Likelihood based inference for high-dimensional extreme value distributions

Alexis BIENVENÜE and Christian Y. ROBERT
Université de Lyon, Université Lyon 1, Institut de Science Financière et d’Assurances,
50 Avenue Tony Garnier, F-69007 Lyon, France

March 1, 2014

Abstract

Multivariate extreme value statistical analysis is concerned with observations on several variables which are thought to possess some degree of tail-dependence. In areas such as the modeling of financial and insurance risks, or as the modeling of spatial variables, extreme value models in high dimensions (up to fifty or more) with their statistical inference procedures are needed. In this paper, we consider max-stable distributions for which the tail-dependence function has a regular Schlather’s representation. We provide quasi-explicit analytical expressions of the full likelihoods for random samples with max-stable distributions and for random samples in the max-domain of attraction of a max-stable distribution. When the full likelihood becomes numerically intractable, it is however necessary to split the variables into subgroups and to consider a composite likelihood approach. The asymptotic properties of the estimators are given and the utility of the methods is examined via simulation. The estimators are also compared with those derived from the pairwise composite likelihood method which has been previously proposed in the spatial extreme value literature. Finally, a real data application on financial extreme events is presented.

Keywords: Composite likelihood methods; High-dimensional extreme value distributions; High thresholding; Likelihood and simulation-based likelihood inference; Spatial max-stable processes

1 Introduction

Technological innovations have had deep impact on scientific research by allowing to collect massive amount of data with relatively low cost. These new data create opportunities for the development of data analysis. In particular the availability of high-dimensional data has significantly challenged the traditional multivariate statistical theory.

Problems concerning financial risks are typically of high-dimensions in character. Financial portfolios often consist of a large number of assets and complex instruments (more than one hundred) and the dependence between the underlying components of the portfolios plays a crucial role in their pricing (see e.g. [38]). Understanding the links between extreme events is also of crucial importance for an efficient quantitative risk management.

Problems involving spatial dependence are also of high-dimensions in character. Extreme environmental phenomena such as major precipitation events or extreme wind speeds manifestly exhibit spatial dependence and are of much interest for engineers because knowledge of the spatial dependence is essential for regional risk assessment.
We focus in this paper on multivariate extreme value distributions in high dimensions. Models and inference methods for these distributions is an exciting research field that is in its beginning of development. One way to characterize the dependence among multivariate extremes is to con-
sider the tail-dependence function or the tail-dependence density derived from known copulas of
multivariate distributions ([34]). Some models based on Archimedean copulas or elliptical copulas
have been proposed (see e.g. [30], [32]) but no general inference method have been associated to
these models. The literature on modeling high-dimensional extreme events mainly concerns spatial
max-stable processes. These processes arise as limits in distribution of component–wise maxima of
independent spatial processes, under suitable centering and normalization. Max-stable processes
are very useful to study extreme value phenomena for which spatial patterns can be discerned.

Likelihood inference for Archimedean copulas in high dimensions has been recently investigated
in [26] and gives an opportunity for the only archimedean extreme value copula, the Logistic (also
known as Gumbel–Hougaard) copula. But this approach is clearly insufficient because it is limited
to exchangeable random variables whose dependence is characterised by this Archimedean copula.

The composite likelihood methods that are based on combinations of valid likelihood objects
related to small subsets of data are rather naturally considered as an appealing way to deal with
inference in high dimensions. The merit of composite likelihood is to reduce the computational
complexity. These methods go back to [3] and have been extensively studied in particular for the
case where the full likelihood is unavailable and is replaced by quantities that combine bivariate
or trivariate joint likelihoods, see e.g. [35], [8], [51]. Recent works concerning the application for
(i) random samples from max-stable processes and for (ii) random samples in the max-domain of
attraction of a max-stable distribution can be found in [39], [19], [27] and [44] for (i), and in [1],
[29] and [33] for (ii).

For a class of processes, commonly known as Brown–Resnick processes (see [4], [31]), [16]
and [54] offer different perspectives for random samples in their max-domains of attraction. [16]
establishes that the multivariate conditional distribution of extremal increments with respect to
a single extreme component is multivariate Gaussian, therefore offering the possibility to perform
high-dimensional inference. [54] exploits the limiting multivariate Poisson process intensities for
inference on vectors exceeding a high marginal threshold in at least one component, employing a
censoring scheme to incorporate information below the marginal threshold. We will follow this last
direction but will consider the multivariate densities of the vectors of exceedances rather than the
intensities of the Poisson process.

In this paper, we consider multivariate max-stable distributions for which the tail-dependence
function may be written in the same integral form as those derived from Schlather’s spectral repre-
sentation of max-stable processes ([45]). With regards to practical application, these models offer
a compromise between flexibility and tractability. For these multivariate max-stable distributions,
we provide quasi-explicit analytical expressions of the full likelihoods for random samples with
max-stable distributions and for random samples in the max-domain of attraction of a max-stable
distribution. When the full likelihood becomes numerically intractable, we gather together compo-
nents of the vector within clusters with bounded sizes for which the likelihood is feasible with a
moderate computational cost and we consider the pseudolikelihood based on the likelihoods of the
clusters.

The paper is organized as follows. Section 2 presents Schlather’s representation for simple
multivariate extreme value distributions (with unit Fréchet margins) and considers several practical
examples. The analytical expressions of the full likelihoods for random samples with max-stable
distributions and the analytical expressions of the asymptotic full likelihoods for (censored) random
samples in the max-domain of attraction of a max-stable distribution are also given. In Sections 3
and 4 we consider respectively parametric max-stable models for random samples with max-stable distributions and for random samples in the max-domain of attraction of a max-stable distribution. We define maximum likelihood estimators and pseudo-maximum likelihood estimators built by clustering components of the vector if necessary. We prove consistency and asymptotic normality for these estimators when the dimension of the vector is fixed and as the number of observations tends to infinity. For some max-stable distributions, it may be necessary to compute parts of the likelihoods via simulation-based methods. In that case we also give the asymptotic distributions of the estimators with respect to the number of simulations. Section 4 illustrates the performance of the methods through a number of simulation studies and gives comparisons with the composite likelihood method. Finally, a real data application on financial extreme events is presented. All proofs have been deferred to the appendix.

2 Full likelihoods with Schlather’s representation for the tail-dependence function

2.1 Introduction

Let \( Y_i = (Y_{i1}, \ldots, Y_{im}) \), for \( i = 1, \ldots, n \), be independent and identically distributed random vectors in dimension \( m \). The analysis of multivariate extreme values has first been based on the weak convergence of the vector of componentwise maxima \( M_n = (M_{n1}, \ldots, M_{nm}) \) where \( M_{nj} = \max_{i=1,\ldots,n} Y_{ij} \). Under general conditions \( M_n \), suitably normalized, converges in distribution to a member of the multivariate extreme value distributions (see e.g. [45]). Suppose that the marginals components of the multivariate extreme value distribution \( G_* \) have a unit Fréchet distribution, then there exists a random vector \( W = (W_1, \ldots, W_m) \) defined on the \((m-1)\)-dimensional simplex \( \Delta_{m-1} = \{ w \in \mathbb{R}_+^m : \sum_{j=1}^m w_j = m \} \), satisfying \( \mathbb{E}[W_j] = 1 \), for \( j = 1, \ldots, m \), and such that, for \( \mathbf{z} = (z_1, \ldots, z_m) \in \mathbb{R}_+^m \),

\[
G_*(\mathbf{z}) = \exp \left( -\mathbb{E} \left[ \max \left( z_1^{-1} W_1, \ldots, z_m^{-1} W_m \right) \right] \right).
\]

This representation is known as Pickands’ representation (see e.g. Remark 6.1.16 in [13]). The function \( V_*(\mathbf{z}) = -\log (G_*(\mathbf{z})) \) is referred to as the tail-dependence function.

Since it may be difficult to build parametric models in high dimensions for \( W \), we rather consider the family of tail-dependence functions of the following form

\[
V_*(\mathbf{z}) = \mathbb{E} \left[ \max \left( z_1^{-1} U_1^+, \ldots, z_m^{-1} U_m^+ \right) \right] \tag{2.1}
\]

where \( \mathbf{U} = (U_1, \ldots, U_m) \) is a random vector on \( \mathbb{R}_+^m \) such that \( \mathbb{E}[U_j^+] = 1 \), for \( j = 1, \ldots, m \), with \( U_j^+ = \max (0, U_j) \). Put \( R = (U_1^+ + \ldots + U_m^+) \) and \( V_j = m U_j^+ / R \) for \( j = 1, \ldots, m \). We deduce that

\[
\mathbb{E} \left[ \max_{j=1,\ldots,m} \left( z_j^{-1} U_j^+ \right) \right] = \mathbb{E} \left[ \mathbb{E} \left[ m^{-1} R | V \right] \max_{j=1,\ldots,m} \left( z_j^{-1} V_j \right) \right] = \mathbb{E} \left[ \max_{j=1,\ldots,m} \left( z_j^{-1} W_j \right) \right]
\]

with \( \Pr(W \in dw) = \mathbb{E}[m^{-1} R | V = w] \Pr(V \in dw) \) (see also [46]).

Such tail-dependence functions appear for the finite dimensional distributions of simple max-stable processes characterized by Schlather’s spectral representation (see [13]): let \( \{ \zeta_j \}_{j \geq 1} \) be the points of a Poisson process on \( \mathbb{R}_+ \) with intensity \( ds/s^2 \), and let \( \{ U_j \}_{j \geq 1} \) be independent replicates of a spatial processes \( U \) on a compact set \( \mathcal{X} \) of \( \mathbb{R}_2 \), with continuous sample paths, independent of the Poisson process and satisfying \( \mathbb{E}[U^+(x)] = 1 \) for all \( x \in \mathcal{X} \). Then

\[
Z(x) = \max_{j \geq 1} \zeta_j U_j(x), \quad x \in \mathcal{X},
\]
is a simple max-stable process and its finite dimensional distributions satisfy
\[
\Pr \left( Z(x_1) \leq z_1, \ldots, Z(x_m) \leq z_m \right) = \exp \left( -\mathbb{E} \left[ \max \left( z_1^{-1}U^+(x_1), \ldots, z_m^{-1}U^+(x_m) \right) \right] \right),
\]
(2.2)
for \( x_1, \ldots, x_m \in \mathcal{X} \) and \( z_1, \ldots, z_m \in \mathbb{R}_+ \). There are very few spatial models for which \( V_s(z) \) is analytically known for dimension \( m \) larger than two. An exception is given by the spatial max-stable for which the process \( U \) is a log-Gaussian random field: [27] proves that \( V_s(z) \) may be written as a linear combination of \((m-1)\)-dimensional multivariate normal probabilities (see also [29]).

When the \( U_j \) are independent and identically distributed random variables, it is however possible to derive several analytical forms for the tail-dependence functions in (2.1). Even if they lead to distributions of exchangeable random variables, such tail dependence functions may offer interest for inference and practical applications. Let \( \eta_j \), for \( j = 1, \ldots, m \), be positive independent and identically distributed random variables with distribution function \( H \), and let \( \alpha > 0 \) be a constant such that \( \mathbb{E}(\eta_j^\alpha) < \infty \). We define \( U_j = \eta_j^\alpha / \mathbb{E}(\eta_j^\alpha) \). In that case, we have
\[
V_s(z) = \sum_{l=1}^m z_l^{-1} \mathbb{E} \left[ \eta^\alpha \left( \eta^{\eta_j^\alpha} \prod_{s \neq l} H \left( \left( \frac{z_s}{z_l} \right)^{1/\alpha} \eta \right) \right) \right]
\]
where \( \eta \) is a positive random variable with distribution \( H \). Let us give some examples by considering different distributions \( H \) and their associated tail-dependence functions.

- Assume that \( H \) is a Bernoulli distribution with parameter \( 0 < p < 1 \). The distribution of \( \eta^\alpha \) is also a Bernoulli distribution with the same parameter and
\[
V_s(z) = p^{m-1} \max (z_1^{-1}, \ldots, z_m^{-1}).
\]
For \( m = 2 \), \( G_s \) is known as the Marshall-Olkin distribution [37].

- Assume that \( H \) is a Uniform distribution on \([0, 1]\). The distribution of \( \eta^\alpha \) is the Beta distribution with parameters \( 1/\alpha \) and 1, and
\[
V_s(z) = \frac{1 + \alpha}{m + \alpha} \left( \prod_{j=1}^m z_j^{1/\alpha} \right) \sum_{l=1}^m z_l^{-1-m/\alpha}.
\]

- Assume that \( \eta = e^X \) where \( X \) has a Gaussian distribution \( \mathcal{N} (\mu, \sigma^2) \). Then \( \eta^\alpha \) has a Lognormal distribution with parameter \( \alpha \mu \) and \( \alpha^2 \sigma^2 \) and
\[
V_s(z) = \sum_{l=1}^m z_l^{-1} \Phi_{\Sigma(m-1)} \left( (\kappa^{-1} + (\kappa \log (z_l^{-1}))) / 2 ; j \neq l \right),
\]
where \( \kappa = \sqrt{2} / (\alpha \sigma) \) and \( \Phi_{\Sigma(m-1)} \) is the \((m-1)\)-variate Gaussian probability distribution function with mean vector equal to zero and correlation matrix \( \Sigma (m-1) = \left( \sigma_{l,l} (m-1) \right) \) given by \( \sigma_{l,l} (m-1) = 1 \) for \( 1 \leq l \leq m-1 \) and \( \sigma_{l,j} (m-1) = 1/2 \) for \( 1 \leq l < j \leq m-1 \). For \( m = 2 \), \( G_s \) is known as the Hüsler-Reiss distribution [28].

- Assume that \( H \) is a Weibull distribution \( \text{Weib}(c, \tau) \), \( c > 0 \) and \( \tau > 0 \), i.e. \( H(x) = \exp (-c x^\tau) \). Then \( \eta^\alpha \) has a Weibull distribution \( \text{Weib}(c, \tau/\alpha) \) and
\[
V_s(z) = \sum_{l=1}^m (-1)^{m-1} \sum_{1 \leq j_1 < \ldots < j_l \leq m} \left( \sum_{j=1}^l z_{j_l}^{-\alpha} \right)^{-1/\kappa},
\]
where \( \kappa = \tau / \alpha \). For \( m = 2 \), \( G_s \) is known as the Galambos distribution [18].
• Assume that $H$ is a Fréchet distribution $(Fré(c, \tau))$, $c > 0$ and $\tau > 0$, i.e. $H(x) = \exp(-cx^{-\tau})$. If $\tau/\alpha > 1$ then $\eta^\alpha$ has a Fréchet distribution $Fré(c, \tau/\alpha)$ with a finite mean and

$$V_*(z) = \left( \sum_{l=1}^{m} z_l^{-\kappa} \right)^{1/\kappa},$$

where $\kappa = \tau/\alpha$. For $m = 2$, $G_*$ is known as the Logistic or the Gumbel distribution [24].

Note that, if $V_*(z)$ is a tail-dependence function, then $V_*^{\beta}(z^{1/\beta})$, for $0 < \beta \leq 1$, is also a tail-dependence function (see [43]).

2.2 Density functions for random samples of high dimensional max-stable distributions

We assume that $Z = (Z_1, \ldots, Z_m)$ has a max-stable distribution whose tail-dependence function is of the form (2.1) and for which the probability density function of the vector $U$ is well-defined. We introduce some notation. Let $I = \{1, \ldots, m\}$. For a set $B \subset I$, we let $U_B = (U_j)_{j \in B}$, $z_B = (z_j)_{j \in B}$ and $\{U_B \leq z_B\} = \cap_{j \in B} \{U_j \leq z_j\}$. Then define

$$\mu(B; z) = \int_{0}^{\infty} \gamma^{|B|} \Pr(U_B^c \leq z_B^c | U_B = z_B^\gamma) f_{U_B}(z_B^\gamma) d\gamma$$

where $|B|$ denotes the cardinality of the set $B$, $B^c = I \setminus B$ and $f_{U_B}$ is the probability density function of $U_B$. Note that, if the conditional distributions of $U_B^c$ given $U_B$ and the density functions $f_{U_B}$ are analytically known, then $\mu(B; z)$ may be easily computed because it is a one-dimensional integral. Finally, we denote by $\Pi$ the set of all partitions of $I$. For a partition $\pi \in \Pi$, $B \in \pi$ means that $B$ is one of the blocks of the partition $\pi$.

**Proposition 1** For $z \in \mathbb{R}_+^m$, let $h(z)$ be the density function of $Z$, i.e. $h(z) = \partial^m \Pr(Z \leq z) / \partial z$. We have

$$\log h(z) = - \sum_{l=1}^{m} z_l \mu(\{l\}; z) + \log \left( \sum_{\pi \in \Pi} \prod_{B \in \pi} \mu(B; z) \right). \quad (2.4)$$

Note in particular that $V_*(z) = \sum_{l=1}^{m} z_l \mu(\{l\}; z)$. Similar expressions involving the partial derivatives of $V_*$ have been given in [49] and in [42]. Here the partial derivatives of $V_*$ are expressed as integrals of the partial derivatives of the probability distribution function of the vector $U$ and are in this way easy to compute.

The previous proposition gives us a quasi-explicit analytical expression of the density function in the sense that it only depends on quantities like $\mu(B; z)$. However high-dimensional densities present an explosion of terms. Indeed the number of partitions of $I$ is given by the $m$-th Bell number and increases dramatically with $m$. For example $m = 7$ (resp. $m = 10$) would require to sum over around 1000 (resp. 116 000) terms. In practise, $m = 7$ is today an upper bound. Let us give some examples where $\mu(B; z)$ may be efficiently computed.

• Assume that $U_j = \eta_j^\alpha / \mathbb{E}(\eta_j^\alpha)$ where $\eta_j$ are positive, independent and identically distributed random variables with distribution function $H$ and density function $h$, then

$$\mu(B; z) = \int_{0}^{\infty} \gamma^{|B|} \prod_{j \in B^c} H \left( (\gamma z_j \mathbb{E}(\eta_j^\alpha))^{1/\alpha} \right) \prod_{j \in B} \left[ \alpha^{-1} z_j^{\gamma z_j \mathbb{E}(\eta_j^\alpha)} \right] d\gamma.$$
Assume that \( U_j = \alpha_j^{-1} V_j \) where \( V_j \) are independent Gamma(\( \alpha_j, 1 \)) random variables with \( \alpha_j > 0 \). Then \( V_s(z) \) gives the Dirichlet model introduced in [5]. In that case

\[
\mu(B; z) = \int_0^\infty \gamma^{(B)} \prod_{j \in B^c} \frac{\Gamma(\alpha_j, \gamma \alpha_j)}{\Gamma(\alpha_j)} \prod_{j \in B} \frac{\alpha_j}{\Gamma(\alpha_j)} e^{-\gamma \alpha_j z_j} d\gamma
\]

where \( \Gamma(a, z) \) is the lower incomplete gamma function.

Assume that \( \Pr(U \leq z) = C(F_{U_1}(z_1), \ldots, F_{U_m}(z_m)) \) where \( F_{U_i} \) is the probability function of \( U_i \) such that \( \mathbb{E}[U_j^+] = 1 \), for \( j = 1, \ldots, m \), and

\[
C(u) = \psi(\psi^{-1}(u_1) + \ldots + \psi^{-1}(u_m))
\]
is an Archimedean copula with a completely monotone generator \( \psi \). Then

\[
\mu(B; z) = \int_0^\infty \gamma^{[B, \psi([B])] (\sum_{i=1}^m \psi^{-1}(F_{U_i}(\gamma z_i))) \prod_{j \in B} \frac{1}{\psi^{-1}(F_{U_i}(\gamma z_i))} f_{U_i}(\gamma z_i) d\gamma
\]

where \( f_{U_i} \) is the probability density function of \( U_i \).

Assume that \( U \) has a multivariate Gaussian distribution with covariance matrix \( \Sigma \) such that \( \mathbb{E}[U_j^+] = 1 \), for \( j = 1, \ldots, m \). Let us denote by \( \Sigma_B \) the covariance matrix of \( U_B \) and \( \Sigma_{B \cap B'} \) the covariance matrix between \( U_{B^c} \) and \( U_B \). Let \( \|z_B\|_{\Sigma_B^{-1}}^2 = \|z_B\|_{\Sigma_B^{-1}} \Sigma_B^{-1} z_B, \tilde{z}_{B^c} = (z_{B^c} - \Sigma_{B^c} B \Sigma_B^{-1} z_B) / \|z_B\|_{\Sigma_B^{-1}} \). Assume that \( V_{B^c|B} \sim \mathcal{N}(0, \Sigma_{B^c|B}) \) with \( \Sigma_{B^c|B} = \Sigma_B - \Sigma_{B^c} B \Sigma_B^{-1} \Sigma_{B^c B} \) and \( \Lambda \sim \mathcal{N}(0, 1) \) independent of \( V_{B^c|B} \). Then define \( Y \) as a random vector of dimension \( |B^c| + 1 \) in the following way: \( Y_{B^c} = V_{B^c|B} + \tilde{z}_{B^c} \Lambda \) and \( Y_{(|B^c|+1)} = \lambda \). It may be shown that

\[
\mu(B; z) = \frac{1}{(2\pi)^{m/2} (\det \Sigma_B)^{1/2} (\det \Sigma_Y)^{1/2}} \int_{-\infty}^0 \ldots \int_{-\infty}^0 |y_{B^c}|^{[B]} \exp \left( -\frac{1}{2} y' \Sigma_Y^{-1} y \right) dy.
\]

Such integrals are easily computed by using the approach developed by Genz (see [20], [21] and [23]).

Consider the spatial Brown-Resnick max-stable process. The process \( U \) in Schlather’s spectral representation is equal to \( \exp\{\sigma \varepsilon - \sigma^2/2 \} \) where \( \sigma > 0, \varepsilon \) is an intrinsically stationary Gaussian process with semivariogram function \( \nu \) and satisfying \( \varepsilon(0) = 0 \) almost surely. For \( x \in X_m \), let \( \varepsilon = (\varepsilon(x_i))_{i=1,\ldots,m} \) and \( \nu = (\nu(x_i))_{i=1,\ldots,m} \). Let us denote by \( \Sigma_B \) the covariance matrix of \( \varepsilon_B \) and \( \Sigma_{B^c B} \) the covariance matrix between \( \varepsilon_{B^c} \) and \( \varepsilon_B \). Let \( \|z_B\|_{\Sigma_B^{-1}} = \|z_B\|_{\Sigma_B^{-1}} \). We define

\[
\log \tilde{z}_B = \log z_B + \nu_{B^c} - \Sigma_{B^c B} \Sigma_B^{-1} (\log z_B + \nu_B) - (e_B - \Sigma_{B^c B} \Sigma_B^{-1} e_B) \|e_B\|_{\Sigma_B^{-1}}^{-2} \langle e_B \| \log z_B \rangle_{\Sigma_B^{-1}}.
\]

Assume that \( V_{B^c|B} \sim \mathcal{N}(0, \Sigma_{B^c|B}) \) with \( \Sigma_{B^c|B} = \Sigma_B - \Sigma_{B^c} B \Sigma_B^{-1} \Sigma_{B^c B} \) and \( \Lambda \sim \mathcal{N}(0, \|e_B\|_{\Sigma_B^{-1}}^{-2}) \) independent of \( V_{B^c|B} \). Then define \( Y \) as a random vector of dimension \( |B'| \) in the following way: \( Y_{B^c} = V_{B^c|B} - (e_{B^c} - \Sigma_{B^c B} \Sigma_B^{-1} e_B) \Lambda \). It may be shown that

\[
\mu(B; z) = \frac{1}{(2\pi)^{|B|/2} (\det \Sigma_B)^{|B|/2}} \exp \left( \frac{1}{2} \|e_B\|_{\Sigma_B^{-1}}^{-2} - \langle e_B \| \log z_B - \nu_B \rangle_{\Sigma_B^{-1}} \right) \times \exp \left( \frac{1}{2} \|e_B\|_{\Sigma_B^{-1}}^{-2} \left( \langle e_B \| \log z_B - \nu_B \rangle_{\Sigma_B^{-1}} \right)^2 - \frac{1}{2} \|\log z_B - \nu_B\|_{\Sigma_B^{-1}}^2 \right) \Phi_{\Sigma_B^{-1}} (\log \tilde{z}_B)
\]

6
where $\Phi_{\Sigma_Y} (\log \tilde{z}_B)$ may be efficiently computed by using the approach developed by Genz (see [20], [21] and [23]). Therefore $\mu (B; z)$ may be written as a quantity proportional to a $|B^c|$-dimensional multivariate normal probability (see also [16] for an equivalent approach).

Figure 1: The top row shows the conditioning locations (empty circles) and the locations where the conditional densities are computed (full circles with numbers) for two simulated paths of the Schlather process used to get the conditioning events (left and right). The bottom row shows the conditional density functions for the numbered locations. The correlation function is given by the Whittle Mattern function with parameter $c = 0.5$ and $\nu = 1.5$.

In [11], the authors introduced a method for conditional simulations of the Brown–Resnick and Schlather processes. For the latter, the spatial process $U$ is a stationary Gaussian process (see [45]). With our density formula (2.4), it is possible to derive the conditional densities of the values of the max-stable process for any location in the space given the values at other locations. Figure 1 shows several conditional density functions for a Schlather max-stable process with a Whittle
Mattern correlation function given by

\[ \rho(h) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \frac{h}{c} \right)^\nu K_\nu \left( \frac{h}{c} \right) \]

(2.5)

where \( c \) and \( \nu \) are the range and the smooth parameters of the correlation function, \( \Gamma \) is the gamma function, \( K_\nu \) is the modified Bessel function of the third kind with order \( \nu \) and \( h \) is the Euclidean distance between two locations.

### 2.3 Asymptotic density functions for (censored) random samples in the max-domain of attraction of a high dimensional max-stable distribution

We assume that \( Y = (Y_1, \ldots, Y_m) \) has marginals with unit Pareto distribution and is in the max-domain of attraction of \( G_* \). Hence we have

\[ \lim_{t \to \infty} t \Pr \left( \bigcup_{j=1}^m \{ Y_j > tz_j \} \right) = -\log G_* (z) = \mathbb{E} \left[ \max_{j=1,\ldots,m} z_j^{-1} U_j^+ \right]. \]

For a given threshold \( t > 0 \), we censor components that do not exceed the threshold and consider the vector \( X_t = t^{-1} Y \vee e \) with \( e = (1, \ldots, 1)' \in \mathbb{R}^m \). We are interested in the asymptotic conditional distribution of \( X_t^* \) with \( \| X_t^* \|_\infty > 1 \) as \( t \) tends to infinity, where \( \| X_t \|_\infty \) is the maximum norm of \( X_t \).

Let \( \mathcal{P} = \mathcal{P}(\mathcal{I}) \) be the power set of \( \mathcal{I} = \{1, \ldots, m\} \) without \( \emptyset \), and, for \( B \in \mathcal{P} \), let \( \mathcal{A}_B = \{ z \in [1, \infty)^m : z_i > 1, i \in B, z_i = 1, i \in B^c \} \). Note that \( X_t^* \) takes its values in \( \mathcal{A} = \cup_{B \in \mathcal{P}} \mathcal{A}_B \). Define

\[ V_B^*(z) = \int_0^\infty P \left( U_B > \gamma z_B, U_{B^c} \leq \gamma z_{B^c} \right) d\gamma. \]

Since \( \sum_{B \in \mathcal{P}} V_B^*(z) = V^*(z) \), we deduce that \( (p_B(z))_{B \in \mathcal{P}} \), where \( p_B(z) = V_B^*(z)/V^*(z) \), defines a discrete probability distribution.

**Proposition 2** As \( t \to \infty \), \( X_t^* \) converges in distribution to \( X^* \) whose density function \( f_{X^*} \) is characterized in the following way: for \( B \in \mathcal{P} \) and \( x = (x_B, e_{B^c}) \in \mathcal{A}_B \)

\[ f_{X^*|A_B}(x) = \frac{\mu(B; x)}{V_B^*(e)} \]

where \( f_{X^*|A_B} \) is the restriction of the density function of \( X^* \) to \( \mathcal{A}_B \). It follows that, for \( x \in \mathcal{A} \),

\[ f_{X^*}(x) = \sum_{B \in \mathcal{P}} p_B(e) f_{X^*|A_B}(x) 1_{\{x \in \mathcal{A}_B\}} = \frac{1}{\sum_{i=1}^m \mu(\{i\}; e)} \sum_{B \in \mathcal{P}} \mu(B; x) 1_{\{x \in \mathcal{A}_B\}}. \]

When considering random samples of max-stable vectors, the density functions may be rather complex. It is indeed necessary to consider all possible partitions of \( \mathcal{I} \) and to take into account the events for which elements of the component wise maxima occur simultaneously in the blocks of a partition. This structure may be substantially simplified by considering the individual extreme events for which at least one component of the vector has exceeded a high threshold. If all data may be modelled, it is preferable to use them and to censored observations which are less than the high threshold, because this simplifies a lot the likelihood. Such analyses have already been done in [49] and [54] for the likelihood of the censored Poisson process associated with exceedances.
3 Likelihood and simulation-based likelihood inference

3.1 Assumptions

We impose that the probability density function of $U$ belongs to some parametric family $\{f_U(\cdot;\theta), \theta \in \Theta\}$ where $\Theta$ is a compact set in $\mathbb{R}^p$, for $p \geq 1$, and let

$$
\mu(\theta; B, z) = \int_0^\infty \int_{-\infty}^{z_B} \gamma^m f_{U^{(B)}, U_B} (u_B^{(\gamma)}, z_B^{(\gamma)}, \theta) \, du_B \, d\gamma.
$$

For the existence of $\nabla_0 \mu(\theta; B, z)$ and $\nabla^2_0 \mu(\theta; B, z)$, we will assume the following conditions:

- **C1:** There exist gradients $\nabla_0 \log f_U (z; \theta)$ and Hessian $\nabla^2_0 \log f_U (z; \theta)$ for any $z \in \mathbb{R}^m$.
- **C2:** It is possible to interchange differentiation (with respect to $\gamma$ and $z_{B^c}$) for

$$
\int_0^\infty \int_{-\infty}^{z_B} \gamma^m f_{U^{(B)}, U_B} (u_B^{(\gamma)}, z_B^{(\gamma)}, \theta) \, du_B \, d\gamma \quad \text{and} \quad \int_0^\infty \int_{-\infty}^{z_B} \gamma^m \nabla_0 f_{U^{(B)}, U_B} (u_B^{(\gamma)}, z_B^{(\gamma)}, \theta) \, du_B \, d\gamma.
$$

If $Y = (Y_1, \ldots, Y_m)$ is in the max-domain of attraction of $G_\ast$, we will further assume that the following condition holds:

- **C3:** There exists $\alpha > 0$ such that, uniformly for $B \in \mathcal{P}$ and $x = (x_B, e_{B^c}) \in \mathcal{A}_B$, as $t \to \infty$,

$$
\Pr(X_1^* \in (e, x]) - F_{X^*_\gamma(A_B)}(x_B) = O \left( t^{-\alpha} \right),
$$

where $F_{X^*_\gamma(A_B)}$ is the multivariate distribution function of $X^*$ (defined in Proposition 2) restricted to $\mathcal{A}_B$.

3.2 Likelihood inference for random samples of high dimensional max-stable distributions

We consider the parametric statistical model: $\mathcal{F} = \{h(\theta; z), \theta \in \Theta, z_i \in \mathbb{R}^m_i, i = 1, \ldots, n\}$ where $z_i$ are observations of a sample of size $n$ distributed as $Z = (Z_1, \ldots, Z_m)$ with density function $h(\theta; \cdot)$ as in Proposition 1. Let denote by $\theta_0$ the true parameter. We assume that $\theta_0$ is interior to the parameter space $\Theta$.

The log-likelihood of an observation $z \in \mathbb{R}^m_+$ is given by

$$
\log \ell_1(\theta; z) = -\sum_{l=1}^m z_l \mu(\theta; \{l\}, z) + \log(\delta(\theta; z))
$$

where $\delta(\theta; z) = \sum_{\pi \in \Pi} \chi(\theta; \pi, z)$ with $\chi(\theta; \pi, z) = \prod_{B \in \pi} \mu(\theta; B, z)$.

As explained previously, the computation of this log-likelihood requires to consider the set $\Pi$ of all possible partitions of the components of $z$. Due to the explosive behavior of the number of partitions, this likelihood may become numerically intractable for a too large number of components. We therefore decide to also consider the partition-composite likelihood

$$
\log \ell_2(\theta; z) = \sum_{B \in \pi} |B| \log \ell_1(\theta; z_B)
$$
for which it is necessary to only compute the likelihoods for the blocks of a partition \( \pi \). By bounding the sizes of the blocks, the computation is feasible in a moderate time. Of course, the partition should be chosen such that the components of the vector in blocks are the most dependent as possible or equivalently such that the (presumed) assumption of independence between the blocks of the partition is the most reasonable as possible.

**Remark 1** In practice, such a partition may be chosen by using a clustering algorithm like the Partitioning Around Medoids (PAM) algorithm as proposed in [2]. Let \( d_{ij} \) be the F-madogram between \( Z_i \) and \( Z_j \), that is

\[
d_{ij} = \frac{1}{2} \mathbb{E}[\exp(-Z_i^{-1}) - \exp(-Z_j^{-1})] = \frac{1}{2} \mathbb{E}[\max(U_i^+, U_j^+)] - 1
\]

(see [6]). It can be easily evaluated with non-parametric estimators. The PAM algorithm divides the \( m \) components into \( p \) clusters in the following way: (i) select randomly an initial set of \( p \) medoids, (ii) form \( p \) clusters by assigning every point to its closest medoid (using the distance matrix \( (d_{ij}) \)), (iii) for each cluster, find the new medoid for which the total intra-cluster distance based on \( d_{ij} \) is minimized, (iv) if at least one medoid has changed, then go back to (ii), otherwise end the algorithm.

To bound the sizes of the clusters, it may be necessary to increase the number of clusters until the maximum size of the clusters is less than a given integer.

Moreover we will also compare the two previous log-likelihoods with the pairwise log marginal likelihood

\[
\log \ell_3 (\theta; z) = \sum_{i<j} \log \ell_1 (\theta; z_{\{i,j\}})
\]

for which only \( m(m - 1)/2 \) bivariate-likelihoods are needed.

Conditions 1 and 2 of Section 3.1 imply the existence of the score functions

\[
\nabla_\theta \log \ell_1 (\theta; z) = -\sum_{i=1}^m z_i \nabla_\theta \mu (\theta; \{i\}, z) + \sum_{\pi \in \Pi} \frac{\chi (\theta; \pi, z)}{\sum_{\pi' \in \Pi} \chi (\theta; \pi', z)} \sum_{B \in \pi} \nabla_\theta \log \mu (\theta; B, z),
\]

\[
\nabla_\theta \log \ell_2 (\theta; z) = \sum_{B \in \pi} |B| \nabla_\theta \log \ell_1 (\theta; z_B),
\]

\[
\nabla_\theta \log \ell_3 (\theta; z) = \sum_{i<j} \nabla_\theta \log \ell_1 (\theta; z_{\{i,j\}}),
\]

with their respective Hessian \( \nabla_\theta^2 \log \ell_j (\theta; z) \) for \( j = 1, 2, 3 \) (see Section 6.7 for their expression). Let, for \( j = 1, 2, 3 \),

\[
I_j (\theta) = \mathbb{E} \left[-\nabla_\theta^2 \log \ell_j (\theta; Z)\right] \quad \text{and} \quad J_j (\theta) = \mathbb{E} \left[\nabla_\theta \log \ell_j (\theta; Z) \nabla_\theta \log \ell_j (\theta; Z')\right].
\]

The maximum likelihood estimator \( \hat{\theta}_n^{(1)} \) (MLE), the maximum partition-composite likelihood estimator \( \hat{\theta}_n^{(2)} \) (MpcLE) and the maximum pairwise marginal likelihood estimator \( \hat{\theta}_n^{(3)} \) (MpmLE) are respectively defined by the conditions

\[
\sum_{i=1}^n \nabla_\theta \log \ell_j (\hat{\theta}_n^{(j)}; z_i) = 0, \quad j = 1, 2, 3.
\]

We get the following proposition (given without proof, see e.g. Section 4.4.2 in [9]):
Proposition 3 Assume that (C1) and (C2) hold. The likelihood estimators are consistent and asymptotically distributed as \( n \) tends to infinity

\[
\sqrt{n}(\hat{\theta}_n^{(j)} - \theta_0) \xrightarrow{d} N(0, I^{-1}_j(\theta_0)) \quad \text{and} \quad \sqrt{n}(\hat{\theta}_n^{(j)} - \theta_0) \xrightarrow{d} N(0, I^{-1}_j(\theta_0) J_j(\theta_0) I^{-1}_j(\theta_0)), \quad j = 2, 3.
\]

It is well-known that the MLE attains the Cramér–Rao lower bound \( I^{-1}_j(\theta) \) and that estimation using the composite likelihood (for the MpmLE and the MpcLE) results in a loss of efficiency. But it is very difficult to assess the level of efficiency from a theoretical point of view (see e.g. [8]).

It has been assumed that the one-dimensional marginal distributions of \( Z \) are unit Fréchet. It is however possible to consider the case when the one-dimensional marginal distributions belong to the class of generalized extreme value (GEV) distributions. The GEV distributions are characterized by location, scale, and shape parameters (see e.g. [13]). Generally these parameters are described through parsimonious regression models which are functions of covariates to avoid computational issues that could arise for large numbers of parameters. In this way likelihood based inference allows simultaneous assessment of the parameters of the tail-dependence function as well as the location, scale, and shape parameters of the marginal distributions (see an example for spatial extremes with the pairwise likelihood in [39]).

To perform model selection between nonnested models, it is possible to use Takeuchi’s information criterion ([50]) that has already been proposed for composite likelihood approaches (see [52]). This criterion selects the model minimizing an estimate of the Kullback–Leibler divergence between the pairwise unknown model and the adopted model. Such an estimate may be given for example by, for \( j = 1, 2, 3, \)

\[
\frac{1}{n} \sum_{i=1}^n \log \ell_j(\hat{\theta}_n^{(j)}; z_i) - \text{tr} \left( \hat{J}_j(\hat{\theta}_n^{(j)})(\hat{I}_j(\hat{\theta}_n^{(j)}))^{-1} \right)
\]

where

\[
\hat{I}_j(\theta) = -\frac{1}{n} \sum_{i=1}^n \nabla^2 \log \ell_j(\theta; z_i) \quad \hat{J}_j(\theta) = \frac{1}{n} \sum_{i=1}^n \nabla \log \ell_j(\theta; z_i) \nabla \log \ell_j(\theta; z_i)'.
\]

If a Monte-Carlo integration is performed as the numerical integration method for computing \( \mu(\theta; B, z) \) and if the number of simulations is not sufficiently large, the asymptotic distributions in Proposition 3 may be modified. Assume that it is possible to write \( \mu(\theta; B, z) \) as an expectation of a function \( a \) with respect to a random variable \( V \) (independent of \( \theta, B \) and \( z \)) such that

\[
\mu(\theta; B, z) = \mathbb{E}[a(V; \theta, B, z)]
\]

and assume that \( \nabla_{\theta} a(v; \theta, B, z) \) exists for any \( v, \theta, B \) and \( z \). For example, one possible way could be to choose \( V \) as a univariate random variable with a unit Pareto distribution and let

\[
a(v; \theta, B, z) = v^{-|B|} \lambda(v^{-1}; \theta, B, z) + v^{2+|B|} \lambda(v; \theta, B, z)
\]

with \( \lambda(v; \theta, B, z) = \Pr(\mathbf{U}_B \leq z_{B'} | \mathbf{U}_B = z_{B'} v) f_{U_B}(z_{B'} v) \). For a unique sample of size \( S \) of \( V \), estimates of \( \mu(\theta; B, z) \) are then given by

\[
\mu_S(\theta; B, z) = \frac{1}{S} \sum_{s=1}^S a(V_s; \theta, B, z)
\]
(we use the same sample \((V_s)_{s=1,...,S}\) for all estimates). The simulated log-likelihood log \(\ell_{1S}\) and the partition-composite log likelihood log \(\ell_{2S}\) are derived by replacing \(\mu\) by \(\mu_S\). The simulated maximum likelihood estimator (SMLE) \(\hat{\theta}_{n,S}^{(1)}\) and the simulated maximum partition-composite likelihood estimator (SMpcLE) \(\hat{\theta}_{n,S}^{(2)}\) then satisfy
\[
\sum_{j=1}^{n} \nabla_\theta \log \ell_{Sj}(\ell_{n,S}; z_i) = 0, \quad j = 1, 2.
\]

For the sake of brevity, the next proposition only states the asymptotic properties of \(\hat{\theta}_{n,S}^{(1)}\). Let
\[
\phi_1 (v; \theta, \pi, z) = -\frac{\nabla_\delta \delta (\theta; z)}{\delta (\theta; z)} \sum_{B \in \pi} \frac{1}{\mu (\theta; B, z)} a (v; \theta, B, z)
\]
\[
\phi_2 (v; \theta, \pi, z) = \left( \sum_{B' \in \pi} \nabla_\theta \log \mu (\theta; B', z) \right) \sum_{B \in \pi} \frac{1}{\mu (\theta; B, z)} a (v; \theta, B, z)
\]
\[
\phi_3 (v; \theta, \pi, z) = \sum_{B \in \pi} \frac{1}{\mu (\theta; B, z)} \left[ \nabla_\theta a (v; \theta, B, z) - \nabla_\theta \log \mu (\theta; B, z) a (v; \theta, B, z) \right]
\]
and
\[
\psi (v; \theta, \pi, z) = \frac{1}{\delta (\theta; z)} \sum_{\pi \in \Pi} \chi (\theta; \pi, z) \sum_{j=1}^{3} \phi_j (v; \theta, \pi, z) - \sum_{l=1}^{m} z_l \nabla_\theta a (v; \theta, \{l\}, z).
\]

**Proposition 4** Assume that \(n\) and \(S\) tend to infinity. Under the regularity conditions of the previous proposition,
- if \(n/S\) tends to zero, then
\[
\sqrt{n}(\hat{\theta}_{n,S}^{(1)} - \theta_0) \xrightarrow{d} \mathcal{N} \left( 0, I_1^{-1}(\theta_0) \right),
\]
- if \(n/S\) tends to infinity, then
\[
\sqrt{S} \left( \hat{\theta}_{n,S}^{(1)} - \theta_0 \right) \xrightarrow{d} \mathcal{N} \left( 0, I_1^{-1}(\theta_0) \Sigma (\theta_0) I_1^{-1}(\theta_0) \right)
\]
with \(\Sigma (\theta_0) = \mathbb{V} (\mathbb{E} [\psi (V; \theta_0, \pi, Z) | V])\),
- if \(n/S\) tends to a positive constant \(\varphi < \infty\), then
\[
\sqrt{n}(\hat{\theta}_{n,S}^{(1)} - \theta_0) \xrightarrow{d} \mathcal{N} \left( 0, I_1^{-1}(\theta_0) (I_1 (\theta_0) + \varphi \Sigma (\theta_0)) I_1^{-1}(\theta_0) \right).
\]

### 3.3 Likelihood inference for (censored) random samples in a max-domain of attraction

Let \((Y_i)_{i=1,...,n}\) be independent an identically vectors distributed as \(Y = (Y_1, \ldots, Y_m)\) whose marginals have unit Pareto distributions and assume that \(Y\) is in the max-domain of attraction of \(G_*\). For \(1 \leq k \leq n\), we define
\[
X_{i,n/k} = \frac{k}{n} Y_i \vee e
\]
and we only consider observations \(X_{i,n/k}\) for which \(\|X_{i,n/k}\|_\infty > 1\). We hence censor components that do not exceed the threshold \(n/k\). The number of censored observations is given by the cardinal of the set
\[
N_k = \left\{ X_{i,n/k} : 1 \leq i \leq n, \|X_{i,n/k}\|_\infty > 1 \right\}.
\]
We consider the (pseudo)-parametric statistical model \( \mathcal{F}_k = \{ \ell_{X} (\theta; x_{i,k}) \mid \theta \in \Theta, x_{i,k} \in \mathcal{A}, i \in \mathcal{N}_k \} \) whose log-likelihood, for \( x \in \mathcal{A} \), is given by

\[
\log \ell_{X} (\theta; x) = \sum_{B \in \mathcal{P}} \log \mu (\theta; B, (x_B, e_{B^c})) \mathbb{I}_{\{x \in A_B\}} - \log \sum_{i=1}^m \mu (\theta, \{i\}, e).
\]

Conditions 1 and 2 of Section 3.1 imply the existence of the score function

\[
\nabla_\theta \log \ell_{X} (\theta; x) = \sum_{B \in \mathcal{P}} \left( \frac{\nabla_\theta \mu (\theta; B, (x_B, e_{B^c}))}{\mu (\theta; B, (x_B, e_{B^c}))} - \frac{\nabla_\theta V_B^*(\theta; e)}{V_B^*(\theta; e)} \right) \mathbb{I}_{\{x \in A_B\}}
\]

and the Hessian (see Section 6.7 for its expression).

The maximum likelihood estimator (MLE) satisfies the condition

\[
\sum_{i \in \mathcal{N}_k} \nabla_\theta \log \ell_{X} \left( \hat{\theta}_k; x_{i,n/k} \right) = 0. \tag{3.8}
\]

**Proposition 5** Assume that \((C1), (C2) and (C3)\) hold. If, as \(n \to \infty, k \to \infty\) such that \(k = o \left( n^{2\alpha/(1+2\alpha)} \right)\), then

\[
\sqrt{k} \left( \hat{\theta}_k - \theta_0 \right) \overset{d}{\to} \mathcal{N} \left( 0, V^{-1}_* (\theta_0) I^{-1}_X (\theta_0) \right)
\]

where \(I_{X} (\theta) = \mathbb{E} \left[ \nabla_\theta \log \ell_{X} (\theta; x^*) \nabla_\theta \log \ell_{X} (\theta; x^*) \right]\).

It has been assumed that the one-dimensional marginal distributions of \(Y\) are unit Pareto distributions. When it is not the case, it is common to transform the marginals to have common unit Pareto could also be used.

Let us consider a sample of random vectors \((Y_i)_{i=1,\ldots,n}\) distributed as \(Y = (Y_1, \ldots, Y_m)\) where \(Y\) is in the max-domain of attraction of an extreme value distribution \(G\) whose marginals are Fréchet distributed with unknown parameters \(\alpha_j\) and such that \(G \left( z^{1/\alpha} \right) = G_z (z)\). For \(1 \leq k \leq n\), we then define

\[
\hat{x}_{i,n/k} = \left( \frac{Y_i}{\hat{U}_Y (n/k)} \right)^{\hat{\alpha}_n} \vee e
\]

where \(\hat{\alpha}_n = (\hat{\alpha}_{1,n}, \ldots, \hat{\alpha}_{m,n})'\) and \(\hat{U}_Y (n/k) = (\hat{U}_Y (1/n/k), \ldots, \hat{U}_Y (n/k))'\) with

\[
\frac{1}{\hat{\alpha}_{j,n}} = \frac{1}{k} \sum_{i=0}^{k-1} \log \frac{Y_{(n-i)j}}{Y_{(n-k)j}}, \quad \hat{U}_Y (n/k) = Y_{(n-k)j} \quad \text{and} \quad Y_{(1)j} \leq Y_{(2)j} \leq \ldots \leq Y_{(n)j}.
\]

The number of censored observations is given by the cardinal of the set

\[
\hat{N}_k = \{ \hat{x}_{i,n/k} : 1 \leq i \leq n, \| \hat{x}_{i,n/k} \|_\infty > 1 \},
\]

and the maximum likelihood estimator (MLE) satisfies the condition

\[
\sum_{i \in \hat{N}_k} \nabla_\theta \log \ell_{X} \left( \hat{\theta}_k; \hat{x}_{i,n/k} \right) = 0.
\]
This estimator is also asymptotically Gaussian, but its variance is modified due to the non-linear transformation of data. Let \((W_B(x_B), x_B \in [e_B, \infty))_B \in \mathcal{P}\) be independent zero-mean Gaussian random fields with covariance functions

\[
\text{Cov}(W_B(x_B), W_B(y_B)) = \nu_B((x_B, \infty] \cap (y_B, \infty])
\]

where

\[
\nu_B((x_B, \infty]) = \int_0^{\infty} \Pr(U_B > \gamma x_B, U_{B^c} \leq \gamma e_B) d\gamma.
\]

Then define

\[
W_j(x_j) = \sum_{B \in \mathcal{P}, B \cap \{j\} \neq \emptyset} W_B((1, \ldots, 1, x_j, 1, \ldots, 1))
\]

and let

\[
\psi_j = \mathbb{E}[\nabla^2_{\theta x_j} \log \ell_{X^*}(\theta_0; X^*) X_*^* \log X_j^* | X_j^* > 1]
\]

\[
\omega_j = \mathbb{E}[\nabla^2_{\theta x_j} \log \ell_{X^*}(\theta_0; X^*) X_*^* | X_j^* > 1].
\]

**Proposition 6** Assume that (C1), (C2) and (C3) hold. If, as \(n \to \infty, k \to \infty\) such that \(k = o(n^{2\alpha/(1+2\alpha)})\), then

\[
\sqrt{k} V_*(\theta_0; e) I_{X^*}(\theta_0) \left( \hat{\theta}_k - \theta_0 \right)
\]

\[
\xrightarrow{d} \sum_{B \in \mathcal{P}} \int_{A_B} \nabla_\theta \log \ell_{X^*}(\theta_0; x) W_B(dx_B) - \sum_{j=1}^{m} (\psi_j - \omega_j) W_j(1) - \sum_{j=1}^{m} \alpha_j \omega_j \int_1^{\infty} W_j(x_j^*) \frac{dx_j}{x_j}.
\]

Note that the random variable \(\sum_{B \in \mathcal{P}} \int_{A_B} \nabla_\theta \log \ell_{X^*}(\theta_0; x) W_B(dx_B)\) has a Gaussian distribution with mean 0 and variance equal to \(V_*(\theta_0; e) I_{X^*}(\theta_0)\).

## 4 Simulation studies and real data example

We design a number of simulation studies to investigate the use of our likelihood based methods for inference on the parameters of high dimensional extreme value distributions. A real data application on financial extreme events is also presented.

### 4.1 Monte Carlo experiments for random samples from the Schlather max-stable process

We first consider random samples from the Schlather max-stable process. The spatial process \(U\) in (2.2) is a stationary Gaussian process. We examine various forms of correlation functions for this spatial process: the Whittle Mattern correlation function given in (2.5), the Cauchy correlation function

\[
\rho(h) = [1 + (h/c)^2]^{-\nu}, \quad c > 0, \nu > 0,
\]

and the Powered Exponential correlation function

\[
\rho(h) = \exp(-(h/c)^{\nu}), \quad c > 0, 0 < \nu \leq 2,
\]

where \(c\) and \(\nu\) are the range and the smooth parameters of the correlation functions, and \(h\) is the distance between two locations.
To minimize the computational times of \( \mu(B; z) \), we used the efficient Fortran code proposed by Genz (see Example 4 in Section 2.2 and [22]) and introduced the additional term \( |y_{B_{|+1}}|^B \) in the multiple integral of the normal probability.

For 100 simulation replicates, we randomly generate \( m = 50 \) (resp. 100) locations uniformly in the square \([0, 2] \times [0, 2]\). We then simulate \( n = 40 \) (resp. 20) max-stable process realizations under each model at the sampled \( m \) locations using the SpatialExtremes package in R ([41]).

Since the intensity of the dependence between two locations is a function of the distance between these locations (isotropic correlation functions), we decided to gather sites for the partition likelihood with a usual K-means algorithm based on the distance between the locations irrespectively of the values of the max-stables processes (we increased the number of clusters until the largest size of the clusters become smaller than or equal to 5).

We compared partition composite likelihood inference to that from pairwise marginal composite likelihood. The maximum likelihoods estimates of \((c, \nu)\) were computed from the 100 simulation replicates and were used to get the empirical standard deviations, the empirical mean square errors and the empirical relative efficiencies \((RE)\) which is equal to the ratio of the mean square errors for the pairwise marginal composite likelihood inference method over the mean square errors for the partition composite likelihood inference method.

The estimates, with their standard deviations (in brackets) and their relative efficiencies, are reported respectively in Tables 1, 2 and 3 for the Whittle Mattern, the Cauchy and the Powered Exponential correlation functions.

We observe very small biases for the MpcLE with the Whittle Mattern and the Powered Exponential correlation functions and moderate biases for the estimates of the smooth parameter with the Cauchy correlation function. On the contrary the MpmLE of the range and the smooth parameters (resp. the smooth parameters) have serious upward biases for the Cauchy (resp. Whittle Mattern) correlation function. It is noticeable that it is strongly linked with issues of convergence of the maximization algorithm of the pairwise marginal composite likelihood. The standard deviations of the MpcLE are always smaller than those of the MpmLE and the differences may be very large for the Whittle Mattern and the Cauchy correlation functions. The pairwise marginal composite likelihood achieves more accurate parameter estimations in the set up \( m = 50 \) and \( n = 40 \) (more spatial observations but less temporal observations) while the partition composite likelihood gives roughly the same precision. We conclude to very significant efficiency gains when using the partition composite likelihood inference.

| Whittle Mattern \( m/n \) | 100/20 | 50/40 |
|--------------------------|-------|-------|
| \( c \) | \( \nu \) | \% | \( RE_c/RE_\nu \) | \( c \) | \( \nu \) | \% | \( RE_c/RE_\nu \) |
| True | 1 | 0 | 0.98(0.13) | 1 | 0 | 1.00/1.00 | 1 | 1 | 1.00(0.06) | 0.13 | 1.00/1.00 |
| MpcLE | 0.98(0.13) | 1.01(0.05) | 0 | 1.00/1.00 | 1.01(0.13) | 1.00(0.06) | 0 | 1.00/1.00 |
| MpmLE | 1.01(0.46) | 1.78(3.65) | 4 | 11.20/4825 | 1.02(0.34) | 1.48(3.93) | 2 | 6.23/4134 |
| True | 1 | 2 | 0 | 1.00/1.00 | 1.00(0.12) | 1.00(0.13) | 0 | 1.00/1.00 |
| MpcLE | 1.01(0.13) | 2.00(0.12) | 0 | 1.00/1.00 | 1.00(0.12) | 2.01(0.14) | 0 | 1.00/1.00 |
| MpmLE | 1.09(0.61) | 5.07(10.18) | 3 | 22.0/7140 | 1.10(0.50) | 4.24(5.52) | 0 | 17.2/3860 |
| True | 0.5 | 1 | 0.5 | 0.51(0.08) | 1.00(0.08) | 1.00(0.13) | 0 | 1.00/1.00 |
| MpcLE | 0.51(0.08) | 1.00(0.08) | 0 | 1.00/1.00 | 5.07(10.18) | 4.24(5.52) | 0 | 8.70/881 |
Table 1: Means of the partition (MpcLE) and pairwise marginal (MpmLE) maximum likelihood estimates with their standard deviations (in brackets) and their relative efficiencies (RE) for the Whittle Mattern correlation function. The statistics are computed over 100 repetitions for each configuration of \(m, n\) and \((c, \nu)\). The column (%) gives the number of times when the maximization algorithm failed to converge.

| Cauchy \(m/n\) | \(100/20\) | \(50/40\) | \(c\) | \(\nu\) | \(\%\) | \(RE_c/RE_\nu\) | \(c\) | \(\nu\) | \(\%\) | \(RE_c/RE_\nu\) |
|----------------|-----------|-----------|-------|-------|------|----------------|-------|-------|------|----------------|
| True           | 1         | 1         | 0     | 1.00/1.00 | 1.03(0.16) | 1.12(0.45) | 0     | 1.00/1.00 |
| MpcLE          | 1.03(0.18)| 1.17(0.56)| 0     | 1.00/1.00 | 1.03(0.16) | 1.12(0.45) | 0     | 1.00/1.00 |
| MpmLE          | 1.71(1.89)| 7.13(18.32)| 9     | 126/1083 | 1.34(1.35) | 3.61(11.35)| 2     | 70.6/620  |
| True           | 1         | 2         | 2     | 1.00/1.00 | 1.07(0.28) | 2.54(1.59) | 2     | 1.00/1.00 |
| MpcLE          | 1.08(0.28)| 2.57(1.73)| 2     | 1.00/1.00 | 1.07(0.28) | 2.54(1.59) | 2     | 1.00/1.00 |
| P.M.           | 1.68(1.68)| 11.83(24.03)| 19    | 38.0/202 | 1.22(1.10) | 5.56(11.53)| 11    | 14.8/78   |
| True           | 0.5       | 1         | 0.5   | 1.00/1.00 | 0.53(0.12) | 1.24(0.75) | 0     | 1.00/1.00 |
| MpcLE          | 0.52(0.11)| 1.16(0.59)| 4     | 46.9/283 | 0.69(0.63) | 3.58(11.00)| 2     | 26.1/208  |
| True           | 0.76(0.69)| 4.35(9.73)| 4     | 46.9/283 | 0.69(0.63) | 3.58(11.00)| 2     | 26.1/208  |

Table 2: Means of the partition (MpcLE) and pairwise marginal (MpmLE) maximum likelihood estimates with their standard deviations (in brackets) and their relative efficiencies (RE) for the Cauchy correlation function. The statistics are computed over 100 repetitions for each configuration of \(m, n\) and \((c, \nu)\). The column (%) gives the number of times when the maximization algorithm failed to converge.

| Powered Exponential \(m/n\) | \(100/20\) | \(50/40\) | \(c\) | \(\nu\) | \(\%\) | \(RE_c/RE_\nu\) | \(c\) | \(\nu\) | \(\%\) | \(RE_c/RE_\nu\) |
|-----------------------------|-----------|-----------|-------|-------|------|----------------|-------|-------|------|----------------|
| True                        | 1         | 1         | 1     | 1.00/1.00 | 1.00(0.08) | 1.04(0.23) | 1.00(0.08) | 0     | 1.00/1.00 |
| MpcLE                       | 1.02(0.26)| 1.00(0.08)| 0     | 1.00/1.00 | 1.04(0.23) | 1.00(0.08) | 0     | 1.00/1.00 |
| MpmLE                       | 1.08(0.45)| 1.00(0.16)| 0     | 3.08/3.83 | 1.03(0.26) | 1.02(0.12) | 0     | 3.55/7.58 |
| True                        | 1         | 2         | 2     | 1.00/1.00 | 1.01(0.11) | 1.50(0.06) | 0     | 1.00/1.00 |
| MpcLE                       | 1.02(0.13)| 1.50(0.05)| 0     | 1.00/1.00 | 1.01(0.11) | 1.50(0.06) | 0     | 1.00/1.00 |
| MpmLE                       | 1.05(0.24)| 1.49(0.19)| 0     | 3.61/12.5 | 1.04(0.20) | 1.50(0.10) | 0     | 3.55/7.58 |
| True                        | 0.5       | 1         | 0.5   | 1.00/1.00 | 0.52(0.09) | 1.00(0.09) | 0     | 1.00/1.00 |
| MpcLE                       | 0.53(0.11)| 1.00(0.08)| 0     | 1.00/1.00 | 0.52(0.09) | 1.00(0.09) | 0     | 1.00/1.00 |
| MpmLE                       | 0.59(0.28)| 0.98(0.21)| 0     | 6.25/6.83 | 0.52(0.13) | 1.00(0.16) | 0     | 2.13/3.26 |

Table 3: Means of the partition (MpcLE) and pairwise marginal (MpmLE) composite maximum likelihood estimates with their standard deviations (in brackets) and their relative efficiencies (RE) for the Powered Exponential correlation function. The statistics are computed over 100 repetitions for each configuration of \(m, n\) and \((c, \nu)\). The column (%) gives the number of times when the maximization algorithm failed to converge.
4.2 Monte Carlo experiments for (censored) random samples in the max-domain of attraction of the Schlather max-stable distribution

We now simulate random samples \((Y_i)_{i=1,...,n}\) distributed as \(Y = (Y_1, \ldots, Y_m)\) in the max-domain of attraction of the Schlather max-stable distribution. We consider the following model: \(Y_j = \Gamma U(x_j)\) for \(j = 1, \ldots, m\) where \(\Gamma\) is a random variable with unit Pareto distribution, \(U\) is the spatial stationary Gaussian process in \((2,2)\), and \(x_1, \ldots, x_m\) are the site locations. We assume moreover that \(\Gamma\) and \((U(x_1), \ldots, U(x_m))\) are independent. As in the previous section, we also examine the cases when the correlation function is given by the Whittle Mattern function, the Cauchy function and the Powered Exponential function.

For 100 simulation replicates, we randomly generate \(m = 20\) locations uniformly in the square \([0,2] \times [0,2]\). We then simulate \(n = 250\) (resp. 1000) realizations of \(\Gamma\) and \((U(x_1), \ldots, U(x_m))\) under each model at the sampled \(m\) locations \((x_i)_{i=1,...,m}\). The (censored) maximum likelihood estimates of \((c, \nu)\) with their empirical standard deviations were computed from these replicates. They are respectively reported in Tables 4, 5 and 6 for the Whittle Mattern, the Cauchy and the Powered Exponential correlation functions.

| WM          | 5%  | 10% | 15% | 20% | 5%  | 10% | 15% | 20% |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|
| \( (1,1) \) | 0.96(0.29) | 0.96(0.24) | 1.01(0.19) | 1.00(0.17) | 1.09(0.21) | 1.07(0.19) | 1.02(0.12) | 1.01(0.10) |
| \( (1,2) \) | 1.02(0.25) | 0.99(0.14) | 0.99(0.11) | 0.99(0.10) | 2.04(0.28) | 2.03(0.18) | 2.02(0.14) | 2.01(0.13) |
| \( (0.5,1) \) | 0.52(0.19) | 0.52(0.12) | 0.51(0.10) | 0.50(0.09) | 1.10(0.36) | 1.04(0.22) | 1.03(0.18) | 1.03(0.16) |

Table 4: Means of the maximum likelihood estimates with their standard deviations (in brackets) for the Whittle Mattern correlation function. The statistics are computed over 100 repetitions for each configuration of \(n, k/n\) and \((c, \nu)\).

| C           | 5%  | 10% | 15% | 20% | 5%  | 10% | 15% | 20% |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|
| \( (1,1) \) | 1.07(0.29) | 1.07(0.22) | 1.06(0.14) | 1.05(0.18) | 1.32(1.11) | 1.27(0.80) | 1.20(0.42) | 1.18(0.61) |
| \( (1,2) \) | 1.10(0.51) | 1.03(0.22) | 1.08(0.28) | 1.04(0.17) | 2.62(2.42) | 2.29(1.21) | 2.39(1.18) | 2.28(0.88) |
| \( (0.5,1) \) | 0.54(0.17) | 0.54(0.14) | 0.52(0.11) | 0.52(0.10) | 1.24(0.73) | 1.26(0.74) | 1.16(0.60) | 1.17(0.53) |

Table 5: Means of the maximum likelihood estimates with their standard deviations (in brackets) for the Cauchy correlation function. The statistics are computed over 100 repetitions for each configuration of \(n, k/n\) and \((c, \nu)\).
Table 6: Means of the maximum likelihood estimates with their standard deviations (in brackets) for the Powered Exponential correlation function. The statistics are computed over 100 repetitions for each configuration of $n$, $k/n$ and $(c, \nu)$.

We observe very small biases for the MLE with the Whittle Mattern and the Powered Exponential correlation functions and moderate biases for the estimates of the smooth parameter with the Cauchy correlation function. Note that it is consistent with the finding that between 1\% and 14\% of the maximizations failed to converge for this correlation function. As expected, efficiency is improved when the numbers of observations $n$ increases and/or the ratio $k/n$ (the reciprocal of the threshold used to censor the observations) increases. From this Monte-Carlo experiment, we may conclude that our inference method performs very well for samples of moderate size generated from Schlater’s spatial processes.

4.3 Monte Carlo experiments for (censored) random samples in the max-domain of attraction of a “clustered” max-stable distribution

Let $(P_i)_{i=1,\ldots,l}$ be a partition of $I = \{1, \ldots, m\}$ and $U = (U_{P_1}, \ldots, U_{P_l})$ be a positive random vector in $\mathbb{R}^m$ with expectation equal to $e$ such that $U_{P_1}, \ldots, U_{P_l}$ are independent sub-vectors and

$$
\Pr(U_{P_i} \leq z_{P_i}) = C_i((F_i(z_j))_{j \in P_i})
$$

where $C_i$ is an Archimedean copula with a completely monotone generator $\psi_i$ and $F_i$ is the common marginal distribution function of $U_{P_i}$. In that case, we have

$$
\mu(B; z) = \int_0^\infty \gamma |B| \prod_{i=1}^l \left[ \psi_i^{(B\cap P_i)} \left( \sum_{j \in P_i} \psi_i^{-1}(F_i(z_j)) \right) \prod_{j \in B \cap P_i} \psi_i^{-1}(F_i(z_j)) f_i(\gamma z_j) \right] d\gamma
$$

(4.9)

where $f_i$ is the probability density function of the $U_j$ for $j \in P_i$. We call the max-stable distribution associated with the tail-dependence function generated by $U$ (see Eq. (2.1)) an homogeneous “clustered” max-stable distribution.

We decided to simulate random samples $(Y_i)_{i=1,\ldots,n}$ distributed as $Y = \Gamma U$ where $\Gamma$ is a random variable with unit Pareto distribution independent of $U$. For 100 simulation replicates, we randomly generate $Y$ with the following assumptions:

- $n = 2500$, $m = 100$ and $I = 3$ with $P_1 = \{1, \ldots, 50\}$, $P_2 = \{51, \ldots, 80\}$ and $P_3 = \{81, \ldots, 100\}$;
- $C_1$ is the Gumbel copula with generator $\psi_1(t) = \exp(-t^{\theta_1})$ and $\theta_1 = 1.7$, $C_2$ is the Clayton copula with generator $\psi_2(t) = (1 + t)^{-\theta_2}$ and $\theta_2 = 0.4$, $C_3$ is the Gumbel copula with parameter $\theta_3 = 1.2$;
• $F_1$ is the lognormal distribution associated to the Gaussian distribution with standard error $\alpha_1 = 0.9$ and mean equal to $-\alpha_1^2/2$, $F_2$ is the Weibull distribution with shape parameter $\alpha_2$ equal to 1.5 and a scale parameter such that its expectation is equal to one, $F_3$ is the Fréchet distribution with shape parameter $\alpha_3$ equal to 1.7 and a scale parameter such that its expectation is equal to one.

We assumed for this Monte-Carlo experiment that the partition $(P_i)_{i=1,...,3}$, the types of copula and the marginal distributions for all the sets of the partition are known (see the next section for a procedure to partition $U$ via a clustering algorithm and to choose the appropriate types of copula and marginal distribution in practice). The statistical procedure we implemented follows two steps. Firstly we estimated separately the pairs $(\theta_i, \alpha_i)$ for each cluster $(i = 1, \ldots, 3)$ of the partition. Secondly we estimated jointly all the parameters by using (4.9) for the likelihood with initial conditions given by estimates of the pairs.

The results are presented in the following table.

| Cluster 1 (size=50) | Cluster 2 (size=30) | Cluster 3 (size=20) |
|---------------------|---------------------|---------------------|
| Copula              | Marg. dist.         | Copula              | Marg. dist.         | Copula              | Marg. dist.         |
| Gumbel              | LogNormal           | Clayton             | Weibull             | Gumbel              | Fréchet             |
| $(\theta_i, \alpha_i)$ | 1.7                  | 0.9                 | 0.4                 | 1.5                 | 1.2                 |
| Means (step 1)      | 1.694                | 0.903               | 0.432               | 1.482               | 1.223               |
| s.e. (step 1)       | 0.055                | 0.033               | 0.039               | 0.031               | 0.066               |
| RMSE (step 1)       | 0.055                | 0.034               | 0.051               | 0.036               | 0.074               |
| Means (step 2)      | 1.706                | 0.915               | 0.435               | 1.460               | 1.222               |
| s.e. (step 2)       | 0.039                | 0.021               | 0.032               | 0.029               | 0.022               |
| RMSE (step 2)       | 0.039                | 0.026               | 0.048               | 0.049               | 0.032               |

Table 7. Means, standard errors and root-mean-square errors of parameter estimates of the copulas and of the marginal distributions for the first and the second step.

We observe very small biases for the estimates of the parameters of the copulas and of the marginal distributions in the first step of the estimation procedure. The second step may increase slightly these biases since the whole distribution is now taken into consideration, but it mainly reduces significantly the standard errors of the estimators which leads to smaller root-mean-square errors (except for the parameter of the Clayton copula). This Monte-Carlo experiment shows that clustered max-stable distributions may be efficiently estimated and therefore that they can be used to model high-dimensional extreme events in practice.

4.4 Real data example

The Capital Asset Pricing Model (CAPM) due to Sharpe and Lintner ([47] and [36]) relates the expected return of an asset to the expected return of the market portfolio. A key parameter of this model is the beta coefficient which is the slope of the market model:

$$(r_{jt} - r_f) = \beta_j (r_{Mt} - r_f) + \varepsilon_{jt}, \quad j = 1, \ldots, m, \quad t = 1, \ldots, T,$$

where $r_{jt}$ and $r_{Mt}$ are the returns of asset $j$ and of the market portfolio at time $t$ respectively, $r_f$ is the risk free rate of the market, and $\varepsilon_{jt}$ is the idiosyncratic risk or diversifiable risk of asset $j$. The beta is thus interpreted as the sensitivity of the asset return to changes in the return of the market portfolio. Commonly, the beta coefficients are estimated by Ordinary Least Squares applied to the
linear regressions for each asset. A lot of empirical studies have shown that the expected returns of assets cannot be perfectly explained by the market factor because betas are not stable over time and should be rather considered as stochastic to predict or explain the cross-section of asset returns (see e.g. the conditional CAPM of Hansen and Richard [25]). There is no agreement about the best process assumption for beta dynamics. Stochastic models for betas assume in general that betas have a constant mean and a random component, or that betas have a Gaussian mean reverting dynamics (conditional (or not) to observed instruments). Beta estimates are essential for many areas of modern finance, including asset pricing, cash flow valuation or performance evaluation. We will see that it is also the case for risk management of extreme events.

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models are used to model volatility of the market portfolio and heavy tailness of its returns. GARCH processes have indeed power law marginal tails and, more generally, regularly varying finite-dimensional distributions (see e.g. [7]). Let us assume that the returns of the market portfolio follows a GARCH process such that

\[ \Pr (r_{Mt} > r) \sim c_{Mt} r^{-\kappa}, \quad \Pr (r_{Mt} < -r) \sim c_{Mt} r^{-\kappa}, \quad r \to \infty, \]

with \( c_{Mt} > 0, c_{Mt} > 0 \) and \( \kappa > 0 \). We then consider a financial model with asymmetric beta in bull and bear market conditions

\[ r_{jt} = \alpha_i + \beta^+_j r_{Mt} \mathbb{I}[r_{Mt} > 0] + \beta^-_j r_{Mt} \mathbb{I}[r_{Mt} \leq 0] + \varepsilon_{jt}, \quad j = 1, \ldots, m, \quad t = 1, \ldots, T, \]

where \((\beta^+_j)_{j=1,\ldots,m} \) and \((\beta^-_j)_{j=1,\ldots,m} \) are independent and identically distributed random vectors for \( t = 1, \ldots, T \) such that \( \mathbb{E}[(\beta^+_j)^\kappa] < \infty \) and \( \mathbb{E}[(\beta^-_j)^\kappa] < \infty \) for all \( j = 1, \ldots, m \), and are independent of the market portfolio. Moreover we assume that the idiosyncratic risks are independent of the market portfolio and that they have lower (positive and negative) tails than the marginal distribution of the market portfolio.

Let us consider the extreme negative returns of the assets. Then it may be shown that \((-r_{1t}, \ldots, -r_{mt})\) belongs to the maximum domain of attraction of an extreme value distribution characterized by

\[ \lim_{r \to \infty} r \Pr \left( \bigcup_{j=1}^m \{r_{jt} < -r \} \right) = \log G_\ast (\mathbf{z}) = \mathbb{E} \left[ \max_{j=1,\ldots,m} z_{jt}^{-1/\kappa} (\beta^+_j)^\kappa \right]. \]

We now evaluate the relevance of this form on financial data. The data are those used by Fan, Liao and Mincheva to estimate high dimensional covariance matrices with a conditional sparsity structure in [17]. These data were obtained from the Center for Research in Security Prices database and consist of \( p = 100 \) (anonymous) stocks with their annualized daily returns for the period January 1st, 2000, to December 31st, 2010 (approximately 2700 observations for each asset). The stocks have been chosen from several different industry sectors.

We focused on the negative returns of the assets and chose to use the family of homogeneous “clustered" max-stable models as introduced in Section (4.3) to explain the extreme values of these data. Note that we transformed the marginal distributions of the negative returns of the assets such that they became approximately unit Pareto distributions by using their respective order statistics.

We began by clustering stocks that have the same extremal behaviors with the PAM algorithm (we performed the algorithm for maxima calculated over blocks of size 50). The average silhouette widths for the entire data set suggested us to choose three clusters \( P_i \) of respective sizes 20, 58 and 22. It is not possible to interpret the composition of the clusters with respect to the characteristics of the assets due to the anonymity of the assets, but it is expected that the extreme returns are
mainly clustered within companies in the same industrial sectors. It is also reasonable to think that assets in the same cluster share the same distribution for their betas.

We then considered for each cluster three copulas with parameter \( \theta \): the Gumbel copula, the Clayton copula and the Frank copula (with generator \( \psi(t) = -\log(1-e^{-t}(1-e^{-\theta})) \)), and three distributions with parameter \( \alpha \): the lognormal distribution, the Weibull distribution and the Fréchet distribution (see the previous subsection). We estimated the pairs \((\theta, \alpha)\) for the nine combinations of copulas and distributions. To choose the most appropriate ones, we used the chi-squared distance between the empirical distribution of \( N \), the number of components of the normalized vector exceeding the threshold 1, and the associated theoretical distribution using the estimated parameters. Indeed, for the cluster \( i \) of size \(|P_i|\), it may be shown that, for \( j = 1, \ldots, |P_i| \),

\[
\Pr(N_i = j) = \binom{|P_i|}{j} \frac{\sum_{k=0}^{j} (-1)^{k-1} \binom{j}{k} \int_0^\infty \left[ 1 - \psi_i \left( \frac{|P_i| - j + k}{\psi_i^{-1}(F_i(\gamma))} \right) \right] d\gamma}{\int_0^\infty \left[ 1 - \psi_i \left( \frac{|P_i|}{\psi_i^{-1}(F_i(\gamma))} \right) \right] d\gamma}.
\]

The values of the chi-squared distances for Cluster 1 are given as examples in Table 8. The Weibull distribution and the Clayton copula appear as the most suitable pair of distribution and copula for Cluster 1. Actually it is also the case for the two other clusters.

| Copula    | Lognormal | Weibull | Fréchet |
|-----------|-----------|---------|----------|
| Gumbel    | 0.199     | 0.104   | 0.554    |
| Clayton   | 0.174     | 0.088   | 0.627    |
| Frank     | 0.179     | 0.130   | 0.632    |

Table 8: For Cluster 1, chi-squared distances between the empirical distribution of the number of components of the normalized vector exceeding the threshold 1, and the associated theoretical distribution using the estimated parameters for the nine combinations of copulas and marginal distributions.

Table 9 gives for the three clusters the maximum likelihood estimates of the parameters \( \theta \) and \( \alpha \) for the combinations of copula and distribution that minimize the chi-squared distance. For these combinations, we re-estimated in a second step jointly all the parameters by using (4.9) for the likelihood with initial conditions given by the estimates of the pairs derived from the first step. Table 9 also provides the values of the joint maximum likelihood estimates.

| Copula    | Marg. dist. | Copula    | Marg. dist. | Copula    | Marg. dist. |
|-----------|-------------|-----------|-------------|-----------|-------------|
| Cluster 1 (size=20) | Cluster 2 (size=58) | Cluster 3 (size=22) |
| 1st step |
| \( \theta_1 = 0.625 \) | \( \hat{\alpha}_1 = 1.109 \) | \( \theta_2 = 1.034 \) | \( \hat{\alpha}_2 = 0.821 \) | \( \theta_3 = 0.669 \) | \( \hat{\alpha}_3 = 1.275 \) |
| s.e. = 0.120 | s.e. = 0.049 | s.e. = 0.056 | s.e. = 0.015 | s.e. = 0.105 | s.e. = 0.045 |
| \( \chi^2 = 0.094 \) | \( \chi^2 = 0.080 \) | \( \chi^2 = 0.028 \) |
| 2nd step |
| \( \theta_1 = 1.111 \) | \( \hat{\alpha}_1 = 0.939 \) | \( \theta_2 = 0.953 \) | \( \hat{\alpha}_2 = 0.814 \) | \( \theta_3 = 1.523 \) | \( \hat{\alpha}_3 = 0.920 \) |
| s.e. = 0.063 | s.e. = 0.018 | s.e. = 0.032 | s.e. = 0.010 | s.e. = 0.088 | s.e. = 0.013 |
| \( \chi^2 = 0.170 \) | \( \chi^2 = 0.083 \) | \( \chi^2 = 0.093 \) |

Table 9: For the first step: maximum likelihood estimates of the parameters \( \theta_i \) and \( \alpha_i \) for each cluster \( i = 1, 2, 3 \) for the combination of copula and marginal distribution that minimizes the chi-squared distance between the empirical distribution of the number of components of the normalized
vector exceeding the threshold 1, and the theoretical one. For the second step: jointly maximum likelihood estimates of the parameters \( \theta_i \) and \( \alpha_i \) based on the choices for the copula and marginal distribution of the first step. The standard errors have been computed by simulations.

Figure 2: Left: Empirical distributions of the number of exceeding components with their theoretical distributions that minimizes the chi-squared distance for the three clusters. Right: Empirical distributions of the number of exceeding components and theoretical distributions with parameters derived from the second step of the estimation procedure.
We observe excellent matches between empirical and theoretical distributions of the number of exceeding components for the Cluster 2 and 3 and a very good match for Cluster 1. This should validate the use of an homogeneous “clustered” max-stable model for these data. However the estimates of the parameters obtained after the second step differ significantly from the estimates of the parameters of the first step for Cluster 1 and 2. Moreover the chi-squared distances for these two clusters increases in an important way after the second step.

Since parameters of Cluster 1 and 3 are quite close for the first and the second steps, we decided to gather these two clusters into a single cluster 1∪3. We then estimated the parameters of the copulas and of the distributions for the two new clusters, separately as well as jointly. The values are given in Table 10.

|                  | Cluster 1∪3 (size=42) | Cluster 2 (size=58) | Cluster 1∪2∪3 (size=100) |
|------------------|-----------------------|---------------------|---------------------------|
| Copula           | Clayon                | Clayon              | Clayon                    |
| Marg. dist.      | Weibull               | Weibull             | Weibull                   |
| 1<sup>st</sup> step |                       |                     |                           |
| θ<sub>1∪3</sub>  | 0.587                 | 1.034               | 0.949                     |
| α<sub>1∪3</sub>  | 1.080                 | 0.821               | 0.825                     |
| s.e.             | 0.040                 | 0.056               | 0.038                     |
| χ<sup>2</sup>    | 0.121                 | 0.079               | 0.138                     |
| 2<sup>nd</sup> step |                       |                     |                           |
| θ<sub>1∪3</sub>  | 0.955                 | 0.950               | 0.950                     |
| α<sub>1∪3</sub>  | 0.912                 | 0.817               |                           |
| s.e.             | 0.045                 | 0.056               | 0.039                     |
| χ<sup>2</sup>    | 0.208                 | 0.079               | 0.083                     |

Table 10: For the first step: maximum likelihood estimates of the parameters θ<sub>i</sub> and α<sub>i</sub> for clusters <i>i</i> =1∪3, 2 and 1∪2∪3. For the second step: jointly maximum likelihood estimates of the parameters θ<sub>i</sub> and α<sub>i</sub> for clusters <i>i</i> =1∪3 and 2. The standard errors have been computed by simulations.

The estimates of the parameters of the new cluster 1∪3 for the first step are roughly the same as those obtained when it was splitted into two clusters 1 and 3. The estimates of the parameters obtained after the second step differ again from the estimates of the parameters of the first step and provide values that are close to the values of Cluster 2. We therefore decided to consider only one cluster 1∪2∪3 for these data. Figure 3 shows that there is a very good match between the empirical distributions of the number of exceeding components and the theoretical distribution. This is confirmed by the Chi-squared distance which is small for distributions with 100 possible values.
Figure 3: Left: Empirical distributions of the number of exceeding components with their theoretical distributions that minimizes the chi-squared distance for clusters $1 \cup 3$, $2$ and $1 \cup 2 \cup 3$. Right: Empirical distributions of the number of exceeding components and theoretical distributions with parameters derived from the second step for clusters $1 \cup 3$ and $2$. 

Cluster #1U3 (C: Clayton M: Weibull)

Cluster #1U3 (C: Clayton M: Weibull)

Cluster #2 (C: Clayton M: Weibull)

Cluster #2 (C: Clayton M: Weibull)

Cluster #1U2U3 (C: Clayton M: Weibull)
5 Conclusion and discussion

We studied the problem of estimating high dimensional extreme value distributions. We considered multivariate extreme value distributions for which the tail-dependence function has a regular Schlather’s representation. Such distributions appear naturally for the finite-dimensional distributions of the max-stable processes which are used to study extreme value phenomena for which spatial patterns can be discerned. But they also offer rich variety of ways for modeling high dimensional distribution by using e.g. copulas toolboxes. For these distributions, we provided quasi-explicit analytical expressions of the full likelihoods for random samples with max-stable distributions and for random samples in the max-domain of attraction of a max-stable distribution. Monte Carlo experiments show that our likelihood-based methods of inference are very efficient. Finally our real data application on financial data has shown that homogeneous clustered max-stable models can be successfully applied to estimate the high-dimensional dependence structure of extreme negative returns. We are convinced that this family of models open new opportunities for the development of high dimensional extreme data analysis.

6 Appendix

6.1 An intermediate lemma

Lemma 7 Let \( \varphi : [0, \infty) \to \mathbb{R}^+ \) be such that \( \int_0^\infty \varphi^{-1}(\gamma) \, d\gamma < \infty \). If \( \Gamma \) is a random variable with a unit Pareto distribution, then

\[
\lim_{n \to \infty} n \mathbb{E} \left[ \varphi^{-1}(\Gamma) \right] = \int_0^\infty \varphi^{-1}(\gamma) \, d\gamma.
\]

Proof: We have

\[
n \mathbb{E} \left[ \varphi^{-1}(\Gamma) \right] = n \int_1^\infty \varphi^{-1}(\gamma) \gamma^{-2} \, d\gamma = \int_0^\infty \varphi(v-1) \, dv \to \int_0^\infty \varphi(v-1) \, dv. \quad \Box
\]

If \( Y = \Gamma U \) where \( U \) is a random variable independent of \( \Gamma \) such that \( \mathbb{E}[|U|] < \infty \), then we deduce from the previous Lemma that

\[
\lim_{n \to \infty} n \mathbb{P}(Y > yn) = y^{-1} \mathbb{E}[U^+]
\]

since \( n \mathbb{P}(Y > yn) = n \mathbb{E}[\hat{F}_U(yn/\Gamma)] \) where \( \hat{F}_U(u) = \mathbb{P}(U > u) \).

6.2 Proof of Proposition 1

Let us assume that \( Y = (Y_1, \ldots, Y_m) \) is such that \( Y_j = \Gamma U_j \) for \( j = 1, \ldots, m \) where \( \Gamma \) is a random variable with unit Pareto distribution and independent of \( U = (U_1, \ldots, U_m) \). Then \( Y \) belongs to the max-domain of attraction of \( G_\ast \). Indeed we have

\[
\log \mathbb{P}(M_n \leq nz) = n \log \mathbb{P} \left( \Gamma \max_{j=1,\ldots,m} z_j^{-1} U_j \leq n \right)
\]

and it follows from Lemma 7 that

\[
\lim_{n \to \infty} \log \mathbb{P}(M_n \leq nz) = - \lim_{n \to \infty} n \mathbb{P} \left( \Gamma \max_{j=1,\ldots,m} z_j^{-1} U_j > n \right) = - \mathbb{E} \left[ \max_{j=1,\ldots,m} z_j^{-1} U_j^+ \right].
\]
Let $F_Y$ be the multivariate distribution function of $Y$, $f_n(v) = v^n$ for $v \in [0, 1]$ and $g_n(z) = F_Y(nz)$. We have
\[
\Pr(M_n \leq nz) = F^n_Y(nz) = f_n(g_n(z)).
\]
The density function of $n^{-1}M_n$ is given by $\partial^m F^n_Y(nz)/\partial z_1 \ldots \partial z_m$ and it follows by de Faà di Bruno’s formula that this density function is equal to
\[
\sum_{\pi \in \Pi} f^{(|\pi|)}_n(g_n(z)) \prod_{B \in \pi} \frac{\partial^{|B|}g_n(z)}{\prod_{j \in B} \partial z_j}.
\]
Note that
\[
\frac{\partial^{|B|}g_n(z)}{\prod_{j \in B} \partial z_j} = \frac{\partial^{|B|}F_Y(nz)}{\prod_{j \in B} \partial z_j} = \mathbb{E} \left[ \left( \frac{n^{|B|}}{\Gamma} \right) \prod_{j \in B} \frac{\partial^{|B|}F_U(nz/\Gamma)}{\partial z_j} \right]
\]
and since
\[
\frac{\partial^{|B|}F_U(nz/\gamma)}{\prod_{j \in B} \partial z_j} = \Pr(U_{B^c} \leq nz_{B^c}/\gamma | U_B = nz_B/\gamma) f_U(nz_B/\gamma),
\]
we deduce that
\[
\frac{\partial^m F^n_Y(nz)}{\partial z_1 \ldots \partial z_m} = \sum_{\pi \in \Pi} n^{n-|\pi|} \frac{(n-|\pi|+1)}{n!} F_Y^{n-|\pi|}(nz) \prod_{B \in \pi} \mathbb{E} \left[ n \left( \frac{n^{|B|}}{\Gamma} \right) \Pr \left( U_{B^c} \leq \frac{nz_{B^c}}{\Gamma} \big| U_B = \frac{nz_B}{\Gamma} \right) f_U \left( \frac{nz_B}{\Gamma} \right) \right].
\]
By Lemma 7, we derive that
\[
\lim_{n \to \infty} n \mathbb{E} \left[ \left( \frac{n^{|B|}}{\Gamma} \right) \Pr \left( U_{B^c} \leq \frac{nz_{B^c}}{\Gamma} \big| U_B = \frac{nz_B}{\Gamma} \right) f_U \left( \frac{nz_B}{\Gamma} \right) \right] = \mu(B; z).
\]
Moreover since $Y$ belongs to the max-domain of attraction of $G_*$, we have
\[
\lim_{n \to \infty} F_Y^{n-|\pi|}(nz) = \exp(-V_*(z))
\]
where
\[
V_*(z) = \mathbb{E} \left[ \max_{j=1,\ldots,m} z_j^{-1} U_j^+ \right] = \mathbb{E} \left[ \left( \max_{j=1,\ldots,m} z_j^{-1} U_j \right)^+ \right] = \int_0^\infty P(M_z > t) dt
\]
with $M_z = \max_{j=1,\ldots,m} z_j^{-1} U_j$. Since $V_*(z)$ is an homogeneous function of order $-1$, Euler’s homogeneous theorem implies that
\[
V_*(z) = - \sum_{i=1}^m z_i \frac{\partial V_*(z)}{\partial z_i} = \sum_{i=1}^m z_i \int_0^\infty \frac{\partial}{\partial z_i} \Pr(U \leq tz) dt = \sum_{i=1}^m z_i \mu(\{i\}; z)
\]
and the result follows.

### 6.3 Proof of Proposition 2

Let us assume that $\tilde{Y} = (\tilde{Y}_1, \ldots, \tilde{Y}_m)$ is such that $\tilde{Y}_j = \Gamma U_j$ for $j = 1, \ldots, m$ where $\Gamma$ is a random variable with unit Pareto distribution and independent of $U = (U_1, \ldots, U_m)$. Then $\tilde{Y}$ belongs to the max-domain of attraction of $G_*$ (see the Proof of Proposition 1). We may therefore assume
without loss of generality that $X_t = t^{-1} \tilde{Y} \vee e$ instead of $t^{-1} Y \vee e$ to establish the weak convergence of $X_t^*$ to $X^*$. We have, for $B \in \mathcal{P}$ and $x \in \mathcal{A}_B$,

$$\Pr (X_t^* \in (e, x]) = \Pr (X_{B,t} \leq x_B | X_{B,t} > e_B, X_{B^c,t} = e_{B^c})$$

and

$$\lim_{t \to \infty} \Pr (X_{B,t} \leq x_B | X_{B,t} > e_B, X_{B^c,t} = e_{B^c}) = \lim_{t \to \infty} \frac{t \Pr (e_B < X_{B,t} \leq x_B, X_{B^c,t} = e_{B^c})}{t \Pr (X_{B,t} > e_B, X_{B^c,t} = e_{B^c})} = \frac{\int_0^\infty \Pr (\gamma e_B < U_B \leq \gamma x_B, U_{B^c} \leq \gamma e_{B^c}) d\gamma}{\int_0^\infty \Pr (U_B > \gamma e_B, U_{B^c} \leq \gamma e_{B^c}) d\gamma}$$

by using again Lemma 7. By differentiating with respect to the components of $x_B$, we derive that

$$f_{X^*(A_B)} (x_B) = \frac{\mu (B; (x_B, e_{B^c}))}{V^*_B (e)}.$$

### 6.4 Proof of Proposition 4

The SMLE $\hat{\theta}_{n,S}^{(1)}$ is solution of

$$\frac{1}{n} \sum_{i=1}^n \nabla_{\theta} \log \ell_{S1}(\hat{\theta}_{n,S}^{(1)}; z_i) = 0$$

where $\log \ell_{1S} (\theta; z) = -\sum_{i=1}^n z_i \mu_S (\theta; \{1\}, z) + \log (\delta_S (\theta; z))$ with

$$\mu_S (\theta; B, z) = \frac{1}{S} \sum_{s=1}^S a (V_s; \theta, B, z), \quad \delta_S (\theta; z) = \sum_{\pi \in \Pi} \chi_S (\theta; \pi, z), \quad \chi_S (\theta; \pi, z) = \prod_{B \in \pi} \mu_S (\theta; B, z).$$

An expansion around $\theta_0$ and a multiplication by $n^{1/2}$ give

$$0 = \frac{1}{n^{1/2}} \sum_{i=1}^n \nabla_{\theta} \log \ell_{S1}(\theta_0; z_i) + \left( \frac{1}{n} \sum_{i=1}^n \nabla^2_{\theta} \log \ell_{S1}(\theta_0; z_i) \right) n^{1/2} (\hat{\theta}_{n,S}^{(1)} - \theta_0) + O_{Pr} \left( n^{1/2} (\hat{\theta}_{n,S}^{(1)} - \theta_0)^2 \right).$$

By the central limit theorem, the Lindeberg theorem (assuming without loss of generality that $S = S_n$ is a function of $n$) and the delta method, we derive that

$$\frac{1}{n} \sum_{i=1}^n \nabla^2_{\theta} \log \ell_{S1}(\theta_0; z_i) = \frac{1}{n} \sum_{i=1}^n \nabla^2_{\theta} \log \ell_1(\theta_0; z_i) + \frac{1}{n} \sum_{i=1}^n \left( \nabla^2_{\theta} \log \ell_{S1}(\theta_0; z_i) - \mathbb{E} \left[ \nabla^2_{\theta} \log \ell_{S1}(\theta_0; z_i) \right] \right)$$

$$+ \frac{1}{n} \sum_{i=1}^n \left( \mathbb{E} \left[ \nabla^2_{\theta} \log \ell_{S1}(\theta_0; z_i) \right] - \nabla^2_{\theta} \log \ell_1(\theta_0; z_i) \right)$$

$$= I_1 (\theta_0) + O_{Pr} (n^{-1/2}) + O_{Pr} (S^{-1}).$$

And it follows that

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^n \nabla_{\theta} \log \ell_{S1}(\theta_0; z_i) + I_1 (\theta_0) n^{1/2} (\hat{\theta}_{n,S}^{(1)} - \theta_0) + O_{Pr} \left( n^{1/2} (\hat{\theta}_{n,S}^{(1)} - \theta_0)^2 \right)$$

$$+ O_{Pr} \left( (n/S)^{1/2} S^{-1/2} (\hat{\theta}_{n,S}^{(1)} - \theta_0) \right).$$

27
Let us now study the difference

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_\theta \log \ell_{S_1}(\theta_0; z_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_\theta \log \ell_1(\theta_0; z_i).
\]

Recall that

\[
\nabla_\theta \log \ell_{1S}(\theta_0; z) = -\sum_{i=1}^{m} z_i \nabla_\theta \mu_{S_1}(\theta_0; \{l\}, z) + \frac{\nabla_\theta \delta_{S}(\theta_0; z)}{\delta_{S}(\theta_0; z)}
\]

with

\[
\nabla_\theta \mu_{S}(\theta_0; \{l\}, z) = \sum_{\pi \in \Pi} \chi_{S}(\theta_0; \pi, z) \sum_{B \in \pi} \nabla_\theta \log \mu_{S}(\theta_0; B, z).
\]

First it holds that

\[
\nabla_\theta \mu_{S}(\theta_0; \{l\}, z) - \nabla_\theta \mu(\theta_0; \{l\}, z) = \int \nabla_\theta a(v; \theta_0, \{l\}, z) ((dF_{V,S}(v) - dF_V(v))
\]

where \(F_{V,S}(v) = S^{-1} \sum_{s=1}^{S} \mathbb{I}_{\{V_s \leq v\}}.\)

Second, by the law of large numbers, we deduce that

\[
\frac{\nabla_\theta \delta_{S}(\theta_0; z)}{\delta_{S}(\theta_0; z)} \frac{\nabla_\theta \delta_{S}(\theta_0; z)}{\delta_{S}(\theta_0; z)} = -\frac{\nabla_\theta \delta_{S}(\theta_0; z)}{\delta_{S}(\theta_0; z)} \sum_{\pi \in \Pi} \chi(\theta_0; \pi, z) \sum_{B \in \pi} (\log \mu_{S}(\theta_0; B, z) - \log \mu(\theta_0; B, z))
\]

\[
+ \frac{1}{\delta(\theta_0; z)} \sum_{\pi \in \Pi} \left( \sum_{B' \in \pi} \nabla_\theta \log \mu(\theta_0; B', z) \right) \chi(\theta_0; \pi, z) \sum_{B \in \pi} (\log \mu_{S}(\theta_0; B, z) - \log \mu(\theta_0; B, z))
\]

\[
+ \frac{1}{\delta(\theta_0; z)} \sum_{\pi \in \Pi} \chi(\theta_0; \pi, z) \sum_{B \in \pi} (\nabla_\theta \log \mu_{S}(\theta_0; B, z) - \nabla_\theta \log \mu(\theta_0; B, z)) + O_P(S^{-1}).
\]

Now note that, by using again the law of large numbers,

\[
\log \mu_{S}(\theta_0; B, z) - \log \mu(\theta_0; B, z) = \frac{1}{\mu(\theta_0; B, z)} \int a(v; \theta_0, B, z) ((dF_{V,S}(v) - dF_V(v))) + O_P(S^{-1})
\]

and

\[
\nabla_\theta \log \mu_{S}(\theta_0; B, z) - \nabla_\theta \log \mu(\theta_0; B, z) = \frac{1}{\mu(\theta_0; B, z)} \int \nabla_\theta a(v; \theta_0, B, z) ((dF_{V,S}(v) - dF_V(v)))
\]

\[
- \frac{\nabla_\theta \log \mu(\theta_0; B, z)}{\mu(\theta_0; B, z)} \int a(v; \theta_0, B, z) ((dF_{V,S}(v) - dF_V(v))) + O_P(S^{-1}).
\]

We therefore deduce that

\[
\nabla_\theta \log \ell_{1S}(\theta_0; z) = \nabla_\theta \log \ell_1(\theta_0; z) + \int \psi(v; \theta_0, \pi, z) ((dF_{V,S}(v) - dF_V(v))) + O_P(S^{-1}).
\]
It turns that

\[
0 = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \nabla_{\theta} \log \ell_1(\theta_0; z_i) + I_1(\theta_0) n^{1/2} (\hat{\theta}_{n,S}^{(1)} - \theta_0) \\
+ \left( \frac{n}{S} \right)^{1/2} \sum_{i=1}^{n} \int \psi(v; \theta_0, \pi, z_i) \left( (dF_{V,S}(v) - dF_V(v)) \right) \\
+ O_P \left( n^{1/2} (\hat{\theta}_{n,S}^{(1)} - \theta_0)^2 \right) + O_P \left( (n/S)^{1/2} S^{-1/2} \right).
\]

Observe that, by the central limit theorem,

\[
\frac{1}{n^{1/2}} \sum_{i=1}^{n} \nabla_{\theta} \log \ell_1(\theta_0; z_i) \overset{d}{\to} \mathcal{N}(0, I_1(\theta_0))
\]

and, by the Lindeberg theorem (assuming without loss of generality that \( n = n_S \) is a function of \( S \)),

\[
\frac{\sqrt{S}}{n} \sum_{i=1}^{n} \int \psi(v; \theta_0, \pi, z_i) \left( (dF_{V,S}(v) - dF_V(v)) \right) \\
= \frac{1}{\sqrt{S}} \sum_{s=1}^{S} \left[ \frac{1}{n} \sum_{i=1}^{n} \left( \psi((v_s; \theta_0, \pi, z_i) - \mathbb{E}[\psi(V; \theta_0, \pi, z_i)]) \right) \right] \\
\overset{d}{\to} \mathcal{N}(0, \mathbb{V} (\mathbb{E}[\psi(V; \theta_0, \pi, Z) | V])).
\]

Therefore

- if \( n/S \) tends to zero, then

\[
\sqrt{n} (\hat{\theta}_{n,S}^{(1)} - \theta_0) \overset{d}{\to} \mathcal{N}(0, I_1^{-1}(\theta_0));
\]

- if \( n/S \) tends to infinity, then

\[
\sqrt{S} (\hat{\theta}_{n,S}^{(1)} - \theta_0) \overset{d}{\to} \mathcal{N}(0, I_1^{-1}(\theta_0) \Sigma(\theta_0) I_1^{-1}(\theta_0));
\]

- if \( n/S \) tends to \( \varphi \), then

\[
\sqrt{n} (\hat{\theta}_{n,S}^{(1)} - \theta_0) \overset{d}{\to} \mathcal{N}(0, I_1^{-1}(\theta_0) (I_1(\theta_0) + \varphi \Sigma(\theta_0)) I_1^{-1}(\theta_0)).
\]

6.5 Proof of Proposition 5

The maximum likelihood estimator \( \hat{\theta}_k \) satisfies the first order condition (3.8). For any \( B \in \mathcal{P} \) and \( A_B \subset (e_B, \infty] \in \mathcal{A}_B \), let us define the empirical measure \( N_{B,k} \) by

\[
N_{B,k}(A_B) = \sum_{i=1}^{n} \mathbb{I}\{X_{i,B,n/k} \in A_B, X_{i,B^c,n/k} = e_{B^c}\}.
\]

By using e.g. Proposition 2.1 in [14], \( k^{-1} N_{B,k} \) converges in the vague topology on \( M_{+}(([e_B, \infty]), \mathbb{P}_{B}) \), the space of positive Radon measures on \( [e_B, \infty] \), and therefore we have for \( x_B \in (e_B, \infty] \)

\[
\frac{1}{k} N_{B,k} ((x_B, \infty]) \overset{Pr}{\to} \lim_{t \to \infty} t \mathbb{P}(\mathbb{X}_{B,t} > x_B, \mathbb{X}_{B^c,t} = e_{B^c}).
\]
By Lemma 7, this limit is equal to
\[ \nu_B((x_B, \infty)) = \int_0^\infty \Pr(U_B > \gamma x_B, U_{B^c} \leq \gamma e_{B^c}) \, d\gamma. \]

By using Condition C3 and since \( k = o\left(n^{2\alpha/(1+2\alpha)}\right) \), it holds that
\[ \lim_{k \to \infty} \sqrt{k} \left( \nu_B((x_B, \infty)) - \frac{n}{k} \Pr(X_{B,n/k} > x_B, X_{B^c,n/k} = e_{B^c}) \right) = 0. \]

By Proposition 3.1 in [14], there exist mean Gaussian random fields \((W_B(x_B), x_B \in [e_B, \infty))\) with covariance functions
\[ \text{Cov}(W_B(x_B), W_B(y_B)) = \nu_B((x_B, \infty) \cap (y_B, \infty)) \]
such that the random fields
\[ W_{B,n}(x_B) = \sqrt{k} \left( \frac{1}{k} N_{B,k}((x_B, \infty)) - \nu_B((x_B, \infty)) \right), \quad x_B \in [e_B, \infty), \]
converge weakly to \( W_B \) in the space of cadlag functions defined on \([e_B, \infty), D([e_B, \infty))\), equipped with Skorohod’s topology:
\[ W_{B,n}(\cdot) \Rightarrow W_B(\cdot). \quad (6.10) \]

Moreover there exist versions of \( W_B \) which are sample continuous. Note that \( W_B \) is independent of \( W_{B'} \) for \( B \neq B' \) because the sets \( A_B \) and \( A_{B'} \) are disjoint.

Now note that the first order condition is equivalent to
\[ 0 = \frac{1}{k} \sum_{B \in \mathcal{P}} \int_{A_B} \nabla_\theta \log \ell_{x^*}(\hat{\theta}_k; x) N_{B,k}(dx_B) \]
where
\[ \nabla_\theta \log \ell_{x^*}(\theta; x) \mid_{x \in \mathcal{A}_B} = \frac{\nabla_\theta \mu(\theta; B, (x_B, e_{B^c}))}{\mu(\theta; B, (x_B, e_{B^c}))} - \frac{\nabla_\theta V_{B}^*(\theta; e)}{V_{B}^*(\theta; e)}. \]

An expansion around \( \theta_0 \) gives
\[ 0 = \frac{1}{k} \sum_{B \in \mathcal{P}} \int_{A_B} \nabla_\theta \log \ell_{x^*}(\theta_0; x) N_{B,k}(dx_B) \]
\[ + \left( \hat{\theta}_k - \theta_0 \right) \frac{1}{k} \sum_{B \in \mathcal{P}} \int_{A_B} \nabla^2_\theta \log \ell_{x^*}(\theta_0; x) N_{B,k}(dx_B) + O_P\left((\hat{\theta}_k - \theta_0)^2\right). \]

First note that
\[ \frac{1}{k} \sum_{B \in \mathcal{P}} \int_{A_B} \nabla^2_\theta \log \ell_{x^*}(\theta_0; x) N_{B,k}(dx_B) \xrightarrow{P} \sum_{B \in \mathcal{P}} \int_{A_B} \nabla^2_\theta \log \ell_{x^*}(\theta_0; x) \nu_B(dx_B) \]
and that this limit is equal to
\[ V_*(\theta_0; e) \sum_{B \in \mathcal{P}} p_B(\theta_0; e) \int_{A_B} \nabla^2_\theta \log \ell_{x^*}(\theta_0; x) \frac{\mu(\theta_0; B, (x_B, e_{B^c}))}{V_{B}^*(\theta_0; e)} \, dx_B \]
\[ = V_*(\theta_0; e) \mathbb{E}\left[ \nabla^2_\theta \log \ell_{x^*}(\theta_0; X^*) \right]. \]
Now observe that
\[
\int_{A_B} \nabla_\theta \log \ell_{\mathbf{X}^*} (\theta_0; \mathbf{x}) \, d\nu_B ((\mathbf{x}_B, \infty)) = \int_{A_B} \nabla_\theta \log \ell_{\mathbf{X}^*} (\theta_0; \mathbf{x}) \mu (\theta_0; B, (\mathbf{x}_B, \mathbf{e}_{B^c})) \, d\mathbf{x}_B = 0
\]
and derive by (6.10) that
\[
\frac{1}{\sqrt{k}} \int_{A_B} \nabla_\theta \log \ell_{\mathbf{X}^*} (\theta_0; \mathbf{x}) \, N_{B,k} (d\mathbf{x}_B) \xrightarrow{d} \mathcal{N} \left( 0, \int_{A_B} \nabla_\theta \log \ell_{\mathbf{X}^*} (\theta_0; \mathbf{x}) \nabla_\theta \log \ell_{\mathbf{X}^*} (\theta_0; \mathbf{x}^\prime) \mu (\theta_0; B, (\mathbf{x}_B, \mathbf{e}_{B^c})) \, d\mathbf{x}_B \right).
\]
The asymptotic variance may be rewritten as
\[
V_\ast (\theta; \mathbf{e}) \mathbb{E} \left[ \nabla_\theta \log \ell_{\mathbf{X}^*} (\theta; \mathbf{X}^*) \nabla_\theta \log \ell_{\mathbf{X}^*} (\theta; \mathbf{X}^*) \right]
\]
by using Proposition 2. Since
\[
I_{\mathbf{X}^*} (\theta_0) = \mathbb{E} \left[ \nabla_\theta \log \ell_{\mathbf{X}^*} (\theta; \mathbf{X}^*) \nabla_\theta \log \ell_{\mathbf{X}^*} (\theta; \mathbf{X}^*) \right] = -\mathbb{E} \left[ \nabla_\theta^2 \log \ell_{\mathbf{X}^*} (\theta_0; \mathbf{X}^*) \right],
\]
we finally have
\[
\sqrt{k} (\hat{\theta}_k - \theta_0) \xrightarrow{d} \mathcal{N} \left( 0, V_\ast^{-1} (\theta_0; \mathbf{e}) I_{\mathbf{X}^*}^{-1} (\theta_0) \right).
\]

6.6 Proof of Proposition 6

Let
\[
\mathbf{X}_{i,n/k} = \left( \frac{\mathbf{Y}_i}{\hat{\mathbf{Y}} (n/k)} \right)^\alpha \vee \mathbf{e}
\]
where \( \mathbf{U}_{\mathbf{Y}} (n/k) = (U_{Y_1} (n/k), \ldots, U_{Y_m} (n/k))^\prime \) with \( U_{Y_j} (n/k) = \hat{F}_{Y_j} (1 - k/n) \).

For \( \mathbf{X}_{i,n/k} \in \mathcal{A}_B \),
\[
\nabla_{\theta} \log \ell_{\mathbf{X}^*} (\hat{\theta}_j; \mathbf{X}_{i,n/k}) - \nabla_{\theta} \log \ell_{\mathbf{X}^*} (\hat{\theta}_j; \mathbf{X}_{i,n/k}) = -\sum_{j \in B} \alpha_j \nabla_{\theta}^2 \log \ell_{\mathbf{X}^*} (\hat{\theta}_j; \mathbf{X}_{i,n/k}) \mathbf{X}_{i,j,n/k} \left[ \left( \frac{\hat{U}_{Y_j} (n/k) - U_{Y_j} (n/k)}{U_{Y_j} (n/k)} \right) \right] + \left( \frac{1}{\alpha_{j,n}} - \frac{1}{\alpha_j} \right) \log \mathbf{X}_{i,j,n/k} + o_{\mathbb{P}_1} (k^{-1/2}).
\]

Let, for \( x_j > 1 \),
\[
N_k (x_j) = \sum_{B \in \mathcal{P}, B \cap \{j\} \neq \emptyset} N_{B,k} ((1, \ldots, 1, x_j, 1, \ldots, 1)) = \sum_{i=1}^n \mathbb{I} \{ X_{i,j,n/k} > x_j \}
\]
and define
\[
W_{j,n} (x_j) = \sqrt{k} \left( \frac{1}{\alpha_{j,n}} - \frac{1}{\alpha_j} \right) \left( \frac{1}{k} \right) N_k (x_j) - x_j^{-1}
\]
As for the proof Proposition 5, it may be shown that \( W_{j,n} (\cdot) \Longrightarrow W_j (\cdot) \), and by using the same arguments as in [14] we get
\[
\sqrt{k} \left( \frac{1}{\alpha_{j,n}} - \frac{1}{\alpha_j} \right) \rightarrow \int_1^\infty W_j (x_j^{\alpha_j}) \frac{dx_j}{x_j} - \frac{1}{\alpha_j} W_j (1)
\]
\[
\sqrt{k} \left( \frac{\hat{U}_{Y_j} (n/k) - U_{Y_j} (n/k)}{U_{Y_j} (n/k)} \right) \rightarrow \frac{1}{\alpha_j} W_j (1).
\]
The maximum likelihood estimator (MLE) satisfies the condition

\[
0 = \frac{1}{k^{1/2}} \sum_{i \in N_k} \nabla_\theta \log \ell_{\mathbf{x}} \left( \hat{\theta}_k; \mathbf{x}_{i,n/k} \right)
\]

\[
= \frac{1}{k^{1/2}} \sum_{i \in N_k} \nabla_\theta \log \ell_{\mathbf{x}} \left( \theta_0; \mathbf{x}_{i,n/k} \right)
+ \frac{1}{k} \sum_{i \in N_k} \sum_{B \in \mathcal{P}} \sum_{\mathbf{x}_{i,n/k} \in A_B} \left( \nabla_\theta \log \ell_{\mathbf{x}} \left( \theta_k; \mathbf{x}_{i,n/k} \right) - \nabla_\theta \log \ell_{\mathbf{x}} \left( \hat{\theta}_k; \mathbf{x}_{i,n/k} \right) \right).
\]

An expansion of the first term around \( \theta_0 \) gives

\[
\frac{1}{k^{1/2}} \sum_{i \in N_k} \nabla_\theta \log \ell_{\mathbf{x}} \left( \theta_0; \mathbf{x}_{i,n/k} \right) = \frac{1}{k^{1/2}} \sum_{i \in N_k} \nabla_\theta \log \ell_{\mathbf{x}} \left( \theta_0; \mathbf{x}_{i,n/k} \right) + \frac{1}{k} \sum_{i \in N_k} \nabla_\theta^2 \log \ell_{\mathbf{x}} \left( \theta_0; \mathbf{x}_{i,n/k} \right) k^{1/2} \left( \theta_k - \theta_0 \right) + O_P \left( k^{1/2} (\hat{\theta}_k - \theta_0)^2 \right)
\]

Note that i) 

\[
\frac{1}{k} \sum_{i \in N_k} \nabla_\theta^2 \log \ell_{\mathbf{x}} \left( \theta_0; \mathbf{x}_{i,n/k} \right) \overset{P}{\to} V_*(\theta_0; \mathbf{e}) I_{\mathbf{x}}(\theta_0)
\]

ii) 

\[
\frac{1}{k^{1/2}} \sum_{i \in N_k} \nabla_\theta \log \ell_{\mathbf{x}} \left( \theta_0; \mathbf{x}_{i,n/k} \right) \overset{d}{\to} \sum_{B \in \mathcal{P}} \int_{A_B} \nabla_\theta \log \ell_{\mathbf{x}} \left( \theta_0; \mathbf{x} \right) W_B \left( d\mathbf{x}_B \right).
\]

Now

\[
\frac{1}{k^{1/2}} \sum_{B \in \mathcal{P}} \sum_{\mathbf{x}_{i,n/k} \in A_B} \left( \nabla_\theta \log \ell_{\mathbf{x}} \left( \theta_0; \mathbf{x}_{i,n/k} \right) - \nabla_\theta \log \ell_{\mathbf{x}} \left( \hat{\theta}_k; \mathbf{x}_{i,n/k} \right) \right)
\]

\[
= - \sum_{B \in \mathcal{P}} \sum_{j \in B} \alpha_j \left( \frac{\hat{U}_{Y_j}(n/k) - U_{Y_j}(n/k)}{U_{Y_j}(n/k)} \right) \left( \sum_{\mathbf{x}_{i,j,n/k} \in A_B} \nabla_\theta^2 \log \ell_{\mathbf{x}} \left( \hat{\theta}_k; \mathbf{x}_{i,n/k} \right) \mathbf{x}_{i,j,n/k} \right)
\]

\[
- \sum_{B \in \mathcal{P}} \sum_{j \in B} \alpha_j \left( \frac{1}{\alpha_{j,k,n}} - \frac{1}{\alpha_j} \right) \left( \sum_{\mathbf{x}_{i,j,n/k} \in A_B} \nabla_\theta^2 \log \ell_{\mathbf{x}} \left( \hat{\theta}_k; \mathbf{x}_{i,n/k} \right) \mathbf{x}_{i,j,n/k} \right) + o_P(1).
\]

Note also that

\[
\sum_{B \in \mathcal{P}} \sum_{j \in B} \alpha_j \left( \frac{\hat{U}_{Y_j}(n/k) - U_{Y_j}(n/k)}{U_{Y_j}(n/k)} \right) \left( \sum_{\mathbf{x}_{i,j,n/k} \in A_B} \nabla_\theta^2 \log \ell_{\mathbf{x}} \left( \hat{\theta}_k; \mathbf{x}_{i,n/k} \right) \mathbf{x}_{i,j,n/k} \right)
\]

\[
= \sum_{j=1}^{m} \alpha_j \left( \frac{\hat{U}_{Y_j}(n/k) - U_{Y_j}(n/k)}{U_{Y_j}(n/k)} \right) \sum_{B \in \mathcal{P}, B \neq \emptyset} \sum_{\mathbf{x}_{i,j,n/k} \in A_B} \nabla_\theta^2 \log \ell_{\mathbf{x}} \left( \hat{\theta}_k; \mathbf{x}_{i,n/k} \right) \mathbf{x}_{i,j,n/k}
\]

\[
= \sum_{j=1}^{m} \alpha_j \left( \frac{1}{\alpha_{j,k,n}} - \frac{1}{\alpha_j} \right) \left( \sum_{\mathbf{x}_{i,j,n/k} \in A_B} \sum_{\mathbf{x}_{i,j,n/k} \in A_B} \nabla_\theta^2 \log \ell_{\mathbf{x}} \left( \hat{\theta}_k; \mathbf{x}_{i,n/k} \right) \mathbf{x}_{i,j,n/k} \mathbf{x}_{i,j,n,k} \right)
\]

\[
= \sum_{j=1}^{m} \alpha_j \left( \frac{1}{\alpha_{j,k,n}} - \frac{1}{\alpha_j} \right) \sum_{B \in \mathcal{P}, B \neq \emptyset} \sum_{\mathbf{x}_{i,j,n/k} \in A_B} \nabla_\theta^2 \log \ell_{\mathbf{x}} \left( \hat{\theta}_k; \mathbf{x}_{i,n/k} \right) \mathbf{x}_{i,j,n/k} \mathbf{x}_{i,j,n,k}.
\]
Moreover
\[ \frac{1}{k} \sum_{B \in \mathcal{P}, B \cap \{j\} \neq \emptyset} \sum_{X_{i,j,n/k} \in A_B} \nabla^2_{\theta x_j} \log \ell (\theta_k; X_{i,n/k}) X_{i,j,n/k} \]
\[ = \frac{N_k^{(j)}}{k} \frac{1}{N_k^{(j)}} \sum_{B \in \mathcal{P}, B \cap \{j\} \neq \emptyset} \sum_{X_{i,j,n/k} \in A_B} \nabla^2_{\theta x_j} \log \ell (\theta_k; X_{i,n/k}) X_{i,j,n/k} \]
\[ \Pr \mathbb{E} \left[ \nabla^2_{\theta x_j} \log \ell (\theta_0; X_{i,n/k}) X^*_j \mid X^*_j > 1 \right] = \omega_j \]
and
\[ \frac{1}{k} \sum_{B \in \mathcal{P}, B \cap \{j\} \neq \emptyset} \sum_{X_{i,j,n/k} \in A_B} \nabla^2_{\theta x_j} \log \ell (\theta_k; X_{i,n/k}) X_{i,j,n/k} \]
\[ \Pr \mathbb{E} \left[ \nabla^2_{\theta x_j} \log \ell (\theta_0; X_{i,n/k}) X^*_j \log X^*_j \mid X^*_j > 1 \right] = \psi_j \cdot \]

It follows that
\[ \sqrt{k} \sum_{j=1}^{m} \alpha_j \left( \frac{\hat{U}_{Y_j} (n/k) - U_{Y,j} (n/k)}{U_{Y,j} (n/k)} \right) \frac{1}{k} \sum_{B \in \mathcal{P}, B \cap \{j\} \neq \emptyset} \sum_{X_{i,j,n/k} \in A_B} \nabla^2_{\theta x_j} \log \ell (\theta_k; X_{i,n/k}) X_{i,j,n/k} \]
\[ \to \sum_{j=1}^{m} \omega_j W_j (1) \]

and
\[ \sqrt{k} \sum_{j=1}^{m} \alpha_j \left( \frac{1}{\alpha_{j,k,n}} - \frac{1}{\alpha_j} \right) \frac{1}{k} \sum_{B \in \mathcal{P}, B \cap \{j\} \neq \emptyset} \sum_{X_{i,j,n/k} \in A_B} \nabla^2_{\theta x_j} \log \ell (\theta_k; X_{i,n/k}) X_{i,j,n/k} \log X_{i,j,n/k} \]
\[ \to \sum_{j=1}^{m} \alpha_j \psi_j \left( \int_1^{\infty} W_j \left( z_j^\alpha \right) \frac{dz_j}{z_j} - \frac{1}{\alpha_j} W_j (1) \right) . \]

This concludes the proof.

6.7 Hessians

i) Hessians of \( \log \ell_1 (\theta; x) \), \( \log \ell_2 (\theta; z) \) and \( \log \ell_3 (\theta; z) \):

\[ \nabla^2_{\theta} \log \ell_1 (\theta; z) = - \sum_{i=1}^{m} z_i \nabla^2_{\theta} \mu (\theta; \{i\}, z) + \sum_{\pi \in \Pi} \sum_{\pi' \in \Pi} \chi (\theta; \pi, z) \sum_{B \in \pi} \nabla^2_{\theta} \log \mu (\theta; B, z) \]
\[ + \sum_{\pi \in \Pi} \sum_{\pi' \in \Pi} \chi (\theta; \pi, z) \sum_{B \in \pi} \nabla_{\theta} \log \mu (\theta; B, z) \sum_{B' \in \pi} \nabla_{\theta} \log \mu (\theta; B', z)' \]
\[ - \left( \sum_{\pi \in \Pi} \sum_{\pi' \in \Pi} \chi (\theta; \pi, z) \sum_{B \in \pi} \nabla_{\theta} \log \mu (\theta; B, z) \right) \left( \sum_{\pi \in \Pi} \sum_{\pi' \in \Pi} \chi (\theta; \pi, z) \sum_{B \in \pi} \nabla_{\theta} \log \mu (\theta; B, z) \right)' \]

and

\[ \nabla^2_{\theta} \log \ell_2 (\theta; z) = \sum_{B \in \pi} |B| \nabla_{\theta} \log \ell_1 (\theta; z_B) , \quad \nabla^2_{\theta} \log \ell_3 (\theta; z) = \sum_{i < j} \nabla^2_{\theta} \log \ell_1 (\theta; z_{i,j}) \]
ii) Hessian of $\log \ell_{X^*}(\theta; x)$:

$$
\nabla_{\theta}^2 \log \ell_{X^*}(\theta; x) = \sum_{B \in \mathcal{P}} \left( \frac{\nabla_{\theta}^2 \mu(\theta; B, (x_B, e_{B^c}))}{\mu(\theta; B, (x_B, e_{B^c}))} - \frac{\nabla_{\theta} \mu(\theta; B, (x_B, e_{B^c})) \nabla_{\theta} \mu(\theta; B, (x_B, e_{B^c}))}{\mu^2(\theta; B, (x_B, e_{B^c}))} \right) \mathbb{I}_{\{x \in A_B\}} \\
- \left( \sum_{i=1}^m \nabla_{\theta} \mu(\theta; \{i\}, e) \sum_{j=1}^m \nabla_{\theta} \mu(\theta; \{j\}, e) \frac{\sum_{i=1}^m \nabla_{\theta} \mu(\theta; \{i\}, e)}{\sum_{i=1}^m \mu(\theta; \{i\}, e)^2} \right).
$$

References

[1] Bacro, J.N. and Gaetan, C. (2013) Estimation of spatial max-stable models using threshold exceedances. arXiv:1205.1107.

[2] Bernard, E., Naveau, P., Vrac, M. and Mestre, O. (2013) Clustering of maxima: Spatial dependencies among heavy rainfall in france. *Journal of Climate*, to appear.

[3] Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc. B*, 36, 192–236.

[4] Brown, B. and Resnick, S. (1977) Extremes values of independent stochastic processes. *Journal of Applied Probability*, 14, 732–739.

[5] Coles, S.G. and Tawn, J.A. (1991) Modelling extreme multivariate events. *J. R. Statist. Soc. B*, 53, 377–92.

[6] Cooley, D., Naveau, P. and Poncet, P. (2006) Variograms for spatial max-stable random fields. Dependence in Probability and Statistics, Lecture Notes In Statistics, 187, 373-320 90.

[7] Davis, R.A. and Mikosch, T. (2009) Extreme value theory for GARCH processes. *Handbook of Financial Time Series*, 187–200.

[8] Cox, D. and Reid, N. (2004) A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91, 729–737.

[9] Davison, A.C. (2003) *Statistical Models*. New York: Cambridge University Press.

[10] Davison, A., Padoan, S. and Ribatet, M. (2012) Statistical modelling of spatial extremes. *Statistical Science*, 27, 161–186.

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

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

[13] de Haan, L. and Ferreira, A. (2006) *Extreme Value Theory: an Introduction*. Springer Science Business Media, LLC, New York.

[14] de Haan, L. and Resnick, S.I. (1993) Estimating the limit distribution of multivariate extremes. *Communications in Statistics. Stochastic Models*, 9, 275–309.
[15] Einmahl, J.H., Krajina, A. and Segers, J. (2012) An M-estimator for tail dependence in arbitrary dimensions. The Annals of Statistics, 40, 1764–1793.

[16] Engelke, S., Malinowski, A., Kabluchko, Z. and Schlather, M. (2014) Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. J. R. Statist. Soc. B, to appear.

[17] Fan, J., Liao, Y. and Mincheva, M. (2013) Large covariance estimation by thresholding principal orthogonal complements. J. R. Statist. Soc. B, 75, 603–680.

[18] Galambos, J. (1975). Order statistics of samples from multivariate distributions. Journal of the American Statistical Association, 70, 674–680.

[19] Genton, M.G., Ma, Y. and Sang, H. (2013) Large covariance estimation by thresholding principal orthogonal complements. J. R. Statist. Soc. B, 75, 603–680.

[20] Genz, A. (1992) Numerical computation of multivariate normal probabilities. J. Computational and Graphical Statist., 1, 141–149.

[21] Genz, A. (1993) Comparison of methods for the computation of multivariate normal probabilities. Computing Science and Statistics, 25, 400–405.

[22] Genz, A.: Fortran code available at http://www.math.wsu.edu/math/faculty/genz/software/fort77/mvn

[23] Genz, A. and Bretz, F. (2009) Computation of Multivariate Normal and t Probabilities. Lecture Notes in Statistics 195, Springer-Verlag, New York.

[24] Gumbel, E.J. (1960) Distributions des valeurs extrêmes en plusieurs dimensions. Publication de l’ISUP, 9, 171–173.

[25] Hansen, L. and Richard, S.F. (1987) The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. Econometrica, 55, 587–613.

[26] Hofert, M., Mächler, M. and McNeil, A.J. (2012) Likelihood inference for Archimedean copulas in high dimensions under known margins. Journal of Multivariate Analysis, 110, 133–150.

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

[28] Hüsler, J. and Reiss, R.D. (1989) Maxima of normal random vectors: between independence and complete dependence. Statistics and Probability Letters, 9, 283–286.

[29] Jeon, S. and Smith, R.L. (2012) Dependence structure of spatial extremes using threshold approach. arXiv:1209.6344.

[30] Joe, H., Li, H. and Nikoloulopoulos, A.K. (2010) Tail dependence functions and vine copulas. Journal of Multivariate Analysis, 101, 252–270.

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

[32] Klüppelberg, C., Kuhn, G. and Peng, L. (2008). Semi-parametric models for the multivariate tail dependence function - the asymptotically dependent. Scandinavian Journal of Statistics, 35, 701–718.
[33] Kiriliouk, A., Segers, J., Einmahl J. and Krajina, A. (2014). Estimation of tail dependence in high-dimensional and spatial models.

[34] Li, H. and Wu, P. (2013). Extremal dependence of copulas: A tail density approach. *Journal of Multivariate Analysis*, **114**, 99–111.

[35] Lindsay, B. (1988) Composite likelihood methods. *Contemp. Math.*, **80**, 221–239.

[36] Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, **47**, 261–290.

[37] Marshall, A.W. and Olkin, I. (1967) A multivariate exponential distribution. *Journal of the American Statistical Association*, **62**, 30–44.

[38] McNeil, A.J., Frey, R. and Embrechts, P. (2005) *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press.

[39] Padoan, S.A., Ribatet, M. and Sisson, S. (2010) Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, **105**, 263–277.

[40] Resnick, S. (1987) *Extreme values, Regular variation, and Point Processes*. Springer, New York.

[41] Ribatet, M. (2011). SpatialExtremes: modelling spatial extremes. R Package Version 1.8-0. URL http://CRAN.R-project.org/package=SpatialExtremes.

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

[43] Robert, C.Y. (2013) Some new classes of stationary max-stable random fields. *Statistics and Probability Letters*, **83**, 1496–1503.

[44] Sang, H. and Genton, M.G. (2014) Tapered composite likelihood for spatial max-stable models. *Spatial Statistics*, to appear.

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

[46] Segers, J. (2012) Max-stable models for extremal dependence. *RevStat Statistical Journal*, **10**, 61–82.

[47] Sharpe, C. (1964) Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance*, **19**, 425–442.

[48] Smith, R.L. (1990) Max-stable processes and spatial extremes. Preprint, University of Surrey, Surrey.

[49] Stephenson, A.G. and Tawn, J.A. (2005) Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika*, **92**, 213–27.

[50] Takeuchi, K. (1976) Distribution of information statistics and criteria for adequacy of models. *Mathematical Science*, **153**, 12–18.

[51] Varin, C. (2007) On composite marginal likelihoods. *Adv Statist. Anal.*, **92**, 1–28.
[52] Varin, C. and Vidoni, P. (2005) Pairwise likelihood inference and model selection. *Biometrika*, **92**, 519–528.

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

[54] Wadsworth, J.L. and Tawn, J.A. (2014) Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. To appear in *Biometrika*. 