [25] for more general LePage series for strictly stable distributions. In practice, continuity of stochastic processes that model physical processes is a standard assumption. The inclusion of a *nugget effect* has been addressed in the Reich–Shaby model, cf. Section 3.4.

2.3 Max-domain of attraction

If the convergence of the maxima process $M_n^X = \{M_n^X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ with suitable normalization as in (1) to a max-stable process $Z = \{Z(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ holds, we say that the process X lies in the *max-domain of attraction* of the max-stable process Z. Let Z be a given max-stable process. We have seen above that we may assume without loss of generality that Z is in standard form

$$Z(\boldsymbol{s}) = \frac{Z^*(\boldsymbol{s})^{\xi(\boldsymbol{s})} - 1}{\xi(\boldsymbol{s})}, \quad \boldsymbol{s} \in \mathcal{S}.$$

Naturally, the question arises if we can find conditions on X, which guarantee the convergence of (1) to Z, but where we can separate the role of marginal features of X (characterized by ξ in the limit Z) from the role of dependence features of X (characterized by Z^* in the limit Z). This is indeed the case, cf. [30] and [29, Section 9.5].

To make this precise, let X^* be the transformation of the process X to standard Pareto margins, that is,

$$X^*(\boldsymbol{s}) = \frac{1}{1 - F_{\boldsymbol{s}}\{X(\boldsymbol{s})\}}, \quad \boldsymbol{s} \in \mathcal{S}.$$

Then, analogously to the multivariate theory, the convergence (1) is equivalent to the convergence of all marginal distributions of (1) to the marginal distributions of Z (uniformly in $\boldsymbol{s} \in \mathcal{S}$) together with the distributional convergence of the standardized maxima process

$$\left\{\frac{1}{n}M_n^{X^*}(\boldsymbol{s})\right\}_{\boldsymbol{s}\in\mathcal{S}},$$

to the simple max-stable process Z^* as $n \to \infty$.

Moreover, we know that in the univariate and multivariate setting convergence of the maxima and convergence of threshold exceedances are equivalent. Such links can also be derived in the functional setting. In particular, the convergence (1) is equivalent to the convergence of

$$\left\{\frac{X(\boldsymbol{s}) - \mu_n(\boldsymbol{s})}{\sigma_n(\boldsymbol{s})}\right\}_{\boldsymbol{s}\in\mathcal{S}} \quad \text{conditional on} \quad \sup_{\boldsymbol{s}\in\mathcal{S}} \frac{X(\boldsymbol{s}) - \mu_n(\boldsymbol{s})}{\sigma_n(\boldsymbol{s})} > 0$$

to a generalized Pareto process as $n \to \infty$, cf. [47] and Chapter 16.

Especially in the heavy-tailed case, when $\xi(\boldsymbol{s}) = \xi = 1/\alpha > 0$ for all $\boldsymbol{s} \in \mathcal{S}$, the max-domain of attraction theory links further conveniently with the concepts of functional regular variation and convergence of point processes [59, 20]. These connections have been neatly summarized in [41, Theorem 1] for heavy-tailed non-negative processes X. The marginally standardized process X^* falls into this framework with $\alpha = 1$. Loosely speaking, the convergence of

the standardized maxima process $M_n^{X^*}/n$ to the simple max-stable process Z^* is equivalent to the convergence of the collection of normalized i.i.d. processes

$$\left\{\frac{1}{n}X_i^* \,:\, i=1,2,\ldots,n\right\}$$

to the Poisson point process of the storm processes $\{\Gamma_k^{-1}V_k \,:\, k\geq 1\}$ as $n\to\infty$.

2.4 Dependence measures & mixing properties

We have seen that the distribution of a max-stable process $Z=\{Z(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ can be decomposed into (i) the marginal GEV distributions as characterized by the extreme value indices $\xi(\boldsymbol{s})\in\mathbb{R}$, $\boldsymbol{s}\in\mathcal{S}$, and (ii) the residual dependence information as given by the simple max-stable process $Z^*=\{Z^*(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$. Still, the space of all possible simple max-stable processes is huge. In practice, guiding summary measures are needed to navigate the information contained in Z^*, cf. Chapter 8.

To begin with, recall the normalization procedure for the spectral representation from Section 2.2. If $r(f)=\sup_{\boldsymbol{s}\in\mathcal{A}} f(\boldsymbol{s})$, the constant c_r from (4) is also called *extremal coefficient*,

$$\theta_{\mathcal{A}}=\mathrm{E}\Big\{\sup_{\boldsymbol{s}\in\mathcal{A}} V(\boldsymbol{s})\Big\},$$

associated with the set $\mathcal{A}\subset\mathcal{S}$. The coefficient $\theta_{\mathcal{A}}$ can be interpreted as the effective number of independent variables among the collection $\{Z^*(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{A}}$. Naturally, $\theta_{\{s\}}=1$ for a one-point set $\{s\}$. For two-point sets $\{\boldsymbol{s}_1,\boldsymbol{s}_2\}$ the extremal coefficient

$$\theta(\boldsymbol{s}_1,\boldsymbol{s}_2)=\theta_{\{\boldsymbol{s}_1,\boldsymbol{s}_2\}}=\mathrm{E}\big[\max\{V(\boldsymbol{s}_1),V(\boldsymbol{s}_2)\}\big]$$

ranges between 1 and 2. If $\theta(\boldsymbol{s}_1,\boldsymbol{s}_2)=1$ this means that $Z^*(\boldsymbol{s}_1)$ and $Z^*(\boldsymbol{s}_2)$ are almost surely equal, while $\theta(\boldsymbol{s}_1,\boldsymbol{s}_2)=2$ means that they are independent.

Even if we knew all extremal coefficients $\theta_{\mathcal{A}}$, say for all compact sets $\mathcal{A}\subset\mathcal{S}$, this information would not suffice in order to retrieve the distribution of Z^*, cf. [80]. However, knowledge of extremal coefficients restricts the potential dependence structures already significantly, a circumstance, which has been used in [147] for building distributionally robust inference for functionals of the spectral process (albeit in a multivariate situation rather than spatial and from the angle of regular variation).

If a spatial process X lies in the max-domain of attraction of a max-stable process Z, whose marginal transformation to a simple max-stable process is Z^*, then the bivariate tail dependence measure $\chi(\boldsymbol{s}_1,\boldsymbol{s}_2)$ of X (cf. Definition 1.1) and the bivariate extremal coefficient $\theta(\boldsymbol{s}_1,\boldsymbol{s}_2)$ are related by

$$\chi(\boldsymbol{s}_1,\boldsymbol{s}_2)+\theta(\boldsymbol{s}_1,\boldsymbol{s}_2)=2.$$

This means $\chi(\boldsymbol{s}_1,\boldsymbol{s}_2)=\mathrm{E}[\min\{V(\boldsymbol{s}_1),V(\boldsymbol{s}_2)\}]$. Alternatively, it can be convenient to express these quantities as $\chi=1-\delta$ and $\theta=1+\delta$, where

$$\delta(\boldsymbol{s}_1,\boldsymbol{s}_2)=\mathrm{E}|V(\boldsymbol{s}_1)-V(\boldsymbol{s}_2)|.$$

In practice, the underlying simple max-stable process Z^* is typically assumed to be *stationary* in space. Then the extremal coefficient $\theta(\boldsymbol{s}_1,\boldsymbol{s}_2)$ and tail dependence coefficient $\chi(\boldsymbol{s}_1,\boldsymbol{s}_2)$ depend only on the difference $\boldsymbol{h}=\boldsymbol{s}_1-\boldsymbol{s}_2\in\mathbb{R}^d$. By a slight abuse of notation we write only the difference argument and write $\theta(\boldsymbol{h})=\theta(\boldsymbol{s}_1-\boldsymbol{s}_2)$ in the *extremal coefficient function* or $\chi(\boldsymbol{h})=\chi(\boldsymbol{s}_1-\boldsymbol{s}_2)$ in the *tail dependence function* if Z^* is stationary. The quantity χ can also be seen as a special case of the spatial extension of the *extremogram* [21].

The (bivariate) extremal coefficient function θ is often used as a diagnostic tool in statistical inference for max-stable processes, cf. Section 6. Within important modelling classes, cf. Section 3, dependence models for Z^* can be identified via their bivariate extremal coefficient function θ. One should however be aware that θ (equivalently χ) cannot distinguish between different modelling classes, cf. [129].

Ergodicity and mixing properties For stationary max-stable processes on $\mathbb{R}^d$ (or $\mathbb{Z}^d$), the tail dependence function χ contains information about ergodicity and mixing properties of the respective max-stable process. Such properties are generally useful in order to derive limit theorems for estimators; they tell us how much one can expect to learn from a few (a single) realization of a process on an observation window.

Ergodicity and mixing properties for stationary max-stable (and more generally max-infinite divisible) processes were studied by [127, 72, 37, 141, 40] under slightly different regularity conditions. A stationary simple max-stable process $Z^*=\{Z^*(\boldsymbol{t})\}_{\boldsymbol{t}\in\mathcal{T}}$, $\mathcal{T}=\mathbb{R}^d$ or $\mathcal{T}=\mathbb{Z}^d$, is *mixing* if and only if

$$\chi(\boldsymbol{h})\to 0 \quad \text{as} \quad \|\boldsymbol{h}\|\to\infty,$$

whereas Z^* is *weakly mixing* if and only if it is *ergodic* if and only if χ converges to 0 in a Cesáro sense, that is,

$$\frac{\int_{W_R}\chi(\boldsymbol{h})\mathrm{d}\boldsymbol{h}}{\int_{W_R}\mathrm{d}\boldsymbol{h}}\to 0 \quad \text{for} \quad R\to\infty,$$

where $W_R=[-R,R]^d$ (or $W_R=\{-R,\ldots,R\}^d$ in the discrete case) and the integration is with respect to the Lebesgue-measure (or counting measure in the discrete case).

Some interesting connections to the time change theory for regularly varying time series have been derived in [101, 68], whilst [87] study implications for long-memory behaviour and statistical properties of tail dependence estimators.

3 Models

When specifying a class of models of max-stable processes, we need to specify (i) a spatial model for the GEV margins and (ii) a dependence model, i.e. a model for the underlying simple max-stable process Z^*. In this section, we will focus on the latter task (ii).

The spectral representation from Theorem 2.1 is again useful in this context, as it leads directly to a universal construction principle. Just pick any stochastic process $V = \{V(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ that satisfies the integrability condition $\mathrm{E}\{V(\boldsymbol{s})\} = 1$ for all $\boldsymbol{s} \in \mathcal{S}$. Let $(\Gamma_k)_{k\geq 1}$ be the unit rate Poisson process on $(0, \infty)$ and consider independent markings $(V_k)_{k\geq 1}$ drawn from V. Then the resulting process

$$Z^*(\boldsymbol{s}) = \max_{k\geq 1} \Gamma_k^{-1} V_k(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{S}, \tag{7}$$

is simple max-stable.

However, will it have other useful properties considered essential or desirable in (geo-)statistical practice? For instance, for which choices of V will the model Z^* for residual extremal dependence be *stationary*?

Stationarity When proposing spatial max-stable models, stationarity of the model is usually a minimal requirement – at least as a starting point, from which more complex models can be built if desired.

Clearly, if V is stationary, so is Z^*. This is the case in the construction of the *extremal-t* process (Section 3.3 below). It leads automatically to some strong forms of dependence across long distances in the model. On the other hand, the converse statement is false. When the max-stable process Z^* is stationary, then the spectral process V does not necessarily have to be stationary. The most popular example of this kind is the class of *Brown–Resnick processes* (Section 3.2 below), which straddles a wider range of mixing and ergodicity behaviours.

Spectral processes V, for which the resulting process Z^* becomes stationary, have been termed *Brown–Resnick stationary* in [73] (albeit on the Gumbel scale therein). [79] provide a geometric interpretation and characterization of this concept. Alternatively, the construction of *mixed moving maxima (M3)* processes (Section 3.1) reminds us that there are other representations of max-stable processes that will directly guarantee stationarity.

3.1 Mixed moving maxima

The following modification of the classical spectral representation that we have encountered above goes back to the pioneering spatial extremes manuscript [123]. It introduces what we nowadays call a *mixed moving maxima (M3)* process

$$Z^*(\boldsymbol{s}) = \max_{k\geq 1} \Gamma_k^{-1} F_k(\boldsymbol{s} - \boldsymbol{S}_k), \quad \boldsymbol{s} \in \mathcal{S}.$$

Here, $(\Gamma_k, \boldsymbol{S}_k)_{k\geq 1}$ is a measurable enumeration of a Poisson point process whose intensity measure is the Lebesgue measure on $(0, \infty) \times \mathbb{R}^d$ and, independently, $(F_k)_{k\geq 1}$ are independent copies of a non-negative stochastic process F satisfying the integrability condition $\mathrm{E}\{\int_{\mathbb{R}^d} F(\boldsymbol{h})\mathrm{d}\boldsymbol{h}\} = 1$. In the special case that $F \equiv f : \mathbb{R}^d \to [0, \infty)$ is deterministic, the resulting process Z^* is a *moving maxima (M2)* process.

By construction M3 processes are stationary with finite-dimensional distributions

$$\begin{aligned} &-\log \mathrm{P}\{Z^*(\boldsymbol{s}_i) \leq u_i \text{ for } i \in \{1, 2, \ldots, m\}\} \\ &\quad = \mathrm{E}\left\{ \int_{\mathbb{R}^d} \max_{i=1,\ldots,m} \frac{F(\boldsymbol{s}_i - \boldsymbol{h})}{u_i} \mathrm{d}\boldsymbol{h} \right\} \end{aligned}$$

for $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m \in \mathcal{S}$ and $u_1, \ldots, u_m > 0$.

In the M2 case, any multivariate density function on $\mathbb{R}^d$ will satisfy the integrability constraint for y. For example, [123] proposed the multivariate normal or multivariate t distributions for f. When f is chosen multivariate normal, the resulting M2 process Z^* is also known as the *Smith process*. In practice, M2 processes have rather smooth sample paths for such straightforward choices for F and are therefore usually not considered as a serious competitor to some alternative suggestions with more practically motivated properties.

Whilst historically M3 processes were the first proposed stationary models for max-stable processes, the interest in M3 processes is nowadays more of a theoretical nature, as they turn out to play a crucial role in the ergodic theory for max-stable processes. Several articles [70, 142, 141, 40] study the *Hopf decomposition* of a stationary max-stable process and show that the purely dissipative part is precisely of M3 type. [40] establish further criteria for the existence of an M3 representation. If we are given a stationary simple max-stable process Z^* with spectral process V in the sense of Theorem 2.1, it has an M3 representation if and only if

$$\int_{\mathbb{R}^d} V(\boldsymbol{h})\mathrm{d}\boldsymbol{h} < \infty \quad \text{almost surely,}$$

a condition, which is surprisingly equivalent to $\lim_{h\to\infty} V(\boldsymbol{h}) = 0$ almost surely in this context.

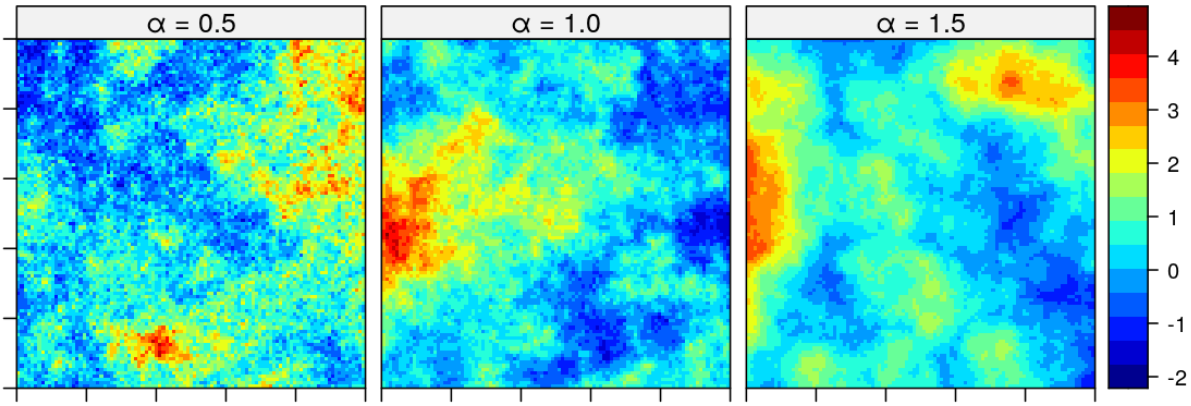

Figure 2: Three simulations of Brown–Resnick processes with variogram $\gamma(\boldsymbol{h}) = \|\boldsymbol{h}\|^\alpha$ on the grid $\{0, 0.05, \ldots, 5\}^2$ for three different values of $\alpha \in \{0.5, 1.0, 1.5\}$. Marginal distributions have been transformed to standard Gumbel margins.

3.2 Brown–Resnick processes

Brown–Resnick processes are often considered an analog of Gaussian processes in the max-stable world, as they come along with many useful properties inherited from their underlying Gaussian structure.

Let $W = \{W(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ be a Gaussian process with zero mean and variance $\{\sigma^2(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$. Then the law of the simple max-stable process Z^* that arises from the spectral process

$$V(\boldsymbol{s}) = \exp\left\{W(\boldsymbol{s}) - \frac{\sigma^2(\boldsymbol{s})}{2}\right\}, \quad \boldsymbol{s} \in \mathcal{S}, \qquad (8)$$

depends only on the *variogram*

$$\gamma(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathrm{E}\big[\{W(\boldsymbol{s}_1) - W(\boldsymbol{s}_2)\}^2\big],$$

cf. [71]. Note that Gaussian processes with different covariance functions may share an identical variogram and therefore lead to the same max-stable process Z^*. Such processes are (up to location and scale normalization) the only possible limit laws of maxima of triangular arrays of Gaussian processes [58, 71]. The multivariate marginal distributions of Z^* are *Hüsler–Reiss distributions* [67], that is,

$$\begin{aligned}&-\log \mathrm{P}\{Z^*(\boldsymbol{s}_i) \le u_i \text{ for } i \in \{1, 2, \ldots, m\}\}\\ &= \sum_{i=1}^{m} \frac{1}{u_i}\, \Phi_{m-1}\big(\mathbf{y}^{(i)}\,;\, \boldsymbol{\Sigma}^{(i)}\big),\end{aligned}$$

where $\Phi_{m-1}(\mathbf{y}^{(i)}, \boldsymbol{\Sigma}^{(i)})$ is the cumulative distribution function of the centered multivariate normal distribution with $(m-1)\times(m-1)$ covariance matrix

$$\boldsymbol{\Sigma}^{(i)}_{jk} = \frac{\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j) + \gamma(\boldsymbol{s}_i, \boldsymbol{s}_k) - \gamma(\boldsymbol{s}_j, \boldsymbol{s}_k)}{2\sqrt{\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j)}\sqrt{\gamma(\boldsymbol{s}_i, \boldsymbol{s}_k)}},$$
$$j, k \in \{1, \ldots, m\} \setminus \{i\}$$

evaluated at

$$\mathbf{y}^{(i)}_j = \frac{\sqrt{\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j)}}{2} + \frac{\log(u_j/u_i)}{\sqrt{\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j)}},$$
$$j \in \{1, \ldots, m\} \setminus \{i\},$$

cf. [83, 60, 75]. Its tail dependence function and extremal coefficient function are given by

$$\chi = 2\overline{\Phi}(\sqrt{\gamma}/2) \quad \text{and} \quad \theta = 2\Phi(\sqrt{\gamma}/2),$$

or, alternatively, $\chi = 1 - \delta$ and $\theta = 1 + \delta$, where $\delta = \mathrm{erf}(\sqrt{\gamma/8})$ and erf denotes the *error function*, cf. also Figure 3.

If the underlying Gaussian process W has *stationary increments*, that is, the law of $\{W(\boldsymbol{h} + \boldsymbol{s}_0) - W(\boldsymbol{s}_0)\}_{\boldsymbol{h}\in\mathbb{R}^d}$ does not depend on the choice of $\boldsymbol{s}_0 \in \mathbb{R}^d$, then the resulting max-stable process Z^* is stationary [73]. In other words, the spectral process (8) is then Brown–Resnick stationary. In fact, it was Brown and Resnick [9] who considered the special case of W being a Brownian motion on the real line already in the 1970s. This explains, why we are calling max-stable processes of this kind nowadays *Brown–Resnick processes*.

More generally, the most prominent examples of Gaussian processes that lead to Brown–Resnick stationary spectral processes (8) are fractional Brownian sheets. Their variogram family is given by

$$\gamma(\boldsymbol{h}) = \|\boldsymbol{h}/c\|^\alpha, \quad \boldsymbol{h} \in \mathbb{R}^d, \quad \alpha \in (0, 2],\, c > 0. \qquad (9)$$

Here, the parameter α controls the local smoothness of Z^*; the larger α, the smoother the sample paths (on piecewise patches). Figure 2 illustrates this phenomenon. The case $\alpha = 2$ is a special case that recovers the (isotropic) Smith process, cf. Section 3.1.

In general, any variogram γ defines a Brown–Resnick process Z^* via the construction above, and different variograms will lead to different laws of

Brown–Resnick processes. A Brown–Resnick process is identified by its variogram. In order to be a valid variogram model, a function $\gamma : \mathcal{S} \times \mathcal{S} \to [0, \infty)$ must satisfy $\gamma(\boldsymbol{s}, \boldsymbol{s}) = 0$ for $\boldsymbol{s} \in \mathcal{S}$ and it needs to be *conditionally negative definite*, that is,

$$\sum_{i=1}^{m} \sum_{j=1}^{m} a_i \gamma(\boldsymbol{s}_i, \boldsymbol{s}_j) a_j \leq 0$$

for all $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m \in \mathcal{S}$, $a_1, \ldots, a_m \in \mathbb{R}$ with $a_1 + \cdots + a_m = 0$, $m \in \mathbb{N}$.

Brown–Resnick processes have become a popular modelling class for extremal phenomena due to a number of desirable properties. They are characterized by a bivariate quantity (its variogram); their sample paths look realistic for environmental processes; they are quite flexible in their behaviour (local smoothness, ergodicity/mixing behaviours) – for instance, [40] construct a stationary Brown–Resnick processes which is ergodic, but not mixing; and it is well-understood how these behaviours are linked to the characterizing variogram.

3.3 Extremal-t processes

Another class that is built on Gaussian processes in the spectral representation, is the class of *extremal-t processes*, whose spectral representation is given by

$$V(\boldsymbol{s}) = c_\nu W(\boldsymbol{s})_+^\nu, \quad \boldsymbol{s} \in \mathcal{S}. \tag{10}$$

Here, $W = \{W(\boldsymbol{s})\}_{\boldsymbol{s} \in \mathcal{S}}$ is a Gaussian process on $\mathcal{S}$ with standard normal margins and $x_+ = \max(0, x)$, $\nu > 0$ is a degrees of freedom parameter for a t-distribution (as we shall see below) and multiplication with the constant

$$c_\nu = \frac{\sqrt{\pi}\, 2^{1-\nu/2}}{\Gamma\{(\nu+1)/2\}},$$

where Γ denotes the *gamma function*, ensures that the resulting max-stable process Z^* has standard Fréchet margins.

In principle, we might allow W to have any correlation function $\rho : \mathcal{S} \times \mathcal{S} \to [-1, 1]$ with $\rho(\boldsymbol{s}, \boldsymbol{s}) = 1$ for $\boldsymbol{s} \in \mathcal{S}$ and which is *positive semi-definite*, i.e.

$$\sum_{i=1}^{m} \sum_{j=1}^{m} a_i \rho(\boldsymbol{s}_i, \boldsymbol{s}_j) a_j \geq 0$$

for all $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m \in \mathcal{S}$, $a_1, \ldots, a_m \in \mathbb{R}$, $m \in \mathbb{N}$. This setup will always lead to a (not necessarily stationary) max-stable process Z^* based on the spectral process (10).

If the underlying Gaussian process W is stationary, so is the resulting spectral process V, and it follows that the resulting max-stable process Z^* must be stationary, too. A popular class of correlation models is, for instance,

$$\rho(\boldsymbol{h}) = \mathrm{Cov}\{W(\boldsymbol{s}+\boldsymbol{h}), W(\boldsymbol{s})\} = \exp\{-\|\boldsymbol{h}/c\|^\alpha\},$$
$$\boldsymbol{h} \in \mathbb{R}^d,\ \alpha \in (0, 2],\ c > 0.$$

More generally, one could work with $\rho(\boldsymbol{h}) = \exp\{-\gamma(\boldsymbol{h})\}$, where γ is any variogram [55] or, alternatively, with compactly supported correlation functions [53].

Originally explored in the language of copulas [33, 83], the extremal-t model was introduced to spatial statistics for extremes by [24], while [96] recognized its max-stable process nature, its spectral representation and its role as max-stable limit for processes with elliptical marginal distributions, which exhibit asymptotic dependence. For $\nu = 1$, the extremal-t process is also known as *Schlather model* or *extremal Gaussian model* [115]. If we set $\rho = \exp(-\gamma/(2\nu)) \sim 1 - \gamma/(2\nu)$ and let $\nu \to \infty$, we retrieve the Brown–Resnick process for the variogram γ [83, 24].

The multivariate marginal distributions of Z^* are given by

$$-\log \mathrm{P}\{Z^*(\boldsymbol{s}_i) \leq u_i \text{ for } i \in \{1, 2, \ldots, m\}\}$$
$$= \sum_{i=1}^{m} \frac{1}{u_i}\, t_{m-1,\nu+1}\big(\mathbf{y}^{(i)}\,;\, \boldsymbol{\Sigma}^{(i)}\big),$$

where $t_{m-1,\nu+1}(\mathbf{y}^{(i)}, \boldsymbol{\Sigma}^{(i)})$ is the cumulative distribution function of the centered multivariate t distribution with $\nu+1$ degrees of freedom and $(m-1)\times(m-1)$ scale matrix

$$\boldsymbol{\Sigma}^{(i)}_{jk} = \frac{\rho(\boldsymbol{s}_j, \boldsymbol{s}_k) - \rho(\boldsymbol{s}_i, \boldsymbol{s}_j)\rho(\boldsymbol{s}_i, \boldsymbol{s}_k)}{\sqrt{1-\rho^2(\boldsymbol{s}_i, \boldsymbol{s}_j)}\sqrt{1-\rho^2(\boldsymbol{s}_i, \boldsymbol{s}_k)}},$$
$$j, k \in \{1, \ldots, m\} \setminus \{i\}$$

evaluated at

$$\mathbf{y}^{(i)}_j = \frac{\sqrt{\nu+1}}{\sqrt{1-\rho^2(\boldsymbol{s}_i, \boldsymbol{s}_j)}} \left\{ \left(\frac{u_j}{u_i}\right)^{1/\nu} - \rho(\boldsymbol{s}_i, \boldsymbol{s}_j) \right\},$$
$$j \in \{1, \ldots, m\} \setminus \{i\},$$

cf. [83, 96]. Hence, its tail dependence function and extremal coefficient function are given by

$$\chi = 2\bar{t}_{\nu+1}\left(\sqrt{(\nu+1)\frac{1-\rho}{1+\rho}}\right)$$

and

$$\theta = 2t_{\nu+1}\left(\sqrt{(\nu+1)\frac{1-\rho}{1+\rho}}\right),$$

or, alternatively, $\chi = 1 - \delta$ and $\theta = 1 + \delta$, where $\delta = I_{(1-\rho)/2}(1/2, (\nu+1)/2)$ and $I_x(a, b)$ denotes the *regularized incomplete beta function*.

Figure 3 illustrates the dependence of θ on the correlation function ρ and degrees of freedom ν. It is worth noting that correlation functions of isotropic random fields on $\mathbb{R}^d$ are bounded from below by the minimum of

$$\Omega_d(r) = \Gamma(d/2)\left(\frac{2}{r}\right)^{(d-2)/2} J_{(d-2)/2}(r), \qquad r \geq 0, \tag{11}$$

where $J_{(d-2)/2}$ is a *Bessel function of the first kind*, as such correlation functions arise as scale mixtures of these functions [117, 54]. Specifically, this leads to $\rho \gtrapprox -0.403$ in dimension $d = 2$ and $\rho \gtrapprox -0.217$ for $d = 3$. Both lower bounds have been added as vertical lines in Figure 3. For a fixed $\nu > 0$ they lead to upper bounds on the extremal coefficient function θ. For the Schlather model ($\nu = 1$), this means $\theta \lessapprox 1.837$ for $d = 2$ and $\theta \lessapprox 1.780$ in dimension $d = 3$ (horizontal lines). Hence, extremal-t processes exhibit dependence even across large distances. They cannot be ergodic, cf. Section 2.4. On the other hand, the bound for θ increases with ν. So one might observe that within practically relevant ranges this kind of long-range dependence will not be very restrictive and can be dealt with by a sufficiently large ν.

After all, both classes of processes (extremal-t and Brown–Resnick) share similar properties as parsimonious models with some realistic features inherited from underlying Gaussian processes and modeled by a bivariate quantity (correlation or variogram function). From a theoretical angle, the class of Brown–Resnick process can be more appealing due to its broader range of ergodicity behaviours. From a practical point of view, the extremal-t process might be preferred as a richer more flexible class of models as it (almost) includes the Brown–Resnick class for large values of ν.

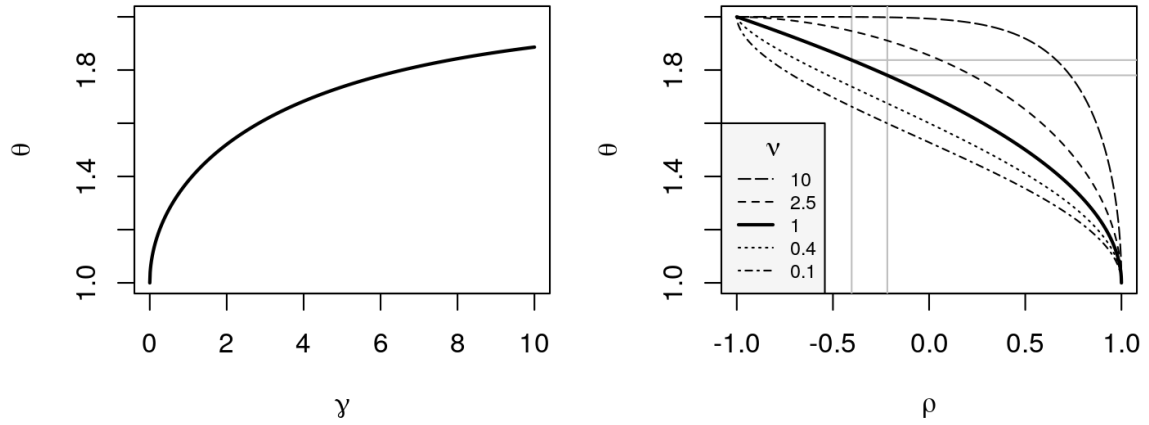

Figure 3: The bivariate extremal coefficient function θ is a function of the variogram γ in the case of the Brown–Resnick process (left, cf. Section 3.2) and a function of the correlation ρ and the degrees of freedom ν for an extremal-t process (right, cf. Section 3.3). The vertical gray lines mark the lower bounds for a correlation function of an isotropic random field in $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively, cf. (11).

3.4 Reich–Shaby model and generalizations

In geostatistics it is common to include a *nugget effect* in the modelling of physical processes representing small-scale noise in the data. If we consider the Gumbel scale to be appropriate for additive noise, this will lead to multiplicative noise on Fréchet scales. Let $V = \{V(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ be a classical spectral process in the sense of de Haan's spectral representation (Theorem 7) and, independently, let $\{\varepsilon(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ be independent copies of a standard Fréchet variable. Then the simple max-stable process

$$Z^{(p)}(\boldsymbol{s}) = \max_{k\geq 1} \Gamma_k^{-1} V_k^{(p)}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{S},$$

that is based on the noisy spectral process

$$V^{(p)}(\boldsymbol{s}) = \frac{\varepsilon(\boldsymbol{s})^{1/p}}{\Gamma(1-1/p)} \cdot V(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{S},$$

for some $p \in (1,\infty)$ can be interpreted as a noisy version of the original process Z^* from (7), where the parameter $p \in (1,\infty)$ determines the strength of the nugget effect. The process $Z^{(p)}$ inherits stationarity, ergodicity and mixing behaviour from its denoised counterpart Z^* [85]. Its marginal distributions are *generalized logistic mixtures* [48, 49]

$$\begin{aligned} &-\log \mathrm{P}\{Z^{(p)}(\boldsymbol{s}_i) \leq u_i \text{ for } i \in \{1,2,\ldots,m\}\} \\ &\quad = \mathrm{E}\left[\sum_{i=1}^{m} \left\{\frac{V(\boldsymbol{s}_i)}{u_i}\right\}^p\right]^{1/p}. \end{aligned}$$

At first sight it may surprise that the process $Z^{(p)}$ allows not only for a classical max-series representation, but also for an *ℓ^p-norm based representation* [85]

$$Z^{(p)}(\boldsymbol{s}) = \frac{\varepsilon(\boldsymbol{s})^{1/p}}{\Gamma(1-1/p)} \cdot \left[\sum_{k\geq 1} \left\{\Gamma_k^{-1} V_k(\boldsymbol{s})\right\}^p\right]^{1/p}, \ \boldsymbol{s} \in \mathcal{S},$$

where, as before, $(\Gamma_k, V_k)_{k\geq 1}$ are the arrival times of a unit rate Poisson process on $(0,\infty)$ with independent markings from the law of V. This representation can be revealed, for instance, by a simple conditioning argument and related interpretations of this model in

terms of latent stable random vectors that have been brought together in [48, 49].

Moreover, $Z^{(p)}$ coincides with a max-stable process proposed by Reich and Shaby [104] if we choose V uniformly distributed on a set of L deterministic non-negative weight functions $Lw_j = \{Lw_j(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$, $j \in \{1, 2, \ldots, L\}$, where $w_1(\boldsymbol{s}) + w_2(\boldsymbol{s}) + \cdots + w_L(\boldsymbol{s}) = 1$. In other words,

$$V(\boldsymbol{s}) = Lw_J(\boldsymbol{s}) = L\sum_{j=1}^{L} w_j(\boldsymbol{s})\mathbf{1}_{J=j}, \quad \boldsymbol{s} \in \mathcal{S},$$

for J uniformly distributed on $\{1, 2, \ldots, L\}$. In this case

$$Z^{(p)}(\boldsymbol{s}) = \varepsilon(\boldsymbol{s})^{1/p} \cdot \left\{\sum_{j=1}^{L} B_j w_j^p(\boldsymbol{s})\right\}^{1/p}, \quad \boldsymbol{s} \in \mathcal{S},$$

where

$$B_j = \left\{\frac{L}{\Gamma(1-1/p)}\right\}^p \sum_{k\geq 1} \Gamma_k^{-p}\mathbf{1}_{J_k=j}, \quad j \in \{1, \ldots, L\},$$

are independent and identically distributed (sum-)stable random variables with Laplace transform $\mathrm{E}\{\exp(-tB_j)\} = \exp(-t^{1/p})$ (as follows from [112, Theorem 1.4.5 and Corollary 3.5.4.]), and

$$\begin{aligned} &-\log \mathrm{P}\{Z^{(p)}(\boldsymbol{s}_i) \leq u_i \text{ for } i \in \{1, 2, \ldots, m\}\} \\ &\quad = \sum_{j=1}^{L}\left[\sum_{i=1}^{m}\left\{\frac{w_j(\boldsymbol{s}_i)}{u_i}\right\}^p\right]^{1/p}. \end{aligned}$$

Besides the flexibility offered through modelling the strength of a nugget effect, the Reich–Shaby model and its generalizations enjoy the advantage of a tractable likelihood and the model can be interpreted via latent stable variables [48, 49]. In particular, this makes them suitable for the inclusion in hierarchical Bayesian models [104, 105, 126].

3.5 Further remarks on model building

Naturally, many more ideas have been proposed and explored, not all of which can be addressed here. For instance, [115, 22, 129] consider the stationary truncation by a random compact set. Spatial deformations such as those arising from anisotropies may be incorporated in the modelling by replacing underlying variograms $\gamma(\boldsymbol{h})$ and correlation functions $\rho(\boldsymbol{h})$ by $\gamma(\mathbf{A}\boldsymbol{h})$ and $\rho(\mathbf{A}\boldsymbol{h})$, where the mapping $\mathbf{A}$ accounts for the desired spatial deformation. While a linear transformation $\mathbf{A}$ still preserves stationarity, more general non-linear transformations can be introduced to allow for non-stationarity [145]. Other approaches for modelling non-stationarity include, for instance, embeddings into climate or latent spaces [8, 16] or the use of the non-stationary dependence model by [97] for extremal-t processes [63]. The heavy reliance of the most common models on (symmetric) Gaussian or t-distributions may be overhauled through the introduction of skewed versions when deemed more appropriate from inspection of the data [6, 5].

4 Statistical inference

The last two sections gave an insight into the rich theory around max-stable processes. In this section we get back to their role in statistical practice. Naturally, such an endeavor can only be indicative here, as each research question comes along with its own context; it will be critical to be guided by domain knowledge. However, to make this more specific, let us recall our introductory illustrative examples from Section 1, in short: *Is there a higher than 1% chance of a joint temperature exceedance for the three regions during one summer ("joint exceedance" meaning "exceedance within the same fortnight")?* and *What does a "one in 1000 years" wind event look like for the bridge?*

A priori, we need to check a few prerequisites to ensure that max-stable processes are an appropriate model to answer the practical question at hand. These include:

- Is it a question about extreme values that involves extrapolation far into the tail of a distribution?
- Is spatial modelling needed? Can the question of interest be expressed in terms of the joint distribution of maxima across space?
- Does the assumption of asymptotic dependence between different sites seem justified (cf. also Chapter 8 for diagnostic tools based on χ and θ)?

If we can answer these questions with "yes", we might proceed with a block maxima approach as follows.

4.1 The block maxima paradigm

As in the univariate and multivariate situation, the key idea is that max-stable distributions approximate the distribution of block maxima, where the blocks are built in the time dimension. Let us assume that $X_1, X_2, \ldots, X_N$ are independent replicates of a spatial stochastic process $X = \{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$, which is in the domain of attraction of a max-stable process. If $N = bn$ for a sufficiently large block size n (for a better readability we omit the possibility of incomplete blocks and remainders), the collection of b block maxima

$$\begin{aligned} M_{n,i}^X(\boldsymbol{s}) = \max\big\{&X_{(i-1)n+1}(\boldsymbol{s}), X_{(i-1)n+2}(\boldsymbol{s}), \\ &\ldots, X_{(i-1)n+n}(\boldsymbol{s})\big\}, \quad \boldsymbol{s} \in \mathcal{S}, \end{aligned}$$

for $i \in \{1, \ldots, b\}$ may be considered a sample of an approximating max-stable process

$$Z_n(\boldsymbol{s}) = \mu_n(\boldsymbol{s}) + \sigma_n(\boldsymbol{s}) \frac{Z^*(\boldsymbol{s})^{\xi(\boldsymbol{s})} - 1}{\xi(\boldsymbol{s})}, \quad \boldsymbol{s} \in \mathcal{S},$$

cf. (2).

The choice of the block size n versus sample size b among N available consecutive time points constitutes a *bias-variance trade-off*. The block size n needs to be large enough in order to justify the approximation with the limiting model Z_n. At the same time the number of blocks $b = N/n$ needs to be large enough in order to have a sufficiently large sample from Z_n for inference on the parameters of the model. The assumption of independence of the $(X_i)_{i=1,\ldots,N}$ can be relaxed, as long as the block maxima $(M^X_{n,i})_{i=1,\ldots,b}$ are independent and the effective block size $\widetilde{n} \leq n$ remains large enough.

Classical environmental applications consider b annual maxima of daily observations ($n \approx 365$) or b annual maxima of daily observations of a specific season only, e.g. [8, 12, 122, 131]. Accordingly, this kind of inference can seem a rather inefficient use of data. Moving to a smaller block size can often be justified and mitigate the inefficiency. For instance, [42] consider three-day maxima of summer wind speeds and [95] 14-day summer temperature maxima for Dutch inland stations. In any case, diagnostic checks for violation of independence between consecutive block maxima, validation of the approximation and sensitivity checks for the block size are imperative. Some elementary ones are included in Section 6.

Specifying a model In practice, the process X (and thereby its associated maxima process $M^X_n \approx Z_n$) will only be measured at a finite number of sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m \in \mathcal{S}$, while questions of interest concern the entire region $\mathcal{S}$, hence also ungauged sites. If the approach above should provide meaningful answers to these, spatial modelling for the marginal GEV parameter functions $\mu_n(\boldsymbol{s})$, $\sigma_n(\boldsymbol{s})$ and $\xi(\boldsymbol{s})$ will need to be included as well. This in itself may lead to many interesting research questions in applications, cf. also Chapter 6 and [146], or more generally, [4] Section 4.3. One of the most basic approaches assumes the extreme value index $\xi(\boldsymbol{s})$ constant across $\mathcal{S}$, while the location and scale functions $\mu_n(\boldsymbol{s})$ and $\sigma_n(\boldsymbol{s})$ are modelled as smooth functions of relevant spatial covariates that are known across the entire region $\mathcal{S}$. Finally, one needs to choose a model for the residual spatial dependence, which occurs in the form of a simple max-stable process Z^* in the above formulation of Z_n. Section 3 accumulates some popular models alongside an interpretation of their parameters.

Answering the original questions – the road ahead Before proceeding with further details on fitting a max-stable model to data, let us briefly reflect on how to close the loop to the original questions (Section 1). Let us assume for a moment that $M^X_n = Z_n$ and that we can perfectly estimate all its parameters correctly.

Joint temperature exceedance Suppose that the block length is a fortnight, i.e. $n = 14$ (days), and that each summer consists of 6 such fortnights (if we have a record of 30 years of data, this would amount to $b = 6 \cdot 30 = 180$ blocks available for the estimation). The probability of a joint temperature exceedance above $\tau = 40°\mathrm{C}$ for the three non-overlapping regions $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3 \subset \mathcal{S}$ in a summer fortnight can then be calculated as

$$p = \mathrm{P}\Big\{ \max_{\boldsymbol{s} \in \mathcal{S}_1} Z_n(\boldsymbol{s}) > \tau\,, \max_{\boldsymbol{s} \in \mathcal{S}_2} Z_n(\boldsymbol{s}) > \tau\,, \max_{\boldsymbol{s} \in \mathcal{S}_3} Z_n(\boldsymbol{s}) > \tau \Big\}, \tag{12}$$

which leads to a probability of a joint exceedance during a given summer (six of such consecutive fortnights) of $1 - (1-p)^6 \approx 6p$, i.e. the answer to our original question.

Wind speeds across a bridge In this example, let us assume that the block length is $n = 7$ days and that each summer consists of 12 such weeks (if we have a record of 25 years of data, this would amount to $b = 12 \cdot 25 = 300$ blocks available for the estimation). Then the interest lies here on the

$$\left(1 - \frac{1}{12 \cdot 1000}\right)\text{-quantile}$$

of the random variable

$$\max_{\boldsymbol{s} \in \mathcal{S}} Z_n(\boldsymbol{s}). \tag{13}$$

In practice, we need to estimate the parameters of Z_n first (see the remainder of this section). Second, since no closed-form expressions exist for the distribution of the (joint) spatial maxima of Z_n, our actual answers will depend on simulations from the fitted model for Z_n on a sufficiently dense grid (Section 5). Naturally, each step of the inference (including the approximation $M^X_n \approx Z_n$) comes along with its own kind of uncertainty leading to uncertainties in the final outputs.

4.2 Likelihood-based inference

In principle, one could follow again the paradigm of treating marginal GEV parameters and dependence parameters of Z^* separately also for the estimation. Even, when using the (wrong) independence likelihood for estimation of the marginal GEV parameters,

the resulting estimators have been shown to be consistent and asymptotically normal for $\xi > -0.5$ [10, 15]; and the 2-step procedure can seem appealing to reduce the computational cost. On the other hand, the 1-step procedure (joint estimation of all parameters) is typically more suitable to assess the uncertainty of the estimates.

Either way, one has to deal with the combinatorial explosion of the *likelihood function* for the distribution of $(Z^*(\boldsymbol{s}_1), Z^*(\boldsymbol{s}_2), \ldots, Z^*(\boldsymbol{s}_m))^{\mathrm{T}}$. If $\mathcal{P}$ denotes the set of partitions $\pi = \{A_1, \ldots, A_{|\pi|}\}$ of $\{1, 2, \ldots, m\}$ and $u_A = \{u_i\}_{i \in A}$, then its m-variate density $f_{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_m}(u_1, \ldots, u_m)$ is given by

$$e^{-V^*(u_1,\ldots,u_m)} \sum_{\pi \in \mathcal{P}} \prod_{A \in \pi} \Big\{ -\frac{\partial}{\partial u_A} V^*(u_1, \ldots, u_m) \Big\}, \tag{14}$$

where the *exponent function* $V^*(u_1, \ldots, u_m)$ equals (3). Even for moderate number of sites m, say $m = 16$ stations, we would arrive already at more than 10 billion (!) summands.

Two main strategies have evolved. On the one hand, *composite likelihood* techniques [78, 135] replace the full likelihood by a combination of likelihood terms of lower order to achieve computational efficiency. Despite the misspecification of the model, this strategy leads to consistent estimators under mild regularity conditions if the parameters are identifiable. The most common choice in this context is a *pairwise likelihood* approach [98], whose implemenation is readily available in `R` [102] in the package `SpatialExtremes` [108]. The pairwise composite log-likelihood for data $u_{ij} = M^X_{n,i}(\boldsymbol{s}_j)$ from b independent blocks at m sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m$ is

$$\sum_{i=1}^{b} \sum_{(j,k)} w(\boldsymbol{s}_j, \boldsymbol{s}_k) \log f_{\boldsymbol{s}_j,\boldsymbol{s}_k}(u_{ij}, u_{ik}), \tag{15}$$

where the second sum goes through all ordered pairs (j, k) in $\{1, \ldots, m\}$ and the function $w : \mathcal{S} \times \mathcal{S} \to [0, \infty)$ may reweigh the contribution of information from each pair $(\boldsymbol{s}_j, \boldsymbol{s}_k)$. Uncertainty quantification can be achieved either via resampling or from a *sandwich variance matrix* [23, 98]. *Higher order composite likelihood* inference and *truncation* strategies have been considered comprehensively in [14], who investigate and discuss practical compromises between computational feasibility and estimator variance.

The second strategy is based on the observation that the complicated form of the likelihood (14) is due to the loss of information on occurrence times of maxima [124]. If we knew for each spatial maximum $M^X_n(\boldsymbol{s}_j) = \max\{X_1(\boldsymbol{s}_j), \ldots, X_n(\boldsymbol{s}_j)\}$ the time within the block when the maximum occurred, then we could retrieve a partition of the spatial sites into sets A whose maxima $\{M^X_n(\boldsymbol{s}_j)\}_{j \in A}$ arose from the same event. In the approximating limiting process this partitioning problem is related to a similar phenomenon in the construction of the spectral representation, cf. Figure 1. For each realization of Z^*, the underlying space $\mathcal{S}$ can be partitioned into (not necessarily connected) sets $\mathcal{A}$, so that on $\mathcal{A}$ only one and the same underlying spatial profile $\Gamma_k^{-1} V_k$ contributes to the final maximum process Z^*. Knowledge of the corresponding partition π for the site labels $\{1, \ldots, m\}$ would then render all other summands for other partitions in the likelihood (14) irrelevant. More precisely, the *augmented joint density* $f_{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_m;\Pi}(u_1, \ldots, u_m; \pi)$ of $(Z^*(\boldsymbol{s}_1), \ldots, Z^*(\boldsymbol{s}_m))^{\mathrm{T}}$ and the random partition Π on $\mathbb{R}^m \times \mathcal{P}$ is of the form

$$e^{-V^*(u_1,\ldots,u_m)} \prod_{A \in \pi} \Big\{ -\frac{\partial}{\partial u_A} V^*(u_1, \ldots, u_m) \Big\}, \tag{16}$$

so that the likelihood (14) arises as marginal likelihood of (16) integrated over all possible partitions. [124] suggest to employ the augmented likelihood (16) making use of the partition obtained from observed discrete occurrence times in block maxima. However, as noted in [61, 139], the resulting maximum likelihood estimator can be heavily biased. As an alternative, (16) can be used for inference based on the full likelihood treating the partition as hidden variable. For instance, maximum likelihood estimation can be performed via an *Expectation-Maximization (EM) algorithm* [62], while a *Bayesian framework* allows for sampling from the marginal posterior distribution [36, 131].

More recently, alternative composite likelihood techniques have been introduced for statistical inference for max-stable processes; [144] show computational gains from *symbolic data analysis*, [65] investigate advantages from the *Vecchia approximation* [137], and [56] propose a hybrid-approach that involves spatial partitioning, local modeling with censored pairwise likelihoods combined with moment-type inference.

4.3 Other approaches

Although most inference for max-stable spatial data has been based on likelihood techniques, other approaches have been explored. For instance, [45] avoid knowledge of the likelihood by considering *approximate Bayesian computation*, or [77, 111] take advantage of *neural networks* for inference for intractable likelihoods such as (14). In practice, we will never encounter data that are *exactly* from a max-stable process. It is more realistic to assume that the underlying data generating process is in the domain of attraction of such a process, cf. Section 2.3. The *M-estimator of spatial tail dependence* [42] is a generic moment-type estimator that has been proven to be consistent and asymptotically normal under quite general domain-of-attraction conditions. Other approaches based on single extreme events or threshold exceedances include

[3, 43, 44, 133, 140].

5 Simulation

Since closed-form expressions for functionals of interest such as (12) or (13) are in most cases not available for max-stable processes, statistical analysis will usually involve simulation from the respective spatial distributions on a sufficiently dense grid $\mathcal{G} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{|\mathcal{G}|}\}$. We distinguish unconditional and conditional simulation.

In the conditional case, information on the values of the max-stable process is already available at a set of conditioning sites $\mathcal{C} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{|\mathcal{C}|}\}$ for $|\mathcal{C}| < |\mathcal{G}|$. Besides this classical setting, where values of the max-stable process are assumed to be available at specific sites, also other information such as spatial averages over subregions could be included, see [86] for an application to downscaling.

5.1 Unconditional simulation

Existing (unconditional) simulation approaches have been compared systematically in [95]. Due the marginal transformation (2) we can restrict our attention to the simulation of simple max-stable processes of the form Z^* as in Theorem 2.1. The literature is governed by two main ideas.

Threshold stopping The first approach goes back to [115] and exploits that the arrival times $(\Gamma_k)_{k\geq 1}$ in de Haan's spectral representation form a renewal process with exponential inter-arrival times. Hence, the sequence of $(\Gamma_k)_{k\geq 1}$ will eventually surpass any given bound, and we would expect the contributions of the spatial profiles $\Gamma_k^{-1} V_k$ to the final maximum process Z^* to decrease in k until it becomes negligible or even irrelevant.

If the spectral process V has an almost-sure upper bound, say τ, on the simulation domain $\mathcal{G}$, then a new spatial profile $\Gamma_{k+1}^{-1} V_{k+1}$ cannot contribute anymore to the final maximum process Z^* (nor can any of the next ones with index $k+2$, $k+3$, ...) if the infimum (over $\mathcal{G}$) of the maximum of the first k contributions,

$$\inf_{\boldsymbol{s}\in\mathcal{G}} \max_{j=1,\ldots,k} \Gamma_j^{-1} V_j(\boldsymbol{s}),$$

lies already above $\Gamma_{k+1}^{-1}\tau$. Using this criterion as a stopping rule, and declaring the first such k as stopping time K, we have (cf. [89, 95])

$$\mathrm{E}(K) = \tau\, \mathrm{E}\big\{1/\inf_{\boldsymbol{s}\in\mathcal{G}} Z^*(\boldsymbol{s})\big\}. \tag{17}$$

Now let us recall that for a given max-stable process Z^*, the spectral process V in de Haan's representation is not uniquely determined by the law of Z^*. In many cases, the most natural process V may not even have an almost-sure upper bounded and we may have to consider alternative spectral processes for Z^*. Two canonical candidates emerge from the normalized spectral processes V^r from (5); if we consider $r(v_1, \ldots, v_{|\mathcal{G}|}) = \max(v_1, \ldots, v_{|\mathcal{G}|})$ (*sup-normalization*), V^r has upper bound $\tau = \theta_{\mathcal{G}}$ (the extremal coefficient of the simulation domain); if we consider $r(v_1, \ldots, v_{|\mathcal{G}|}) = v_1 + \cdots + v_{|\mathcal{G}|}$ (*sum-normalization*), V^r has upper bound $\tau = |\mathcal{G}|$. Sup- and sum-normalizations have been introduced in this context by [92] and [34, 35], respectively, where the focus of [34] lies on Brown–Resnick processes.

If one has to decide among a range of spectral processes V that are in principle available to simulate from for a given process Z^*, the general trade-off lies in weighing the computational cost of simulating from a specific spectral process V against the cost of obtaining Z^* from simulations of V. For instance, although $\theta_{\mathcal{G}} \leq |\mathcal{G}|$, it may be more difficult to simulate from the sup-normalized spectral process than from the sum-normalized one. In the case of Brown–Resnick processes, if one uses the generic rejection sampling approach of [26] to obtain samples from the sup-normalized spectral process, one can expect a similar overall effort for simulation from either sum- or sup-normalized representation to obtain samples from Z^*. On the other hand, efficiency gains to obtain samples from the sup-normalized V^r have been proposed in [57, 91].

In practice, it can pay off to deliberately terminate a simulation with the sum-normalized threshold stopping procedure earlier at a significantly lower threshold than $\tau = |\mathcal{G}|$. In the case of Brown–Resnick and extremal-t processes this strategy has been found to lead to a significantly reduced computational cost with almost negligible loss of accuracy [94, 95]. However, one should then conduct a range of diagnostic checks on the collection of obtained samples before using them to answer questions of interest.

Extremal functions When building the max-stable process Z^* as in de Haan's spectral representation, typically only very few spatial profiles $\varphi_k = \Gamma_k^{-1} V_k$ will actually contribute to the final maximum process, cf. Figure 1. The contributing φ_k have been termed *extremal functions* and the non-contributing ones *subextremal*. Based on theoretical results from [38] on conditional independencies among the involved point processes, [35] show that is is possible to derive an exact simulation algorithm that goes iteratively through all sites in the simulation domain $\mathcal{G}$ and simulates in each iteration the site-specific extremal function.

For $\boldsymbol{s}_j \in \mathcal{G}$ let $\varphi_+^{(j)}$ be the extremal function whose value at $\boldsymbol{s}_j$ coincides with $Z^*(\boldsymbol{s}_j)$, and let $V^{[j]} = V^r$ be the point-evaluation-normalized spectral process with V^r as in (5) for $r(v_1, \ldots, v_{|\mathcal{G}|}) = v_j$ (which necessarily satisfies $V^{[j]}(\boldsymbol{s}_j) = 1$ almost surely due to marginal standardization). Then the following recursive proce-

dure can be employed to obtain a sample from Z^*:

- First, simulate $\varphi_+^{(1)}$ as $\Gamma_1^{-1}V^{[1]}$.
- If the first j extremal functions $\varphi_+^{(1)}$, …, $\varphi_+^{(j)}$ are already given and

$$Z_+^{(j)}(\boldsymbol{s}_{j+1}) = \max\left\{\varphi_+^{(1)}(\boldsymbol{s}_{j+1}), \ldots, \varphi_+^{(j)}(\boldsymbol{s}_{j+1})\right\}, \tag{18}$$

sampling from the conditional law $\mathcal{L}(\varphi_+^{(j+1)} \mid \varphi_+^{(1)}, \ldots, \varphi_+^{(j)})$ is similar to the threshold stopping procedure above, but with the site-specific spectral process $V^{[j+1]}$ (instead of any other V) and using an additional rejection rule and an adapted stopping rule: A new proposal $\Gamma_k^{-1}V_k^{[j+1]}$ is accepted as extremal function $\varphi_+^{(j+1)}$ only if

$$\Gamma_k^{-1}V_k^{[j+1]}(\boldsymbol{s}_i) \leq \varphi_+^{(i)}(\boldsymbol{s}_i), \quad i = 1, \ldots, j,$$

i.e., if it is in accordance with all previously sampled extremal functions. Otherwise, the proposal is rejected. The procedure stops once a proposal is accepted or $\Gamma_k^{-1} = \Gamma_k^{-1}V_k^{[j+1]}(\boldsymbol{s}_{j+1})$ falls below (18). In the latter case, the extremal function $\varphi_+^{(j+1)}$ coincides with one of the previous ones.

The expected number of spectral processes $V \in \{V^{[1]}, \ldots, V^{[|\mathcal{G}|]}\}$ that need to be simulated for one sample of Z^* is $|\mathcal{G}|$ in this case. For Brown–Resnick and extremal-t processes this translates into the same expected number of Gaussian processes, and so the extremal functions algorithm outperforms the sum- and sup-normalized threshold stopping procedures in this case, as realizations from $\{V^{[1]}, \ldots, V^{[|\mathcal{G}|]}\}$ are easy to obtain. Similarly to the threshold stopping approaches, an early termination of an adaptive sampling strategy for the extremal functions approach will typically lead to efficiency gains at almost negligible cost of simulation accuracy, cf. [35]. In general, each of these strategies (of threshold stopping or extremal functions type) can outperform the others, depending on the desired Z^*, desired efficiency and accuracy. A more detailed discussion can be found in [95].

5.2 Conditional simulation

When the values $Z^*(\boldsymbol{s}_1) = z_1, \ldots, Z^*(\boldsymbol{s}_{|\mathcal{C}|}) = z_{|\mathcal{C}|}$ of the max-stable process Z^* are already available, this information should be taken into account. Similarly to the extremal functions approach for unconditional simulation, conditional simulation approaches are based on the theoretical results in [38] supplemented by algorithmic aspects developed in [39]. A key element is the random partition Π of $\{1, \ldots, |\mathcal{C}|\}$ that is induced by the extremal functions. With this in mind, conditional simulations can be obtained by the following three-step procedure:

Step 1. *Sample a partition π from $\mathcal{L}(\Pi \,|\, Z^*(\boldsymbol{s}_1) = z_1, \ldots, Z^*(\boldsymbol{s}_{|\mathcal{C}|}) = z_{|\mathcal{C}|})$:* Up to a normalizing constant, its probability mass function is obtained from

$$\begin{aligned}\mathrm{P}\big\{\Pi = \pi \,\big|\, Z^*(\boldsymbol{s}_1) = z_1, \ldots, Z^*(\boldsymbol{s}_{|\mathcal{C}|}) = z_{|\mathcal{C}|}\big\} \\ \propto f_{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_{|\mathcal{C}|};\Pi}(z_1, \ldots . z_{|\mathcal{C}|}; \pi),\end{aligned} \tag{19}$$

where $f_{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_{|\mathcal{C}|};\Pi}$ is the augmented likelihood in (16). [39] suggest using a Gibbs sampler.

Step 2. *Simulate the extremal functions $\{\varphi_*^{(A)}\}_{A\in\pi}$ conditional on the data from the previous step*: For $A \in \pi$ the extremal function $\varphi_*^{(A)}$ needs to satisfy $\varphi_*^{(A)}(\boldsymbol{s}_i) = z_i$ for all $i \in A$. Such a function can be obtained by conditional sampling from the spectral process $V^{[i]}$.

Step 3. *Simulate the remaining extremal functions that are subextremal for the conditioning sites $\mathcal{C}$:*

Due to conditional independence of extremal and subextremal functions, all relevant subextremal functions that could potentially contribute to the maximum outside the conditioning points can be obtained from an independent process $\{\widetilde{Z}(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ given by

$$\widetilde{Z}(\boldsymbol{s}) = \max_{k\in\mathcal{I}_-} \Gamma_k^{-1}V_k(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{S},$$

where

$$\mathcal{I}_- = \{k \in \mathbb{N} : \ \Gamma_k^{-1}V_k(\boldsymbol{s}_i) \leq z_i, \ i = 1, \ldots, |\mathcal{C}|\}.$$

Samples from the process $\widetilde{Z}$ can be obtained in the same way as unconditional samples from Z^* (either via threshold stopping or the extremal functions approach) rejecting those functions $\Gamma_k^{-1}V_k$ that do not satisfy the conditions defining $\mathcal{I}_-$. A conditional sample from Z^* is then obtained as the pointwise maximum of the extremal functions and the process $\widetilde{Z}$, cf. also [88] for more details.

6 Application

Besides numerous examples that are mostly concerned with hydrological extremes, e.g. [1, 7, 12, 46, 103, 113, 114, 120, 118, 121, 125, 132, 134, 143], statistical inference that involves max-stable processes can meanwhile be found across a broader spectrum of the environmental literature. They appear, for instance, in spatial models of temperature minima and maxima [45, 76, 105, 119, 131], droughts [93], forest fire danger [126], air pollution extremes [138], extreme ocean environments [69], snowfalls and snowdepths [8, 50, 82], or wind gusts [17, 51, 90, 107].

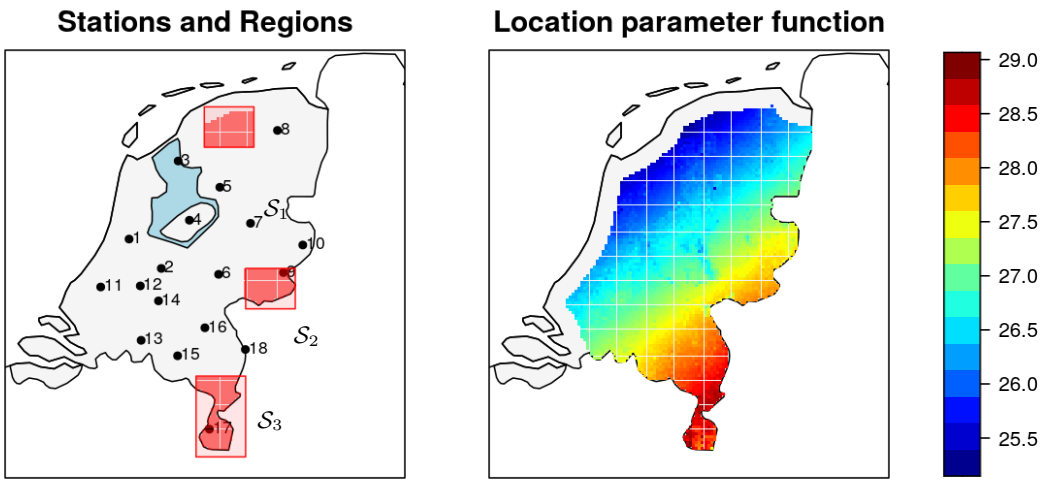

Figure 4: Map of gauging stations 1–18 for Dutch temperature data and three regions $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$ of interest (left); estimated GEV location function $\widehat{\mu}(\boldsymbol{s})$ for sites s on an inland grid that is 15 km away from the Dutch coast with grid distance approximately 2.5 km (right).

To provide a simple illustration of the steps of a spatial data analysis as outlined in Section 4, let us reconsider here the first hypothetical situation from Section 1: We assume that health officials have identified three regions $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$ (Figure 4) to be of particular concern, and we seek to determine the probability of the following event: *During one summer the temperature is rising above* 40°C *in all three regions during the same fortnight.* A similar analysis has been conducted in [95]. Our analysis here is built on the same publicly available data set of Dutch summer temperatures, which can be accessed in a preprocessed form at

`https://github.com/strokorb/`
`max-stable-spatial-inference`

alongside all of the related code for this analysis. Specifically, we use the same 180 block maxima over 14 day-periods over 30 summers (each summer consists of 6 such blocks) at 18 stations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{18} \in \mathcal{S}$ (Figure 4), and we assume a max-stable process of the following form as underlying data-generating mechanism[1]:

$$Z(\boldsymbol{s}) = \mu(\boldsymbol{s}) + (\sigma/\xi) \cdot \big\{ Z^*(\boldsymbol{s})^\xi - 1 \big\}, \quad \boldsymbol{s} \in \mathcal{S},$$

where the location function μ absorbs geographical information as

$$\begin{aligned} \mu(\boldsymbol{s}) = \mu_0 &+ \mu_1 \, \texttt{longitude}(\boldsymbol{s}) \\ &+ \mu_2 \, \texttt{latitude}(\boldsymbol{s}) \\ &+ \mu_2 \, \texttt{altitude}(\boldsymbol{s}), \end{aligned}$$

and Z^* is an isotropic max-stable Brown–Resnick process based on the power-variogram family (9), which adds two more dependence parameters to the model. Complementing the route taken in [95], where marginals and dependence structure were estimated in two steps using the GEV independence likelihood and the M-estimator approach, we follow here a one-step approach estimating marginal and dependence parameters jointly using the pairwise composite likelihood approach with its implementation in the package `SpatialExtremes` (cf. Section 4). Weight factors in the composite likelihood (15) were chosen such that they down-weigh contributions of pairs with large geographical distance.

The resulting marginal GEV estimates are of Weibull-type ($\widehat{\xi} \approx -0.126$ with standard error 0.032). Figure 5 shows three QQ-plots of station-specific block maxima after location-scale standardization against the model quantiles of $\mathrm{GEV}(0, 1, \widehat{\xi})$. Note that we cannot expect the data to be as well aligned with the diagonal as if we had fitted the marginal GEV model separately based on the independence likelihood. The estimated location function $\widehat{\mu}(\boldsymbol{s})$ is displayed in Figure 4. As one might expect more northerly locations or those closer to the coast experience generally cooler summer temperatures than the southern inland regions.

Beyond marginal diagnostics, the next natural step is a comparison of pairwise (or higher order) features of the dependence model against non-parametric estimates of such dependence features. Figure 6 displays the theoretical extremal coefficient function $\theta(\boldsymbol{s}_i - \boldsymbol{s}_j)$ obtained from the fitted model alongside pairwise estimates of bivariate extremal coefficients $\theta(\boldsymbol{s}_i, \boldsymbol{s}_j)$ for the station pairs $(\boldsymbol{s}_i, \boldsymbol{s}_j)$. As non-parametric estimator the F-madogram [19] has been used here. Alternatively, one might consider for instance the λ-madogram [81] or the Capéraà–

[1]For notational convenience we suppress the block size index n here that we used to keep track of earlier in (2).

Fougères–Genest procedure of [13]; [116] suggest an additional sequential correction procedure to ensure self-consistency of such estimates. The overall rather low range of extremal coefficients (below 1.3) confirms our initial assumption of a strong extremal dependence among high temperatures even across large distances of 250 km.

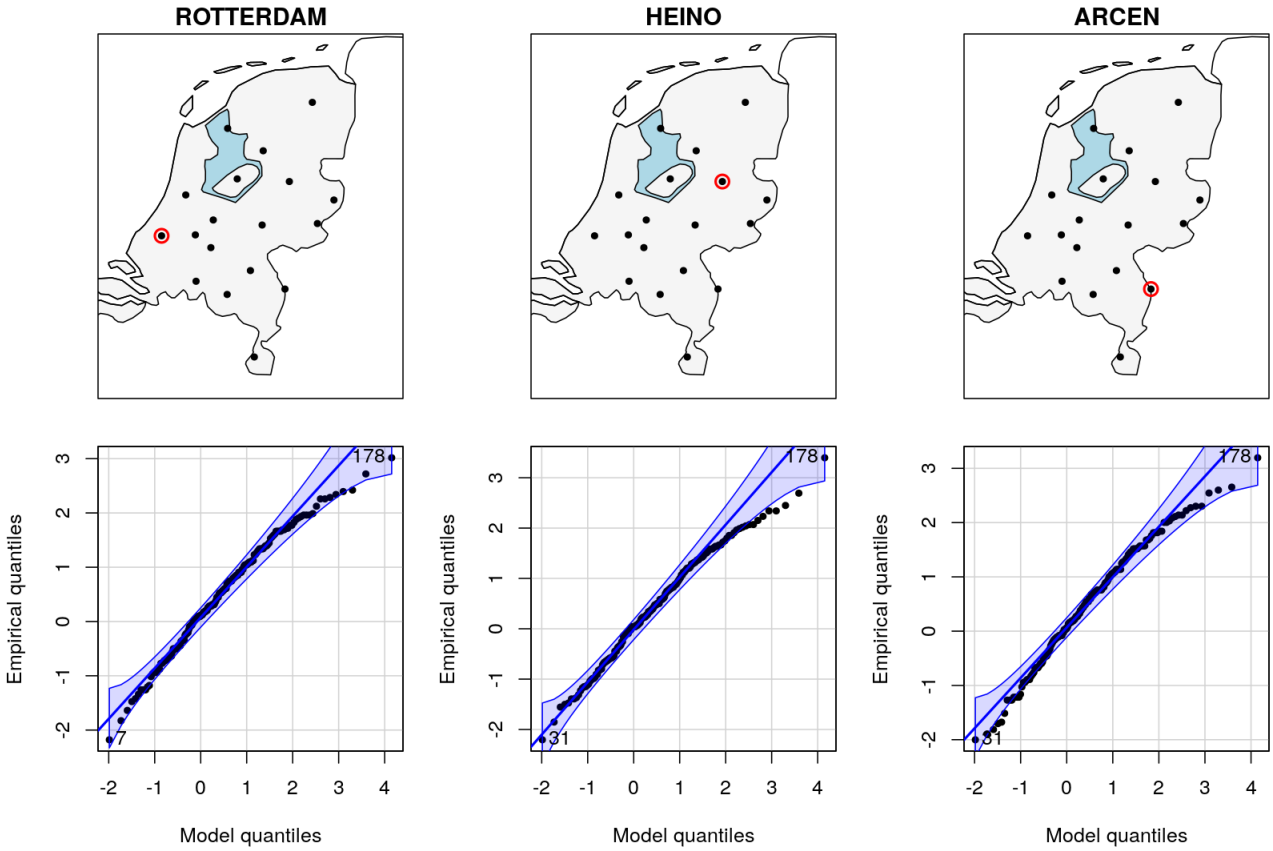

Figure 5: QQ-plots of location-scale transformed empirical quantiles versus model quantiles of GEV$(0, 1, \widehat{\xi})$ for three stations.

Finally, in order to estimate

$$p = \mathrm{P}\Big\{ \max_{\boldsymbol{s}\in\mathcal{S}_1} Z(\boldsymbol{s}) > \tau, \max_{\boldsymbol{s}\in\mathcal{S}_2} Z(\boldsymbol{s}) > \tau, \max_{\boldsymbol{s}\in\mathcal{S}_3} Z(\boldsymbol{s}) > \tau \Big\},$$

for $\tau = 40°\mathrm{C}$, we simulate 30 000 realizations from the fitted max-stable process for Z on the same inland grid as in [95]. Numerical efficiency is achieved by using the Dieker–Mikosch algorithm [34] with a relative threshold of 10%. Within this procedure we simulate Gaussian processes based on an incomplete Cholesky-decomposition of the relevant kernel matrix [74]. Since each summer consists of six 14-day blocks, this corresponds to the simulation of data for $5\,000 = 30\,000/6$ summers. Figure 7 displays three such realizations on the entire grid. In view of our original question, one can gain even more efficiency by simulating the max-stable processes only on the grid points of the three regions $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$ in order to obtain 30 000 samples from our estimated model for

$$\Big\{ \max_{\boldsymbol{s}\in\mathcal{S}_1} Z(\boldsymbol{s}) \,,\, \max_{\boldsymbol{s}\in\mathcal{S}_2} Z(\boldsymbol{s}) \,,\, \max_{\boldsymbol{s}\in\mathcal{S}_3} Z(\boldsymbol{s}) \Big\}.$$

Scatterplots of the pairs from this vector are displayed as well in Figure 7. Among these 30 000 sampled events there are 153 with a joint exceedance of 40°C. Hence, we estimate $\widehat{p} = 0.0051$ for the probability of such an event and obtain $1 - (1 - \widehat{p})^6 \approx 6\widehat{p} \approx 3\,\%$ for the chance of a summer, during which a temperature record of 40°C is broken simultaneously in all three regions during a 14-day interval. If we interpret the result for the first hypothetical situation from Section 1, contingency plans would need to be adapted, as the estimated chance exceeds the 2.5% level.

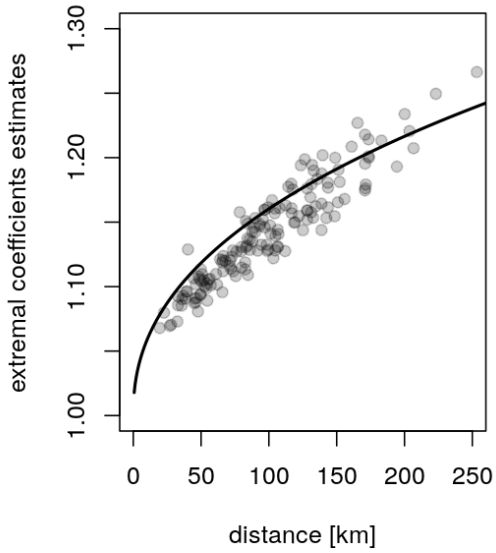

Figure 6: Estimated extremal coefficient function of the fitted model (solid line) alongside non-parametric estimates of extremal coefficients from station pairs based on the F-madogram (points).

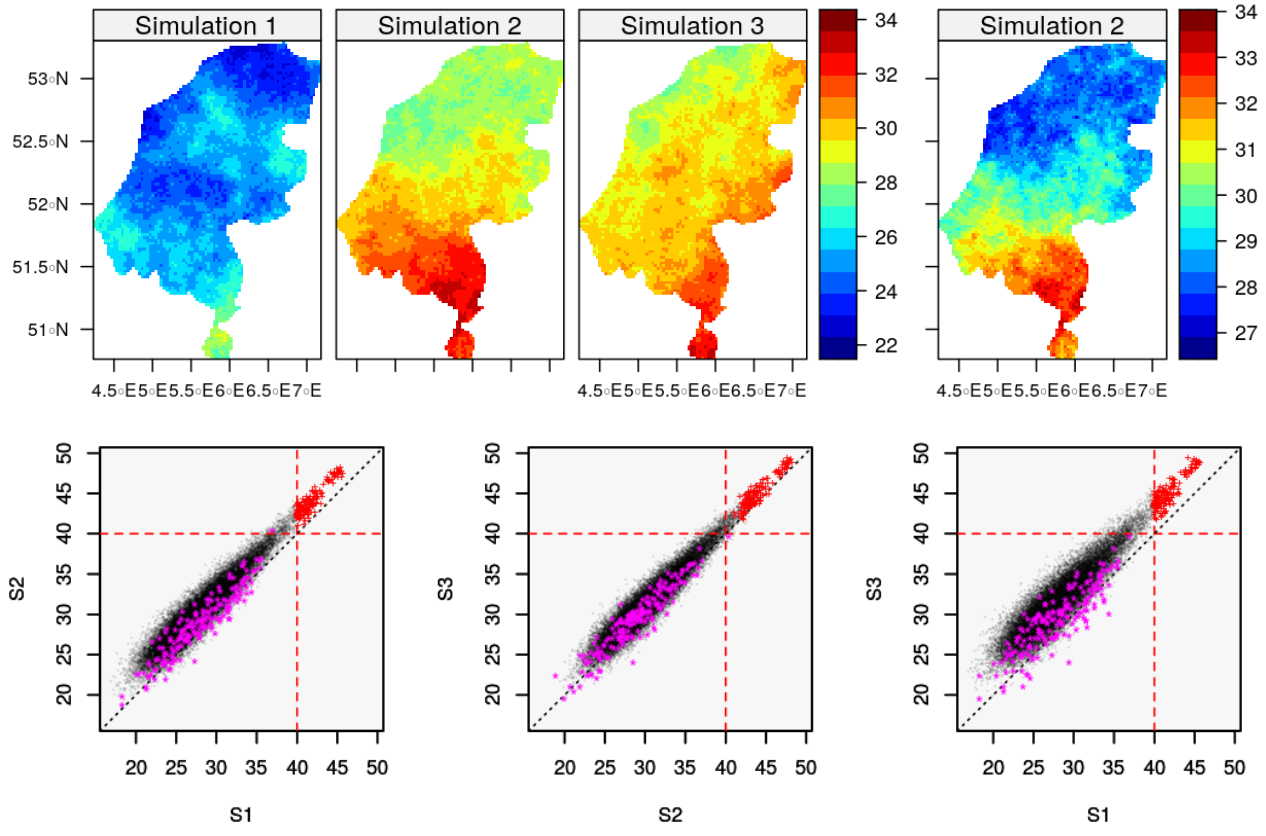

Figure 7: Top row: Three samples from the fitted max-stable process; each simulation represents a field of 14-day maxima of summer temperatures; Simulation 2 is shown separately on a different color scale. Bottom row: Scatterplots of pairs of 30 000 simulated areal 14-day summer temperature maxima across the regions $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$; data points that correspond to a joint exceedance of 40°C across all three regions have been marked with a red plus (+). Stars (∗) in magenta are taken from the original temperature data if we let stations 8, 9, 17 represent regions $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$ (cf. Figure 4).

Warning! Due to the indicative nature of this analysis, we have omitted several steps that are imperative in any decision-relevant data analysis. This includes proper uncertainty quantification, a more thorough model validation, sensitivity checks for results and, last but not least, more attention to domain knowledge with relevance for the original question. Here we have only considered a hypothetical situation for illustrative purposes.

7 Concluding remarks

Max-stable processes are keeping the scientific community busy since several decades. In the 1980s and early 1990s they entered the stage of probability theory as the only limiting stochastic processes for maxima of stochastic processes under site-wise location-scale normalization (Section 2). With the arrival of suitable computational capabilities to simulate from such models since the early 2000s (Section 5) and the proposal of parsimonious spatial models that were able to incorporate more and more realistically looking features (Section 3), they have become increasingly attractive as a geostatistical model (Section 6) to answer questions about spatial tail extrapolation as we have seen in Section 1. Most of the estimation methodology has been introduced and refined since the early 2010s (Section 4). In parallel, their ergodic properties have been revealed (Section 2.4). Due to their intimite links to regular variation and point process theory further connections to time series analysis or long-range dependence are still being explored, cf. e.g. [68, 87, 100].

This chapter about max-stable processes puts an emphasis on the knowledge that has been built surrounding their use as a spatial geostatistical model for tail regions. We have explored indicative questions (Section 1) and associated inference (Sections 4 and 6) to give an idea, when such a model may be considered and how to carry out the steps involved. Naturally, as with any statistical tool or modelling paradigm, this does not go without warnings, nor do we want to suggest to use the block maxima framework as a default methodology. Rather, we consider it an important benchmark, especially in situations when data availability may be restricted to block maxima. Wherever possible, we have tried to emphasize critical points concerning inference with max-stable process. Beyond that we would like to draw attention to the recent opinion piece [64], which is a good reminder why a statistical modeller should not fall in love with just one model, and instead reflect on its limitations and suitability. At the same time it is a call to carry out extensive diagnostic checks.

To begin with, inference with and simulation from max-stable models can be computationally challenging as we have seen in Sections 4 and 5 despite a range of ideas how to navigate these difficulties. Often (not only) the computational issues are better eased by turning to a modelling paradigm based on a risk functional exceeding a high threshold rather than sticking with the block maxima framework (Chapter 16). The theoretical grounds of such a threshold approach are at least closely connected to the maxima convergence (cf. Section 2.3). Further practical reasons to turn to such an approach are summarised in [27]. On the other hand, for good or for bad, a peaks-over-threshold approach will also require the modeller to answer a larger number of questions about the underlying data generating processes that are linked to event detections and related choices of risk functionals and thresholds. In the univariate case [11] provide a comprehensive comparison between maxima and threshold approaches. Several of their considerations transfer to a spatial context. At the same time spatial settings are much more complex, threshold-based approaches still under-explored, and we may encounter fitness for purpose questions on an even more fundamental level.

After all, situations where we might involve spatial models for extreme values are about extrapolation into the tail regions of a stochastic process (cf. our guiding questions from Section 1). For max-stable models (this chapter) or threshold-stable models (Chapter 16) such extrapolation is guided by the respective stability properties of these models that amount to an underlying regular variation assumption (Section 2.3). However, in many environmental data examples, still observable levels of data in the tail regions seem to exhibit less rigid tail behaviour. They deviate from the hypothesized stability properties in the sense that extreme events appear more localized the more extreme they are. Spatial statistical models which behave in this way have been loosely termed sub-asymptotic models for spatial extremes; in this book they are introduced in Chapter 17.

In practice, it may be difficult to distinguish these different modelling regimes. For instance, it has been noted in [46] that even simulated data from max-stable models seem to exhibit asymptotic independent behaviour, at least at large enough distances between sites. To a certain extent the choice of model will always depend on either beliefs about the underlying processes or risk preferences. Asymptotic dependent models such as the stable ones will often lead to more conservative estimates of tail probabilities. We anticipate further research along the lines of these practically relevant decisions in years to come.

References

[1] P. Asadi, A. C. Davison, and S. Engelke. Extremes on river networks. *Ann. Appl. Stat.*, 9(4):2023–2050, 2015.

[2] J.-N. Bacro and C. Gaetan. A review on spatial extreme modelling. In E. Porcu, J. M. Montero, and M. Schlather, editors, *Advances and Challenges in Space-time Modelling of Natural Events*, pages 103–124, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[3] J.-N. Bacro and C. Gaetan. Estimation of spatial max-stable models using threshold exceedances. *Stat. Comput.*, 24(4):651–662, 2014.

[4] L. R. Belzile, C. Dutang, P. J. Northrop, and T. Opitz. A modeler's guide to extreme value software. *Extremes*, 26(4):595–638, 2023.

[5] B. Beranger, S. A. Padoan, and S. A. Sisson. Models for extremal dependence derived from skew-symmetric families. *Scand. J. Stat.*, 44(1):21–45, 2017.

[6] B. Beranger, A. G. Stephenson, and S. A. Sisson. High-dimensional inference using the extremal skew-t process. *Extremes*, 24(3):653–685, 2021.

[7] J. Blanchet, C. Aly, T. Vischel, G. Panthou, Y. Sané, and M. Diop Kane. Trend in the co-occurrence of extreme daily rainfall in West Africa since 1950. *Journal of Geophysical Research: Atmospheres*, 123(3):1536–1551, 2018.

[8] J. Blanchet and A. C. Davison. Spatial modeling of extreme snow depth. *Ann. Appl. Stat.*, 5(3):1699–1725, 2011.

[9] Bruce M. Brown and Sidney I. Resnick. Extreme values of independent stochastic processes. *J. Appl. Probability*, 14(4):732–739, 1977.

[10] A. Bücher and J. Segers. On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes*, 20(4):839–872, 2017.

[11] A. Bücher and C. Zhou. A horse race between the block maxima method and the peak-over-threshold approach. *Statist. Sci.*, 36(3):360–378, 2021.

[12] T. A. Buishand, L. de Haan, and C. Zhou. On spatial extremes: with application to a rainfall problem. *Ann. Appl. Stat.*, 2(2):624–642, 2008.

[13] P. Capéraà, A.-L. Fougères, and C. Genest. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3):567–577, 1997.

[14] S. Castruccio, R. Huser, and M. G. Genton. High-order composite likelihood inference for max-stable distributions and processes. *J. Comput. Graph. Statist.*, 25(4):1212–1229, 2016.

[15] R. E. Chandler and S. Bate. Inference for clustered data using the independence loglikelihood. *Biometrika*, 94(1):167–183, 2007.

[16] C. Chevalier, O. Martius, and D. Ginsbourger. Modeling nonstationary extreme dependence with stationary max-stable processes and multidimensional scaling. *J. Comput. Graph. Statist.*, 30(3):745–755, 2021.

[17] S. G. Coles and D. Walshaw. Directional modelling of extreme wind speeds. *J. R. Stat. Soc. Ser. C*, 43(1):139–157, 1994.

[18] D. Cooley, J. Cisewski, R. J. Erhardt, S. Jeon, E. Mannshardt, B. O. Omolo, and Y. Sun. A survey of spatial extremes: measuring spatial dependence and modeling spatial effects. *REVSTAT*, 10(1):135–165, 2012.

[19] D. Cooley, P. Naveau, and P. Poncet. Variograms for spatial max-stable random fields. In *Dependence in Probability and Statistics*, volume 187 of *Lect. Notes Stat.*, pages 373–390. Springer, New York, 2006.

[20] R. A. Davis and T. Mikosch. Extreme value theory for space-time processes with heavy-tailed distributions. *Stochastic Process. Appl.*, 118(4):560–584, 2008.

[21] R. A. Davis and T. Mikosch. The extremogram: a correlogram for extreme events. *Bernoulli*, 15(4):977–1009, 2009.

[22] A. C. Davison and M. M. Gholamrezaee. Geostatistics of extremes. *Proc. R. Soc. Lond. Ser. A*, 468(2138):581–608, 2012.

[23] A. C. Davison, R. Huser, and E. Thibaud. Spatial extremes with application to climate and environmental exposure. In A. E. Gelfand, M. Fuentes, J. A. Hoeting, and R. L. Smith, editors, *Handbook of Environmental and Ecological Statistics*. CRC Press, 2018.

[24] A. C. Davison, S. A. Padoan, and M. Ribatet. Statistical modeling of spatial extremes. *Stat. Sci.*, 27(2):161–186, 2012.

[25] Y. Davydov, I. Molchanov, and S. Zuyev. Strictly stable distributions on convex cones. *Electron. J. Probab.*, 13:no. 11, 259–321, 2008.

[26] R. de Fondeville and A. C. Davison. High-dimensional peaks-over-threshold inference. *Biometrika*, 105(3):575–592, 2018.

[27] R. de Fondeville and A. C. Davison. Functional peaks-over-threshold analysis. *J. R. Stat. Soc. Ser. B*, 84(4):1392–1422, 2022.

[28] L. de Haan. A spectral representation for max-stable processes. *Ann. Probab.*, 12(4):1194–1204, 1984.

[29] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer, Berlin, 2006.

[30] L. de Haan and T. Lin. On convergence toward an extreme value distribution in $C[0,1]$. *Ann. Probab.*, 29(1):467–483, 2001.

[31] L. de Haan and J. Pickands, III. Stationary min-stable stochastic processes. *Probab. Theory Relat. Fields*, 72(4):477–492, 1986.

[32] L. de Haan and S. I. Resnick. Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 40(4):317–337, 1977.

[33] S. Demarta and A. J. McNeil. The t copula and related copulas. *International Statistical Review / Revue Internationale de Statistique*, 73(1):111–129, 2005.

[34] A. B. Dieker and T. Mikosch. Exact simulation of Brown–Resnick random fields at a finite number of locations. *Extremes*, 18(2):301–314, 2015.

[35] C. Dombry, S. Engelke, and M. Oesting. Exact simulation of max-stable processes. *Biometrika*, 103(2):303–317, 2016.

[36] C. Dombry, S. Engelke, and M. Oesting. Bayesian inference for multivariate extreme value distributions. *Electron. J. Stat.*, 11(2):4813–4844, 2017.

[37] C. Dombry and F. Eyi-Minko. Strong mixing properties of max-infinitely divisible random fields. *Stochastic Process. Appl.*, 122(11):3790–3811, 2012.

[38] C. Dombry and F. Eyi-Minko. Regular conditional distributions of continuous max-infinitely divisible random fields. *Electron. J. Probab.*, 18:1–21, 2013.

[39] C. Dombry, F. Eyi-Minko, and M. Ribatet. Conditional simulation of max-stable processes. *Biometrika*, 100(1):111–124, 2013.

[40] C. Dombry and Z. Kabluchko. Ergodic decompositions of stationary max-stable processes in terms of their spectral functions. *Stochastic Process. Appl.*, 127(6):1763–1784, 2017.

[41] C. Dombry and M. Ribatet. Functional regular variations, Pareto processes and peaks over threshold. *Stat. Interface*, 8(1):9–17, 2015.

[42] J. H. J. Einmahl, A. Kiriliouk, A. Krajina, and J. Segers. An M-estimator of spatial tail dependence. *J. R. Stat. Soc. Ser. B*, 78(1):275–298, 2016.

[43] S. Engelke, A. Malinowski, Z. Kabluchko, and M. Schlather. Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *J. R. Stat. Soc. Ser. B*, 77(1):239–265, 2015.

[44] S. Engelke, A. Malinowski, M. Oesting, and M. Schlather. Statistical inference for max-stable processes by conditioning on extreme events. *Adv. in Appl. Probab.*, 46(2):478–495, 2014.

[45] R. J. Erhardt and R. L. Smith. Approximate Bayesian computing for spatial extremes. *Comput. Statist. Data Anal.*, 56(6):1468–1481, 2012.

[46] L. Fawcett and D. Walshaw. Estimating the probability of simultaneous rainfall extremes within a region: a spatial approach. *J. Appl. Stat.*, 41(5):959–976, 2014.

[47] A. Ferreira and L. de Haan. The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4):1717–1737, 2014.

[48] A.-L. Fougères, C. Mercadier, and J. P. Nolan. Dense classes of multivariate extreme value distributions. *J. Multivariate Anal.*, 116:109–129, 2013.

[49] A.-L. Fougères, J. P. Nolan, and H. Rootzén. Models for dependent extremes using stable mixtures. *Scand. J. Stat.*, 36(1):42–59, 2009.

[50] J. Gaume, N. Eckert, G. Chambon, M. Naaim, and L. Bel. Mapping extreme snowfalls in the French Alps using max-stable processes. *Water Resour. Res.*, 49(2):1079–1098, 2013.

[51] M. G. Genton, S. A. Padoan, and H. Sang. Multivariate max-stable spatial processes. *Biometrika*, 102(1):215–230, 02 2015.

[52] E. Giné, M. G. Hahn, and P. Vatan. Max-infinitely divisible and max-stable sample continuous processes. *Probab. Theory Related Fields*, 87(2):139–165, 1990.

[53] T. Gneiting. Compactly supported correlation functions. *J. Multivariate Anal.*, 83(2):493–508, 2002.

[54] T. Gneiting and Z. Sasvári. The characterization problem for isotropic covariance functions. *Math. Geol.*, 31(1):105–111, 1999.

[55] T. Gneiting, Z. Sasvári, and M. Schlather. Analogies and correspondences between variograms and covariance functions. *Adv. in Appl. Probab.*, 33(3):617–630, 2001.

[56] E. C. Hector and B. J. Reich. Distributed inference for spatial extremes modeling in high dimensions. *J. Amer. Statist. Assoc.*, 119(546):1297–1308, 2024.

[57] Z. W. O. Ho and C. Dombry. Simple models for multivariate regular variation and the Hüsler–Reiß Pareto distribution. *J. Multivariate Anal.*, 173:525–550, 2019.

[58] T. Hsing, J. Hüsler, and R.-D. Reiss. The extremes of a triangular array of normal random variables. *Ann. Appl. Probab.*, 6(2):671–686, 1996.

[59] H. Hult and F. Lindskog. Extremal behavior of regularly varying stochastic processes. *Stochastic Process. Appl.*, 115(2):249–274, 2005.

[60] R. Huser and A. C. Davison. Composite likelihood estimation for the Brown–Resnick process. *Biometrika*, 100(2):511–518, 2013.

[61] R. Huser, A. C. Davison, and M. G. Genton. Likelihood estimators for multivariate extremes. *Extremes*, 19(1):79–103, 2016.

[62] R. Huser, C. Dombry, M. Ribatet, and M. G. Genton. Full likelihood inference for max-stable data. *Stat*, 8:e218, 14, 2019.

[63] R. Huser and M. G Genton. Non-stationary dependence structures for spatial extremes. *J. Agric. Biol. Environ. Stat.*, 21(3):470–491, 2016.

[64] R. Huser, T. Opitz, and J. Wadsworth. Modeling of spatial extremes in environmental data science: Time to move away from max-stable processes. *Environ. Data Sci.*, 2024+. To appear.

[65] R. Huser, M. L. Stein, and P. Zhong. Vecchia likelihood approximation for accurate and fast inference with intractable spatial max-stable models. *J. Comput. Graph. Stat.*, pages 1–22, 2023.

[66] R. Huser and J. L. Wadsworth. Advances in statistical modeling of spatial extremes. *Wiley Interdiscip. Rev. Comput. Stat.*, 14(1):e1537, 2022.

[67] J. Hüsler and R.-D. Reiss. Maxima of normal random vectors: between independence and complete dependence. *Statist. Probab. Lett.*, 7(4):283–286, 1989.

[68] A. Janßen. Spectral tail processes and max-stable approximations of multivariate regularly varying time series. *Stochastic Process. Appl.*, 129(6):1993–2009, 2019.

[69] P. Jonathan and K. Ewans. Statistical modelling of extreme ocean environments for marine design: A review. *Ocean Engineering*, 62:91–109, 2013.

[70] Z. Kabluchko. Spectral representations of sum- and max-stable processes. *Extremes*, 12(4):401–424, 2009.

[71] Z. Kabluchko. Extremes of independent Gaussian processes. *Extremes*, 14(3):285–310, 2011.

[72] Z. Kabluchko and M. Schlather. Ergodic properties of max-infinitely divisible processes. *Stochastic Process. Appl.*, 120(3):281–295, 2010.

[73] Z. Kabluchko, M. Schlather, and L. de Haan. Stationary max-stable fields associated to negative definite functions. *Ann. Probab.*, 37(5):2042–2065, 2009.

[74] A. Karatzoglou, A. Smola, and K. Hornik. *kernlab: Kernel-Based Machine Learning Lab*, 2023. R package version 0.9-32.

[75] P. Krupskii, H. Joe, D. Lee, and M. G. Genton. Extreme-value limit of the convolution of exponential and multivariate normal distributions: link to the Hüsler–Reiß distribution. *J. Multivariate Anal.*, 163:80–95, 2018.

[76] Y. Lee, S. Yoon, M. S. Murshed, M.-K. Kim, C. Cho, H.-J. Baek, and J.-S. Park. Spatial modeling of the highest daily maximum temperature in Korea via max-stable processes. *Adv. Atmos. Sci.*, 30(6):1608–1620, Nov 2013.

[77] A. Lenzi, J. Bessac, J. Rudi, and M. L. Stein. Neural networks for parameter estimation in intractable models. *Comput. Statist. Data Anal.*, 185:107762, 2023.

[78] B. G. Lindsay. Composite likelihood methods. In *Statistical Inference from Stochastic Processes (Ithaca, NY, 1987)*, volume 80 of *Contemp. Math.*, pages 221–239. Amer. Math. Soc., Providence, RI, 1988.

[79] I. Molchanov, M. Schmutz, and K. Stucki. Invariance properties of random vectors and stochastic processes based on the zonoid concept. *Bernoulli*, 20(3):1210–1233, 2014.

[80] I. Molchanov and K. Strokorb. Max-stable random supmeasures with comonotonic tail dependence. *Stochastic Process. Appl.*, 126(9):2835–2859, 2016.

[81] P. Naveau, A. Guillou, D. Cooley, and J. Diebolt. Modelling pairwise dependence of maxima in space. *Biometrika*, 96(1):1–17, 2009.

[82] G. Nicolet, N. Eckert, S. Morin, and J. Blanchet. Assessing climate change impact on the spatial dependence of extreme snow depth maxima in the French alps. *Water Resour. Res.*, 54(10):7820–7840, 2018.

[83] A. K. Nikoloulopoulos, H. Joe, and H. Li. Extreme value properties of multivariate t copulas. *Extremes*, 12(2):129–148, 2009.

[84] T. Norberg. Random capacities and their distributions. *Probab. Theory Related Fields*, 73(2):281–297, 1986.

[85] M. Oesting. Equivalent representations of max-stable processes via ℓ^p-norms. *J. Appl. Probab.*, 55(1):54–68, 2018.

[86] M. Oesting, L. Bel, and C. Lantuéjoul. 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, 2018.

[87] M. Oesting and A. Rapp. Long memory of max-stable time series as phase transition: asymptotic behaviour of tail dependence estimators. *Electron. J. Stat.*, 17(2):3316–3336, 2023.

[88] M. Oesting, M. Ribatet, and C. Dombry. Conditional simulation of max-stable processes. In D. K. Dey and J. Yan, editors, *Extreme Value Modeling and Risk Analysis: Methods and Applications*, pages 215–238. CRC Press, Boca Raton, 2016.

[89] M. Oesting, M. Ribatet, and C. Dombry. Simulation of max-stable processes. In D. K. Dey and J. Yan, editors, *Extreme Value Modeling and Risk Analysis: Methods and Applications*, pages 195–214. CRC Press, Boca Raton, 2016.

[90] M. Oesting, M. Schlather, and P. Friederichs. Statistical post-processing of forecasts for extremes using bivariate Brown–Resnick processes with an application to wind gusts. *Extremes*, 20(2):309–332, 2017.

[91] M. Oesting, M. Schlather, and C. Schillings. Sampling sup-normalized spectral functions for Brown–Resnick processes. *Stat*, 8(1):e228, 2019.

[92] M. Oesting, M. Schlather, and C. Zhou. Exact and fast simulation of max-stable processes on a compact set using the normalized spectral representation. *Bernoulli*, 24(2):1497–1530, 2018.

[93] M. Oesting and A. Stein. Spatial modeling of drought events using max-stable processes. *Stoch. Environ. Res. Risk Assess.*, 32:63–81, 2018.

[94] M. Oesting and K. Strokorb. Efficient simulation of Brown–Resnick processes based on variance reduction of Gaussian processes. *Adv. in Appl. Probab.*, 50(4):1155–1175, 2018.

[95] M. Oesting and K. Strokorb. A comparative tour through the simulation algorithms for max-stable processes. *Statist. Sci.*, 37(1):42–63, 2022.

[96] T. Opitz. Extremal t processes: elliptical domain of attraction and a spectral representation. *J. Multivariate Anal.*, 122:409–413, 2013.

[97] C. J. Paciorek and M. J. Schervish. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17(5):483–506, 2006.

[98] S. A. Padoan, M. Ribatet, and S. A. Sisson. Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.*, 105(489):263–277, 2010.

[99] M. D. Penrose. Semi-min-stable processes. *Ann. Probab.*, 20(3):1450–1463, 1992.

[100] H. Planinić. Palm theory for extremes of stationary regularly varying time series and random fields. *Extremes*, 26(1):45–82, 2023.

[101] H. Planinić and P. Soulier. The tail process revisited. *Extremes*, 21(4):551–579, 2018.

[102] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023.

[103] C. R. Rajulapati and P. P. Mujumdar. Dependence structure of urban precipitation extremes. *Adv. Water Resour.*, 121:206–218, 2018.

[104] B. J. Reich and B. A. Shaby. A hierarchical max-stable spatial model for extreme precipitation. *Ann. Appl. Stat.*, 6(4):1430–1451, 2012.

[105] B. J. Reich, B. A. Shaby, and D. Cooley. A hierarchical model for serially-dependent extremes: a study of heat waves in the western US. *J. Agric. Biol. Environ. Stat.*, 19(1):119–135, 2014.

[106] S. I. Resnick and R. Roy. Random usc functions, max-stable processes and continuous choice. *Ann. Appl. Probab.*, 1(2):267–292, 1991.

[107] M. Ribatet. Spatial extremes: Max-stable processes at work. *Journal de la Société Française de Statistique*, 154(2):156–177, 2013.

[108] Mathieu Ribatet. *SpatialExtremes: Modelling Spatial Extremes*, 2022. R package version 2.1-0.

[109] H. Rootzén and R. W. Katz. Design life level: Quantifying risk in a changing climate. *Water Resour. Res.*, 49(9):5964–5972, 2013.

[110] A. Sabourin and J. Segers. Marginal standardization of upper semicontinuous processes. With application to max-stable processes. *J. Appl. Probab.*, 54(3):773–796, 2017.

[111] M. Sainsbury-Dale, A. Zammit-Mangion, and R. Huser. Likelihood-free parameter estimation with neural Bayes estimators. *The American Statistician*, 78(1):1–14, 2024.

[112] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes*. Stochastic Modeling. Chapman & Hall, New York, 1994. Stochastic Models with Infinite Variance.

[113] H. Sang and M. G. Genton. Tapered composite likelihood for spatial max-stable models. *Spat. Stat.*, 8:86–103, 2014.

[114] K. R. Saunders, A. G. Stephenson, and D. J. Karoly. A regionalisation approach for rainfall based on extremal dependence. *Extremes*, 24(2):215–240, 2021.

[115] M. Schlather. Models for stationary max-stable random fields. *Extremes*, 5(1):33–44, 2002.

[116] M. Schlather and J. A. Tawn. A dependence measure for multivariate and spatial extreme values: properties and inference. *Biometrika*, 90(1):139–156, 2003.

[117] I. J. Schoenberg. Metric spaces and completely monotone functions. *Ann. of Math. (2)*, 39(4):811–841, 1938.

[118] Q. Sebille, A.-L. Fougères, and C. Mercadier. Modeling extreme rainfall: a comparative study of spatial extreme value models. *Spat. Stat.*, 21:187–208, 2017.

[119] B. A. Shaby and B. J. Reich. 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, 2012.

[120] H. Shang, J. Yan, and X. Zhang. El Niño–Southern Oscillation influence on winter maximum daily precipitation in California in a spatial model. *Water Resour. Res.*, 47(11), 2011.

[121] H. Shang, J. Yan, and X. Zhang. A two-step approach to model precipitation extremes in California based on max-stable and marginal point processes. *Ann. Appl. Stat.*, 9(1):452–473, 2015.

[122] E. L. Smith and A. G. Stephenson. An extended Gaussian max-stable process model for spatial extremes. *J. Statist. Plann. Inference*, 139(4):1266–1275, 2009.

[123] R. L. Smith. Max-stable processes and spatial extremes. *Unpublished Manuscript*, 1990.

[124] A. Stephenson and J. Tawn. Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika*, 92(1):213–227, 2005.

[125] A. G. Stephenson, E. A. Lehmann, and A. Phatak. A max-stable process model for rainfall extremes at different accumulation durations. *Weather and Climate Extremes*, 13:44–53, 2016.

[126] A. G. Stephenson, B. A. Shaby, B. J. Reich, and A. L. Sullivan. Estimating spatially varying severity thresholds of a forest fire danger rating system using max-stable extreme-event modeling. *Journal of Applied Meteorology and Climatology*, 54(2):395 – 407, 2015.

[127] S. A. Stoev. On the ergodicity and mixing of max-stable processes. *Stochastic Process. Appl.*, 118(9):1679–1705, 2008.

[128] S. A. Stoev and M. S. Taqqu. Extremal stochastic integrals: a parallel between max-stable processes and α-stable processes. *Extremes*, 8(4):237–266, 2005.

[129] K. Strokorb, F. Ballani, and M. Schlather. Tail correlation functions of max-stable processes: construction principles, recovery and diversity of some mixing max-stable processes with identical TCF. *Extremes*, 18(2):241–271, 2015.

[130] J. A. Tawn. Modelling multivariate extreme value distributions. *Biometrika*, 77(2):245–253, 06 1990.

[131] E. Thibaud, J. Aalto, D. S. Cooley, A. C. Davison, and J. Heikkinen. Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures. *Ann. Appl. Stat.*, 10(4):2303–2324, 2016.

[132] E. Thibaud, R. Mutzner, and A. C. Davison. Threshold modeling of extreme spatial rainfall. *Water Resour. Res.*, 49(8):4633–4644, 2013.

[133] E. Thibaud and T. Opitz. Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102(4):855–870, 2015.

[134] H. Tyralis and A. Langousis. Estimation of intensity–duration–frequency curves using max-stable processes. *Stochastic Environmental Research and Risk Assessment*, 33(1):239–252, Jan 2019.

[135] C. Varin, N. Reid, and D. Firth. An overview of composite likelihood methods. *Statist. Sinica*, 21(1):5–42, 2011.

[136] P. Vatan. Max-infinite divisibility and max-stability in infinite dimensions. In *Probability in Banach Spaces, V (Medford, Mass., 1984)*, volume 1153 of *Lecture Notes in Math.*, pages 400–425. Springer, Berlin, 1985.

[137] A. V. Vecchia. Estimation and model identification for continuous spatial processes. *J. Roy. Statist. Soc. Ser. B*, 50(2):297–312, 1988.

[138] S. Vettori, R. Huser, and M. G. Genton. Bayesian Modeling of Air Pollution Extremes Using Nested Multivariate Max-Stable Processes. *Biometrics*, 75(3):831–841, 04 2019.

[139] J. L. Wadsworth. On the occurrence times of componentwise maxima and bias in likelihood inference for multivariate max-stable distributions. *Biometrika*, 102(3):705–711, 2015.

[140] J. L. Wadsworth and J. A. Tawn. Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, 101(1):1–15, 2014.

[141] Y. Wang, P. Roy, and S. A. Stoev. Ergodic properties of sum- and max-stable stationary random fields via null and positive group actions. *Ann. Probab.*, 41(1):206–228, 2013.

[142] Y. Wang and S. A. Stoev. On the structure and representations of max-stable processes. *Adv. in Appl. Probab.*, 42(3):855–877, 2010.

[143] S. Westra and S. A. Sisson. Detection of non-stationarity in precipitation extremes using a max-stable process model. *Journal of Hydrology*, 406(1):119–128, 2011.

[144] T. Whitaker, B. Beranger, and S. A. Sisson. Composite likelihood methods for histogram-valued random variables. *Stat. Comput.*, 30(5):1459–1477, 2020.

[145] B. D. Youngman. Flexible models for nonstationary dependence: Methodology and examples, 2020. https://arxiv.org/abs/2001.06642.

[146] B. D. Youngman. evgam: An R package for generalized additive extreme value models. *J. Stat. Softw.*, 103(3):1–26, 2022.

[147] R. Yuen, S. Stoev, and D. Cooley. Distributionally robust inference for extreme Value-at-Risk. *Insurance Math. Econom.*, 92:70–89, 2020.

Index