

Spatial extremes: Max-stable processes at work

Titre: Extrêmes spatiaux : les processus max-stables en pratique

Ribatet Mathieu¹

Abstract: Since many developments to the functional extreme value theory have been made during the last decades, this paper reviews recent results on max-stable processes and covers a large range of themes such as finite dimensional distributions, parametric models, dependence measure, inferential procedure, model selection and (conditional) simulations. An application to the spatial modeling of wind gusts in Netherlands is given.

Résumé : De nombreux progrès ont été accomplis ces dernières décennies sur la théorie des valeurs extrêmes fonctionnelle. Dans ce papier nous regroupons les résultats principaux concernant les processus max-stables. Ainsi cette revue de littérature couvre une gamme variée de domaines : lois fini-dimensionnelles, modèles paramétriques, mesures de dépendance, procédure inférentielles, sélection de modèles et simulations (conditionnelles). Une application à la modélisation spatiale des rafales de vents aux Pays-bas est donnée.

Keywords: Max-stable process, Extremal coefficient function, Composite likelihood, Simulation

Mots-clés : Processus max-stable, Fonction du coefficient extrême, Vraisemblance composite, Simulation

AMS 2000 subject classifications: 60G70, 60E05, 62P12

1. Introduction

During the last decades, extreme value theory has seen many developments on both theoretical and practical levels. Functional extreme value theory, i.e., extremes of stochastic processes, was especially active with the spectral characterization of max-stable processes and the construction of parametric and tractable max-stable models.

This paper gathers several recent results on max-stable processes and try to emphasize their usefulness for statistical modelling of spatial extremes. Starting with the definition of max-stable processes in Section 2, Section 3 introduces the general form of the finite dimensional distributions of max-stable process and lists several widely used parametric max-stable models. Some summary measures of the spatial dependence are given in Section 4 while Section 5 introduces inferential procedures. Section 6 is devoted to model selection. Sections 7 and 8 deals with unconditional and conditional simulation of max-stable processes. The paper ends with the modeling of extreme wind gusts in the Netherlands with some details on how to do it from the R package `SpatialExtremes` Ribatet et al. (2013).

2. Asymptotics and spectral characterization

Let \mathcal{X} be a compact subset of \mathbb{R}^d , $d \geq 1$. Throughout this paper we will work with $\mathbb{C}(\mathcal{X})$, the space of continuous (random) functions f defined on \mathcal{X} , endowed with the uniform norm

¹ Department of Mathematics, Université Montpellier 2, 4 place Eugène Bataillon, 34095 cedex 2 Montpellier, France.
E-mail: mathieu.ribatet@math.univ-montp2.fr

$\|f\| = \sup_{x \in \mathcal{X}} |f(x)|$. Recall that with this setting, any stochastic processes living in $\mathbb{C}(\mathcal{X})$ is completely characterized by its finite dimensional distributions. Throughout this paper we will use componentwise algebra, e.g., $(f - g)(x) = f(x) - g(x)$ or $\{\max(f, g)\}(x) = \max\{f(x), g(x)\}$ for some functions f and g and $x \in \mathcal{X}$.

A stochastic process Z is said to be max-stable if there exist continuous normalizing functions $\{a_n > 0\}$ and $\{b_n \in \mathbb{R}\}$ such that

$$Z \stackrel{d}{=} \max_{i=1, \dots, n} \frac{Z_i - b_n}{a_n},$$

where Z_i are independent copies of Z .

It is often more convenient to work with arbitrarily fixed margins and a common choice is to have unit Fréchet margins, i.e., $\Pr\{Z(x) \leq z\} = \exp(-1/z)$ for all $x \in \mathcal{X}$ and $z > 0$. With this choice the normalizing functions are $a_n \equiv n$ and $b_n \equiv 0$.

Max-stable processes are widely used processes for modeling spatial extremes since they arise as the pointwise maxima taken over an infinite number of (appropriately rescaled) stochastic processes. More precisely, given independent copies $\{X_i: i \in \mathbb{N}\}$ of a stochastic process X defined (at least) on \mathcal{X} , it can be shown that if there exist normalizing functions $\{c_n > 0\}$ and $\{d_n \in \mathbb{R}\}$ such that

$$\max_{i=1, \dots, n} \frac{X_i(x) - d_n(x)}{c_n(x)} \longrightarrow Z(x), \quad n \rightarrow \infty, \quad x \in \mathcal{X}, \quad (1)$$

then either Z is degenerate or Z is a max-stable process [de Haan \(1984\)](#). If Z is non degenerate, univariate extreme value theory says that the pointwise distributions of Z have to be generalized extreme value distributions [de Haan and Ferreira \(2006\)](#), i.e.,

$$\Pr\{Z(x) \leq z\} = \exp \left\{ - \left(1 + \xi \frac{z - \mu}{\sigma} \right)^{-1/\xi} \right\}, \quad 1 + \xi \frac{z - \mu}{\sigma} > 0, \quad x \in \mathcal{X}, \quad (2)$$

for some location $\mu \in \mathbb{R}$, scale $\sigma > 0$ and shape $\xi \in \mathbb{R}$ parameters. For theoretical purposes, it is often more convenient to fix the pointwise distributions of Z to a suitable (univariate) distribution and hence restrict our attention to the spatial dependence structure. Throughout this paper we will assume that the limiting process Z is simple, that is, non degenerate with unit Fréchet margins [de Haan and Ferreira \(2006\)](#), i.e.,

$$\Pr\{Z(x) \leq z\} = \exp \left(-\frac{1}{z} \right), \quad z > 0, \quad x \in \mathcal{X}.$$

Similarly to the results for univariate extremes, (1) justifies the use of max-stable processes for modelling pointwise maxima by assuming that the limiting process is met for finite but large enough $n \in \mathbb{N}$.

Any simple max-stable process Z has a nice representation, known as the spectral characterization. The first representation [de Haan \(1984\)](#) says that there exists a family of non negative continuous functions $\{f(x, y): x, y \in \mathbb{R}^d\}$ such that $\int_{\mathbb{R}^d} f(x, y) dy = 1$ for all $x \in \mathbb{R}^d$ and for any compact set $K \subset \mathcal{X}$ we have

$$\int_{\mathbb{R}^d} \sup_{x \in K} f(x, y) dy < \infty$$

and for which the max-stable process Z has the same distribution as

$$\max_{i \geq 1} \zeta_i f(x, U_i), \quad x \in \mathcal{X}, \quad (3)$$

where $\{(\zeta_i, U_i) : i \in \mathbb{N}\}$ are the points of a Poisson process on $(0, \infty) \times \mathbb{R}^d$ with intensity $d\Lambda(\zeta, u) = \zeta^{-2} d\zeta du$. To build parametric max-stable models based on (3), an especially convenient family of functions is $\{f(x-y) : x, y \in \mathbb{R}^d\}$ where f is any valid probability density function on \mathbb{R}^d .

A closely related second spectral characterization enables the use of random functions instead of a family of deterministic functions Schlather (2002); Penrose (1992). Let $\{\zeta_i : i \in \mathbb{N}\}$ be the points of a Poisson process on $(0, \infty)$ with intensity $d\Lambda(\zeta) = \zeta^{-2} d\zeta$. There exists a non negative stochastic process Y with continuous sample paths such that $\mathbb{E}\{Y(x)\} = 1$ for all $x \in \mathbb{R}^d$ and for which Z has the same distribution as

$$\max_{i \geq 1} \zeta_i Y_i(x), \quad x \in \mathcal{X}, \quad (4)$$

where Y_i are independent copies of Y . Clearly representation (3) can be retrieved from (4) by taking $Y_i(x) = f(x, U_i)$. Therefore throughout this paper we will restrict our attention to characterization (4).

3. Finite dimensional distributions

The finite dimensional distributions for a max-stable process Z are easily derived from the spectral characterization (4) and the counting measure N of the associated Poisson point process. More precisely for $z_1, \dots, z_k > 0$ and $x_1, \dots, x_k \in \mathcal{X}$, let

$$\begin{aligned} A &= \{(\zeta, f) \in (0, \infty) \times \mathbb{C}^+(\mathcal{X}) : \zeta f(x_j) > z_j \text{ for some } j \in \{1, \dots, k\}\} \\ &= \left\{ (\zeta, f) \in (0, \infty) \times \mathbb{C}^+(\mathcal{X}) : \zeta > \min_{j=1, \dots, k} \frac{z_j}{f(x_j)} \right\}, \end{aligned}$$

where $\mathbb{C}^+(\mathcal{X})$ denotes the set of non negative continuous functions defined on \mathcal{X} . We have

$$\begin{aligned} \Pr\{Z(x_1) \leq z_1, \dots, Z(x_k) \leq z_k\} &= \Pr[\text{no atoms of } \{(\zeta_i, Y_i) : i \in \mathbb{N}\} \text{ lie in } A] \\ &= \exp \left[- \int \int_0^\infty 1_{\{(\zeta, f) \in A\}} \zeta^{-2} d\zeta dP(f) \right] \\ &= \exp \left\{ - \int \max_{j=1, \dots, k} \frac{f(x_j)}{z_j} dP(f) \right\} \\ &= \exp \left[- \mathbb{E} \left\{ \max_{j=1, \dots, k} \frac{Y(x_j)}{z_j} \right\} \right], \end{aligned} \quad (5)$$

where dP denotes the probability measure of the stochastic process Y .

Based on (3) or (4), several models have been proposed. A first possibility, known as the Smith model Smith (1990), consists in taking $f(x, U_i) = \varphi(x - U_i; \Sigma)$ in (3) where $\varphi(\cdot; \Sigma)$ is the d -variate probability density function of a centered Gaussian random vector with covariance matrix Σ . Its bivariate cumulative distribution function is

$$\Pr\{Z(x_1) \leq z_1, Z(x_2) \leq z_2\} = \exp \left[-\frac{1}{z_1} \Phi \left(\frac{a}{2} + \frac{1}{a} \log \frac{z_2}{z_1} \right) - \frac{1}{z_2} \Phi \left(\frac{a}{2} + \frac{1}{a} \log \frac{z_1}{z_2} \right) \right], \quad (6)$$

where $a^2 = (x_1 - x_2)^T \Sigma^{-1} (x_1 - x_2)$ is the Mahalanobis distance between locations x_1 and x_2 and Φ denotes the standard normal cumulative distribution function. Although closed forms are appreciably more complicated, expressions for the k -variate cumulative distribution functions are given in [Nikoloulopoulos et al. \(2009\)](#).

A second possibility, known as the Schlather model [Schlather \(2002\)](#), takes $Y_i(x) = \sqrt{2\pi} \max\{0, \varepsilon_i(x)\}$ in (4) where ε_i are independent copies of a standard Gaussian process with correlation function ρ . The bivariate cumulative distribution function is

$$\Pr\{Z(x_1) \leq z_1, Z(x_2) \leq z_2\} = \exp \left[-\frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left(1 + \sqrt{1 - \frac{2\{1 + \rho(x_1 - x_2)\} z_1 z_2}{(z_1 + z_2)^2}} \right) \right]. \quad (7)$$

A third possibility, known as the Brown–Resnick model [Brown and Resnick \(1977\)](#); [Kablichko et al. \(2009\)](#), takes $Y_i(x) = \exp\{\varepsilon_i(x) - \sigma^2(x)/2\}$ in (4) where ε_i are independent copies of a centered Gaussian process with stationary increments and such that $\text{Var}\{Y(x)\} = \sigma^2(x)$ for all $x \in \mathcal{X}$. Its bivariate cumulative distribution function is given by (6) with $a^2 = \text{Var}\{Y(x_1 - x_2)\}$. Although the process Y is not stationary, the associated max-stable process is stationary [Kablichko et al. \(2009\)](#).

Finally a fourth possibility that generalizes the Schlather model and known as the extremal- t [Opitz \(2012\)](#); [Ribatet and Sedki \(2013\)](#), assumes in (4) that

$$Y_i(x) = c_v \max\{0, \varepsilon_i(x)\}^v, \quad c_v = \sqrt{\pi} 2^{-(v-2)/2} \Gamma\left(\frac{v+1}{2}\right)^{-1}, \quad v \geq 1,$$

where ε is a standard Gaussian process with correlation function ρ and Γ is the Gamma function. Its bivariate cumulative distribution is

$$\Pr\{Z(x_1) \leq z_1, Z(x_2) \leq z_2\} = \exp \left[-\frac{1}{z_1} T_{v+1} \left\{ -\frac{\rho(x_1 - x_2)}{b} + \frac{1}{b} \left(\frac{z_2}{z_1} \right)^{1/v} \right\} - \frac{1}{z_2} T_{v+1} \left\{ -\frac{\rho(x_1 - x_2)}{b} + \frac{1}{b} \left(\frac{z_1}{z_2} \right)^{1/v} \right\} \right], \quad (8)$$

where T_v is the cumulative distribution function of a Student random variable with v degrees of freedom and $b^2 = \{1 - \rho(x_1 - x_2)^2\}/(v+1)$.

It is important to mention that several choices for Y in (4) yield to the same max-stable process. To illustrate, consider the process

$$\tilde{Z}(x) = \max_{i \geq 1} \zeta_i S_i Y_i(x), \quad x \in \mathcal{X},$$

where $\{(\zeta_i, Y_i) : i \in \mathbb{N}\}$ are as in (4) and S_i are independent copies of a non negative random variable S such that $\mathbb{E}(S) = 1$, S independent from the points $\{(\zeta_i, Y_i) : i \in \mathbb{N}\}$. From (5) we have

$$\begin{aligned} \Pr\{\tilde{Z}(x_1) \leq z_1, \dots, \tilde{Z}(x_k) \leq z_k\} &= \exp \left[-\mathbb{E} \left\{ \max_{j=1, \dots, k} \frac{SY(x_j)}{z_j} \right\} \right] \\ &= \exp \left[-\mathbb{E} \left\{ \max_{j=1, \dots, k} \frac{Y(x_j)}{z_j} \right\} \right] \\ &= \Pr\{Z(x_1) \leq z_1, \dots, Z(x_k) \leq z_k\}, \end{aligned}$$

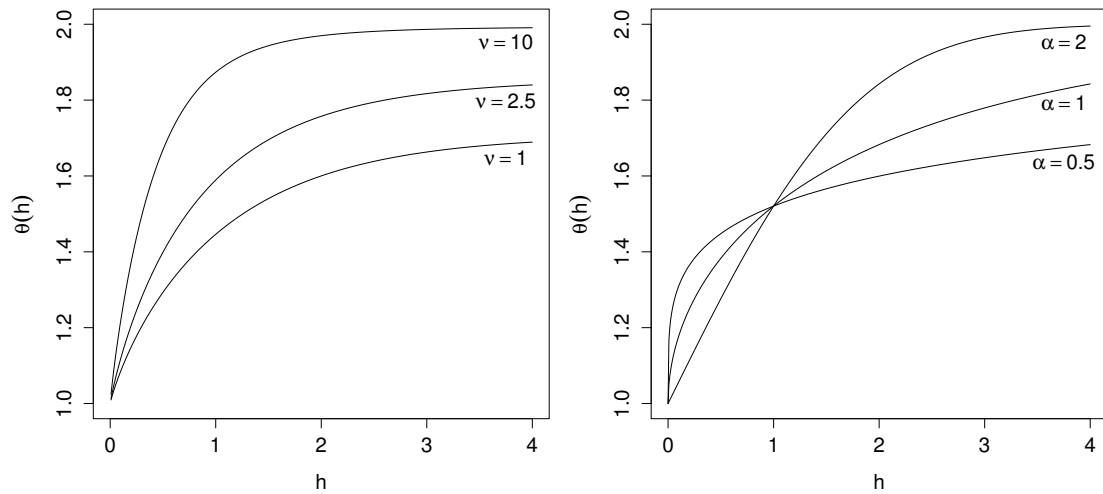


FIGURE 1. Extremal coefficient functions. Left: Extremal- t process with $\rho(h) = \exp(-h)$ and degrees of freedom $v = 1, 2.5, 10$. Right: Brown-Resnick process with $\gamma(h) = h^\alpha$ and $\alpha = 0.5, 1, 2$. For the extremal- t model the case $v = 1$ corresponds to a Schlather process. For the Brown-Resnick model the case $\alpha = 2$ corresponds to a Smith process with a covariance matrix equal to the identity matrix.

i.e., \tilde{Z} and Z are identically distributed. As an example, taking $Y(x) = \sqrt{2\pi} \max\{0, T(x)\}$ in (4) where T is a standard Student process with v degrees of freedom yields to the usual Schlather process (7) since any Student processes with v degrees of freedom has the representation $T(x) = S\varepsilon(x)$, $x \in \mathcal{X}$, where ε is a standard Gaussian process and $v/S^2 \sim \chi_v^2$, $v \geq 1$.

4. Spatial dependence

In Section 3 we introduced several max-stable models with different dependence structures. For concrete applications, it is essential to be able to assess if a fitted max-stable model is able to capture the spatial dependence structure of extreme events.

One convenient way to summarize the dependence structure of max-stable process Z is through its extremal coefficient function Schlather and Tawn (2003)

$$\theta(x_1 - x_2) = -z \log \Pr\{Z(x_1) \leq z, Z(x_2) \leq z\} = \mathbb{E}[\max\{Y(x_1), Y(x_2)\}]. \quad (9)$$

The extremal coefficient function takes values in $[1, 2]$; the lower bounds indicates complete dependence while the upper bound independence. The function $h \mapsto 2 - \theta(h)$ is semi definite positive and not differentiable at the origin unless $\theta(h) \equiv 1$ Schlather and Tawn (2003); Cooley et al. (2006). Figure 1 plots the extremal coefficient function for isotropic versions of the max-stable models introduced in Section 3 and different parameter values. For the extremal- t process, we can see that the degrees of freedom v controls the upper bound of the extremal coefficient since for this model we have

$$\theta(h) = 2T_{v+1} \left\{ \sqrt{\frac{v+1}{1-\rho(h)^2}} - \sqrt{\frac{1-\rho(h)^2}{v+1}} \rho(h) \right\} \rightarrow 2T_{v+1}(\sqrt{v+1}), \quad h \rightarrow \infty.$$

The Brown–Resnick process has extremal coefficient function $\theta(h) = 2\Phi\{\sqrt{\gamma(h)}/2\}$ and its upper bound is related to that of the semi variogram γ . In particular if γ is unbounded, e.g., $\gamma(h) \propto h^\alpha$, $\theta(h) \rightarrow 2$ as $h \rightarrow \infty$.

In geostatistics it is common to assess the spatial dependence using the semi variogram. Unfortunately when working with extreme values, it might happen, e.g., with simple max-stable processes, that the variogram does not exist. As a substitute to the semi variogram, one can use the F -madogram [Cooley et al. \(2006\)](#)

$$v_F(x_1 - x_2) = \frac{1}{2} \mathbb{E}[|F_{x_1}\{Z(x_1)\} - F_{x_2}\{Z(x_2)\}|], \quad x_1, x_2 \in \mathcal{X}, \quad (10)$$

where F_x denotes the cumulative distribution function of the random variable $Z(x)$. Using the relation $|x - y| = 2 \max(x, y) - x - y$ and the max-stability property of Z , it can be shown that

$$\theta(x_1 - x_2) = \frac{1 + 2v_F(x_1 - x_2)}{1 - 2v_F(x_1 - x_2)}, \quad x_1, x_2 \in \mathcal{X}. \quad (11)$$

From a practical point of view, to estimate the extremal coefficient function one need to estimate the above expectation and F_{x_1} and F_{x_2} . This can be done non parametrically by using the empirical version of (10) and plugin it into (11), i.e.,

$$\hat{v}_F(x_1 - x_2) = \frac{1}{2n(n+1)} \sum_{i=1}^n |R_i(x_1) - R_i(x_2)|,$$

where $R_i(x_j)$ denotes the rank of $Z_i(x_j)$, i.e., $R_i(x_j) = \sum_{\ell=1}^n 1_{\{Z_\ell(x_j) \leq Z_i(x_j)\}}$ with $i = 1, \dots, n$ and $j = 1, 2$. Instead of using empirical distribution functions for F_{x_j} , $j = 1, 2$, another possibility is to use the fitted generalized extreme value distributions for locations x_1 and x_2 .

Similarly to the estimation of the empirical variogram in geostatistics, if the process has been observed at several location $x_1, \dots, x_k \in \mathcal{X}$, it is often useful to average over all pairs of stations whose pairwise distances lie in $C_h = (h - \Delta, h + \Delta)$ for some suitable $\Delta > 0$. The binned F -madogram estimator is

$$\hat{v}_{F,b}(h) = \frac{1}{|C_h|} \sum_{i=1}^{k-1} \sum_{j=i+1}^k \hat{v}_F(x_i - x_j) 1_{\{\|x_i - x_j\| \in C_h\}}, \quad |C_h| = \sum_{i=1}^{k-1} \sum_{j=i+1}^k 1_{\{\|x_i - x_j\| \in C_h\}}.$$

5. Inference

Due to the specific form of the finite dimensional distributions (5), fitting max-stable processes to spatial data sets is complicated. The aim of a spatial analysis is usually to enable prediction at unobserved location and, although non parametric approaches are possible [Buishand et al. \(2008\)](#), parametric inferences seems more appropriate. Likelihood based inferences are widely used either in a frequentist context [Genton et al. \(2011\)](#); [Davison et al. \(2012\)](#); [Padoan et al. \(2010\)](#) or in a Bayesian framework [Ribatet et al. \(2012\)](#); [Erhardt and Smith \(2012\)](#).

Unfortunately the likelihood is intractable even when the process has been observed at a moderate number $k \geq 1$ of locations $x_1, \dots, x_k \in \mathcal{X}$. Since the finite dimensional distributions are

$$\Pr\{Z(x_1) \leq z_1, \dots, Z(x_k) \leq z_k\} = \exp\{-V(z_1, \dots, z_k)\}, \quad V(z_1, \dots, z_k) = \mathbb{E} \left\{ \max_{j=1, \dots, k} \frac{Y(x_j)}{z_j} \right\},$$

the associated density is

$$f(z_1, \dots, z_k; \psi) = \exp\{-V(z_1, \dots, z_k)\} \sum_{\tau \in \mathcal{P}_k} w(\tau), \quad (12)$$

where the functions V and w depend on some unknown parameter ψ , \mathcal{P}_k denotes the set of all possible partition of the set $\{x_1, \dots, x_k\}$, $\tau = (\tau_1, \dots, \tau_\ell)$, $|\tau| = \ell$ is the size of the partition τ and

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

where $\partial^{|\tau_j|} / \partial z_{\tau_j}$ denotes the mixed partial derivatives with respect to the elements of the j -th element of the partition τ . For instance when $\tau = \{(x_1, x_2), (x_3)\} \in \mathcal{P}_3$, we have

$$\frac{\partial^{|\tau_1|}}{\partial z_{\tau_1}} V(z_1, \dots, z_k) = \frac{\partial^2}{\partial z_1 \partial z_2} V(z_1, z_2, z_3), \quad \frac{\partial^{|\tau_2|}}{\partial z_{\tau_2}} V(z_1, \dots, z_k) = \frac{\partial}{\partial z_3} V(z_1, z_2, z_3).$$

As a consequence and even if we were able to compute explicitly $\mathbb{E}\{\max_{j=1, \dots, k} Y(x_j) / z_j\}$, the number of weights $w(\tau)$ in (12) corresponds to the k -th Bell number and the likelihood is intractable at least numerically. For instance when $k = 10$ this would require to sum over more than 115000 terms.

To bypass this computational burden, the use of composite likelihoods has been introduced Padoan et al. (2010); Genton et al. (2011). In this section we focus on the pairwise likelihood but composite likelihoods based on triplets are sometimes possible Genton et al. (2011); Huser and Davison (2013). Given a single observation $\mathbf{z} = (z_1, \dots, z_k)$, the (weighted) pairwise log-likelihood is

$$\ell_p(\psi; \mathbf{z}) = \sum_{i=1}^{k-1} \sum_{j=i+1}^k \omega_{i,j} \log f(z_i, z_j; \psi), \quad (13)$$

where $\omega_{i,j}$ are suitable non negative weights and $f(\cdot, \cdot; \psi)$ is the bivariate density, i.e., f in (12) with $k = 2$. One has to remind that composite likelihoods are linear combination of log-likelihoods but, apart from trivial cases, are not genuine likelihoods. Composite likelihood is a simple strategy to build unbiased estimating equations.

With the same regularity conditions required for the maximum likelihood estimator and provided that the parameter ψ is identifiable from the bivariate densities in (13), the maximum pairwise likelihood estimator

$$\hat{\psi}_p = \arg \max_{\psi \in \Psi} \ell_p(\psi; \mathbf{z}),$$

satisfies

$$\sqrt{n}(\hat{\psi}_p - \psi_0) \longrightarrow N\{0, H^{-1}(\psi_0)J(\psi_0)H^{-1}(\psi_0)\}, \quad n \rightarrow \infty, \quad (14)$$

where

$$H(\psi_0) = -\mathbb{E}\{\nabla^2 \ell_p(\psi_0; \mathbf{Z})\}, \quad J(\psi_0) = \text{Var}\{\nabla \ell_p(\psi_0; \mathbf{Z})\}.$$

Due to the model under specification, (14) shows that there is a loss in efficiency compared to the maximum likelihood estimator. To mitigate the loss in efficiency, a pragmatic approach

consists in defining suitable weights to improve efficiency [Padoan et al. \(2010\)](#); [Varin et al. \(2011\)](#) but unfortunately defining optimal weights remains an open question.

Provided we have independent replications $\mathbf{z}_1, \dots, \mathbf{z}_n$, estimation of $H(\psi_0)$ and $J(\psi_0)$ is not too complicated [Varin et al. \(2011\)](#). Since the regularity conditions imply that $\sum_{m=1}^n \nabla \ell_p(\hat{\psi}_p; \mathbf{z}_m) = 0$, the $J(\psi_0)$ matrix can be estimated by

$$\hat{J}(\psi_0) = \frac{1}{n} \sum_{m=1}^n \{\nabla \ell_p(\hat{\psi}_p; \mathbf{z}_m)\} \{\nabla \ell_p(\hat{\psi}_p; \mathbf{z}_m)\}^T.$$

Assuming that the bivariate densities in (13) are correctly specified we can use the second Bartlett identity to have

$$H(\psi_0) = - \sum_{i < j} \omega_{i,j} \mathbb{E} \{ \nabla^2 \log f(Z_i, Z_j; \psi_0) \} = \sum_{i < j} \omega_{i,j} \text{Var} \{ \nabla \log f(Z_i, Z_j; \psi_0) \},$$

and the matrix $H(\psi_0)$ can be estimated by

$$\hat{H}(\psi_0) = \frac{1}{n} \sum_{k=1}^n \sum_{1 \leq i < j \leq k} \omega_{i,j} \nabla \log f(z_{m,i}, z_{m,j}; \hat{\psi}_p) \{ \nabla \log f(z_{m,i}, z_{m,j}; \hat{\psi}_p) \}^T.$$

6. Model selection

Although the inference procedure introduced in Section 5 is non standard, model selection using (adapted) information criterion [Varin and Vidoni \(2005\)](#) or (composite) likelihood ratio tests [Chandler and Bate \(2007\)](#) is possible.

Because we are working with composite likelihood, one can define a composite Kullback–Leibler divergence based on (13) between two statistical models g and f [Varin and Vidoni \(2005\)](#)

$$D_p(f_\psi; g) = \sum_{i=1}^{k-1} \sum_{j=i+1}^k w_{i,j} \mathbb{E} \left\{ \log \frac{g(Z_i, Z_j)}{f(Z_i, Z_j; \psi)} \right\}, \quad (15)$$

where $\mathbf{Z} = (Z_1, \dots, Z_k) \sim g$. In (15) composite likelihoods are used for both the true model g and the one under consideration f . Since the composite Kullback–Leibler divergence consists in a linear combination of (usual) Kullback–Leibler divergences, it can be shown using standard computations on likelihoods that one can perform model selection by minimizing the composite information criterion

$$\text{TIC}(f_\psi) = -2\ell_p(\hat{\psi}_p; \mathbf{z}) + 2 \text{tr} \{ \hat{J}(\psi_0) \hat{H}(\psi_0)^{-1} \}, \quad (16)$$

where $\hat{H}(\psi_0)$ and $\hat{J}(\psi_0)$ are consistent estimator of the matrices $H(\psi_0)$ and $J(\psi_0)$. Equation (16) is a generalization of the Takeuchi's information criterion to the case of the maximum composite likelihood estimator. Similarly to the Bayesian information criterion, one can prefer to use the composite Bayesian information criterion

$$\text{BIC}(f_\psi) = -2\ell_p(\hat{\psi}_p; \mathbf{z}) + \text{tr} \{ \hat{J}(\psi_0) \hat{H}(\psi_0)^{-1} \} \log n. \quad (17)$$

Similarly to the likelihood ratio test, one can use the composite likelihood ratio to test if the data are consistent with the hypothesis that the parameter $\psi = (\phi, \gamma)$ takes the particular form (ϕ, γ_*) . It can be useful when comparing nested models, for instance whether it is more appropriate to use an exponential correlation function $\rho(h) = \exp(-h/\lambda)$ or a powered exponential correlation function $\rho(h) = \exp\{-(h/\lambda)^\kappa\}$ in (7), i.e., $H_0: \kappa = 1$ against $H_1: \kappa \neq 1$.

Because the maximum composite likelihood estimator $\hat{\psi}_p$ is asymptotically normal, it can be shown using standard arguments Kent (1982) that

$$W(\gamma_*) = 2 \{ \ell_p(\hat{\psi}_p; \mathbf{z}) - \ell_p(\hat{\phi}_{\gamma_*}, \gamma_*) \} \longrightarrow \sum_{i=1}^p \lambda_i X_i, \quad n \rightarrow \infty, \quad (18)$$

where p is the dimension of the parameter ψ , $\hat{\phi}_{\gamma_*}$ is the maximum pairwise likelihood estimator subject to the constraint $\gamma = \gamma_*$, X_i are independent χ_1^2 random variables and λ_i are the eigenvalues of

$$\{H^{-1}(\psi_0)J(\psi_0)H^{-1}(\psi_0)\}_\gamma [\{H^{-1}(\psi_0)\}_\gamma]^{-1},$$

where A_γ denotes the submatrix of the matrix A corresponding to the elements of γ . Although the limiting distribution of $W(\gamma_*)$ has no closed form, it can be modified so that the usual χ_p^2 limiting distribution is (approximately) true Chandler and Bate (2007); Rotnitzky and Jewell (1990).

7. Unconditional simulation

Simulation from max-stable processes seems at first glance impossible since it consists in computing the pointwise maxima over an infinite number of random functions. Fortunately, as we will see later, the special form of the intensity measure Λ of the Poisson point process $\{\zeta_i: i \in \mathbb{N}\}$ in (4) makes the impossible possible.

A naive approach consists in finding the normalizing functions a_n and b_n associated to the process X in (1) and in computing the maxima over n normalized random functions from (1) for n large enough. This approach is not recommended since convergence is usually slow and n has to be very large to ensure good approximations.

A second approach relies on the spectral characterizations (3) or (4) Schlather (2002) and is detailed in Algorithm 1. This approach is often much more efficient. Without loss of generality, we can assume that the points $\{\zeta_i: i \in \mathbb{N}\}$ are sorted in decreasing order so that $\zeta_i \downarrow 0$ as $i \rightarrow \infty$. If the process Y is uniformly bounded by a positive constant $B < \infty$, then there exists almost surely a finite number of atoms $\{(\zeta_i, Y_i)\}$ that contribute to the pointwise maxima Z —see line 8 of Algorithm 1. Usually Y will not be uniformly bounded and one can use a “pseudo uniform bound” B such that $\Pr(\|Y\| > B)$ is small enough.

Figure 2 illustrates the procedure with a one dimensional Smith model on $\mathcal{X} = [-1, 1]$. Clearly for this model the process Y is uniformly bounded by $B = (2\pi\sigma^2)^{-1/2}$ where σ^2 is the variance of the univariate normal density. We can see that the simulated max-stable process was obtained from only 22 random functions.

Algorithm 1 is likely to fail if the process Y in (4) is not stationary but the underlying max-stable process is. A particularly well known example is the case of Brown–Resnick processes and, although more sophisticated strategies are possible Oesting et al. (2012), a simple yet efficient

Algorithm 1: Algorithm for simulating max-stable processes with unit Fréchet margins.

Input : An upper bound $B > 0$.
Output : One realization of a max-stable process.

```

1 Initialization;
2  $i \leftarrow 1$ , flag  $\leftarrow$  true,  $S \leftarrow 0$ ,  $Z \equiv 0$ ;
3 while flag is true do
4    $S \leftarrow S + E$ ,  $E \sim \text{Exp}(1)$ ;
5    $\zeta_i \leftarrow 1/S$ ;
6    $Y_i \sim Y$ ;
7    $Z \leftarrow \max(Z, \zeta_i Y_i)$ ;
8   if  $\zeta_i B < \inf(Z)$  then
9     flag  $\leftarrow$  false;
10  else
11    flag  $\leftarrow$  true;
12  end
13   $i \leftarrow i + 1$ ;
14 end
15 return  $Z$ ;
```

possibility is to use random translations to mitigate the impact of the non stationarity of Y . Instead of using (4), the idea consists in sampling from

$$\tilde{Z}(x) = \max_{i \geq 1} \zeta_i Y_i(x - U_i), \quad x \in \mathcal{X}, \quad (19)$$

where $\{(\zeta_i, Y_i) : i \in \mathbb{N}\}$ are as in (4) and U_i are independently sampled from an arbitrary distribution F defined (at least) on \mathcal{X} .

The process \tilde{Z} has the same distribution as Z since

$$\begin{aligned} \Pr\{\tilde{Z}(x_1) \leq z_1, \dots, \tilde{Z}(x_k) \leq z_k\} &= \exp \left[- \int \mathbb{E} \left\{ \max_{j=1, \dots, k} \frac{Y(x_j - u)}{z_j} \right\} dF(u) \right] \\ &= \exp \left[\int \log \Pr\{Z(x_1 - u) \leq z_1, \dots, Z(x_k - u) \leq z_k\} dF(u) \right] \\ &= \Pr\{Z(x_1) \leq z_1, \dots, Z(x_k) \leq z_k\}, \end{aligned}$$

since the process Z is stationary.

Figure 3 shows two realizations of a Brown–Resnick process on $\mathcal{X} = [-1, 1]$ with semivariogram $\gamma(h) = (5h)^{1.9}$. It is clear that Algorithm 1 yields to a non stationary sample path while the use of independent random translations $U_i \sim U(-1.5, 1.5)$ gives a much more plausible sample path.

8. Conditional simulation

Recently conditional simulation from max-stable processes has gained some interests with the pioneering work of Wang and Stoev (2011) followed by that of Dombry and Éyi-Minko (2013); Dombry et al. (2013). In Wang and Stoev (2011), a first solution was proposed for max-linear

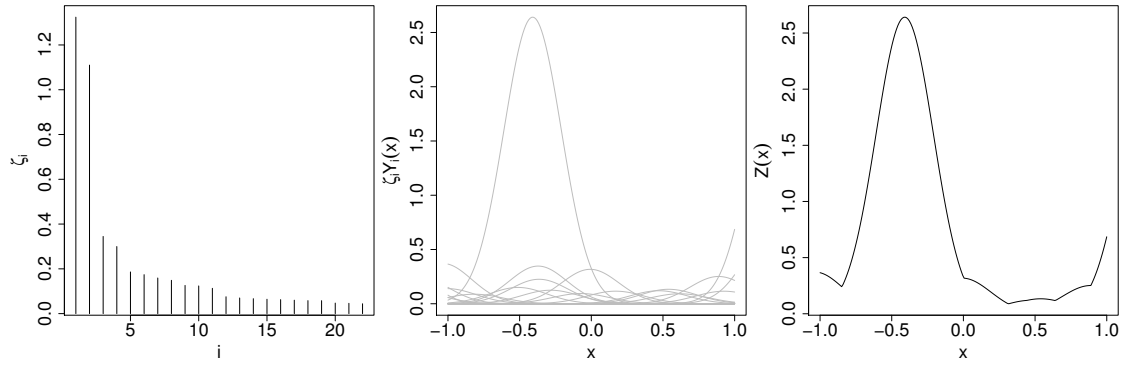


FIGURE 2. Illustration of the sampling procedure for max-stable processes based on the spectral characterizations (3) or (4). The illustration uses a one dimensional Smith model with $\mathcal{X} = [-1, 1]$, $\sigma = 0.2$ and $B = 5/\sqrt{2\pi}$. The left panel plots the ordered sequence $\{\zeta_i: i \in \mathbb{N}\}$. The middle panel shows the simulated random functions $\{\zeta_i Y_i: i \in \mathbb{N}\}$. The right panel shows the simulated max-stable process.

processes but in this paper we will restrict our attention to max-stable processes with spectral representation (4).

Given some fixed locations $x_1, \dots, x_k \in \mathcal{X}$ and fixed values $z_1, \dots, z_k > 0$, our goal is to sample from

$$Z(x) \mid \{Z(x_1) = z_1, \dots, Z(x_k) = z_k\}, \quad x \in \mathcal{X}, \quad (20)$$

where Z is a simple max-stable process on \mathcal{X} .

Let $\mathbf{x} = (x_1, \dots, x_k)$, $\mathbf{z} = (z_1, \dots, z_k)$ and, to ease the notations, we write $f(\mathbf{x}) = \{f(x_1), \dots, f(x_k)\}$ for any function f . Let $\Phi = \{\varphi_i: i \in \mathbb{N}\}$ be a Poisson point process on $\mathbb{C}^+(\mathcal{X})$ with $\varphi_i = \zeta_i Y_i$, (ζ_i, Y_i) as in (4). Provided Φ is regular, i.e., its intensity measure is absolutely continuous with respect to the Lebesgue measure (at least) on $(0, \infty)$, it is straightforward to show that the intensity measure of the Poisson point process $\{\zeta_i Y_i(\mathbf{x}): i \in \mathbb{N}\}$ defined on $(0, \infty)^k$ is

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

for all Borel set $A \subset \mathbb{R}^k$.

Given $Z(\mathbf{x}) = \mathbf{z}$, it can be shown Dombry et al. (2013) that Φ can be decomposed into two independent point processes

$$\Phi^- = \{\varphi \in \Phi: \varphi(x_i) < z_i, i = 1, \dots, k\}, \quad \Phi^+ = \bigcup_{i=1}^k \{\varphi \in \Phi: \varphi(x_i) = z_i\}.$$

The atoms of Φ^+ , called the extremal functions, correspond to random functions contributing to the conditioning event $Z(\mathbf{x}) = \mathbf{z}$ while the atoms of Φ^- , called the sub-extremal functions, are random functions that do not contribute to the process Z at the locations \mathbf{x} but possibly at other locations as illustrated by Figure 4.

Clearly the conditional event $Z(\mathbf{x}) = \mathbf{z}$ is realized by ℓ extremal functions with $\ell \in \{1, \dots, k\}$ and these extremal functions form a partition $\tau = (\tau_1, \dots, \tau_\ell) \in \mathcal{P}_k$ such that there exists a unique

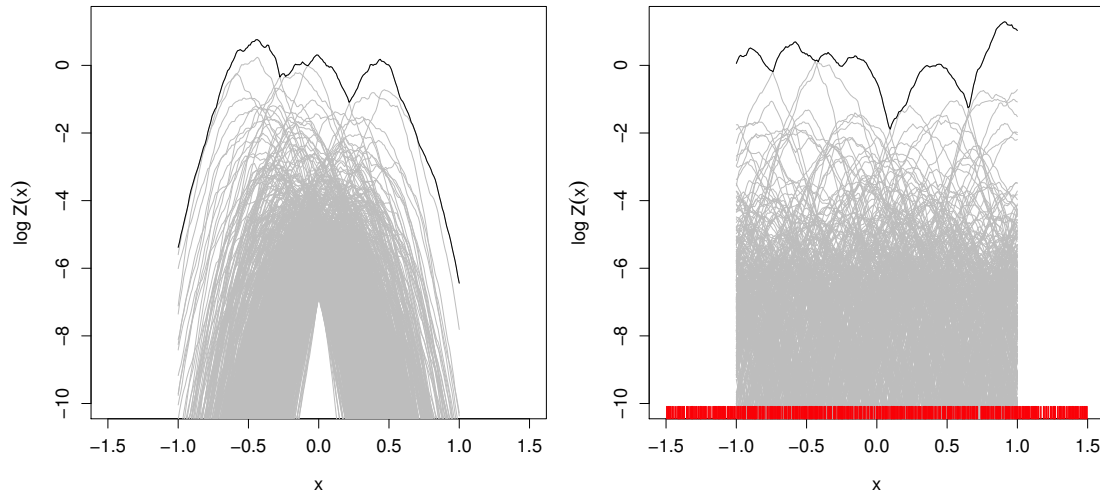


FIGURE 3. Illustration of the problem for simulating Brown-Resnick processes on $\mathcal{X} = [-1, 1]$. The plot shows two realizations of a Brown-Resnick process with semivariogram $\gamma(h) = (5h)^{1.9}$ using Algorithm 1 (left) and random translations. The black lines correspond to the simulated Brown-Resnick processes and the grey lines to the random function $\zeta_i Y_i$. The red tick marks corresponds to the translated origins $U_i \sim U(-1.5, 1.5)$ in (19).

extremal function $\varphi^+ \in \Phi^+$ satisfying $\varphi^+(x_{\tau_j}) = z_{\tau_j}$ and $\varphi^+(x_{\tau_j^c}) < z_{\tau_j^c}$ where $x_{\tau_j} = \{x_i : i \in \tau_j\}$ and $x_{\tau_j^c} = \{x_i : i \notin \tau_j\}$.

Therefore conditional simulation of max-stable processes consists in a three-step procedure:

Step 1 Sample a random partition $\vartheta \in \mathcal{P}_k$ from Dombry and Éyi-Minko (2013); Dombry et al. (2013)

$$\Pr\{\vartheta = \tau \mid Z(\mathbf{x}) = \mathbf{z}\} = \frac{1}{C(\mathbf{x}, \mathbf{z})} \prod_{j=1}^{|\tau|} \int_{\{\mathbf{u}_j < \mathbf{z}_{\tau_j^c}\}} \lambda_{(\mathbf{x}_{\tau_j}, \mathbf{x}_{\tau_j^c})}(\mathbf{z}_{\tau_j}, \mathbf{u}_j) d\mathbf{u}_j, \quad (21)$$

where $\tau \in \mathcal{P}_k$ and the normalizing constant is

$$C(\mathbf{x}, \mathbf{z}) = \sum_{\tau \in \mathcal{P}_k} \prod_{j=1}^{|\tau|} \int_{\{\mathbf{u}_j < \mathbf{z}_{\tau_j^c}\}} \lambda_{(\mathbf{x}_{\tau_j}, \mathbf{x}_{\tau_j^c})}(\mathbf{z}_{\tau_j}, \mathbf{u}_j) d\mathbf{u}_j.$$

Step 2 Given $\vartheta = \tau$ of size ℓ , sample independently the extremal functions $\varphi_1^+, \dots, \varphi_\ell^+$ from the distributions

$$\Pr\left\{\varphi_j^+(\mathbf{s}) \in d\mathbf{v} \mid Z(\mathbf{x}) = \mathbf{z}, \vartheta = \tau\right\} = \frac{1}{C_j} \left\{ \int 1_{\{\mathbf{u} < \mathbf{z}_{\tau_j^c}\}} \lambda_{(\mathbf{s}, \mathbf{x}_{\tau_j^c})|\mathbf{x}_{\tau_j}, \mathbf{z}_{\tau_j}}(\mathbf{v}, \mathbf{u}) d\mathbf{u} \right\} d\mathbf{v},$$

where the normalizing constant is

$$C_j = \int 1_{\{\mathbf{u} < \mathbf{z}_{\tau_j^c}\}} \lambda_{(\mathbf{s}, \mathbf{x}_{\tau_j^c})|\mathbf{x}_{\tau_j}, \mathbf{z}_{\tau_j}}(\mathbf{v}, \mathbf{u}) d\mathbf{u} d\mathbf{v},$$

and the conditional intensity function

$$\lambda_{(\mathbf{s}, \mathbf{x}_{\tau_j^c})|\mathbf{x}_{\tau_j}, \mathbf{z}_{\tau_j}}(\mathbf{v}, \mathbf{u}) = \frac{\lambda_{(\mathbf{s}, \mathbf{x}_{\tau_j^c}, \mathbf{x}_{\tau_j})}(\mathbf{v}, \mathbf{u}, \mathbf{z}_{\tau_j})}{\lambda_{\mathbf{x}_{\tau_j}}(\mathbf{z}_{\tau_j})}, \quad \mathbf{s} \in \mathcal{X}^m, \quad (\mathbf{v}, \mathbf{u}) \in \mathbb{R}^{m+|\tau_j^c|}$$

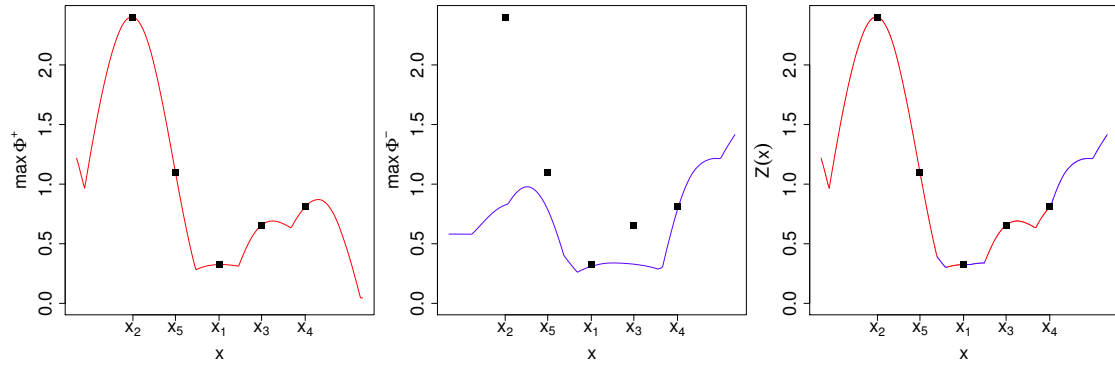


FIGURE 4. Illustration of the decomposition of the Poisson point process $\Phi = \{\zeta_i Y_i : i \in \mathbb{N}\}$ into the two point Processes Φ^+ and Φ^- . From left to right are shown the pointwise maxima of the extremal functions, the subextremal functions and of all atoms of Φ . The squares denotes the conditioning events $Z(\mathbf{x}) = \mathbf{z}$. The decomposition was obtained from a realization of a Schlather process with a Gaussian correlation function, i.e., $\rho(h) = \exp(-h^2)$.

is the distribution of an atom $\varphi \in \Phi$ such that $\varphi(\mathbf{x}_{\tau_j}) = \mathbf{z}_{\tau_j}$. Let

$$Z^+ = \max(\varphi_1^+, \dots, \varphi_\ell^+).$$

Step 3 Independently draw a “thinned” max-stable process, i.e.,

$$Z^- = \max_{i \geq 1} \zeta_i Y_i 1_{\{\zeta_i Y_i(\mathbf{x}) < \mathbf{z}\}},$$

with $\{(\zeta_i, Y_i) : i \in \mathbb{N}\}$ as in (4).

Finally the random vector $\max\{Z^-(\mathbf{s}), Z^+(\mathbf{s})\}$ follows the conditional distribution of $Z(\mathbf{s})$ | $\{Z(\mathbf{x}) = \mathbf{z}\}$, $\mathbf{s} \in \mathcal{X}^m$.

In the above three-step procedure, the most difficult stage is Step 1 as it amounts to sample from a discrete distribution whose state space \mathcal{P}_k is huge. One possibility is to use a Gibbs sampler as proposed in Dombry et al. (2013). Recall that a Gibbs sampler requires to sample from the full conditional distributions. For our purpose because the full conditional distribution satisfies

$$\Pr\{\vartheta = \tau^* \mid Z(\mathbf{x}) = \mathbf{z}, \vartheta_{-j} = \tau_{-j}\} \propto \frac{\Pr\{\vartheta = \tau^* \mid Z(\mathbf{x}) = \mathbf{z}\}}{\Pr\{\vartheta = \tau \mid Z(\mathbf{x}) = \mathbf{z}\}} 1_{\{\tau_{-j}^* = \tau_{-j}\}},$$

where τ_{-j} denotes the restriction of the partition τ to the set $\{x_1, \dots, x_k\} \setminus \{x_j\}$, many terms will cancel out due to the product form of (21). It makes the Gibbs sampler especially convenient.

For practical purposes and given a parametric max-stable model, one need to get closed forms for $\lambda_{\mathbf{x}}$ and $\lambda_{\mathbf{s}|\mathbf{x}, \mathbf{z}}$. If the derivation of the former is usually not too complicated, the derivation of the latter is more interesting since it says from which stochastic processes the extremal functions are drawn.

It can be shown that the extremal functions for the Brown–Resnick model are log-normal processes, i.e.,

$$\lambda_{\mathbf{s}|\mathbf{x}, \mathbf{z}}(\mathbf{u}) = (2\pi)^{-m/2} |\Sigma_{\mathbf{s}|\mathbf{x}}|^{-1/2} \exp\left\{-\frac{1}{2} (\log \mathbf{u} - \mu_{\mathbf{s}|\mathbf{x}, \mathbf{z}})^T \Sigma_{\mathbf{s}|\mathbf{x}}^{-1} (\log \mathbf{u} - \mu_{\mathbf{s}|\mathbf{x}, \mathbf{z}})\right\} \prod_{i=1}^m u_i^{-1},$$

where $\mathbf{s} \in \mathcal{X}^m$, $\mathbf{u} \in \mathbb{R}^m$ and closed forms for the mean vector $\mu_{\mathbf{s}|\mathbf{x},\mathbf{z}}$ and the covariance matrix $\Sigma_{\mathbf{s}|\mathbf{x}}$ are given in Dombry et al. (2013).

For the Schlather model, the extremal functions are Student processes Dombry et al. (2013), i.e.,

$$\lambda_{\mathbf{s}|\mathbf{x},\mathbf{z}}(\mathbf{u}) = \pi^{-m/2} (k+1)^{-m/2} |\tilde{\Sigma}|^{-1/2} \left\{ 1 + \frac{(\mathbf{u} - \mu)^T \tilde{\Sigma}^{-1} (\mathbf{u} - \mu)}{k+1} \right\}^{-(m+k+1)/2} \frac{\Gamma\left(\frac{k+m+1}{2}\right)}{\Gamma\left(\frac{k+1}{2}\right)},$$

where the mean vector is $\mu = \Sigma_{\mathbf{s}:\mathbf{x}} \Sigma_{\mathbf{x}}^{-1} \mathbf{z}$ and the dispersion matrix is

$$\tilde{\Sigma} = \frac{a_{\mathbf{x}}(\mathbf{z})}{k+1} (\Sigma_{\mathbf{s}} - \Sigma_{\mathbf{s}:\mathbf{x}} \Sigma_{\mathbf{x}}^{-1} \Sigma_{\mathbf{x}:\mathbf{s}}), \quad \Sigma_{(\mathbf{s},\mathbf{x})} = \begin{pmatrix} \Sigma_{\mathbf{s}} & \Sigma_{\mathbf{s}:\mathbf{x}} \\ \Sigma_{\mathbf{x}:\mathbf{s}} & \Sigma_{\mathbf{x}} \end{pmatrix}, \quad (22)$$

and $\Sigma_{(\mathbf{s},\mathbf{x})} = \text{Var}[\varepsilon\{(\mathbf{s},\mathbf{x})\}]$ where ε is the standard Gaussian process appearing in the spectral representation of the Schlather model.

Using the same lines as in Dombry et al. (2013), it is not difficult to show (see Appendix A) that the conditional intensity function for the extremal- t model is the density of $T(\mathbf{s})^v$ where T is a Student process with $k+v$ degrees of freedom and respective mean vector and scale matrix

$$\mu = \Sigma_{\mathbf{s}:\mathbf{x}} \Sigma_{\mathbf{x}}^{-1} \mathbf{z}^{1/v}, \quad \tilde{\Sigma} = (k+v)^{-1} a_{\mathbf{x}}(\mathbf{z}, v) (\Sigma_{\mathbf{s}} - \Sigma_{\mathbf{s}:\mathbf{x}} \Sigma_{\mathbf{x}}^{-1} \Sigma_{\mathbf{x}:\mathbf{s}}),$$

where $a_{\mathbf{x}}(\mathbf{z}, v) = (\mathbf{z}^{1/v})^T \Sigma_{\mathbf{x}} \mathbf{z}^{1/v}$ and the matrices $\Sigma_{\mathbf{s}}, \Sigma_{\mathbf{x}}, \Sigma_{\mathbf{x}:\mathbf{s}}$ and $\Sigma_{\mathbf{s}:\mathbf{x}}$ are as in (22)—see Appendix A for more details.

9. Application

In this section we will see how we can apply the theory introduced in the previous sections to model extreme wind gusts in Netherland using the R package `SpatialExtremes` Ribatet et al. (2013). This package collects various function to model spatial extremes using max-stable processes, Bayesian hierarchical models or copula. Simulation routines from various statistical models are available. If one aim at simulating max-stable processes, an excellent alternative package is `RandomFields` Schlather et al. (2013).

The data consist in wind speed series (km/h) observed at 35 weather stations located in the Netherlands and provided by the Royal Netherland Meteorological Institute (http://www.knmi.nl/index_en.html). Annual maxima were extracted from the raw data for the time period 1971–2012. Figure 5 shows the locations of these 35 weather stations as well as time frames where annual maxima were available for each station. We can see that most of the time series start around 1990 and ends in 2012. The right panel of Figure 5 plots the deviation of the sample mean (of annual maxima) at each station from the overall sample mean, i.e., the sample mean computed over all weather stations. It seems that there is a North–West to South–East gradient where, as expected, the largest wind gusts are observed along the coastline.

Although for theoretical purposes there is no loss of generality focusing only on simple max-stable models, i.e., with unit Fréchet margins, for concrete applications one has to allow for (pointwise) marginal distributions varying in space. This can be done by defining trend surfaces

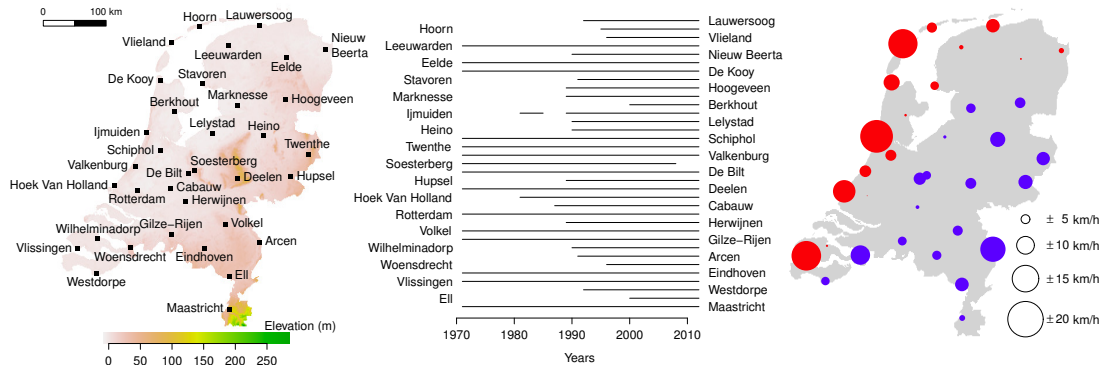


FIGURE 5. The wind gust data. The left panel shows the locations of the 35 Dutch weather stations. The middle panel shows the time period where annual maxima are available at each station. The right panel plots circles at each weather stations whose radii are proportional to the deviation from the overall mean annual maxima computed over all weather stations. A red (resp. blue) circle indicates that the deviation of the corresponding weather station is positive (resp. negative).

for the generalized extreme value parameters μ, σ and ξ in (2). For our application, the right panel of Figure 5 suggests the use of

$$\mu(x) = \beta_{\mu,0} + \beta_{\mu,1}\text{lon}(x) + \beta_{\mu,2}\text{lat}(x), \quad (23)$$

with similar expressions for $\sigma(x)$ and $\xi(x)$ and where $\mu(x), \sigma(x), \xi(x), \text{lon}(x)$ and $\text{lat}(x)$ are respectively the location, scale and shape parameters of the generalized extreme value distribution and the longitude and latitude for location $x \in \mathcal{X}$. With the `SpatialExtremes` package, this is done by invoking in R

```
loc.form <- scale.form <- shape.form <- y ~ lon + lat
```

With such trend surfaces, the pairwise log-likelihood (13) has to be modified to take into account for the (pointwise) switch of generalized extreme value margins to unit Fréchet ones.

The trend surface (23) was just a guess based on Figure 5 and we need to perform model selection to define the most accurate and parsimonious trend surfaces using the theory introduced in Section 6. At this stage and because we aim at defining sensible trend surfaces for the generalized extreme value parameters, it is more convenient to omit for a moment the spatial dependence, i.e., assuming that the weather stations are mutually independent. It amounts to use a special composite likelihood known as the independence likelihood Varin et al. (2011). This is done by invoking

```
M0 <- fitspatgev(data, coord, loc.form, scale.form, shape.form)
```

where `data` is a matrix containing the annual maxima at each location and `coord` is a matrix formed by some suitable covariates used in the trend surfaces. It is often a good idea to center the covariates to avoid numerical instabilities when maximizing the log-likelihood and to check that the optimum found by the numerical optimizer is suitable.

The trend surface for the shape parameter is likely to be too complex because this parameter has usually large uncertainties, and apart from specific situations, it is not too restrictive to assume a constant shape parameter, i.e.,


```
shape.form <- y ~ 1
M1 <- fitspatgev(data, coord, loc.form, scale.form, shape.form)
```

To decide whether one should prefer model M1 over M0 one could use the Takeuchi information criterion by invoking `TIC(M0,M1)` who suggests the use of M1. Because these two models are nested, one can use a composite likelihood ratio test to confirm this model selection. This is done by invoking `anova(M0,M1)` whose output is

```
Eigenvalue(s): 1.07 0.66
```

```
Analysis of Variance Table
```

```
      MDf Deviance Df  Chisq Pr(> sum lambda Chisq)
M1      7   8541.3
M0      9   8539.1  2 2.2225          0.2748
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The composite likelihood ratio test clearly states in favor of M1 with a p -value around 0.30. After comparing various trend surfaces, we found that the following trend surfaces was the most suitable

$$\mu(x) = \beta_{\mu,0} + \beta_{\mu,1}\text{lon}(x) + \beta_{\mu,2}\text{lat}(x), \quad \sigma(x) = \beta_{\sigma,0} + \beta_{\sigma,1}\text{lon}(x), \quad \xi(x) = \beta_{\xi,0}, \quad x \in \mathcal{X}.$$

Although at this stage it is possible to get quantile predictions at any new location $x_* \in \mathcal{X}$, the spatial dependence has not been taken into account yet and we need to choose the most appropriate max-stable model for our data. We will consider the ones introduced in Section 3, namely the Schlather, Brown–Resnick and extremal- t families. Maximizing the associated pairwise likelihoods is challenging and it is often a good idea to help the numerical optimizer in doing so. For example using the built in `nlm` optimizer of R, one can invoke

```
start.trend <- as.list(M2$fitted)
start <- c(list(range = 150, smooth = 0.2), start.trend)
schlat <- fitmaxstab(data, coord, "powexp", loc.form, scale.form,
                    shape.form, nugget = 0, method = "nlm",
                    start = start, typsize = unlist(start))

start <- c(list(range = 13, smooth = 0.24), start.trend)
brown <- fitmaxstab(data, coord, "brown", loc.form, scale.form,
                   shape.form, start = start, method = "nlm",
                   typsize = unlist(start))

start <- c(list(range = 150, smooth = 0.2, DoF = 1), start.trend)
extt <- fitmaxstab(data, coord, "tpowexp", loc.form, scale.form,
                  shape.form, nugget = 0, method = "nlm",
                  start = start, typsize = unlist(start))
```

where M2 is the best fitted model obtained using the `fitspatgev` function.

Table 1 summarizes the fitted max-stable models. Since the models under consideration are not nested, we have resort to the Takeuchi information criterion to select the best model, i.e., by

TABLE 1. Fits of several max-stable models to Dutch wind speed data. The Schlather and extremal- t models used the powered exponential correlation function family $\rho(h) = \exp\{-(h/\lambda)^\kappa\}$ and the Brown-Resnick model has semi variogram $\gamma(h) = (h/\lambda)^\kappa$. ($*$) denotes that the parameter is held fixed. h_+ is the estimated distance at which $\theta(h) = 1.7$. NoP is the number of parameters.

	ν	λ	κ	h_+ (km)	NoP	$\ell_p(\hat{\psi}_p)$	TIC
Schlather	—	51 (16)	0.58 (0.15)	531	8	-114340	229376
Brown-Resnick	—	13 (9)	0.24 (0.02)	318	8	-114453	229638
Extremal- t	2.56 (0.49)	531 (556)	0.39 (0.06)	381	9	-114206	229150
Extremal- t	2.53 (0.24)	500 ($*$)	0.40 (0.07)	372	8	-114206	229100

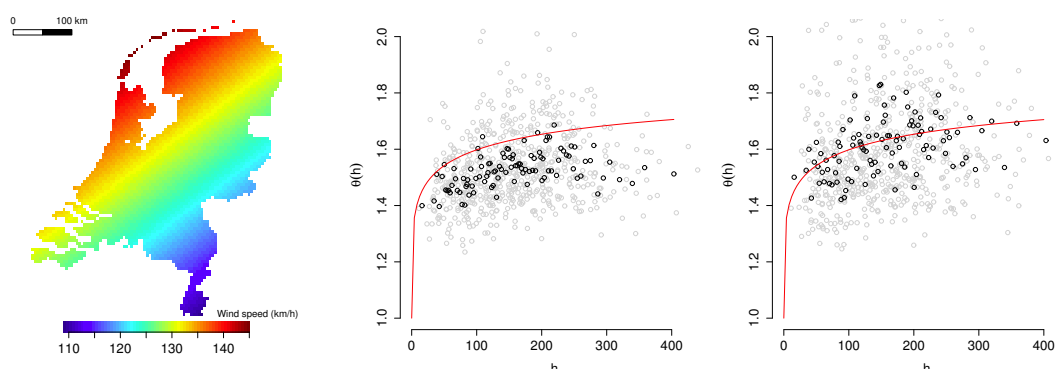


FIGURE 6. Inference for the wind gust data set from the fitted extremal- t model. The left panel plots the predicted pointwise 25 years return levels (km/h) obtained from the fitted extremal- t model. The two rightmost panels compare the empirical F -madogram cloud to the fitted extremal coefficient function. Grey points are pairwise estimates; black ones are binned estimates (with 100 bins). The middle panel uses the empirical distribution function. The right panel uses the generalized extreme value distributions defined from the fitted trend surfaces.

invoking `TIC(schlat, brown, exttt)`. According to this information criterion, the extremal- t model appears to be the most competitive model. The estimation of the range parameter λ is associated to large uncertainties as it appear to be difficult to jointly estimate ν, λ and κ . Similarly to the case study given in Davison et al. (2012), it might indicate difficulties in estimating the upper bound of the extremal coefficient function. To bypass this hurdle, one possibility consists in fixing one of this three parameters, preferably λ as in Table 1 since it does not put any restriction on the upper bound of $\theta(h)$ nor on the sample path regularity of the max-stable process.

The fitted trend surfaces of the extremal- t model are

$$\begin{aligned}\hat{\mu}(x) &= 95.7_{(0.9)} - 0.050_{(0.003)} \log(x) + 0.070_{(0.004)} \text{lat}(x), \\ \hat{\sigma}(x) &= 11.7_{(0.4)} - 0.010_{(0.001)} \log(x), \\ \hat{\xi}(x) &= -0.06_{(0.02)},\end{aligned}$$

where the subscripts give the associated standard errors. In accordance with the North-West to South-East gradient detected in Figure 5, the trend surface parameter estimates $\hat{\beta}_{\mu,1}$ and $\hat{\beta}_{\mu,2}$ are negative and positive respectively.

The left panel of Figure 6 plots the pointwise 25 year return level based on the fitted extremal- t model. As expected we can see that the largest wind speeds occur along the coastline. The two

TABLE 2. Frequencies (in %) of the partition size for the wind gust data.

Partition size	1	2	3	4	5	6	7–35
Frequencies (%)	0	43.5	43.2	12.0	1.1	0.2	0

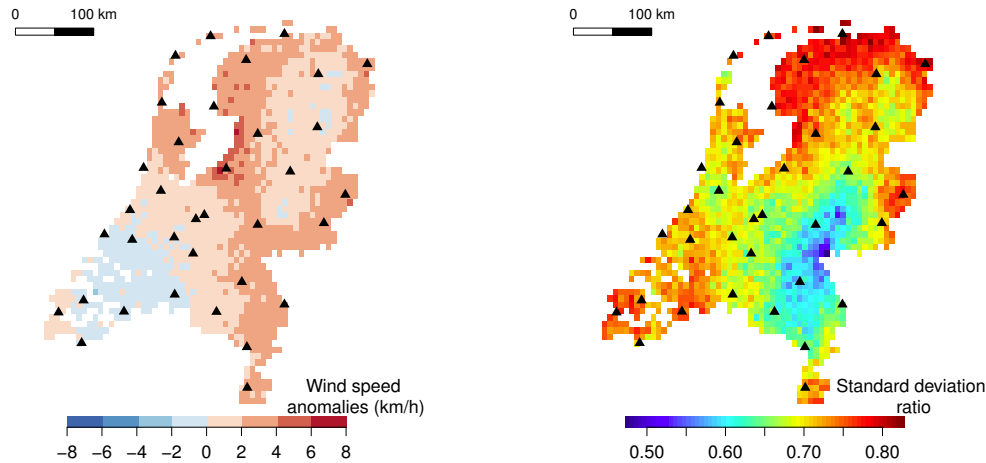


FIGURE 7. Summary plots obtained from 1000 conditional simulations for the wind gust data based on the observed annual maxima for year 2002. The left panel shows the 25 year wind speed pointwise anomalies, i.e., the difference between the pointwise conditional 25 year return levels and the unconditional ones. The right panel plots the ratio of the conditional pointwise standard deviations and the unconditional ones.

rightmost panels of Figure 6 compare the fitted extremal coefficient function to the empirical F -madogram cloud. The difference between these two plots is that the former makes the comparison using the raw annual maxima while the latter uses transformed data using the trend surfaces. We can see that the use of trend surfaces induces a bias in the estimation of the dependence structure probably owing to the use of too simple trend surfaces. However apart from this slight underestimation, overall the fitted spatial dependence structure seems reasonable. This fact is well known in the copula framework where most researchers advocate the use of a two-step fitting procedure: first fit the margins, then the dependence parameters. Although the bias occurring in the dependence structure will be removed, the price to pay with this two-step procedure is that it is likely that standard errors for parameter estimates are likely to be underestimated and hence could potentially impact model selection.

To push the analysis forward we perform 1000 conditional simulations based on the fitted extremal- t model. We set as conditioning values the annual maxima observed in 2002 since this year was found to be especially severe.

Table 2 shows the distribution of the size of the random partition ϑ in (21) conditionally on the annual maxima observed in 2002. We can see that the size of ϑ , i.e., the number of extremal functions, is between 2 and 3 with a probability around 0.85. By having a closer look at the daily data, we can see that the annual maxima of 2002 occurred either on February, 26th or October, 27th.

Figure 7 illustrates the differences between the annual maxima for year 2002 and a common

year based on 1000 conditional simulations from the fitted extremal- t model. The left panel plots the difference between the conditional pointwise 25 year return levels and the unconditional ones. Apart from the South-West part of Netherlands, we can see that the year 2002 seems to be most severe than a typical year. The largest differences are located around Lelystad—see Figure 5 for a precise location.

The right panel of Figure 7 plots the ratio of the pointwise standard deviation calculated from the 1000 conditional simulations and the standard deviation derived from the fitted extremal- t model. As expected the conditional standard deviation is smaller than the unconditional ones. The largest differences are located around Deelen—see Figure 5 for a precise location.

10. Discussion

Many progresses have been made since the derivation by Laurens de Haan of the spectral representation of max-stable processes de Haan (1984). Although max-stable processes are increasingly more considered in the statistical modelling of spatial extremes, there are still some open questions and difficulties.

One difficulty is related to the fitting of max-stable processes. The maximum composite likelihood estimator was introduced by Padoan et al. (2010) but implies a loss in efficiency and shows typically numerical instabilities. Some efforts should be made to propose better inferential procedures.

Simulation of max-stable process is feasible but some specific processes are more difficult to simulate from, e.g., Brown-Resnick processes. Although a theoretical framework exists for conditional simulation of max-stable processes, it is usually very CPU demanding and some efforts should be made to develop more efficient algorithm.

Paralleling the generalized Pareto distribution in the univariate case, little is known about generalized Pareto processes. Although some simple representations have been found Buishand et al. (2008); Aulbach et al. (2012); de Haan and Ferreira (2012); Dombry and Ribatet (2013), we are far from being able to use generalized Pareto processes for concrete applications.

Acknowledgment

M. Ribatet was partly funded by the MIRACCLE-GICC and McSim ANR projects.

Appendix A: Conditional distribution of the extremal- t process

For our purposes it is more convenient to use the following equivalent representation of the extremal- t process

$$Z(x) = c_v \max_{i \geq 1} \zeta_i \varepsilon_i(x)^v, \quad x \in \mathcal{X},$$

where $c_v = \sqrt{\pi} 2^{-(v-2)/2} \Gamma\{(v+1)/2\}^{-1}$, ε is a standard Gaussian process with correlation function ρ and with the convention that $y^v = -\infty$ when $y < 0$.

For all $\mathbf{x} \in \mathcal{X}^k$ and Borel set $A \subset \mathbb{R}^k$,

$$\Lambda_{\mathbf{x}}(A) = \int_0^\infty \Pr\{c_v \zeta \varepsilon(\mathbf{x})^v \in A\} \zeta^{-2} d\zeta = \int_0^\infty \int_{\mathbb{R}^k} 1_{\{c_v \zeta \mathbf{y}^v \in A\}} f_{\mathbf{x}}(\mathbf{y}) d\mathbf{y} \zeta^{-2} d\zeta,$$

where $f_{\mathbf{x}}$ is the density of the random vector $\varepsilon(\mathbf{x})$, i.e., a centered Gaussian random vector with covariance matrix $\Sigma_{\mathbf{x}}$. The change of variable $\mathbf{z} = c_v \zeta \mathbf{y}^v$ gives

$$\begin{aligned}\Lambda_{\mathbf{x}}(A) &= \int_0^\infty \int_A f_{\mathbf{x}} \left\{ \left(\frac{\mathbf{z}}{c_v \zeta} \right)^{1/v} \right\} v^{-k} c_v^{-k/v} \prod_{j=1}^k z_j^{1/v-1} d\mathbf{z} \zeta^{-k/v-2} d\zeta \\ &= (2\pi)^{-k/2} v^{-k} |\Sigma_{\mathbf{x}}|^{-1/2} c_v^{-k/v} \int_A \int_0^\infty \exp \left\{ -\frac{(\mathbf{z}^{1/v})^T \Sigma_{\mathbf{x}}^{-1} \mathbf{z}^{1/v}}{2c_v^{2/v}} \zeta^{2/v} \right\} \zeta^{k/v} d\zeta \prod_{j=1}^k z_j^{1/v-1} d\mathbf{z} \\ &= 2^{-k/2} \pi^{-k/2} v^{-k+1} |\Sigma_{\mathbf{x}}|^{-1/2} c_v^{-(k-2)/v} \int_A a_{\mathbf{x}}(\mathbf{z}, v)^{-1} \mathbb{E} \left\{ X^{(k+v-2)/v} \right\} \prod_{j=1}^k z_j^{1/v-1} d\mathbf{z} \\ &= \int_A \lambda_{\mathbf{x}}(\mathbf{z}) d\mathbf{z},\end{aligned}$$

where

$$\lambda_{\mathbf{x}}(\mathbf{z}) = c_v v^{-k+1} 2^{(v-2)/2} \pi^{-k/2} |\Sigma_{\mathbf{x}}|^{-1/2} a_{\mathbf{x}}(\mathbf{z}, v)^{-(k+v)/2} \Gamma \left(\frac{k+v}{2} \right) \prod_{j=1}^k z_j^{(1-v)/v},$$

with $a_{\mathbf{x}}(\mathbf{z}, v) = (\mathbf{z}^{1/v})^T \Sigma_{\mathbf{x}}^{-1} \mathbf{z}^{1/v}$ and where the third equality is obtained from the $(k+v-2)/v$ raw moment of a Weibull random variable with shape $2/v$ and scale $2^{v/2} c_v a_{\mathbf{x}}(\mathbf{z}, v)^{-v/2}$.

For all $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{s} \in \mathcal{X}^m$, the conditional intensity function is

$$\lambda_{\mathbf{s}|\mathbf{x}, \mathbf{z}}(\mathbf{u}) = v^{-m} \pi^{-m/2} \frac{|\Sigma_{(\mathbf{s}, \mathbf{x})}|^{-1/2}}{|\Sigma_{\mathbf{x}}|^{-1/2}} \frac{a_{(\mathbf{s}, \mathbf{x})} \{(\mathbf{u}, \mathbf{z}), v\}^{-(m+k+v)/2}}{a_{\mathbf{x}}(\mathbf{z}, v)^{-(m+k+v)/2}} a_{\mathbf{x}}(\mathbf{z}, v)^{-m/2} \frac{\Gamma \left(\frac{m+k+v}{2} \right)}{\Gamma \left(\frac{k+v}{2} \right)} \prod_{j=1}^m u_j^{(1-v)/v},$$

and, since following the lines of [Dombry et al. \(2013\)](#), we can show that

$$\frac{|\Sigma_{(\mathbf{s}, \mathbf{x})}|}{|\Sigma_{\mathbf{x}}|} = |\Sigma_{\mathbf{s}} - \Sigma_{\mathbf{s}:\mathbf{x}} \Sigma_{\mathbf{x}}^{-1} \Sigma_{\mathbf{x}:\mathbf{s}}| = \left\{ \frac{k+v}{a_{\mathbf{x}}(\mathbf{z}, v)} \right\}^m |\tilde{\Sigma}|, \quad \frac{a_{(\mathbf{s}, \mathbf{x})} \{(\mathbf{u}, \mathbf{z}), v\}}{a_{\mathbf{x}}(\mathbf{z}, v)} = 1 + \frac{(\mathbf{u}^{1/v} - \mu)^T \tilde{\Sigma}^{-1} (\mathbf{u}^{1/v} - \mu)}{k+v},$$

where $\mu = \Sigma_{\mathbf{s}:\mathbf{x}} \Sigma_{\mathbf{x}}^{-1} \mathbf{z}^{1/v}$ and $\tilde{\Sigma} = (k+v)^{-1} a_{\mathbf{x}}(\mathbf{z}, v) (\Sigma_{\mathbf{s}} - \Sigma_{\mathbf{s}:\mathbf{x}} \Sigma_{\mathbf{x}}^{-1} \Sigma_{\mathbf{x}:\mathbf{s}})$, we finally get

$$\begin{aligned}\lambda_{\mathbf{s}|\mathbf{x}, \mathbf{z}}(\mathbf{u}) &= \pi^{-m/2} (k+v)^{-m/2} |\tilde{\Sigma}|^{-1/2} \left\{ 1 + \frac{(\mathbf{u}^{1/v} - \mu)^T \tilde{\Sigma}^{-1} (\mathbf{u}^{1/v} - \mu)}{k+v} \right\}^{-(m+k+v)/2} \frac{\Gamma \left(\frac{m+k+v}{2} \right)}{\Gamma \left(\frac{k+v}{2} \right)} \times \\ &\quad \left\{ v^{-m} \prod_{j=1}^m u_j^{-(v-1)/v} \right\}.\end{aligned}$$

The last term in bracket in the previous equation corresponds to the Jacobian of the mapping $\mathbf{u} \mapsto \mathbf{u}^{1/v}$. Hence we recognize that the conditional intensity function is the density of the random vector T^v where T is a Student random vector with mean μ and dispersion matrix $\tilde{\Sigma}$.

References

- Aulbach, S., Falk, M., and Hofmann, M. (2012). On max-stable processes and the functional D-norm. *Extremes*, pages 1–29.
- Brown, B. M. and Resnick, S. I. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14:732–739.
- Buishand, T. A., de Haan, L., and Zhou, C. (2008). On spatial extremes: With application to a rainfall problem. *Annals Of Applied Statistics*, 2(2):624–642.
- Chandler, R. E. and Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, 94(1):167–183.
- 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., 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.
- de Haan, L. and Ferreira, A. (2006). *Extreme value theory: An introduction*. Springer Series in Operations Research and Financial Engineering.
- de Haan, L. and Ferreira, A. (2012). The Generalized Pareto process; with application. *Submitted*. Preprint <http://arxiv.org/abs/1203.2551>, 2012.
- 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.
- Dombry, C. and Ribatet, M. (2013). Functional regular variations, Pareto processes and peaks over threshold. *Submitted*. Preprint www.math.univ-montp2.fr/~ribatet/docs/Dombry2012.pdf.
- Erhardt, R. J. and Smith, R. L. (2012). Approximate Bayesian computing for spatial extremes. *Computational Statistics and Data Analysis*, 56(6):1468–1481.
- 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.
- 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.
- Kent, J. T. (1982). Robust properties of likelihood ratio tests. *Biometrika*, 69:19–27.
- Nikoloulopoulos, A., Joe, H., and Li, H. (2009). Extreme values properties of multivariate t copulas. *Extremes*, 12:129–148.
- Oesting, M., Kabluchko, Z., and Schlather, M. (2012). Simulation of Brown–Resnick processes. *Extremes*, 15:89–107. 10.1007/s10687-011-0128-8.
- Opitz, T. (2012). Extremal- t process: Elliptical domain of attraction and a spectral representation. *Submitted*. Preprint <http://arxiv.org/abs/1207.2296>.
- 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., Cooley, D., and Davison, A. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 22:813–845.
- Ribatet, M. and Sedki, M. (2013). Extreme value copulas and max-stable processes. *Journal de la Société Française de Statistique*, 154(138–150).
- Ribatet, M., Singleton, R., and Team, R. C. (2013). *SpatialExtremes: Modelling Spatial Extremes*. R package version 2.0-1.
- Rotnitzky, A. and Jewell, N. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika*, 77:495–497.
- Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44.
- Schlather, M., Menck, P., Singleton, R., and Team, R. C. (2013). *RandomFields: Simulation and Analysis of Random*

- Fields*. R package version 2.0.66.
- Schlather, M. and Tawn, J. (2003). A dependence measure for multivariate and spatial extremes: Properties and inference. *Biometrika*, 90(1):139–156.
- Smith, R. L. (1990). Max-stable processes and spatial extreme. *Unpublished manuscript*.
- Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(5–42).
- Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528.
- Wang, Y. and Stoev, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Advances in Applied Probability*, 443:461–483.