Efficient inference and simulation for elliptical Pareto processes

Emeric Thibaud\textsuperscript{1} and Thomas Opitz\textsuperscript{2}

\textsuperscript{1}Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
\textsuperscript{2}Laboratoire d’Informatique, de Robotique et de Microélectronique de Montpellier, Université Montpellier II, 34095 Montpellier, France

Abstract

Recent advances in spatial extreme value theory have established Pareto processes as the natural limits for threshold exceedances of spatial processes. $\ell$-Pareto processes are obtained by considering exceedances of a risk functional $\ell$, defined as for instance the spatial supremum. Here we introduce elliptical $\ell$-Pareto processes, which arise as the limit of threshold exceedances of certain elliptical processes and provide a flexible dependence model characterized by a correlation function and a shape parameter. These processes correspond to extremal-$t$ limit processes for rescaled maxima. We introduce an efficient inference method for them based on maximizing a full likelihood with partial censoring of components falling below a high marginal threshold and we develop exact conditional and unconditional simulation algorithms. These ideas are illustrated by modelling precipitation extremes in a region of Switzerland.

Keywords: Censored likelihood, Elliptical extremes, Extremal-$t$ process, Pareto process, Simulation.

1 Introduction

There has recently been increasing development of methodologies for modelling extremes in high dimensions, motivated by numerous applications ranging from finance to climatology. Classical extreme value theory relies on max-stable processes, which extend the univariate generalized extreme-value distribution to stochastic processes. Such processes are the only possible non-degenerate limits for rescaled maxima of spatial processes (de Haan & Ferreira, 2006) and provide a natural modelling framework for asymptotically dependent extremes. Various parametric max-stable processes have been constructed (e.g., Davison et al., 2012, 2013). One of the first models proposed was the Gaussian extreme value process (Smith, 1990), but often it is not realistic for modelling complex phenomena; see Thibaud et al. (2013) for instance. More realistic models such as the extremal Gaussian model (Schlather, 2002) or the Brown–Resnick-type models (Brown & Resnick, 1977; Kabluchko et al., 2009) have proven to be well-suited for modelling extremal dependence of environmental data (Davison et al., 2012, 2013; Thibaud et al., 2013).

The extremal-$t$ process is the max-stable limit of $t$ processes characterized by multivariate elliptical $t$ distributions. Opitz (2013) derived a representation of this process which is readily interpreted and enables its simulation. This model provides a flexible dependence structure for multivariate and spatial extremes based on a correlation structure and an additional shape parameter. The extremal Gaussian process and processes
of Brown–Resnick-type are special cases, the latter arising asymptotically from certain configurations of the dependence structure. The extremal-\(t\) process is the maximum attractor for all asymptotically dependent elliptical processes (Opitz, 2013).

Modelling of spatial extremes is typically based on max-stable processes, using pointwise maxima or threshold exceedances, but we shift the focus to Pareto processes (Ferreira & de Haan, 2012; Dombry & Ribatet, 2013), which are natural models for threshold exceedances of spatial processes, since they generalize the peaks-over-threshold stability of multivariate generalized Pareto distributions (Rootzén & Tajvidi, 2006; Falk & Guillou, 2008) to infinite dimensions.

Until recently, inference for max-stable processes was primarily based on composite likelihood methods (Padoan et al., 2010; Genton et al., 2011; Huser & Davison, 2013a), leading to a loss in efficiency compared to full likelihood estimation. Engelke et al. (2012) developed full likelihood methods for Brown–Resnick processes, based either on the distribution of increments with respect to a fixed extreme component, or on the multivariate spectral measure. Wadsworth & Tawn (2013) calculated a partially censored likelihood for Brown–Resnick processes that focuses on exceedances of a thresholding field while censoring the part of the observation vector falling below this field. Since it does not assume that the limiting model is valid below the threshold, this procedure is more robust than the approach of Engelke et al. (2012). Our paper adapts it to the extremal-\(t\) dependence structure and the associated elliptical Pareto process. The resulting inferential procedures, novel and potentially more efficient than composite likelihood methods, are discussed, and efficiency gains over a pairwise likelihood are assessed in a simulation study.

It is often possible to obtain exact simulations from extremal-\(t\) processes and their associated Pareto processes. Ribatet (2013) shows how to obtain conditional simulations from extremal-\(t\) processes, but the approach is computationally very demanding. By contrast, our approach based on the Pareto process allows conditional simulations of extremes very easily.

We develop our results for processes with continuous sample paths defined on a nonempty compact domain \(K \subset \mathbb{R}^m, m \geq 1\); when \(K = \{1, \ldots, D\}\), this boils down to the multivariate setting. The assumption of continuity is natural in applications and ensures that Pareto processes are well-defined. We further rely on functional limit relations which ensure that convergence to limit processes is uniform over the study region. We illustrate the advantages of modelling with elliptical Pareto processes in an application to extreme precipitation in Switzerland.

2 Functional extreme value theory

2.1 \(\ell\)-Pareto processes

We first recall and clarify notions of functional convergence, with particular attention to the notion of an \(\ell\)-Pareto process (Dombry & Ribatet, 2013). We let \(C(K)\) denote the space of continuous functions over \(K\), endowed with the supremum norm \(\|f\|_{\infty} = \sup_{s \in K} |f(s)|\). The restriction of \(C(K)\) to non-negative functions is denoted by \(C_+(K)\). Prior to the characterization of max-stable convergence in \(C(K)\) due to de Haan & Lin (2001), convergence was typically defined in terms of multivariate distributions. Complementary results were provided by Hult & Lindskog (2005) and Davis & Mikosch (2008), using functional
regular variation. In univariate and multivariate theory, a generalized Pareto limit is obtained by conditioning on the exceedance of a high threshold in at least one component (Beirlant et al., 2004; Rootzén & Tajvidi, 2006). This idea is extended to infinite dimensional spaces by conditioning on an exceedance of the largest observation observed over the space, leading to the notion of a generalized Pareto limit process as defined by Ferreira & de Haan (2012). A useful link between the convergence of maxima and threshold exceedences is provided by a point process limit, the “points” being functions in the space $C(K)$. The initial definition of a Pareto process was generalized by Dombry & Ribatet (2013) to so-called $\ell$-Pareto processes.

A functional $\ell : C_+(K) \rightarrow [0, \infty)$ is called a risk functional (or cost functional by Dombry & Ribatet, 2013) if it is continuous and homogeneous, i.e., $\ell(tf) = t \ell(f)$ for $t \geq 0$. In what follows, the “unit sphere” in $C_+(K)$ with respect to $\ell$ is written $S_\ell = \{ f \in C_+(K) : \ell(f) = 1 \}$. Suppose that we are given a risk functional $\ell$ and a probability measure $\rho$ on $S_\ell$. We call any process $Y = \{ Y(s) \}_{s \in K}$ a standard $\ell$-Pareto process with $\ell$-spectral distribution $\rho$ if it can be represented as

$$Y(s) = R f_0(s), \quad R \sim \text{Par}(1),$$

i.e., $\Pr(R > y) = 1/y, y \geq 1$, with $R$ independent of the spectral function $f_0 = \{ f_0(s) \} \sim \rho$.

For continuous real functions $\sigma(s) > 0$, $\mu(s)$, $\xi(s)$ defined over $K$, the process

$$
\begin{cases}
    \mu(s) + \sigma(s) \{ Y(s) \xi(s) - 1 \} / \xi(s), & \xi(s) \neq 0, \\
    \mu(s) + \sigma(s) \log Y(s), & \xi(s) = 0,
\end{cases}
$$

$s \in K$, (2)

is termed a generalized $\ell$-Pareto process. To avoid confusion between processes associated to different risk functionals $\ell$, we will sometimes write $Y_{\ell}$ instead of $Y$. $\ell$-Pareto processes $Y_{\ell}$ satisfy peaks-over-threshold stability: for any $u \geq 1$, the renormalized threshold-exceeding process $[u^{-1} Y_{\ell} \mid \ell(Y_{\ell}) \geq u]$ is equal to $Y_{\ell}$ in distribution (Dombry & Ribatet, 2013). The spectral function of a standard $\ell$-Pareto process $Y_{\ell}$ is $f_0 = Y_{\ell} / \ell(Y_{\ell})$. The variable $R$ in (1) characterizes the intensity of an extreme event in terms of the risk functional $\ell$, whereas $f_0$ describes the corresponding spatial profile.

In applications, different choices of $\ell$ can help answer different questions. If $\ell(f) = \max_{j=1,...,D} f(s_j) / u_j$ for certain sites $s_j \in K$, $j = 1, \ldots, D$, we focus on processes with at least one exceedance of the thresholds $u_j > 0$ at one of the sites. By contrast, $\ell(f) = \min_{j=1,...,D} f(s_j) / u_j$ requires exceedances at each of the $D$ sites. The original definition (Ferreira & de Haan, 2012) uses $\ell(f) = \sup_{s \in K} f(s)$, but conditioning on the supremum over $K$ is not natural in applications where data are only observed at a finite number of sites.

### 2.2 Limiting processes of extremes

We use the notion of a functional maximum domain of attraction and recall equivalent convergence results in terms of threshold exceedances and point processes. Throughout, the symbol “$\Longrightarrow$” indicates weak convergence of random elements from the univariate, multivariate or functional domain. For independent and identically distributed copies $X_1, X_2, \ldots$ of a stochastic process $X = \{ X(s) \}_{s \in K}$ with continuous sample paths, we say that $X$ is in the maximum domain of attraction of a max-stable process $Z = \{ Z(s) \}_{s \in K}$ if there exist sequences of normalizing continuous functions $a_n(s) > 0$ and $b_n(s)$ such that

$$\{ M_n(s) \} = \left\{ \max_{i=1,...,n} a_n(s)^{-1} \{ X_i(s) - b_n(s) \} \right\} \Longrightarrow \{ Z(s) \}, \quad n \to \infty,$$

(3)
in \( C(K) \), with the limit process \( Z \) having non-degenerate univariate distributions \( F_s \). Then \( F_s \) is a generalized extreme value distribution (Coles, 2001). A standard max-stable process \( Z^* \), which has univariate unit Fréchet distributions, is defined by \( Z^*(s) = -\log F_s \{ Z(s) \}^{-1} \). The distribution of the process \( Z^* \) is characterized by a so-called exponent measure \( \Lambda \) on \( C(K) \setminus \{0\} \), which is defined as (Giné et al., 1990)

\[
\Lambda \left[ \bigcup_{j=1}^D \left\{ f \in C(K) : \sup_{s \in K_j} f(s) \geq z_j \right\} \right] = -\log \Pr \left\{ \sup_{s \in K_1} Z^*(s) \leq z_1, \ldots, \sup_{s \in K_D} Z^*(s) \leq z_D \right\}
\]

for any collection of compact sets \( K_j \subset K \) and \( z_j > 0, j = 1, \ldots, D \). The measure \( \Lambda \) is unique if we impose the constraint \( \Lambda(C(K) \setminus C_+(K)) = 0 \), which can be enforced by defining a unique version, \( \Lambda_+(B) = \Lambda(\{ f \in C(K) : f_+ \in B \}) \), for measurable \( B \subset C_+(K) \setminus \{0\} \), with \( f_+(s) = \max\{f(s),0\} \). We will write \( \Lambda_+ \) instead of \( \Lambda \) whenever the restriction to non-negative values is important.

A representation of \( \Lambda_+ \) in terms of pseudo-polar coordinates \( (r,f_0) = T(f) = (\ell(f), f/\ell(f)) \) with some risk functional \( \ell \) is useful. If \( \kappa_\ell(K) = \Lambda_+[\{\ell(f) \geq 1\}] > 0 \), we get

\[
\Lambda_+ \circ T^{-1}\{d(r,f_0)\} = \kappa_\ell(K) r^{-2} dr \times \rho_\ell(df_0), \quad r > 0,
\]

with \( \rho_\ell \) an \( \ell \)-spectral distribution on \( \mathbb{S}_\ell \), which allows us to define the corresponding standard \( \ell \)-Pareto process in (1). When \( \ell(f) = \sup_{s \in K} f(s) \), then \( \kappa_\ell(K) \) is known as the areal extremal coefficient of \( K \) (Lantuéjoul et al., 2011). When the sets \( K_j = \{s_j\} \) are singletons, expression (4) is called the exponent function and denoted \( V(z_1, \ldots, z_D) \), and \( V(1, \ldots, 1) \) is known as the extremal coefficient of the sites \( s_1, \ldots, s_D \). Finite-dimensional projections of \( \Lambda \) relative to \( D \) sites \( s = (s_1, \ldots, s_D) \) are written \( \Lambda_s \), i.e.,

\[
\Lambda_s(\times_{j=1}^D [a_j, b_j]) = \Lambda_{\bigcap_{j=1}^D \{ f \in C(K) : f(s_j) \in [a_j, b_j] \}} \quad \text{for } 0 < a_j < b_j, \ j = 1, \ldots, D.
\]

In particular, \( V(z_1, \ldots, z_D) = \Lambda_s \{ (s_{j=1}^D \leq z_j)^C \} \).

Convergence of the dependence structure and of marginal distributions can be viewed separately in (3). Such separation is often convenient in estimation and allows us to state equivalent conditions that shift the focus to threshold exceedances. Therefore, we define a normalized process \( X^* \) with marginal standard Pareto distributions by \( X^*(s) = 1/[1 - F_s \{ X(s) \}] \). Below we state a number of assumptions and subsequently relate them to the convergence in (3). Consider the following assumptions:

(A1) **Marginal convergence**: sequences of normalizing continuous functions \( a_n(s) > 0 \) and \( b_n(s) \) exist such that

\[
\max_{i=1, \ldots, n} a_n(s)^{-1} \{ X_i(s) - b_n(s) \} \Rightarrow Z(s), \quad n \to \infty,
\]

with a nondegenerate limit \( Z(s) \), for each \( s \) and uniformly in \( s \in K \).

(A2.i) **Standard max-stable convergence**:

\[
\max_{i=1, \ldots, n} n^{-1} X_i^*(s) \Rightarrow \{ Z^*(s) \}, \quad n \to \infty,
\]

where the max-stable limit process is characterized by its exponent measure \( \Lambda_+ \).

(A2.ii) **Standard point process convergence**:

\[
\{ n^{-1} X_i^*(s), i = 1, \ldots, n \} \Rightarrow \mathcal{P}, \quad n \to \infty,
\]
where $\mathcal{P} = \{P_i(s), i = 1, 2, \ldots\}$ is a Poisson process with intensity measure $\Lambda_+$. 

(A2.iii) **Convergence of standard sup-exceedances:**

$$\left\{ n^{-1}X^*(s) : \sup_{s \in K} X^*(s) > n \right\} \implies \{Y^*_\text{sup}(s)\}, \quad n \to \infty,$$

where $Y^*_\text{sup} = \{Y^*_\text{sup}(s)\}$ is a standard sup-Pareto process with sup-spectral distribution $\rho_{\text{sup}}$ related to the exponent measure $\Lambda_+$ of $Z^*$ in (7) through (5).

The convergences in Assumptions (A2.i), (A2.ii) and (A2.iii) take place in $C_+(K)$. From de Haan & Lin (2001), Ferreira & de Haan (2012) and Dombry & Ribatet (2013), Assumptions (A1) and (A2.i) are together equivalent to the max-stable convergence in (3), and Assumptions (A2.i), (A2.ii) and (A2.iii) are equivalent. The convergence of $\ell$-exceedances,

$$\left\{ n^{-1}X^*(s) : \ell(X^*) > n \right\} \implies \{Y^*_\ell(s)\}, \quad n \to \infty,$$

where $Y^*_\ell = \{Y^*_\ell(s)\}$ is a standard $\ell$-Pareto process characterized by its $\ell$-spectral distribution $\rho_\ell$ according to the exponent measure $\Lambda_+$, see (5), follows from each of the Assumptions (A2.i), (A2.ii) and (A2.iii) (Dombry & Ribatet, 2013, Theorem 3). Essentially, these convergences convey that $\Pr\{\ell(X^*) \geq n\} \sim \kappa_\ell(K)/n$ as $n \to \infty$ and we will assume that $\Pr\{\ell(X^*) \geq u_0\} = \kappa_\ell(K)/u_0$ for a sufficiently large threshold $u_0$. Then, $\ell$-exceedances of the threshold $n\kappa_\ell(K)$ correspond to an event with a return period $n$ which we can model by $n\kappa_\ell(K)Y^*_\ell$, with $Y^*_\ell$ the corresponding standard $\ell$-Pareto limit.

The transformation to a generalized $\ell$-Pareto process restores the original marginal scale. Standard theory (Coles, 2001) provides various alternative characterizations of Assumption (A1) in terms of univariate threshold exceedances or point process convergence. It is convenient to fix a high threshold function $u(s)$ and to assume that

$$\Pr\{X(s) > x\} = \left[1 + \xi(s)\{x - \mu(s)\}/\sigma(s)\right]_+^{-1/\xi(s)}, \quad x > u(s),$$

for real parameters $\mu(s), \sigma(s) > 0$ and $\xi(s)$ which correspond to the parameters in (2).

### 2.3 Elliptical extremes

A random vector $X$ defined on $\mathbb{R}^D$ is said to follow an elliptical distribution if it can be written as

$$X = RAU + \mu \tag{12}$$

with $R$ a non-negative random variable, $A$ a $D \times D$ deterministic non-singular matrix defining the dispersion matrix $\Sigma = AA'$, $U$ a random vector independent of $R$ and distributed uniformly on the Euclidean unit sphere $\{x \in \mathbb{R}^D : x'x = 1\}$ and $\mu \in \mathbb{R}^D$ a deterministic shift vector. A point process $\{X_i, i = 1, 2, \ldots\}$ on $\mathbb{R}^D$ is called elliptical if its points can be represented as $X_i = R_iAU_i$ with a non-negative point process $\{R_i, i = 1, 2, \ldots\}$ independent of $\{U_i, i = 1, 2, \ldots\}$ as in (12). We restrict attention to non-singular matrices $A$, excluding some special cases of minor importance in practice. Typical examples of elliptical distributions are the multivariate Gaussian and the multivariate $t$ distributions. Functional max-stable limits according to...
(3) appear from variance mixtures \( X = R W \) of a centered continuous-sample-path Gaussian process \( W \) with a non-negative regularly varying random variable \( R \) such that \( \Pr(R \geq nr)/\Pr(R \geq n) \to r^{-\alpha} \) for some \( \alpha > 0 \) as \( n \to \infty \) (Davis & Mikosch, 2008). The limit corresponds to the extremal-\( t \) process whose dependence structure is characterized by \( \alpha \), which takes the role of a shape parameter, and by the correlation function of \( W \). We call this limit an extremal-\( t \) process because choosing \( R \) such that \( (\nu/R)^2 \sim \Gamma(\nu/2, 2) \), with \( \nu > 0 \), leads to a centered \( t \) process \( X \) with \( \nu \) degrees of freedom and \( \alpha = \nu \) (Demarta & McNeil, 2005; Reisslen & Omre, 2006). We will use the notation \( t_\nu(\mu, \Sigma) \) for the multivariate \( t \) distribution with \( \nu \) degrees of freedom, whose radial variable \( R \) in (12) is characterized by \( D^{-1} R^2 \sim F_{D, \nu} \) with \( F \) the Fisher distribution. The standard max-stable limit process \( Z^* \) of \( X^* \) with exponent measure \( \Lambda_+ \) can be represented in terms of a spectral construction as

\[
Z^*(s) = m_\alpha \max_{i=1,2,\ldots} W_i(s)_{+}^{\alpha}/Q_i, \quad m_\alpha = \sqrt{\pi} 2^{1-\alpha/2} \Gamma\{(\alpha + 1)/2\}^{-1},
\]

where \( 0 < Q_1 < Q_2 < \cdots \) are the points of a unit-rate Poisson process on the positive half-line, and \( W_i = \{W_i(s)\} \) are independent replicates of some standard Gaussian process with continuous sample paths (Opitz, 2013). In particular, \( \alpha = 1 \) yields the extremal Gaussian process (Schlather, 2002). By interpreting the processes \( W_i \) as independent marks of the points of the Poisson process \( \{Q_i\} \), we see that the point process \( \{P_i\} = \{m_\alpha(W_i)_{+}^{\alpha}/Q_i\} \) is a Poisson process with intensity measure \( \Lambda_+ \). We remark that \( (Z^*)^{1/\alpha} \) arises as the pointwise maximum of an elliptical Poisson process with points \( P_i^{1/\alpha} = m_\alpha W_i/Q_i^{1/\alpha} \subset C(K) \), the truncation of \( W_i \) at 0 in (13) being of no relevance since \( Z^* \) is constituted from pointwise maxima that are positive with probability 1. From Poisson process theory (Daley & Vere-Jones, 2007) and the peaks-over-threshold stability of \( \ell \)-Pareto processes, we can conclude that, given a fixed threshold \( u_0 > 0 \), functions \( u_0^{-1} P_i \) with \( \ell(P_i) \geq u_0 \) are realizations of the corresponding \( \ell \)-Pareto process. We use the term elliptical \( \ell \)-Pareto process since the tails of its finite-dimensional distributions correspond to an elliptical distribution with a Pareto-distributed radial variable \( R \) in (12). Based on this observation, we introduce a simulation procedure for elliptical \( \ell \)-Pareto processes in Section 3.3. The finite-dimensional structure associated to \( D \) sites \( s = (s_1, \ldots, s_D) \) is characterized by the exponent function

\[
V_\alpha,\Sigma(z) = -\log \Pr\{Z^*(s_1) \leq z_1, \ldots, Z^*(s_D) \leq z_D\} = \sum_{j=1}^{D} z_j^{-1} t_{\alpha+1} \left\{ (z_j/z_j)^{1/\alpha}; 0, \Sigma_{-j,j}, (\alpha + 1)^{-1} (\Sigma_{j,j} - \Sigma_{j,j}'\Sigma_{j,j}) \right\},
\]

with the correlation matrix \( \Sigma = \{\rho(s_j, s_j')\}_{1 \leq j, j' \leq D} \) related to a correlation function \( \rho \) (Nikolouloupolou et al., 2009), and where \( t_\alpha(\cdot; \mu, \Sigma) \) denotes the cumulative distribution function of a multivariate \( t \) distribution with \( \alpha \) degrees of freedom and parameters \( \mu \) and \( \Sigma \).

Dependence structures of Brown–Resnick type can be interpreted as a special case of extremal-\( t \) dependence, arising asymptotically when \( \alpha \) tends to infinity. Let \( \tilde{W}_i \) be independent and identically distributed copies of a centered Gaussian process with stationary increments, characterized by its variance function \( \tilde{\sigma}^2(s) = \mathbb{E}\{\tilde{W}_1(s)^2\} \) and its variogram \( 2\gamma(s) = \mathbb{E}\{(\tilde{W}_1(s) - \tilde{W}_1(s_0))^2\} \) with an arbitrary fixed reference point \( s_0 \). By analogy with (13), the spectral construction of a Brown–Resnick process is \( Z^*_{BR}(s) = \exp\{\tilde{W}_i(s) - \tilde{\sigma}^2(t)/2\}/Q_i \) (Kabluchko et al., 2009). The distribution of \( Z^*_{BR} \) is characterized solely by
γ. For processes $W_1$ whose correlation function $\rho_\alpha$ depends on $\alpha$ such that the limit $\gamma(s_2 - s_1) = \lim_{\alpha \to \infty} \alpha \{1 - \rho_\alpha(s_1, s_2)\}$ exists and satisfies 0 < $\gamma(s_2 - s_1) < \infty$ for all sites $s_1, s_2$ with $s_1 \neq s_2$, the extremal-$t$ process $Z^*$ converges to $Z^*_{BR}$ as $\alpha$ tends to infinity (Nikoloulopoulos et al., 2009; Kabluchko, 2011; Engelke et al., 2012). For instance, the stable correlation $\rho_\alpha(s_1, s_2) = \exp[-\{\|s_1 - s_2\|/(\alpha^{1/\kappa})\}^\kappa]$ with shape $\kappa \in (0, 2]$ and scale $\alpha^{1/\kappa} \lambda > 0$ yields the variogram $2\{\|s_1 - s_2\|/\lambda\}^\kappa$ of a fractional Brownian motion $\tilde{W}_1$ with Hurst parameter $\kappa/2$. As pointed out by Davison et al. (2012), the extremal-$t$ process generalizes the Brown-Resnick process in terms of the additional parameter $\alpha$ and of different correlation functions $\rho_\alpha$ leading to the same limiting variogram $2\gamma$.

3 Inference

3.1 Estimation approaches

We now consider a collection $s = (s_1, \ldots, s_D)$ of $D$ sites in $K$. The finite-dimensional correlation matrix according to the correlation function $\rho$ of the extremal-$t$ dependence structure is written $\Sigma_s = \{\rho(s_{j_1}, s_{j_2})\}_{1 \leq j_1, j_2 \leq D}$. We suppose that Assumption (A1) and the equivalent Assumptions (A2.i), (A2.ii) and (A2.iii) are satisfied for the finite-dimensional vector $X_s = (X(s_1), \ldots, X(s_D))$ relative to a process $\{X(s)\}_{s \in K}$. Further, we consider a sample $X_{s,1}, \ldots, X_{s,n}$ of independent replications of $X_s$. The finite-dimensional vectors of the related standardized process $X^*$ are denoted by $X^*_s = (X^*(s_1), \ldots, X^*(s_D))$. Often we estimate the marginal parameters $\mu(s), \sigma(s)$ and $\xi(s)$ in (11) before estimating dependence parameters from the normalized vector $X^*_s$ separately. Here we describe this second step, using the convergence of maxima (7), the limiting Poisson point process (8) or the limiting $\ell$-Pareto processes (10). An important role is assigned to exceedances above a high multivariate threshold $u = (u_1, \ldots, u_D)$, with $u_j > 0$ for $j = 1, \ldots, D$ and $\ell(f) = \max_{j=1,\ldots,D} f(s_j)/u_j$, leading to an exceedance at at least one of the sites. In practice, using the full information from an observation $X^*_{s,i}$ with $\ell(X^*_{s,i}) \geq 1$ might not be efficient since the asymptotic distribution might fit the non-exceeding components badly and thus induce bias in the estimators. Hence a partial censoring approach (Smith, 1994) is advisable. Moreover, the extremal-$t$ dependence leads to a singular mass on the boundary of the first hyper-octant in the projection $\Lambda_{s,+}$ of $\Lambda_+$ with respect to the sites $s$, i.e., $\Lambda_{s,+}\{z \in [0, \infty)^D \setminus (0, \infty)^D\} > 0$, complicating any alternative estimation approach without partial censoring.

We first consider the classical limit relation for componentwise maxima. By assuming equality in the convergence (7) for large values of $n$,

$$
\Pr\left(\max_{i=1,\ldots,n} n^{-1} X^*_{s,i} \leq y\right) = \exp\{-V(y)\}, \quad y > 0, \quad (15)
$$

with the exponent function $V$ given by (14). In practice, inference based on the corresponding full likelihood is not feasible for large $D$ because the derivative of (15) involves too many terms. Hence, inference is usually made using a less efficient composite likelihood approach (Padoan et al., 2010; Genton et al., 2011; Huser & Davison, 2013a). In addition, by solely using information from maxima, we discard valuable information from the original observations, so more efficient estimation can be expected from considering threshold exceedances. Fitting a max-stable model to threshold exceedances can be justified by the same asymptotic
joint tail behavior of max-stable and Pareto processes (Ledford & Tawn, 1997). Indeed, using max-stable models for modelling of spatial extremes based on threshold exceedances has become a useful approach (Thibaud et al., 2013; Huser & Davison, 2013b). Instead, in the following, we propose to use the Pareto and Poisson process limits.

The point process model has been used for full likelihood estimation (Coles & Tawn, 1991), as for instance with Brown–Resnick processes (Wadsworth & Tawn, 2013). Consider \( \{ n^{-1}X_{s,1}^*, \ldots, n^{-1}X_{s,n}^* \} \) as a point process and assume equality with the limit in (8) for large values of \( n \). Using a sufficiently high threshold \( u \), we obtain the partially censored Poisson process likelihood, based on the \( N_u \) points \( \{ n^{-1}X_{s,1}^*, \ldots, n^{-1}X_{s,n}^* \} \) falling in the region complementary to \( \times_{j=1,\ldots,D}[{-\infty, n^{-1}u_j}] \), as

\[
L_1(\psi) \propto \exp\{-nV(u)\} \prod_{k=1}^{N_u} \left[-V_{I_k}\{\max(X_{s,k}^*, u)\}\right], 
\]

where \( V_{I_k} \) denotes the partial derivative of \( V \) with respect to the indices \( I_k \subset \{1, \ldots, D\} \) associated to the components that exceed their corresponding marginal thresholds; \( \max(X_{s,k}^*, u) \) is the vector of component-wise maxima. Wadsworth & Tawn (2013) show that a likelihood based on componentwise maxima and their occurrence times (Stephenson & Tawn, 2005) represents a special case of the point process approach.

By analogy with the point process approach, we can consider the convergence of \( \ell \)-exceedances to the \( \ell \)-Pareto process with \( \ell(f) = \max_{j=1,\ldots,D} f(s_j)/u_j \) for some high threshold \( u \). Assuming equality in the convergence (10) yields

\[
\Pr\{X_s^* \leq y \mid \ell(X_s^*) \geq 1\} = 1 - \frac{V(y)}{V(u)}, \quad y \geq u.
\]

As before, we construct a partially censored likelihood from \( N_u \) exceedances \( \ell(X_{s,k}^*) \geq 1 \) and exploit the additional information given by the number of exceedances \( N_u \). The resulting likelihood is

\[
L_2(\psi) = \left\{1 - V(u)\right\}^{n-N_u} \prod_{k=1}^{N_u} \left[-V_{I_k}\{\max(X_{s,k}^*, u)\}\right].
\]

The threshold vector \( u \) must be high enough to yield a valid likelihood function, requiring \( V(u) \leq 1 \). In their treatment of estimation procedures for the Brown–Resnick process, Engelke et al. (2012) propose using \( \ell(f) = \sum_{j=1}^{D} f(s_j) \), leading to the multivariate sum spectral measure, or \( \ell(f) = f(s_0) \) for some fixed site \( s_0 \), leading to the notion of extremal increments. Due to the singularities in \( \Lambda_+ \), the corresponding distributions for the extremal-\( t \) process seem difficult to exploit, so we do not pursue such approaches.

The censored likelihoods \( L_1 \) and \( L_2 \) can be used for inference if the function \( V \) and its partial derivatives are known. In (16), \( N_u \) is considered to be Poisson distributed with intensity \( nV(u) \), while in (18) it is considered to be binomially distributed with parameters \( n \) and \( V(u) \). Since typically \( n \) is large and \( V(u) \) is small, \( L_1 \) and \( L_2 \) are interchangeable in practice.

Whereas the use of exceedance-based likelihoods is motivated from domain-of-attraction arguments in the context of max-stable models, they represent an exact likelihood for the point process and Pareto models. Furthermore, owing to censoring above a threshold \( u > 0 \), the distinction between \( \Lambda \) and \( \Lambda_+ \) does not affect the estimation procedure.
A practical inconvenience is the need to calculate the multivariate \( t \) probabilities in the exponent function (14) when \( D > 2 \). A naive Monte-Carlo or quasi-Monte-Carlo approximation is possible, but may be too slow or too inaccurate. Implementations of faster algorithms exist for integer-valued \( \alpha \) with (roughly) \( D \leq 1000 \) (Genz & Bretz, 2009), but may be inaccurate when \( D \) is large. In large dimensions, one could define a partition of sample sites into moderately large groups and use the composite likelihood based on the full likelihood contribution from each group. For modelling purposes, we propose to restrict \( \alpha \) to integer values. Hence, we get different models for fixed values of \( \alpha = 1, 2, \ldots, \) allowing us to estimate the parameters of the correlation function \( \rho \) and then select the optimal \( \alpha \) from the resulting profile likelihood.

### 3.2 Densities and partial derivatives for extremal-\( t \) processes

A key role in likelihood inference is played by the density \( \lambda_s \) of the finite-dimensional exponent measure \( \Lambda_s \), derived by Ribatet (2013) as

\[
\lambda_s(y) = \alpha^{1-D} \pi^{-(1-D)/2} |\Sigma_s|^{-1/2} \Gamma\{(\alpha + 1)/2\}^{-1} \Gamma\{(\alpha + D)/2\} \times \left( \prod_{j=1}^{D} |y_j| \right)^{1/\alpha - 1} \{T_{1/\alpha}(y)^{-1} \Sigma^{-1}_{s} T_{1/\alpha}(y)\}^{-(\alpha + D)/2}, \quad y \in \mathbb{R}^D
\]

with \( T_a(x) = \text{sign}(x)|x|^a \) for real \( a \). The corresponding unique exponent measure \( \Lambda_{s,+}^{(\alpha)} \) defined on \( [0, \infty)^D \) has positive mass on \( \{ y \in [0, \infty)^D : \min_{j=1, \ldots, D} y_j = 0 \} \). The density \( \lambda_{s,+}^{(\alpha)} \) of \( \Lambda_{s,+,}^{(\alpha)} \) exists on \( (0, \infty)^D \) and is equal to \( \lambda_s \), but it is difficult to describe the behaviour of \( \Lambda_{s,+}^{(\alpha)} \) on the boundary \( \{ y \in [0, \infty)^D : \min_{j=1, \ldots, D} y_j = 0 \} \) of the first hyper-octant because of its singularity.

The partial derivatives \( V_i \) involved in \( L_1 \) and \( L_2 \) are calculated by integrating the intensity \( \lambda_s \) with respect to the components in the set complementary to \( I_i \). The integration is carried out using conditional intensities (Wadsworth & Tawn, 2013). Given a collection \( s_0 = (s_{0,1}, \ldots, s_{0,k}) \) of \( k \) locations, the conditional intensity \( \lambda_{s|s_0,y_0}(y) = \lambda_{(s,s_0)}(y,y_0)/\lambda_{s_0}(y_0) \) equals (Ribatet, 2013)

\[
\lambda_{s|s_0,y_0}(y) = \alpha^{-D} \pi^{-D/2} (k + \alpha)^{-D/2} |\Sigma|^{-1/2} \Gamma\{(\alpha + k)/2\}^{-1} \Gamma\{(\alpha + D + k)/2\} \times \left( \prod_{j=1}^{D} |y_j| \right)^{1/\alpha - 1} \left[ 1 + \frac{\{T_{1/\alpha}(y)^{-1} \Sigma^{-1}_{s} T_{1/\alpha}(y)\}^{-(\alpha + D + k)/2}}{k + \alpha} \right], \quad (19)
\]

with

\[
\bar{\mu} = \Sigma_{s:s_0}^{-1} T_{1/\alpha}(y_0)^{1/\alpha},
\]

\[
\tilde{\Sigma} = \frac{T_{1/\alpha}(y_0)^{1/\alpha} \Sigma_{s_0}^{-1} T_{1/\alpha}(y_0)^{1/\alpha}}{k + \alpha} \left( \Sigma_s - \Sigma_{s:s_0} \Sigma_{s_0}^{-1} \Sigma_{s_0:s_0} \right),
\]

where \( \Sigma_{s:s_0} \) denotes the matrix of covariances between the random vectors corresponding to the location vectors \( s \) and \( s_0 \). Expression (19) is the density of a random vector \( T_\alpha(X) \), where \( X \) follows a \( D \)-dimensional elliptical \( t \) distribution with \( k + \alpha \) degrees of freedom and parameters \( \bar{\mu}, \tilde{\Sigma} \).

Without loss of generality, we consider the partial derivative \( V_{i,k}(y) \) of \( V \) with respect to the indices 1 to \( k \) such that \( I_i = \{1, \ldots, k\} \). It is obtained by calculating the integral of \( \lambda_{s(k+1):D|s_{1:k}:y_{1:k}}(y(k+1):D) \)
under a constraint on the integration constant and by multiplying the resulting expression by \( \lambda_{x_1:k}(y_{1:k}) \). The required integral of the conditional density is 

\[
-V_{1:k}(y) = t_{k+\alpha}(y_{(k+1):D}^{1/\alpha} - \bar{\mu}; \bar{\Sigma}) \sim \pi^{1-k}(1-k)/2|\Sigma_{1:k}|^{-1/2} \Gamma\{(\alpha + 1)/2\}^{-1} \\
\times \Gamma\{(\alpha + k)/2\} \left( \prod_{j=1}^{k} \gamma_j \right)^{1/\alpha-1} \{y_{1:k}^- \Sigma_{1:k}^{-1} y_{1:k}^+ \}^{-(\alpha+k)/2},
\]

with \( \bar{\mu} = \Sigma_{(k+1):D,1:k} \Sigma_{1:k}^{-1} y_{1:k}^{1/\alpha} \) and \( \bar{\Sigma} = (k+\alpha)^{-1}(y_{1:k}^+)^{1/\alpha} \Sigma_{1:k}^{-1} y_{1:k}^+ (\Sigma_{1:k}:D - \Sigma_{(k+1):D,1:k} \Sigma_{1:k}^{-1} \Sigma_{1:k}(k+1):D) \).

3.3 Exact simulation algorithms

Due to the elliptical structure of the points \( P_{i}^{1/\alpha} \) from the point process \( \{ P_i \} \) in (13), an equivalent representation of its finite-dimensional projection relative to \( D \) sites \( s = (s_1, \ldots, s_D) \) is obtained by setting

\[
P_{s,i} = \left\{ \mathbb{E}(U_{1,i}^+)^{-1} \right\}^{-1} (A_s U_i)^{\alpha}/Q_i
\]

with \( A_s \) the Cholesky root of \( \Sigma_s = A_s A'_s \) and \( U_i \) independent and identically distributed copies of some vector \( U \) uniformly distributed on the Euclidean unit sphere. This representation is a special case of Theorem 3.2 of Opitz (2013) and allows exact simulation of both max-stable and Pareto processes due to the boundedness \( \| (A_s U_i)^{\alpha}_+ \|_\infty \leq 1 \) of the spectral vector.

In practice, max-stable processes are simulated using only a finite number of \( P_{s,i} \). When a finite boundary \( b < \infty \) exists for the components of \( Q_i P_{s,i}, \) i.e., \( \text{Pr}\{\max_{i=1,2,...,D} Q_i P_{s,i}(s_j) \leq b \} = 1 \) for \( j = 1, ..., D \), exact simulation of \( Z_s \) can be achieved from a finite number of points \( P_{s,i} \), cf. Theorem 4 of Schlather (2002). In this case, a finite stopping time \( \tau_b \) exists such that \( \max_{1 \leq i \leq \tau_b} P_{s,i}(s_j) \geq \max_{i > \tau_b} P_{s,i}(s_j) \) for \( j = 1, ..., D \). Since the components of \( \{ \mathbb{E}(U_{1,i}^+)^{-1} (A_s U_i)^{\alpha} \} \) in (20) are always bounded by \( b = \{ \mathbb{E}(U_{1,i}^+)^{-1} \} \), exact simulation of extremal-t processes is possible. Two numerical limitations may restrict the applicability of this simulation approach: first, standard algorithms for determining the Cholesky root \( A_s \) of \( \Sigma_s \) require \( O(D^3) \) basic operations; second, \( b \) may be large if \( \alpha \) or \( D \) are large, requiring the simulation of a very large number of points \( P_{s,i} \). More precisely,

\[
b = b(D) = \left[ \mathbb{E} \left\{ (U_{1,i}^+)^{\alpha} \right\} \right]^{-1} = 2 \sqrt{\pi} \frac{\Gamma\{(D + \alpha)/2\}}{\Gamma\{(\alpha + 1)/2\} \Gamma(D/2)},
\]

and applying Stirling’s formula yields \( b(D) \approx 2^{1-\alpha/2} \sqrt{\pi} \Gamma\{(\alpha + 1)/2\}^{-1}(D + \alpha - 2)^{\alpha/2} \) as \( D \to \infty \). In certain situations, notably when \( D \) indexes a fine spatial grid of points, these limitations are too restrictive. Then the conventional approach for approximate simulation is possible, using directly the representation (13) and a pseudo-boundary \( b \) for the random variables \( m_\alpha W(s_j)^{\alpha}_+ \), leading to an approximation error in the simulated max-stable process. Since tails of \( W(s_j)^{\alpha}_+ \) become heavier when \( \alpha \) increases, the approximation error also increases.

Moreover, the simulation of the points \( P_{s,i} \) in (20) yields an algorithm for the simulation of \( \ell \)-Pareto processes. Fix a constant \( u_0 > 0 \). If \( Y^{\alpha}_\ell \) is distributed according to \( \Lambda_+ \{ \{ f : \ell(f) \geq u_0 \} \cap (\cdot) \}/\Lambda_+ \{ f : \ell(f) \geq u_0 \} \), then \( u_0^{-1} Y^{\alpha}_\ell \) is a standard \( \ell \)-Pareto process. Hence the points \( u_0^{-1} P_{s,i} \) falling into the set
\( \{ f : \ell(f) \geq u_0 \} \) are realizations of the standard \( \ell \)-Pareto process. Suppose that we have simulated a finite number \( \tau_0 \) of Poisson process points \( P_{s,i}, i = 1, \ldots, \tau_0 \), as in (20). If the algorithm ensures that \( \tau_0 \) satisfies \( \max_{i > \tau_0} \ell(P_i) \leq u_0 \) for the functions \( P_i \) associated to \( P_{s,i} = P_i(s) \), then the subsample of points \( P_{s,i_k}, k = 1, \ldots, \tau_1 \), with \( \ell(P_{s,i_k}) \geq u_0 \), represents \( \tau_1 \) realizations of \( Y^*_\ell(s) \), where \( \tau_1 \) follows the Poisson distribution with intensity \( \Lambda + \{ f : \ell(f) \geq u_0 \} \). For instance, if \( \ell(f) = \max_{j \in J} f(s_j) \) or \( \ell(f) = \min_{j \in J} f(s_j) \) for some subset of indices \( J \subset \{ 1, \ldots, D \} \), we can exploit the upper bound \( b \) and choose \( \tau_0 \) such that \( Q_{\tau_0} > b/u_0 \). Similarly, if \( \ell(f) = \sum_{j \in J} \omega_j f(s_j) \) with weights \( \omega_j > 0 \), then \( \tau_0 \) such that \( Q_{\tau_0} \geq b \sum_{j \in J} \omega_j/u_0 \) is an admissible choice.

As an alternative to simulating points of the extremal-\( t \) Poisson process, an acceptance-rejection algorithm could be used to generate realisations of \( Y^*_\ell(s) \) without dealing with a random number \( \tau_1 \) of realisations. Fix \( u_0 > 0 \) such that \( \max_{j=1,...,D} f(s_j) = 1 \) whenever \( \ell(f) \geq u_0 \). Since \( \max_{j=1,...,D} (A_s U_j) \leq 1 \), one can easily show that the conditional random vector \( u_0^{-1} R(A_s U_j) | \ell(RA_s U) \geq u_0 \) with standard Pareto distributed \( R \), independent of \( U_s \), is equal to \( Y^*_\ell(s) \) in distribution. Hence, simulating the elliptical random vector \( u_0^{-1} RA_s U \) and retaining vectors fulfilling the condition \( \ell(RA_s U) \geq u_0 \) yields a sample of \( Y^*_\ell(s) \). To minimize the frequency of rejections, \( u_0 \) should be chosen as small as possible.

It is possible to condition not on an \( \ell \)-exceedance, but on a specific value of \( \ell \): we obtain a realization of \( \{ Y^*_\ell \mid \ell(Y^*_s) = u_0 \} \) by simulating an \( \ell \)-Pareto process \( Y^*_\ell \), determining its spectral function \( f_0 = Y^*_\ell / \ell(Y^*_\ell) \) and multiplying it by \( u_0 \). Then \( u_0 f_0 \) is equal in distribution to \( Y^*_\ell \mid \ell(Y^*_\ell) = u_0 \).

Practical difficulties arise when the computation of \( \ell(f) \) requires information from an infinite number of sites \( s_j \), for instance in the case of \( \ell(f) = \sup_{s \in K} f(s) \) with non-finite \( K \). Then it is impossible to determine the value \( \ell(f) \) from a finite collection of values \( \{ f(s_j), j = 1, \ldots, D \} \). However, identifying the \( \ell \)-Pareto process with its restriction to a fine grid of points spanned over \( K \) can yield a satisfactory approximation in practice. In place of \( \ell(f) = \sup_{s \in K} f(s) \), one would then use \( \ell(f) = \max_{s \in G} f(s) \) where \( G \) denotes the set of grid points.

Instead of conditioning \( \Lambda \) on exceedances of the threshold \( u_0 = 1 \) for \( \ell(f) \) to obtain the distribution of the \( \ell \)-Pareto process, one might be interested in the distribution obtained from conditioning on values \( y_0 > 0 \) for a collection of sites \( s_0 = (s_{0,1}, \ldots, s_{0,k}) \). The corresponding finite-dimensional conditional distribution for the sites \( s = (s_1, \ldots, s_D) \), disjoint with \( s_0 \), has density (19). The conditional process defined on \( K \setminus \{ s_{0,1}, \ldots, s_{0,k} \} \) corresponds to a transformed \( t \) process that can easily be simulated on a large number of points.

### 4 Simulation study

We used simulation to investigate the relative efficiency of the full and the pairwise likelihood approaches to estimating the parameters of elliptical Pareto processes using the partial censoring scheme discussed in Section 3.1. Based on the algorithm detailed in Section 3.3, we simulated 1000 elliptical \( \ell \)-Pareto processes, using the exact procedure, with \( \ell(f) = \max_{s \in S} f(s) \) at 16 locations given by \( S = \{(i/3, j/3)\}_{0 \leq i,j \leq 3} \). We chose the exponential correlation function \( \rho(h) = \exp(-\|h\|/\lambda) \), where \( h \) is the lag vector between two locations and \( \lambda > 0 \) is a range parameter. Different combinations of values for the degrees of freedom \( \alpha \) and for \( \lambda \), covering small to strong dependence, were considered. We fixed \( \alpha \) and estimated \( \lambda \) based
on marginal thresholds equal to the 95% percentile. For each parameter configuration, estimates for 5000 replicates were calculated using the two approaches. Table 1 shows the resulting standard deviations of $\hat{\lambda}/\lambda$ and their relative efficiency, defined via the ratio of the two variances. Unreported results showed that the estimation was essentially unbiased for both methods. Throughout, the full likelihood estimator was found to be more efficient than the pairwise one, and the gain is larger for processes with weak dependence. Overall, the relative reduction in variance is around 50%. Unreported results indicated that the gain of using the full likelihood over the pairwise likelihood is larger for larger grids: using a similar framework, we found a reduction of variance around 10% for four locations and around 35% for nine locations.

To investigate the effects of the convergence to a limiting elliptical Pareto process, we further simulated samples composed of 1000 $t$ processes. To focus on the convergence of the dependence structure, marginal distributions were transformed to the standard Pareto scale. We fitted elliptical Pareto processes to threshold exceedances over the marginal 95%, 98% and 99% percentiles. Table 2 reports the bias and standard deviations of $\hat{\lambda}/\lambda$ calculated from 5000 replicates and the ratio of the root mean squared errors (RMSE) in $\hat{\lambda}/\lambda$ for the two estimators. For the three thresholds, both estimators of $\lambda$ have a small bias when $\alpha = 1$, but the bias increases as $\alpha$ increases due to the slower convergence to the limiting dependence structure for larger $\alpha$; the strength of dependence decreases very slowly in the extremes and stronger dependence is still present at finite levels. The bias is reduced by increasing the thresholds such that the exceedance distribution is closer to the asymptotic model; standard deviations increase accordingly. Standard deviations are always smaller for the full likelihood estimator than for the pairwise one, but the bias of the full likelihood estimator is larger in most cases. In terms of RMSE for the 95% threshold, the full likelihood estimator outperforms the pairwise one for $\alpha = 1, 2$, but not for $\alpha > 2$. Using higher thresholds, the bias of the full likelihood estimator approaches that of the pairwise estimator, and the full likelihood estimator generally has a smaller RMSE than the pairwise estimator owing to its smaller variance. Hence for large values of $\alpha$, very high thresholds are needed for the full likelihood estimator to outperform the pairwise estimator in terms of RMSE.

We conclude that the full likelihood inference improves estimation efficiency when the model is correctly specified, but that the pairwise approach appears more robust to certain kinds of model misspecification.
Table 2: Estimation of elliptical Pareto processes based on exceedances of $t$ processes. Bias and standard deviation of the full likelihood and pairwise likelihood approaches for the estimation of the range $\lambda$ of an elliptical Pareto process with exponential correlation function, $\alpha = 1, \ldots, 10$, and thresholds chosen at the 95%, 98% and 99% percentiles. For each $\alpha$, the values of $\lambda$ were chosen to yield the pairwise extremal coefficient $\theta = 1.4$ at a distance 0.5. For each threshold, we report bias/standard deviation of $\hat{\lambda}/\lambda$, estimated from 5000 replicates. The last row reports the ratio of the root mean squared errors $\varrho = \text{RMSE}_{\text{full}}/\text{RMSE}_{\text{pw}}$ of the two methods. All numbers have been multiplied by $10^2$.

| $\alpha$ | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 95%     | full| 1/9 | 10/9| 24/9| 40/11|57/12|73/13|90/14|107/16|124/17|140/18|
| pw     | 1/12| 8/12| 21/13| 36/14| 52/16| 68/17| 85/18|102/20|118/22|135/23|
| $\varrho$ | 50  | 79  | 110 | 115 | 112 | 111 | 110 | 108 | 108 | 107 |
| 98%     | full| 1/14| 4/13| 12/13| 23/14| 35/15| 47/16| 60/18| 72/19| 86/20| 98/22|
| pw     | 1/19| 3/18| 11/18| 21/19| 33/21| 44/22| 57/24| 70/25| 83/28| 96/29|
| $\varrho$ | 51  | 51  | 71  | 89  | 98  | 101 | 102 | 101 | 101 | 101 |
| 99%     | full| 0/20| 2/18| 8/18| 15/18| 24/19| 35/20| 46/22| 56/24| 67/25| 77/27|
| pw     | 2/29| 2/25| 8/25| 15/25| 23/27| 34/29| 44/31| 55/32| 66/34| 75/36|
| $\varrho$ | 47  | 51  | 56  | 65  | 77  | 82  | 89  | 92  | 94  | 96  |

5 Application

We illustrate the concepts introduced earlier through the modelling of extreme precipitation in the region of Zürich, using daily cumulative rainfall data at 44 locations provided by MétéoSuisse, see Figure 1. These stations have elevations varying from 327 to 718 m. Similar data were used by Davison et al. (2012) for max-stable estimation and by Dombry et al. (2013) for conditional max-stable simulation. Our analysis is based on daily summer data recorded from 1 June to 31 August for the years 1962–2012. A preliminary study showed no signs of non-stationarity in the time series and only weak day-to-day dependence in exceedances over the 95% percentiles, leading us to model the daily data as independent and identically distributed. We selected 25 stations for the fit of the spatial model, see Figure 1; the other stations are kept for its validation. First, we fitted a spatial model for the univariate marginal distributions (11) by regressing parameters $\mu(s)$ and $\sigma(s)$ linearly on longitude and latitude while keeping the shape parameter $\xi$ constant. Using threshold exceedances over marginal thresholds taken to be the 95% percentiles at each of the 25 stations, we estimated parameters by maximizing the independence likelihood (Chandler & Bate, 2007), in which the data at each location are wrongly assumed to be independent. This simple regression model did not capture spatial variability among marginal distributions in a satisfactory manner, so we extended it to a Bayesian hierarchical model allowing for spatial random effects in $\mu(s)$ and $\sigma(s)$, similar to the “latent variable model” of Davison et al. (2012); see Appendix A for more details. QQ-plots suggested that the marginal model resulting from posterior means was adequate. The shape parameter $\xi$ was estimated to be 0.11 with 95% confidence interval (0.08, 0.14), corresponding to heavy-tailed marginal distributions. We then transformed the original data at each location to the standard Pareto scale by using the fitted marginal distributions above the thresholds and the empirical...
Pareto processes. Standardized marginal thresholds correspond to the same percentiles as those used for the
(2013) found similar results for extreme rainfall elsewhere in Switzerland.

Whereas the Brown–Resnick process provides a much better fit than the extremal-Gaussian model corre-
model can be selected from a direct comparison of maximized likelihoods; results are shown in Table 3. The models under consideration have the same number of parameters, so the best
process with stable variogram
\(2^\gamma(h) = 2(||h||/\lambda)^\kappa\), corresponding to the limiting model when \(\alpha \to \infty\) as described in Section 2.3. The models under consideration have the same number of parameters, so the best model can be selected from a direct comparison of maximized likelihoods; results are shown in Table 3. Whereas the Brown–Resnick process provides a much better fit than the extremal Gaussian model corre-
responding to \(\alpha = 1\), the best model is an elliptical Pareto process with \(\alpha = 6\). The parameter estimates for
our best model are \(\hat{\lambda} = 483\, (35)\) km and \(\hat{\kappa} = 0.64\, (0.01)\), which corresponds to a process with realizations that are continuous but not differentiable. These results are coherent with those found by Davison et al. (2012) who identified an extremal-\(t\) model with \(\alpha = 5.5\) as the best max-stable model for yearly maxima of rainfall on the same region and found similar parameter values. Thibaud et al. (2013) and Davison et al. (2013) found similar results for extreme rainfall elsewhere in Switzerland.

We validate the accuracy of our model for modelling spatial extremes using the data from the other

---

**Figure 1:** Modelling extreme precipitation in the region of Zürich. Left: region of Zürich. Black dots correspond to the stations used for the fit, and white dots to the stations used for validation. Right: empirical pairwise extremal coefficients for the data used to fit the model (in red) and for the validation data (in black), with the fitted extremal coefficient curve for the best elliptical Pareto model in blue.

**Table 3:** Negative log-likelihoods for Pareto \(t\) and Brown–Resnick models, in terms of differences with respect to the negative log-likelihood of the best model (\(\alpha = 6\)).

| Degrees of freedom \(\alpha\) | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10  | B.–R. |
|-----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|-------|
| Negative log-likelihood     | 1915  | 426   | 134   | 42    | 9     | 0     | 2     | 8     | 18    | 30  | 432   |
19 stations. The right panel of Figure 1 shows estimates of pairwise extremal coefficients related to these stations. The estimates for validation stations are slightly more variable. Overall, the extremal dependence for validation data is adequately represented by the model. Using the conditional distribution (19), we further considered simulation at the 19 validation stations conditional on the values observed at the other stations when at least one of the 25 components exceeded the marginal 95% threshold. We compare the observed extreme values at the 19 validation locations with those predicted by the model by measuring the proportion of true values falling within the 95% confidence bounds of the simulated values. Over the 19 locations, approximately 85% of the observed values were in the confidence bounds. We do not attain 95% corresponding to a perfect prediction. This can be explained by the fact that the simulations did not take into account the uncertainty of the estimates of the fitted model; confidence intervals taking this uncertainty into account should be wider and have a better coverage probability.

Finally, we illustrate the ability of the Pareto process approach to easily simulate conditional rainfall given observed extremes at some set of locations. For a particular day and given the observed data at the 25 locations used for the fit, we simulated conditional values of rainfall over the region using the transformed $t$ process characterized in (19); the left panel of Figure 2 shows the mean of these simulated rainfall fields and the right panel its standard deviations.

### 6 Discussion

Pareto models are theoretically appealing because they generalize peaks-over-threshold stability to the spatial context. We have illustrated their use in extreme value analysis by presenting efficient inference and simulation procedures for the elliptical Pareto model, which stands out for its flexibility and its close relation to classical geostatistics via a correlation function.

The use of the full likelihood instead of a pairwise likelihood provides a large gain in efficiency, but our simulations show that pairwise likelihood inference may be less biased when estimating the limiting model
of a t process, illustrating that composite likelihood may be more robust to some model misspecification. Taking higher thresholds favors the use of full likelihood but is impossible in some cases. The robustness of the pairwise likelihood is an interesting result that necessitates further investigation.

In an application example, a two-steps approach was employed to model extreme events using the elliptical Pareto process. First, we used a Bayesian model that allows for complex modelling of the marginal distributions. Second, a full likelihood approach with partial censoring was applied to dependence modelling. This procedure permits mixing of complex marginal modelling with efficient fitting of an asymptotically dependent model for the dependence. Exact simulations and conditional simulations of extremes based on the elliptical Pareto process are easily obtained using the algorithms we introduced, providing a real improvement compared to models based on extremal-t max-stable processes. Our two-steps approach could be replaced by full Bayesian modelling of generalized elliptical Pareto processes.

Acknowledgments

We thank A. C. Davison and R. Huser for their comments on the manuscript. This study was funded by the Swiss National Science Foundation. The second author is further grateful for financial support from the Mc SIM (“Multisupport conditional simulation of max-stable processes”) and MIRACCLE (“Mesures et Indicateurs de Risques Adaptés au Changement CLimatique”) projects. The rainfall data were provided by MétéoSuisse.

Appendix A

We describe the model applied to the spatial modelling of the marginal distributions using the tail parametrisation (11). We used a latent variable model similar to that of Davison et al. (2012). Given n vectors of D-variate observations \( x_i = (x_{i1}, \ldots, x_{iD}) \), we consider the independent censored likelihood function derived from model (11),

\[
\ell(\mu, \sigma, \xi) = \prod_{j=1}^{D} \left[ \left\{ 1 - V(u_j; \mu_j, \sigma_j, \xi) \right\}^{n-j(u_j)} \prod_{i=1}^{N_j(u_j)} \left\{ -V'(x_{ij}; \mu_j, \sigma_j, \xi) \right\} \right],
\]

for \( \mu = (\mu_1, \ldots, \mu_D), \sigma = (\sigma_1, \ldots, \sigma_D) \) and \( V(x; \mu, \sigma, \xi) = \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi} \). We chose a spatial model for \( \mu \) and \( \sigma \) with latitude and longitude covariates,

\[
\begin{align*}
\mu &= \beta_{\mu,0} + \beta_{\mu,1}\text{lat} + \beta_{\mu,2}\text{long} + \omega_{\mu} = X_{\mu} \beta_{\mu} + \omega_{\mu}, \\
\sigma &= \beta_{\sigma,0} + \beta_{\sigma,1}\text{lat} + \beta_{\sigma,2}\text{long} + \omega_{\sigma} = X_{\sigma} \beta_{\sigma} + \omega_{\sigma},
\end{align*}
\]

where \( \omega_{\mu} \) and \( \omega_{\sigma} \) are two centered normal vectors whose distributions are defined through their covariance functions \( a_{\mu} \exp(-\|h\|/\lambda_{\mu}) \) and \( a_{\sigma} \exp(-\|h\|/\lambda_{\sigma}) \) respectively, with unknown parameters \( a_{\mu}, a_{\sigma}, \lambda_{\mu} \) and \( \lambda_{\sigma} \).

This spatial model for the marginal parameters was fitted to data using Markov Chain Monte Carlo methods. Prior distributions for the parameters \( \beta_{\mu}, \beta_{\sigma}, a_{\mu}, a_{\sigma}, \lambda_{\mu}, \lambda_{\sigma} \) and \( \xi \) are chosen to yield, when
possible, tractable full conditional distributions, thus permitting the use of a Gibbs sampler. For $\beta_\mu$, we chose a normal prior distribution and for $a_\mu$ we chose an inverse-gamma distribution. These two choices lead to conjugate priors, corresponding respectively to normal and inverse-gamma full conditional distributions. We made the same choices for $\beta_\sigma$ and $a_\sigma$. No conjugate priors exist for $\lambda_\mu$, $\lambda_\sigma$ and $\xi$. For the two range parameters we chose gamma priors, and for $\xi$ we chose a normal prior; we used a Metropolis–Hastings algorithm to update these three parameters.

The Markov chain was run for 300,000 iterations after a burn-in period of 5000 iterations. Visual diagnostics indicate that convergence is attained. Parameter estimates are taken to be the empirical mean of the chain. Confidence intervals are obtained from the full conditional distributions. Although there is some long-range dependence in the chain, the number of iterations is enough to provide good estimates.

References

Beirlant, J., Goengeheur, Y., Segers, J. & Teugels, J. (2004) Statistics of Extremes: Theory and Applications. New York: Wiley.

Brown, B. M. & Resnick, S. I. (1977) Extreme values of independent stochastic processes. Journal of Applied Probability 14, 732–739. doi:10.2307/3213346.

Chandler, R. E. & Bate, S. (2007) Inference for clustered data using the independence loglikelihood. Biometrika 94, 167–183. doi:10.1093/biomet/asm015.

Coles, S. (2001) An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics. London: Springer. ISBN 1-85233-459-2.

Coles, S. G. & Tawn, J. A. (1991) Modelling extreme multivariate events. Journal of the Royal Statistical Society. Series B (Methodological) 53, 377–392.

Daley, D. J. & Vere-Jones, D. (2007) An Introduction to the Theory of Point Processes: Elementary Theory and Methods, volume 1. Springer.

Davis, R. A. & Mikosch, T. (2008) Extreme value theory for space-time processes with heavy-tailed distributions. Stoch. Processes Appl. 118, 560–584. doi:10.1016/j.spa.2007.06.001.

Davison, A. C., Huser, R. & Thibaud, E. (2013) Geostatistics of Dependent and Asymptotically Independent Extremes. Mathematical Geosciences 45, 511–529. doi:10.1007/s11004-013-9469-y.

Davison, A. C., Padoan, S. A. & Ribatet, M. (2012) Statistical modelling of spatial extremes (with Discussion). Statistical Science 27, 161–186. doi:10.1214/11-STS376.

de Haan, L. & Ferreira, A. (2006) Extreme Value Theory: An Introduction. Springer Series in Operations Research and Financial Engineering. New York, NY, USA: Springer. ISBN 0387239464.

Demarta, S. & McNeil, A. J. (2005) The $t$ Copula and Related Copulas. International Statistical Review 73, 111–129. doi:10.1111/j.1751-5823.2005.tb00254.x.
Dombry, C., Éyi-Minko, F. & Ribatet, M. (2013) Conditional simulation of max-stable processes. *Biometrika* **100**, 111–124. doi:10.1093/biomet/ass067.

Dombry, C. & Ribatet, M. (2013) Functional regular variations, Pareto processes and peaks over threshold. *Submitted*.

Engelke, S., Malinowski, A., Kabluchko, Z. & Schlather, M. (2012) Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. ArXiv:1207.6886v1 [stat.ME].

Falk, M. & Guillou, A. (2008) Peaks-over-threshold stability of multivariate generalized Pareto distributions. *Journal of Multivariate Analysis* **99**, 715–734. doi:10.1016/j.jmva.2007.03.009.

Ferreira, A. & de Haan, L. (2012) The Generalized Pareto process; with a view towards application and simulation. *Bernoulli* .

Genton, M. G., Ma, Y. & Sang, H. (2011) On the likelihood function of Gaussian max-stable processes. *Biometrika* **98**, 481–488. doi:10.1093/biomet/asr020.

Genz, A. & Bretz, F. (2009) *Computation of Multivariate Normal and t Probabilities*. New York: Springer.

Giné, E., Hahn, M. G. & Vatan, P. (1990) Max-infinitely divisible and max-stable sample continuous processes. *Probability Theory and Related Fields* **87**, 139–165. doi:10.1007/BF01198427.

de Haan, L. & Lin, T. (2001) On convergence toward an extreme value distribution in C[0,1]. *The Annals of Probability* **29**, 467–483. doi:10.1214/aop/1008956340.

Hult, H. & Lindskog, F. (2005) Extremal behavior of regularly varying stochastic processes. *Stochastic Processes and their Applications* **115**, 249–274. doi:10.1016/j.spa.2004.09.003.

Huser, R. & Davison, A. C. (2013a) Composite likelihood estimation for the Brown–Resnick process. *Biometrika* **100**, 511–518. doi:10.1093/biomet/ass089.

Huser, R. & Davison, A. C. (2013b) Space-time modelling of extreme events. *Journal of the Royal Statistical Society, Series B* doi:10.1111/rssb.12035.

Kabluchko, Z. (2011) Extremes of independent Gaussian processes. *Extremes* **14**, 285–310. doi:10.1007/s10687-010-0110-x.

Kabluchko, Z., Schlather, M. & de Haan, L. (2009) Stationary max-stable fields associated to negative definite functions. *The Annals of Probability* **37**, 2042–2065. doi:10.1214/09-AOP455.

Lantuéjoul, C., Bacro, J. & Bel, L. (2011) Storm processes and stochastic geometry. *Extremes* **14**, 413–428. doi:10.1007/s10687-010-0121-7.

Ledford, A. W. & Tawn, J. A. (1997) Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society. Series B (Methodological)* **59**, 475–499. doi:10.1111/1467-9868.00080.
Nikoloulopoulos, A., Joe, H. & Li, H. (2009) Extreme value properties of multivariate $t$ copulas. *Extremes* **12**, 129–148. doi:10.1007/s10687-008-0072-4.

Opitz, T. (2013) Extremal $t$ processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis* **122**, 409–413. doi:10.1016/j.jmva.2013.08.008.

Padoan, S. A., Ribatet, M. & Sisson, S. A. (2010) Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association* **105**, 263–277. doi:10.1198/jasa.2009.tm08577.

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

Røislien, J. & Omre, H. (2006) $t$-Distributed random fields: a parametric model for heavy-tailed well-log data. *Mathematical Geology* **38**, 821–849. doi:10.1007/s11004-006-9050-z.

Rootzén, H. & Tajvidi, N. (2006) Multivariate generalized Pareto distributions. *Bernoulli* **12**, 917–930. doi:10.3150/bj/1161614952.

Schlather, M. (2002) Models for stationary max-stable random fields. *Extremes* **5**, 33–44. doi:10.1023/A:1020977924878.

Smith, R. L. (1990) Max-Stable Processes and Spatial Extremes. Unpublished manuscript, University of Surrey, Guildford GU2 5XH, England. http://www.stat.unc.edu/postscript/rs/spatex.pdf.

Smith, R. L. (1994) Multivariate threshold methods. In *Extreme Value Theory and Applications*, eds. J. Galambos, J. Lechner & E. Simiu, pp. 225–248. Kluwer. ISBN 0792328655.

Stephenson, A. & Tawn, J. A. (2005) Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika* **92**, 213–227. doi:10.1093/biomet/92.1.213.

Thibaud, E., Mutzner, R. & Davison, A. C. (2013) Threshold modeling of extreme spatial rainfall. *Water Resources Research* **49**, 4633–4644. doi:10.1002/wrcr.20329.

Wadsworth, J. L. & Tawn, J. A. (2013) Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. doi:10.1093/biomet/ast042. Biometrika, to appear.