1

Spatial Extremes and max-stable processes

Mathieu Ribatet

University of Montpellier 2 / University of Lyon 1

Clément Dombry

University of Besançon

Marco Oesting

INRA/AgroParisTech

CONTENTS

Abstract

This chapter aims at being a crash course on max-stable processes with an emphasis on their use for modeling spatial extremes. We will see how max-stable processes are defined through a simple spectral representation and how it is possible to derive the finite dimensional distributions from it. Because the goal of this crash course is also to be of practical interest, existing parametric max-stable models will be introduced and discussed. A useful measure of spatial dependence, the extremal coefficient function, will be introduced as well as several approaches to fit max-stable processes to spatial data. Finally we

open the discussion with other alternatives to max-stable processes to model spatial extremes.

1.1   Introduction

In Chapter 1 the extreme value theory was introduced in a finite dimensional setting, i.e., extremes of random variables or vectors. In this chapter we go a bit further by investigating infinite dimensional extremes, i.e., extremes of stochastic processes. Although it is possible to work with weaker assumption, we will work with sample path continuous stochastic processes to ensure that random variables such as $\sup\{Y(x)\colon x \in \mathcal{X}\}$ where $\mathcal{X}$ is a compact subset of $\mathbb{R}^d$, $d \geq 1$, are well defined.

Similarly to univariate extreme value analysis, the aim of modelling spatial extremes is typically related to risk assessments which in a spatial context can take several forms. For example if one observes a precipitation field $Y(x)$ over a given catchment $\mathcal{X} \subset \mathbb{R}^2$, one could be interested in evaluating the probability that the total rainfall amount over this catchment exceeds a given critical quantity $z_{\mathrm{crit}} > 0$, i.e.,

$$\Pr\left\{\int_{\mathcal{X}} Y(x)\mathrm{d}x > z_{\mathrm{crit}}\right\}.$$

One could also be interested in characterizing the distribution of the largest "pointwise" rainfall amount in this catchment, i.e., evaluating probabilities of the form

$$\Pr\left\{\sup_{x\in\mathcal{X}} Y(x) > z\right\}, \qquad z > 0.$$

Clearly evaluating the above probabilities is even more challenging than what we usually do in a univeriate or multivariate setting since, for instance, it requires the knowledge of the distribution of the random variable $Y(x)$ for all $x \in \mathcal{X}$ as well as capturing the spatial dependence of the process $\{Y(x)\colon x \in \mathcal{X}\}$.

1.2   Max-stable processes

1.2.1   Spectral representation

Before introducing the spectral representation of max-stable processes, it seems necessary to motivate their use for modelling spatial extremes. Similarly to the asymptotic arguments justifying the use of the generalized ex-

treme value and the generalized Pareto distributions, see Part I, max-stable processes appear to be a sensible choice for modelling pointwise maxima.

Definition 1. Let $Z_1, Z_2, \ldots,$ be a sequence of independent copies of a stochastic process $\{Z(x) \colon x \in \mathcal{X}\}$. If for each $n \geq 1$ there exist normalizing functions $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\frac{\max_{i=1,\ldots,n} Z_i - b_n}{a_n} \stackrel{\mathrm{d}}{=} Z, \tag{1.1}$$

then $\{Z(x) \colon x \in \mathcal{X}\}$ is said to be max-stable. Recall that equality in distribution for continuous sample path stochastic processes on a compact set means that all finite dimensional distributions are identical.

Based on (1.1) it is not clear why max-stable processes are especially relevant as far as spatial extremes are of concern. Similarly to the univariate case, their use is based on asymptotic arguments.

Theorem 1. *(de Haan, 1984) Let $Y_1, Y_2, \ldots$ be a sequence of independent copies of a stochastic process $\{Y(x) \colon x \in \mathcal{X}\}$ with continuous sample paths. If there exist continuous functions $c_n > 0$ and $d_n \in \mathbb{R}$ such that the limiting process $\{Z(x) \colon x \in \mathcal{X}\}$ defined by*

$$\frac{\max_{i=1,\ldots,n} Y_i(x) - d_n(x)}{c_n(x)} \longrightarrow Z(x), \qquad x \in \mathcal{X}, \quad n \to \infty, \tag{1.2}$$

is non degenerate, then the process $\{Z(x) \colon x \in \mathcal{X}\}$ has to be a max-stable process. Note that the convergence in (1.2) refers to weak convergence in the space of continuous functions on $\mathcal{X}$.

Remark. To be consistent with the univariate extreme value theory, Theorem 1 implies that the marginal distribution of $\{Z(x) \colon x \in \mathcal{X}\}$ have to be generalized extreme value distributed.

The statistical motivation for using max-stable processes for modelling spatial extremes is the following. Based on n independent replicates, we will assume that the limiting process $\{Z(x) \colon x \in \mathcal{X}\}$ is likely to be a good candidate for modeling the partial maxima process $\{\max_{i=1,\ldots,n} Y_i(x) \colon x \in \mathcal{X}\}$, as far as n is large enough. The logic beyond this is exactly the same as for univariate extreme value analysis where one prefers to work directly with the asymptotic distribution of block maxima, i.e., the generalized extreme value distribution, rather than estimating the distribution of $\{Y(x) \colon x \in \mathcal{X}\}$ and raising it to the power n to estimate the distribution of the partial maxima.

So far, neither Definition 1 nor Theorem 1 give a precise description of max-stable processes, and it would be nice to get a better picture them as for the univariate and multivariate cases. This description is known as the spectral representation of max-stable processes. Since we know that the marginal distributions of $\{Z(x) \colon x \in \mathcal{X}\}$ have to be generalized extreme value distributed,

it is more convenient to set the margins to a given distribution. A widely used choice is to use unit Fréchet margins, i.e., $\Pr\{Z(x) \leq z\} = \exp(-1/z)$ for all $x \in \mathcal{X}$ and $z > 0$, and with these specific margins $\{Z(x)\colon x \in \mathcal{X}\}$ is said to be a *simple max-stable process*.

Theorem 2. *(de Haan, 1984; Penrose, 1992) Any non degenerate simple max-stable process $\{Z(x)\colon x \in \mathcal{X}\}$ defined on a compact set $\mathcal{X} \subset \mathbb{R}^d$, $d \geq 1$, with continuous sample paths satisfies*

$$Z(x) \stackrel{\mathrm{d}}{=} \max_{i \geq 1} \zeta_i f_i(x), \qquad x \in \mathcal{X}, \tag{1.3}$$

where $\{(\zeta_i, f_i)\colon i \geq 1\}$ are the points of a Poisson process on $(0, \infty) \times \mathcal{C}$ with intensity $\zeta^{-2} d\zeta \nu(df)$ for some locally finite measure ν defined on the space $\mathcal{C}$ of non-negative continuous functions on $\mathcal{X}$ such that

$$\int f(x)\nu(df) = 1, \qquad x \in \mathcal{X}.$$

Before giving more details on Theorem 2, some comments worth to be mentioned. First the spectral characterization (1.3) is not unique in the sense that different measures ν can lead to the same max-stable process $\{Z(x)\colon x \in \mathcal{X}\}$. Second the restriction on non-negative function is only required for convenience and, since the spectral characterization consists in taking pointwise maxima over an infinite number of such functions, one can consider real functions as long as $\nu\{f(x) > 0\} > 0$ for all $x \in \mathcal{X}$. Lastly an important special case of (1.3) is when ν is a probability measure since in that case the spectral representation may be rewritten as

$$Z(x) \stackrel{\mathrm{d}}{=} \max_{i \geq 1} \zeta_i Y_i(x), \qquad x \in \mathcal{X},$$

where $\{\zeta_i\colon i \geq 1\}$ are the points of a Poisson process on $(0, \infty)$, $Y_1, Y_2, \ldots$ a sequence of independent copies of a non-negative stochastic process $\{Y(x)\colon x \in \mathcal{X}\}$ with continuous sample paths and such that $\mathbb{E}\{Y(x)\} = 1$ for all $x \in \mathcal{X}$.

Similarly to the radial / (pseudo) angular decomposition for multivariate extremes, the points $\{\zeta_i\colon i \geq 1\}$ in (1.3) play the role of the radius while the stochastic processes $\{Y_i(x)\colon i \geq 1\}$ the role of the angle. The spectral representation suggests a rainfall storm based interpretation due to Smith (1990) which, even though has no theoretical justification, has the merits of clarifying things to readers not familiar with point processes and stochastic processes. Think about a rainfall storm impacting a region $\mathcal{X}$ which has an overall intensity ζ and spatial extent driven by $\{Y(x)\colon x \in \mathcal{X}\}$, i.e., $\zeta Y(x)$ corresponds to the amount of rain for this storm at location $x \in \mathcal{X}$. With this conceptualization, max-stable processes appear as the pointwise maxima over an infinite number of storms $\{(\zeta_i, \{Y_i(x)\colon x \in \mathcal{X}\})\colon i \geq 1\}$.

The left panel of Figure 1.1 is an illustration of this rainfall storm interpretation where the simulated max-stable process was obtained from 100 storms.
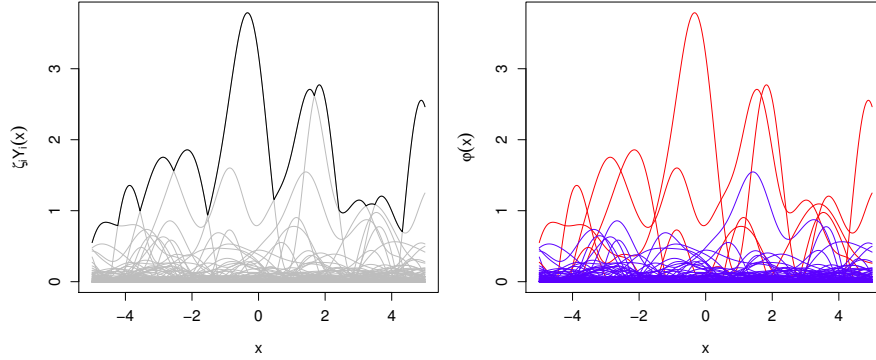
FIGURE 1.1
Simulation of a max-stable process on $\mathcal{X} = [-5, 5]$ from its spectral charac-terization. For this example we take $Y(x) = \sqrt{2\pi}\max\{0, W(x)\}$ where W is a standard Gaussian process with a Gaussian correlation function, i.e., $\rho(h) = \exp(-h^2)$. Left: The grey curves corresponds to $\{\zeta_i Y_i(x)\colon i = 1, \ldots, 100\}$ and the black one to the pointwise maxima. Right: Decomposition of the spectral functions as extremal (red curves) and non extremal (blue curves) functions.

We can see that some storms did not contribute to the pointwise maxima at any location $x \in \mathcal{X}$. This is illustrated by the right panel of Figure 1.1 which decomposed the Poisson process $\Phi = \{\varphi_i = \zeta_i Y_i \colon i \geq 1\}$ into two sub–point processes

$$\Phi^+ = \{\varphi \in \Phi \colon \exists x \in \mathcal{X}, \ \varphi(x) = Z(x)\},$$
$$\Phi^- = \{\varphi \in \Phi \colon \varphi(x) < Z(x), \ x \in \mathcal{X}\}.$$

According to Dombry et al. (2013) and Dombry and Éyi-Minko (2013), the atoms of the sub–point processes Φ^+ and Φ^- are called the extremal and sub–extremal functions respectively. We will see later in Section **??** that this de-composition of Φ into extremal and sub–extremal functions will be especially convenient for likelihood based inference as well as for deriving the regular conditional distributions of max-stable processes in the related chapter.

1.2.2 Parametric max-stable process families

In this section we introduced existing parametric max-stable process families and will do so in an historical order. The first model that appeared in the literature is what is called the *Smith process* (Smith, 1990) also sometimes referred to the *Gaussian extreme value process* (Schlather, 2002). This process is based on a particular mixed moving maxima representation (see Section **??** for more details), i.e.,

$$Z(x) = \max_{i \geq 1} \zeta_i \varphi(x - U_i; 0, \Sigma), \qquad x \in \mathcal{X}, \tag{1.4}$$

where $\{(\zeta_i, U_i) : i \geq 1\}$ are the points of a Poisson process on $(0, \infty) \times \mathbb{R}^d$ with intensity measure $\zeta^{-2}\mathrm{d}\zeta\mathrm{d}u$ and $\varphi(\cdot; 0, \Sigma)$ denotes the d–variable Gaussian density with mean 0 and covariance matrix Σ. Although this process is interesting for historical reasons, it is rarely useful because of a lack of flexibility—the shape of multivariate Gaussian densities being too restrictive.

The second model was introduced about 10 years after in the seminal paper of Schlather (2002) who introduced what is called the *Schlather process*, also sometimes referred to as the *extremal Gaussian process*. This model is defined by

$$Z(x) = \sqrt{2\pi} \max_{i \geq 1} \zeta_i \max\{0, W_i(x)\}, \qquad x \in \mathcal{X}, \tag{1.5}$$

where $\{W_i(x) : x \in \mathcal{X}\}$ are independent copies of a stationary Gaussian process with correlation function ρ. Note that the scaling factor $\sqrt{2\pi}$ is required to have $\sqrt{2\pi}\mathbb{E}[\max\{0, W(x)\}] = 1$ for all $x \in \mathcal{X}$.

The third model historically introduced is the *Brown–Resnick process* (Brown and Resnick, 1977) and, having a look at the date of publication, should have been introduced first. But this model was well known to be difficult to work with and, apart from Husler and Reiss (1989), no further work was done until that of Kabluchko et al. (2009). This process is defined by

$$Z(x) = \max_{i \geq 1} \zeta_i \exp\left\{W_i(x) - \gamma(x)\right\}, \qquad x \in \mathcal{X}, \tag{1.6}$$

where $\{W_i(x) : x \in \mathcal{X}\}$ are independent copies of a zero mean Gaussian process with stationary increments and semi-variogram $\gamma(h) = \mathrm{Var}\{W(x + h) - W(x)\}/2$. It is interesting to note that the Smith model is a special case of (1.6) with

$$W(x) = x^T \Sigma^{-1} X, \qquad X \sim N(0, \Sigma),$$

and whose semi-variogram therefore satisfies $2\gamma(x) = x^T\Sigma^{-1}\mathrm{Var}(X)\Sigma^{-1}x = x^T\Sigma^{-1}x$.

Finally the last model introduced so far is the extremal–t process. It was first introduced in a multivariate setting by Nikoloulopoulos et al. (2009), in a spatial context by Davison et al. (2012) and Ribatet and Sedki (2013). Finally Opitz (2013) derived a spectral characterization for this process

$$Z(x) = c_\nu \max_{i \geq 1} \zeta_i \max\{0, W_i(x)\}^\nu, \qquad x \in \mathcal{X}, \tag{1.7}$$

where $\nu \geq 1$, $\{W_i(x) : x \in \mathcal{X}\}$ are independent copies of a stationary Gaussian process with correlation function ρ and

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

where Γ is the Gamma function. As is seen clearly in (1.7), the Schlather process is a special case of the extremal–t model with $\nu = 1$.

Figure 1.2 plots one realization from each of the max-stable process introduced above. As expected we can see that the Smith process produces artificial
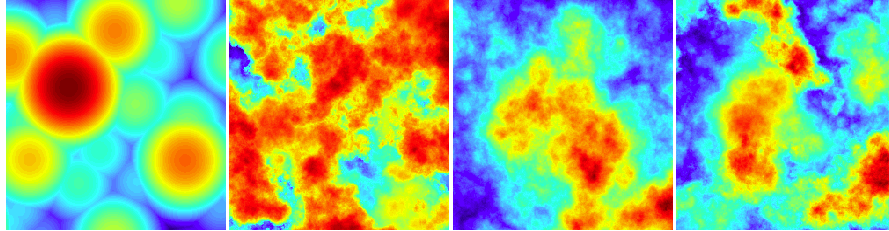
FIGURE 1.2
One realization on a 250×250 grid from each of the max-stable model introduced in Section 1.2.2. From left to right: Smith where Σ is the identity matrix; Schlather with $\rho(h) = \exp\{-(h/2)^{1.5}\}$; Brown–Resnick with $\gamma(h) = (h/2)^{1.5}$; and Extremal–t with $\nu = 5$ and $\rho(h) = \exp\{-(h/4)^{1.5}\}$. Note that for visualization purposes the margins were transformed to standard Gumbel margins.

surfaces; thus confirming its lack of flexibility already mentioned earlier. The other processes produce sample surfaces that are much more wiggly—though we set the dependence parameters such that the sample surfaces are smooth enough. Compared to the Brown–Resnick and the extremal–t processes, we can see that the Schlather process tends to give larger areas where the largest values occur. We will see later that is a consequence of this model that does not allow for spatial independence.

1.2.3 Finite dimensional distributions

From (1.3), it is possible to derive the finite dimensional distribution of $\{Z(x): x \in \mathcal{X}\}$. More precisely for any $\mathbf{x} = (x_1, \ldots, x_k) \in \mathcal{X}^k$, $k \geq 1$, and $\mathbf{z} = (z_1, \ldots, z_k) \in (0, \infty)^k$, we have

$$\Pr\{Z(\mathbf{z}) \leq \mathbf{z}\} = \Pr\left[\text{no atom } (\zeta, Y) \in \Phi \colon \zeta Y(x_j) > z_j \text{ for some } j \in \{1, \ldots, k\}\right]$$

$$= \exp\left[-\int_0^\infty \Pr\left\{\zeta > \min_{j=1,\ldots,k} \frac{z_j}{Y(x_j)}\right\} \zeta^{-2} \mathrm{d}\zeta\right]$$

$$= \exp\left\{-V_{\mathbf{x}}(z_1, \ldots, z_k)\right\},$$

where the function

$$V_{\mathbf{x}}(z_1, \ldots, z_k) = \mathbb{E}\left\{\max_{j=1,\ldots,k} \frac{Y(x_j)}{z_j}\right\}, \tag{1.8}$$

called the *exponent function*, fully characterizes the joint distribution of $Z(\mathbf{x})$.

Working a bit with (1.8), it is straightforward to see that

$$V_{\mathbf{x}}(z, \ldots, z) = \frac{\theta(\mathbf{x})}{z}, \qquad \theta(\mathbf{x}) = \mathbb{E}\left\{\max_{j=1,\ldots,k} Y(x_j)\right\}. \tag{1.9}$$

In the above equation $\theta(\mathbf{x})$ is called the (k-dimensional) *extremal coefficient*

TABLE 1.1

Bivariate distributions for each parametric max-stable model introduced in Section 1.2.2. Smith model: $a^2(h) = h^T \Sigma^{-1} h$. Schlather: ρ is the correlation function of $\{W(x) \colon x \in \mathcal{X}\}$ in (1.5). Brown–Resnick: γ is the semivariogram of $\{W(x) \colon x \in \mathcal{X}\}$ in (1.6). Extremal–t: ρ is the correlation function of $\{W(x) \colon x \in \mathcal{X}\}$ and $b^2 = \{1 - \rho(h)^2\}/(\nu + 1)$ in (1.7).

Model	$-\log \Pr\{Z(x) \leq z_1, Z(x+h) \leq z_2\}$
Smith	$\frac{1}{z_1}\Phi\left\{\frac{a(h)}{2} + \frac{1}{a(h)}\log\frac{z_2}{z_1}\right\} + \frac{1}{z_2}\Phi\left\{\frac{a(h)}{2} + \frac{1}{a(h)}\log\frac{z_1}{z_2}\right\}$
Schlather	$\frac{1}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left\{1 + \sqrt{1 - \frac{2\{1+\rho(h)\}z_1 z_2}{(z_1+z_2)^2}}\right\}$
Brown–Resnick	$\frac{1}{z_1}\Phi\left\{\sqrt{\frac{\gamma(h)}{2}} + \sqrt{\frac{1}{2\gamma(h)}}\log\frac{z_2}{z_1}\right\} + \frac{1}{z_2}\Phi\left\{\sqrt{\frac{\gamma(h)}{2}} + \sqrt{\frac{1}{2\gamma(h)}}\log\frac{z_1}{z_2}\right\}$
Extremal–t	$\frac{1}{z_1}T_{\nu+1}\left\{\frac{1}{b}\left(\frac{z_2}{z_1}\right)^{1/\nu} - \frac{\rho(h)}{b}\right\} + \frac{1}{z_2}T_{\nu+1}\left\{\frac{1}{b}\left(\frac{z_1}{z_2}\right)^{1/\nu} - \frac{\rho(h)}{b}\right\}$

and is a summary measure of dependence across the element of the random vector $Z(\mathbf{x})$. Because of the independence property between the radial and angular components of multivariate extremes, the extremal coefficient is as expected independent of the radius, i.e., the level z appearing in $V_{\mathbf{x}}(z, \ldots, z)$, and focus only on the dependence.

In a spatial context, it is more convenient to restrict our attention to the bivariate case and define what is known as the *extremal coefficient function*, i.e.,

$$\theta \colon h \longmapsto \mathbb{E}\left[\max\{Y(x), Y(x+h)\}\right]. \tag{1.10}$$

The extremal coefficient function takes values in $[1, 2]$ where the lower bound corresponds to perfect dependence and the upper one to independence. Indeed for these two limiting cases we have respectively

$$\Pr\{Z(x+h) \leq z \mid Z(x) \leq z\} = \Pr\{Z(x+h) \leq z\}^{\theta(h)-1}$$
$$= \begin{cases} 1, & \text{perfect dependence,} \\ \Pr\{Z(x+h) \leq z\}, & \text{independence.} \end{cases}$$

It has to be noted that (1.8) can be difficult to evaluate explicitly when $k > 2$ and this is why the finite dimensional distribution of max-stable processes were first restricted to the bivariate case only. These bivariate distributions are given in Table 1.1. However working a bit with (1.8), one can derive an expression for $V_{\mathbf{x}}$ much easier to work with when $k > 2$.

It is straightforward to see that the exponent function is homogeneous of order -1, i.e.,

$$V_{\mathbf{x}}(cz_1, \ldots, cz_k) = c^{-1}V_{\mathbf{x}}(z_1, \ldots, z_k), \qquad c > 0.$$

In particular, since $V_{\mathbf{z}}$ is positive and provided it is continuously differentiable, Euler's homogeneous theorem implies that

$$V_{\mathbf{x}} = \sum_{j=1}^{k} z_j^{-1} \Pr\left\{\frac{Y(x_\ell)}{z_\ell} \leq \frac{Y(x_j)}{z_j},\ \ell \neq j\right\}. \tag{1.11}$$

As an example, (1.11) is more tractable than (1.8) for Brown–Resnick processes since in that case we have (Huser and Davison, 2013)

$$\Pr\left\{\frac{Y(x_\ell)}{z_\ell} \leq \frac{Y(x_j)}{z_j},\ \ell \neq j\right\} = \Pr\left\{W(x_\ell) - W(x_j) \leq \gamma(x_\ell) - \gamma(x_j) + \log\frac{z_\ell}{z_j},\ \ell \neq j\right\}$$

$$= \Pr\left\{W(x_\ell - x_j) \leq \gamma(x_\ell) - \gamma(x_j) + \log\frac{z_\ell}{z_j},\ \ell \neq j\right\}$$

$$= \Phi\left(\log\frac{z_\ell}{z_j};\ \ell \neq j; \mu_j, \Sigma_j\right),$$

where the second equality used the fact that W has stationary increments and where $\mu_j = \{\gamma(x_\ell) - \gamma(x_j),\ \ell \neq j\}$, $\Sigma_j = \{\gamma(x_m - x_j) + \gamma(x_n - x_j) - \gamma(x_m - x_n)\}_{m,n\neq j}$ and $\Phi(\cdot; \mu_j, \Sigma_j)$ denotes the $(k-1)$-variate cumulative normal distribution with mean vector μ_j and covariance matrix Σ_j.

Closed forms for the k–variate distribution for the extremal–t can be found using the same lines as for the Brown–Resnick process but by using the conditional distribution instead of the stationary increment property and is left as an exercise.

1.3 Dependence measure for spatial extremes

In this section we investigate how the dependence between extreme events evolves across space. In what we shall call *conventional geostatistics*, to be opposed to what we aim for, i.e., a *geostatistics of extremes*, the underlying statistical model is usually based on a Gaussian process W whose semi variogram

$$\gamma(h) = \frac{1}{2}\mathbb{E}\left[\{W(x) - W(x+h)\}^2\right], \qquad x, h \in \mathcal{X}, \tag{1.12}$$

plays a fundamental role in driving how dependence evolve in space.

Unfortunately the semi-variogram γ is not suitable for spatial extremes are of interest. For instance since for simple max-stable processes the margins are unit Fréchet, $\mathbb{E}\{Z(x)\} = \infty$ for all $x \in \mathcal{X}$ and the semi-variogram function does not exist. Therefore there is a pressing need to define a summary dependence measure similar to the semi-variogram but devoted to extreme
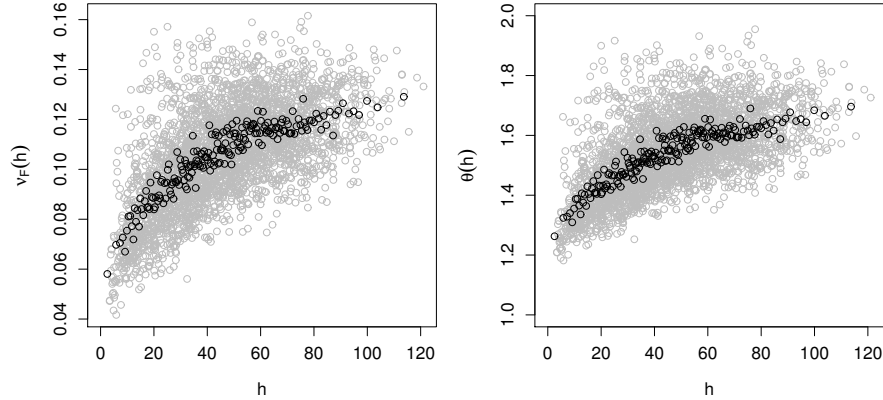
FIGURE 1.3
Empirical F–madogram (left) and corresponding extremal coefficients (right) for the Swiss precipitation data of the `SpatialExtremes` package (Ribatet et al., 2013). The grey points are pairwise estimates while the black ones are binned estimates obtained from 200 bins.

values. Among the various tools introduced, one of the most relevant ones is the F–madogram (Cooley et al., 2006)

$$\nu_F(h) = \frac{1}{2} \mathbb{E}\left[|F\{Z(x+h)\} - F\{Z(x)\}|\right], \qquad x, h \in \mathcal{X}, \qquad (1.13)$$

where F denotes the cumulative distribution function of $Z(x)$, $x \in \mathcal{X}$. Contrary to the semi-variogram, the F–madogram is well defined since $F\{Z(x)\} \sim U(0,1)$ and therefore has expectation $1/2$. The F–madogram is particularly convenient for spatial extremes since it has strong connections with the extremal coefficient function. Indeed using the fact that $|a - b| = 2\max(a, b) - a - b$, it is not difficult to show that

$$\theta(h) = \frac{1 + 2\nu_F(h)}{1 - 2\nu_F(h)}, \qquad h \in \mathcal{X}. \qquad (1.14)$$

Clearly the F–madogram is easily estimated by its empirical counterpart, i.e.,

$$\hat{\nu}_F(h) = \frac{1}{2n(n+1)} \sum_{i=1}^{n} |R_i(x) - R_i(x+h)|, \qquad R_i(x) = \sum_{\ell=1}^{n} 1_{\{Z_\ell(x) \leq Z_i(x)\}},$$
$$(1.15)$$

where $Z_1, \ldots, Z_n$ are independent replicates of a max-stable process (not necessarily simple) Z. To reduce sample variability, it is often a good idea to use a binned version of (1.15), i.e., to average $\hat{\nu}_F$ over suitable classes of distance,

i.e.,

$$\tilde{\nu}_F(h) = \frac{1}{2n(n+1)|\mathcal{N}_h|} \sum_{x,y \in \mathcal{N}_h} \sum_{i=1}^{n} |R_i(x) - R_i(y)|,$$

where

$$\mathcal{N}_h = \{x, y \in \{x_1, \ldots, x_k\} \colon |\|x - y\| - h| < \delta\},$$

and some suitable binning radius $\delta > 0$.

Figure 1.3 plots the empirical estimates for the F–madogram and the related pairwise extremal coefficients for the Swiss precipitation data that are freely available from the `SpatialExtremes` package (Ribatet et al., 2013). As expected we can see that the spatial dependence decreases as the distance lag increases but appears to be still persistent beyond 100km.

1.4 Inference

Although theory for max-stable processes was well developed, several decades have passed to find a suitable framework to use max-stable processes for modeling spatial extremes. The main reason for this was the lack of closed form for the likelihood of such processes. To see this recall that the k–variate distribution function is given by

$$\Pr\{Z(x_1) \leq z_1, \ldots, Z(x_k) \leq z_k\} = \exp\{-V_\mathbf{x}(z_1, \ldots, z_k)\},$$

and hence the related probability density function is

$$f_\mathbf{x}(z_1, \ldots, z_k) = \exp\{-V_\mathbf{x}(z_1, \ldots, z_k)\} \sum_{\tau \in \mathscr{P}_k} w(\tau), \tag{1.16}$$

with

$$w(\tau) = (-1)^{|\tau|} \prod_{j=1}^{|\tau|} \frac{\partial^{|\tau_j|}}{\partial z_{\tau_j}} V(z_1, \ldots, z_k),$$

and where $\mathscr{P}_k$ is the set of all possible partition of the set $\{x_1, \ldots, x_k\}$, $\tau = (\tau_1, \ldots, \tau_\ell)$, $|\tau| = \ell$ is the size of the partition τ and $\partial^{|\tau_j|}/\partial z_{\tau_j}$ denotes the mixed partial derivatives with respect to the elements of the j-th element of the partition τ. As emphasized by Dombry et al. (2013); Ribatet (2013), the number of possible partitions of $\{x_1, \ldots, x_k\}$ corresponds to Bell numbers and hence yield a combinatorial explosion even for moderate values of k—when $k = 10$ there is around 115,000 partitions.

1.4.1 Pairwise likelihood

Due to this computational burden, the maximum likelihood estimator cannot be used and other approaches have been proposed. A first attempt was to

use least squares to estimate the dependence parameters from the empirical extremal coefficient estimates (Smith, 1990)—using the F-madogram for instance. This approach is not completely satisfactory since it focuses only on the dependence parameters and hence prediction at unobserved location is not possible. In some sense a likelihood based approach was missing and this is exactly what was proposed with the work of Padoan et al. (2010). Padoan et al. (2010) propose to use the maximum pairwise likelihood estimator, i.e., maximizing

$$\ell_{\mathrm{p}}\{\psi; z_1(\mathbf{x}), \ldots, z_n(\mathbf{x})\} = \sum_{\ell=1}^{n} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} w_{i,j} \log f\{z_\ell(x_i), z_\ell(x_j); \psi\}, \quad (1.17)$$

where ψ are the parameter to be estimated and $\{w_{i,j},\ 1 \leq i < j \leq k\}$ are suitable weights that can be used either to improve efficiency or to reduce the computational cost.

As mentioned in Part III, composite likelihoods belong to *mis-specified models* or more accurately *under-specified models* as we typically assume that the bivariate densities appearing in (1.17) are correct.

Remark. It is important to remember that composite likelihoods are not genuine likelihoods nor they give an approximation of the likelihood. Composite likelihoods provide a framework to define estimators based on unbiased estimating equations.

As any maximum composite likelihood estimator, and under mild regularity conditions, the maximum pairwise likelihood estimator share similar properties to that of the maximum likelihood estimator. More precisely, it is consistent, asymptotically normal— see Part III for more details.

The advantage of a likelihood based approach is that it is straightforward to extend it to more complex statistical models. For instance so far we restricted our attention to simple max-stable processes, i.e., with unit Fréchet margins, and this assumption is clearly unrealistic for concrete applications where it is expected that the intensity of extreme events may vary spatially. Up to slight modification of (1.17), one can allow for varying marginal parameters. More precisely let $\{\tilde{Z}(x),\ x \in \mathcal{X}\}$ be a max-stable process such that $\tilde{Z}(x) \sim \mathrm{GEV}\{\mu(x), \sigma(x), \xi(x)\}$ for all $x \in \mathcal{X}$ and define the mapping

$$t_x \colon \tilde{z} \longmapsto \left\{1 + \xi(x)\frac{\tilde{z} - \mu(x)}{\sigma(x)}\right\}^{1/\xi(x)}.$$

Then the pointwise transformed stochastic process $\{t_x\{\tilde{Z}(x)\},\ x \in \mathcal{X}\}$ is a simple max-stable process and the log-pairwise likelihood for $\tilde{Z}$ is therefore

$$\ell_p\{\psi; \tilde{z}_1(\mathbf{x}), \ldots, \tilde{z}_n(\mathbf{x})\} = \ell_{\mathrm{p}}\{\psi; t_{\mathbf{x}}\{\tilde{z}_1(\mathbf{x})\}, \ldots, t_{\mathbf{x}}\{\tilde{z}_n(\mathbf{x})\}\} +$$
$$\sum_{\ell=1}^{n} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \log |J\{\tilde{z}_\ell(x_i)\} J\{\tilde{z}_\ell(x_i)\}|,$$

where the additional term correspond to the logarithm of the Jacobian coming from the mapping t_x.

To gain efficiency some authors try to use the triplet-wise likelihood in place of the pairwise likelihood (Genton et al., 2011; Huser and Davison, 2013) but simulation studies indicate that the gain in efficiency is small compared to the large computational increase unless the spatial process is really smooth—which is unfortunately not the case for many environmental processes.

1.4.2   Full likelihood

Recently some authors tried to performed a full likelihood inference for max-stable processes Wadsworth and Tawn (2014) whose framework relies on the derivation of the conditional distributions of max-stable processes given by Dombry and Éyi-Minko (2013) and Dombry et al. (2013)—see the chapter on conditional simulations for more details. Although (1.17) appears rather complicated it is possible to get a better understanding of this expression in terms of the extremal and non extremal point processes Φ^+ and Φ^- introduced in Section 1.2.1.

Having observed one single realization of $\{Z(x)\colon x \in \mathcal{X}\}$ at locations $x_1, \ldots, x_k \in \mathcal{X}$, we have $|\Phi^+| \in \{1, \ldots, k\}$ since a single extremal function might contribute to $\{Z(x_1), \ldots, Z(x_k)\}$ at more than one location. It thus defines a random partition θ of the set $\{x_1, \ldots, x_k\}$ and using the terminology of Wang and Stoev (2011) and Dombry and Éyi-Minko (2013), each of these partitions define a *hitting scenario*. Using the conditional independence property of the point processes Φ^- and Φ^+ and that of the extremal functions (Dombry and Éyi-Minko, 2013), the contribution to the full likelihood of a single realization from any simple max-stable processes with spectral characterization (1.3) is

$$\exp\left\{-\int_{\{z(\mathbf{x}),\boldsymbol{\infty}\}} \lambda_{\mathbf{x}}(\mathbf{u})\mathrm{d}\mathbf{u}\right\} \sum_{\tau \in \mathscr{P}_k} \prod_{j=1}^{|\tau|} \lambda_{\mathbf{x}_{\tau_j}}(\mathbf{x}_{\tau_j}) \int_{\{\mathbf{0}, z(\mathbf{x}_{\tau_j^c})\}} \lambda_{\mathbf{x}_{-\tau_j}|z(\mathbf{x}_{\tau_j}), \mathbf{x}_{\tau_j}}(\mathbf{u})\mathrm{d}\mathbf{u},$$

(1.18)

where $\mathbf{x}_{\tau_j} = \{x \in \{x_1, \ldots, x_k\}\colon x \in \tau_j\}$, $\mathbf{x}_{\tau_j^c} = \{x \in \{x_1, \ldots, x_k\}\colon x \notin \tau_j\}$, $\lambda_{\mathbf{x}}$ is the intensity function of the Poisson point process $\Phi_{\mathbf{x}} = \{\varphi(\mathbf{x}) \in (0, \infty)^k\colon \varphi \in \Phi\}$, i.e., for any Borel set $A \subset (0, \infty)^k$,

$$\Lambda_{\mathbf{x}}(A) = \int_0^\infty \mathrm{Pr}\{\zeta Y(\mathbf{x}) \in A\}\zeta^{-2}\mathrm{d}\zeta = \int_A \lambda_{\mathbf{x}}(\mathbf{u})\mathrm{d}\mathbf{u},$$

and where $\lambda_{\mathbf{x_1}|\mathbf{z},\mathbf{x_2}}(\mathbf{u})$, $x_1, x_2 \in \mathcal{X}$, $\mathbf{z}, \mathbf{u} > 0$, is the conditional intensity function, i.e.,

$$\lambda_{\mathbf{x_1}|\mathbf{z},\mathbf{x_2}}(\mathbf{u}) = \frac{\lambda_{(\mathbf{x_1},\mathbf{x_2})}\{(\mathbf{u},\mathbf{z})\}}{\lambda_{\mathbf{x_2}}(\mathbf{z})}.$$

Dombry et al. (2013) and Ribatet (2013) give closed forms for the intensity

and conditional intensity functions for the max-stable processes introduced in Section 1.2.2.

As expected (1.16) and (1.18) share the same structure and we now understand from the point process theory that

$$f_{\mathbf{x}}(z_1, \ldots, z_k) = \underbrace{\exp\{-V_{\mathbf{x}}(z_1, \ldots, z_k)\}}_{\substack{\text{contribution of the} \\ \text{sub-extremal functions}}} \times \underbrace{\sum_{\tau \in \mathscr{P}_k} w(\tau).}_{\substack{\text{contribution of each} \\ \text{hitting scenario}}}$$

Apart from getting a better understanding of (1.16), (1.18) is still not tractable and maximizing the log-likelihood appears to be numerically impossible. However if the hitting scenario is known to be $\tau = (\tau_1, \ldots, \tau_\ell)$, then the contribution to the full likelihood is much more tractable and becomes

$$\exp\left\{-\int_{\{z(\mathbf{x}), \boldsymbol{\infty}\}} \lambda_{\mathbf{x}}(\mathbf{u})\mathrm{d}\mathbf{u}\right\} \prod_{j=1}^{\ell} \lambda_{\mathbf{x}_{\tau_j}}(\mathbf{x}_{\tau_j}) \int_{\{\mathbf{0}, z(\mathbf{x}_{\tau_j^c})\}} \lambda_{\mathbf{x}_{\tau_j^c}|z(\mathbf{x}_{\tau_j}), \mathbf{x}_{\tau_j}}(\mathbf{u})\mathrm{d}\mathbf{u}.$$

This is what suggest Wadsworth and Tawn (2014) since when daily observations are available, one knows which partition yields to the annual maxima.

1.5   Discussion

The aim of this chapter was to introduce the basic foundations for using max-stable processes for modelling spatial extremes. Although max-stable processes are asymptotically justified models, other modelling strategies are possible and we will cover them briefly in this discussion.

Probably one of the most famous competitors to max-stable process are latent variable models. These models relies on a univariate extreme value argument only and assume that

$$\tilde{Z}(x) \mid \mu(x), \sigma(x), \xi(x) \sim \mathrm{GEV}\{\mu(x), \sigma(x), \xi(x)\},$$

i.e., for each location $x \in \mathcal{X}$, the random variable $Y(x)$ has a generalized extreme value distribution with location, scale and shape parameters equal to $\{\mu(x), \sigma(x), \xi(x)\}$. Typically a *conditional independence* assumption is supposed, i.e., conditionally on the marginal parameters, $\tilde{Z}(x_1)$ is independent of $\tilde{Z}(x_2)$ for any $x_1, x_2 \in \mathcal{X}^2$. As $\{\mu(x), \sigma(x), \xi(x)\}$ are allowed to vary in space, it is typically assumed that $[\{\mu(x), \sigma(x), \xi(x)\} \colon x \in \mathcal{X}]$ is a Gaussian process whose mean function depends on some relevant covariates and whose covariance function is chosen among available parametric covariance function families, e.g., Whittle–Matern. Since the likelihood of this model involves intractable integrals, to bypass this hurdle one often has resort to the Bayesian

paradigm combined with Monte–Carlo Markov chains algorithms—see Davison et al. (2012) for more details. The main drawback of this model is that, because of the conditional independence assumption, the spatial dependence is completely ignored and that, for any $x \in \mathcal{X}$, $\tilde{Z}(x)$ is not extreme value distributed anymore. However in practice if the aim of the study is to characterize the pointwise distribution of extremes then it is so far probably one of the best approaches.

To take into account the spatial dependence, one could think of using copulas (Sang and Gelfand, 2009). For instance one could use the Gaussian copula, i.e., having observed $\{\tilde{Z}(x)\colon x \in \mathcal{X}\}$ at locations $x_1, \ldots, x_k \in \mathcal{X}$, we have

$$\Pr\{\tilde{Z}(x_1) \leq \tilde{z}_1, \ldots, \tilde{Z}(x_k) \leq \tilde{z}_k\} = \Phi_\rho\left\{\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_k)\right\},$$

where Φ_ρ is the multivariate normal distribution with zero mean and correlation matrix $\{\rho(x_i - x_j)\colon i, j = 1, \ldots, k\}$, ρ a parametric correlation function, Φ^{-1} the quantile function of a standard normal distribution and

$$u_j = \exp\left\{-\left(1 + \xi(x_j)\frac{\tilde{z}_j - \mu(x_j)}{\sigma(x_j)}\right)^{-1/\xi(x_j)}\right\}, \qquad 1 + \xi(x_j)\frac{y_j - \mu(x_j)}{\sigma(x_j)} > 0,$$

for $j = 1, \ldots, k$. As for the latent variable model, one typically assume that the marginal parameters $[\{\mu(x), \sigma(x), \xi(x)\}\colon x \in \mathcal{X}]$ is a Gaussian process.

Although this type of models might seem relevant, they have the same weaknesses than the use of copula for multivariate extremes: most often this modelling strategy is not able to capture the spatial dependence of extremes (Davison et al., 2012). One exception needs to be mentioned though: the use of extreme value copulas but in that case it is equivalent to the use of max-stable processes.

Finally one could be tempted to use asymptotic independent models for modelling spatial extremes. If from a strict asymptotic point of view such models might seem irrelevant as, in the tails, spatial extremes will behave as a pure noise process with generalized extreme valued margins, they possess interesting properties that might be true for environmental processes and that max-stable processes cannot have (Wadsworth and Tawn, 2012; Davison et al., 2013). More precisely as shown by (1.9) the dependence structure of max-stable process is independent of the level z. Asymptotically independent models allow that the spatial dependence structure becomes increasingly weaker as we go far in the tail.

References

Brown, B. M. and Resnick, S. I. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14:732–739.

Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. In Bertail, P., Soulier, P., Doukhan, P., Bickel, P., Diggle, P., Fienberg, S., Gather, U., Olkin, I., and Zeger, S., editors, *Dependence in Probability and Statistics*, volume 187 of *Lecture Notes in Statistics*, pages 373–390. Springer New York.

Davison, A., Huser, R., and Thibaud, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences*, 45(5):511–529.

Davison, A., Padoan, S., and Ribatet, M. (2012). Statistical modelling of spatial extremes. *Statistical Science*, 7(2):161–186.

de Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability*, 12(4):1194–1204.

Dombry, C. and Éyi-Minko, F. (2013). Regular conditional distributions of max infinitely divisible random fields. *Electronic Journal of Probability*, 18(7):1–21.

Dombry, C., Éyi-Minko, F., and Ribatet, M. (2013). Conditional simulations of max-stable processes. *Biometrika*, 100(1):111–124.

Genton, M. G., Ma, Y., and Sang, H. (2011). On the likelihood function of Gaussian max-stable processes. *Biometrika*, 98(2):481–488.

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

Husler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters*, 7(4):283–286.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *Annals of Probability*, 37(5):2042–2065.

Nikoloulopoulos, A., Joe, H., and Li, H. (2009). Extreme values properties of multivariate t copulas. *Extremes*, 12:129–148.

Opitz, T. (2013). Extremal-t process: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis*, 122:409–413.

Padoan, S., Ribatet, M., and Sisson, S. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association (Theory & Methods)*, 105(489):263–277.

Penrose, M. D. (1992). Semi-min-stable processes. *Annals of Probability*, 20(3):1450–1463.

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

Ribatet, M. and Sedki, M. (2013). Extreme value copulas and max-stable processes. *Journal de la Société Française de Statistique*, 154(1):138–150.

Ribatet, M., Singleton, R., and Team, R. C. (2013). *SpatialExtremes: Modelling Spatial Extremes*. R package version 2.0-1.

Sang, H. and Gelfand, A. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, 16(3):407–426.

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

Smith, R. L. (1990). Max-stable processes and spatial extreme. *Unpublished manuscript*.

Wadsworth, J. L. and Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika*, 99(2):253–272.

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

Wang, Y. and Stoev, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Advances in Applied Probability*, 443:461–483.

References

Brown, B. M. and Resnick, S. I. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14:732–739.

Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. In Bertail, P., Soulier, P., Doukhan, P., Bickel, P., Diggle, P., Fienberg, S., Gather, U., Olkin, I., and Zeger, S., editors, *Dependence in Probability and Statistics*, volume 187 of *Lecture Notes in Statistics*, pages 373–390. Springer New York.

Davison, A., Huser, R., and Thibaud, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences*, 45(5):511–529.

Davison, A., Padoan, S., and Ribatet, M. (2012). Statistical modelling of spatial extremes. *Statistical Science*, 7(2):161–186.

de Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability*, 12(4):1194–1204.

Dombry, C. and Éyi-Minko, F. (2013). Regular conditional distributions of max infinitely divisible random fields. *Electronic Journal of Probability*, 18(7):1–21.

Dombry, C., Éyi-Minko, F., and Ribatet, M. (2013). Conditional simulations of max-stable processes. *Biometrika*, 100(1):111–124.

Genton, M. G., Ma, Y., and Sang, H. (2011). On the likelihood function of Gaussian max-stable processes. *Biometrika*, 98(2):481–488.

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

Husler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters*, 7(4):283–286.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *Annals of Probability*, 37(5):2042–2065.

Nikoloulopoulos, A., Joe, H., and Li, H. (2009). Extreme values properties of multivariate *t* copulas. *Extremes*, 12:129–148.

Opitz, T. (2013). Extremal-*t* process: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis*, 122:409–413.

Padoan, S., Ribatet, M., and Sisson, S. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association (Theory & Methods)*, 105(489):263–277.

Penrose, M. D. (1992). Semi-min-stable processes. *Annals of Probability*, 20(3):1450–1463.

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

Ribatet, M. and Sedki, M. (2013). Extreme value copulas and max-stable processes. *Journal de la Société Française de Statistique*, 154(1):138–150.

Ribatet, M., Singleton, R., and Team, R. C. (2013). *SpatialExtremes: Modelling Spatial Extremes*. R package version 2.0-1.

Sang, H. and Gelfand, A. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, 16(3):407–426.

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

Smith, R. L. (1990). Max-stable processes and spatial extreme. *Unpublished manuscript*.

Wadsworth, J. L. and Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika*, 99(2):253–272.

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

Wang, Y. and Stoev, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Advances in Applied Probability*, 443:461–483.