Likelihood-based inference for max-stable processes

S. A. Padoan[*], M. Ribatet[†] and S. A. Sisson[‡]

February 23, 2009

Abstract

The last decade has seen max-stable processes emerge as a common tool for the statistical modelling of spatial extremes. However, their application is complicated due to the unavailability of the multivariate density function, and so likelihood-based methods remain far from providing a complete and flexible framework for inference. In this article we develop inferentially practical, likelihood-based methods for fitting max-stable processes derived from a composite-likelihood approach. The procedure is sufficiently reliable and versatile to permit the simultaneous modelling of joint and marginal parameters in the spatial context at a moderate computational cost. The utility of this methodology is examined via simulation, and illustrated by the analysis of U.S. precipitation extremes.

Keywords: Composite likelihood; Extreme value theory; Max-stable processes; Pseudo-likelihood, Rainfall; Spatial Extremes.

1 Introduction

A common objective of spatial analysis is to quantify and characterise the behavior of environmental phenomena such as precipitation levels, windspeed or daily temperatures. A number of generic approaches to spatial modelling have been developed (e.g. Barndorff-Nielsen et al. (1998); Cressie (1993); Ripley (2004)), but these are not necessarily ideal for handling extremal aspects given their focus on mean process levels. Analyses of spatial extremes are useful devices for understanding and predicting extreme events such as hurricanes, storms and floods. In light of recent concerns

---

[*]Laboratory of Environmental Fluid Mechanics and Hydrology, Ecole Polytechnique Fédérale de Lausanne, Switzerland. Email: Simone.Padoan@epfl.ch

[†]Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne, Switzerland. Email: Mathieu.Ribatet@epfl.ch

[‡]School of Mathematics and Statistics, University of New South Wales, Australia. Email: Scott.Sisson@unsw.edu.au

over climate change, the use of robust mathematical and statistical methods for such analyses has grown in importance

While the theory and statistical practice of univariate extremes is well developed, there is much less guidance for the modelling of spatial extremes. This is problematic as many environmental processes have a natural spatial domain. We consider a temporal series of componentwise maxima of process measurements recorded at $k = 1, \ldots, K$ locations, within a contiguous region. Observations $\{y_{n,k}\}$ each denote the maximum of m samples over $n = 1, \ldots, N$ temporal blocks. For example, for daily observations, $m = 366$ implies the $\{y_{n,k}\}$ describe process annual maxima.

The spatial analogue of multivariate extreme value models is the class of max-stable processes (de Haan, 1984; de Haan and Pickands, 1986; Resnick, 1987). Max-stable processes have a similar asymptotic motivation to the univariate Generalised Extreme Value (GEV) (von Mises, 1954; Jenkinson, 1955)), providing a general approach to modelling process extremes incorporating temporal or spatial dependence. Statistical methods for max-stable processes and data analysis of practical problems are discussed by Smith (1990), Coles (1993), Coles and Walshaw (1994) and Coles and Tawn (1996). Standard likelihood methods for such models are complicated by the intractability of the multivariate density function in all but the most trivial cases. This presents an obstacle in the use of max-stable processes for spatial extremes.

There is a lack of a proper inferential framework for the analysis of spatial extremes (although De Haan and Pereira (2006) describe some non-parametric estimators). In this article we develop flexible and inferentially practical methods for the fitting of max-stable processes to spatial data based on non-standard, *composite* likelihood-based methods Lindsay (1988). An appealing feature of this approach is that the estimation of GEV marginal parameters can be performed jointly with the dependence parameters in a unified framework. Accordingly, there is no need for separate estimation procedures. With highly-structured problems such as max-stable processes, this approach produces flexible and reliable results with a moderate computational cost.

The article is organised as follows: Section 2 reviews the theory of max-stable processes and its relationship to spatial extremes. Our composite likelihood approach is developed in Section 3 and Section 4 evaluates the method's performance through a number of simulation studes. We conclude with an illustration of a real extremal data analysis of U.S. precipitation levels.

2 Max-stable processes and spatial extremes

2.1 Max-stable processes

Max-stable processes provide a natural generalisation of extremal dependence structures in continuous spaces. From this, closed-form bivariate distributions can be derived.

Definition. Let T be an index set and $\{\tilde{Y}_i(t)\}_{t\in T}$, $i = 1, \ldots, n$ be n independent replications of a continuous stochastic process. Assume that there are sequences of continuous functions $a_n(t) > 0$ and $b_n(t) \in \mathbb{R}$ such that

$$Y(t) = \lim_{n\to\infty} \frac{\max_{i=1}^n \tilde{Y}_i(t) - b_n(t)}{a_n(t)}, \quad t \in T.$$

If this limit exists, the limit process $Y(t)$ is *a max-stable process* de Haan (1984).

Two properties follow from the above definition De Haan and Resnick (1977). Firstly, the one-dimensional marginal distributions belong to the class of generalised extreme value distributions (GEV), $Y \sim \text{GEV}(\mu, \lambda, \xi)$ with distribution function

$$F(y; \mu, \lambda, \xi) = \exp\left[-\left\{1 + \frac{\xi(y-\mu)}{\lambda}\right\}_+^{-1/\xi}\right], \quad -\infty < \mu, \quad \xi < \infty, \quad \lambda > 0,$$

where $a_+ = \max(0; a)$ and μ, λ and ξ are respectively location, scale and shape parameters Fisher and Tippett (1928). Secondly, for any $K = 2, 3, \ldots$, the K-dimensional marginal distribution belongs to the class of multivariate extreme value distributions.

W.l.o.g. if $a_n(t) = n$, $b_n(t) = 0$ $\forall t$, then the corresponding process, $\{Z(t)\}_{t\in T}$, has unit Fréchet margins, with distribution function $F(z) = \exp(-1/z), z > 0$. This process is obtainable as standardisation of $\{Y(t)\}_{t\in T}$ through

$$\{Z(t)\}_{t\in T} \equiv \left[\left\{1 + \frac{\xi(t)(Y(t) - \mu(t))}{\lambda(t)}\right\}_+^{1/\xi(t)}\right]_{t\in T},$$

where $\mu(t)$, $\xi(t)$ and $\lambda(t) > 0$ are now continuous functions. The process Z is still a max-stable process. If Z is also stationary, the process may be expressed through it's *spectral representation* de Haan and Pickands (1986).

In detail, let $\{X_j, U_j\}_{j\geq 1}$ be a Poisson process, Π, on $\mathbb{R}^n \times \mathbb{R}_+$, with counting measure $\Pi(\cdot) := \Sigma_j \mathbb{I}_{(X_j, U_j)}(\cdot)$ and intensity measure $\nu(dx) \times u^{-2}du$, where $\mathbb{I}_{(X_j, U_j)}(A)$ is the indicator function of the random number of points falling in a bounded set $A \subset \mathbb{R}^n \times \mathbb{R}_+$ and ν is a positive measure. For a nonnegative measurable (for fixed $t \in T$) function $f(x-t)$ such that $\int_{\mathbb{R}^n} f(x-t)\nu(dx) = 1, \forall t \in T$ the stochastic process

$$Z(t) := \max_{j=1,2,\ldots} \{U_j f(X_j - t)\}, \quad t \in T, \tag{1}$$

is a stationary max-stable process de Haan (1984). Smith (1990) reinterprets this process in terms of environmental episodes such as storm phenomena, in which U, X and f represent respectively storm magnitude, the center and the shape. Schlather and Tawn (2003) term this the *storm profile model*.

For a finite set of indexes $t_1, \ldots, t_K \in T$ and positive thresholds $z_1, \ldots, z_K$ for $K \in \mathbb{N}$, the distribution of the random vector $Z(t_1), \ldots, Z(t_K)$ is de Haan (1984)

$$\Pr\{Z(t_k) \leq z_k,\, k = 1, \ldots, K\} = \exp\left[-\int_{\mathbb{R}^n} \max_{1 \leq k \leq K} \left\{\frac{f(x - t_k)}{z_k}\right\} \nu(dx)\right]. \quad (2)$$

It then follows that the marginal distributions are unit Fréchet:

$$\Pr\{Z(t) \leq z\} = \exp\left(-z^{-1} \int_{\mathbb{R}^n} f(x - t)\nu(dx)\right) = \exp(-1/z).$$

Alternative spectral representations of max-stable processes exist (Schlather, 2002).

2.2 Extremal coefficients

Given n independent realisations of a random vector $Y \in \mathbb{R}^d$, the joint distribution of componentwise maxima satisfies (De Haan and Resnick, 1977; Resnick, 1987)

$$\Pr\left\{\max_k \max_{j=1,\ldots,n} Y_k^{(j)}/n \leq z\right\} = \Pr\left\{\max_{j=1,\ldots,n} Y_1^{(j)}/n \leq z\right\}^\theta = \exp(-\theta/z), \quad z > 0,$$

for $k = 1, 2, \ldots, K$ and common threshold z, where the rightmost term is a Fréchet(θ) distribution. The parameter $1 \leq \theta \leq K$ is the *extremal coefficient* and it measures the extremal dependence between the margins, an important practical quantity in applications Smith (1990). The information in the extremal coefficient reflects the practical number of independent variables. If K is finite then $\theta = 1$ indicates complete dependence, whereas $\theta = K$ demonstrates full independence.

In the max-stable process framework, from (2), for all $z > 0$ we have

$$\Pr\{Z(t_k) \leq z,\, k = 1, \ldots, K\} = \exp(-\theta/z),$$

and so

$$\theta = \int_{\mathbb{R}^n} \max_{1 \leq k \leq K} \{f(x - t_k)\} \nu(dx).$$

where θ again represents the effective number of independent variables. Schlather and Tawn (2003) discuss the extremal coefficient within a max-stable context.

2.3 Spatial models

Suppose now that $T \subseteq \mathbb{R}^2$ and that $\{X_j\}_{j \geq 1}$ are random points in $\mathbb{R}^2$. While for $K > 2$, the general K-dimensional distribution function under the max-stable process representation (1) permits no analytically tractable form, a class of bivariate spatial models is available when the storm profile model, f, is a bivariate Gaussian density

and μ is a Lebesgue measure (Smith, 1990; De Haan and Pereira, 2006). In this case, for locations $\mathbf{t}_i$ and $\mathbf{t}_j$ the bivariate distribution function of $\{Z(\mathbf{0}), Z(\mathbf{h})\}$ is

$$
\begin{aligned}
&\Pr\{Z(\mathbf{0}) \leq z_i, Z(\mathbf{h}) \leq z_j\} \\
&= \exp\left[-\frac{1}{z_i}\Phi\left(\frac{a(\mathbf{h})}{2} + \frac{1}{a(\mathbf{h})}\log\frac{z_j}{z_i}\right) - \frac{1}{z_j}\Phi\left(\frac{a(\mathbf{h})}{2} + \frac{1}{a(\mathbf{h})}\log\frac{z_i}{z_j}\right) \right],
\end{aligned} \tag{3}
$$

where $\mathbf{h} = (\mathbf{t}_j - \mathbf{t}_i)^\top$, $\mathbf{0}$ is the origin, Φ is the standard Gaussian distribution function, $a(\mathbf{h}) = (\mathbf{h}^T \boldsymbol{\Sigma}^{-1}\mathbf{h})^{1/2}$ and $\boldsymbol{\Sigma}$ is the covariance matrix of f, with covariance σ_{12} and standard deviations $\sigma_1, \sigma_2 > 0$. A derivation of (3) is in Appendix A.2. A general max-stable process with a Gaussian storm profile model, f, is termed a *Gaussian extreme value process* Smith (1990), whereas the specific model (3) is the *Gaussian extreme value model* Coles (1993).

Second-order partial derivatives of (3) yield the 2-dimensional density function

$$
\begin{aligned}
f(z_i, z_j) = \exp &\left\{ -\frac{\Phi(w(\mathbf{h}))}{z_i} - \frac{\Phi(v(\mathbf{h}))}{z_j} \right\} \left\{ \left(\frac{\Phi(w(\mathbf{h}))}{z_i^2} + \frac{\varphi(w(\mathbf{h}))}{a(\mathbf{h})z_i^2} - \frac{\varphi(v(\mathbf{h}))}{a(\mathbf{h})z_i z_j} \right) \right. \\
&\left. \left(\frac{\Phi(v(\mathbf{h}))}{z_j^2} + \frac{\varphi(v(\mathbf{h}))}{a(\mathbf{h})z_j^2} - \frac{\varphi(w(\mathbf{h}))}{a(\mathbf{h})z_i z_j} \right) + \left(\frac{v(\mathbf{h})\,\varphi(w(\mathbf{h}))}{a(\mathbf{h})^2 z_i^2 z_j} + \frac{w(\mathbf{h})\,\varphi(v(\mathbf{h}))}{a(\mathbf{h})^2 z_i z_j^2} \right) \right\},
\end{aligned} \tag{4}
$$

where φ is the standard Gaussian density function, $w(\mathbf{h}) = a(\mathbf{h})/2 + \log(z_j/z_i)/a(\mathbf{h})$ and $v(\mathbf{h}) = a(\mathbf{h}) - w(\mathbf{h})$. The derivation of (4) is in Appendix A.3.

Observe that $a(\mathbf{h})$ measures the strength of extremal dependence: $a(\mathbf{h}) \to 0$ represents complete dependence, and (in the limit) $a(h) \to \infty$ indicates complete independence. In accordance with spatial models, the extreme dependence between $Z(\mathbf{0})$ and $Z(\mathbf{h})$ decreases monotonically and continuously with $\mathbf{h} = \|\mathbf{t}_j - \mathbf{t}_i\|$ De Haan and Pereira (2006), and for fixed $\mathbf{h}$ the dependence decreases monotonically as $a(\mathbf{h})$ increases. Accordingly, characterisation of extremal dependence is determined by the covariance, $\boldsymbol{\Sigma}$, which is therefore of interest for inference.

Due to high-dimensional distributional complexity the study of extremal dependence is commonly limited to pairwise components through the extremal coefficients

$$
\theta(\mathbf{h}) = \int_{\mathbb{R}^2} \max\{f(\mathbf{x}), f(\mathbf{x} - \mathbf{h})\}d\mathbf{x}, \qquad 1 \leq \theta(\mathbf{h}) \leq 2.
$$

The dependence on $\mathbf{h}$ is explicit. Specifically for the Gaussian extreme value model, $\theta(\mathbf{h}) = 2\Phi(a(\mathbf{h})/2)$, following an argument along the lines of Appendix A.2. Alternative models result by considering e.g. exponential or t storm profile models De Haan and Pereira (2006), or stationary Gaussian process profile models Schlather (2002).

3 Likelihood-based inference

The analysis of spatial extremes is concerned with the joint modelling of a spatial process at large numbers of data-recording stations in a fixed region. As discussed in Section 2, the lack of closed-form distribution for max-stable processes in greater than $K = 2$ dimensions precludes straightforward use of standard maximum likelihood methods for this class of models. We now develop inferentially practical, likelihood-based classes of max-stable processes derived from a *composite-likelihood* approximation (Lindsay, 1988; Varin, 2008). The procedure is sufficiently reliable and versatile to permit the simultaneous and consistent modelling of joint and marginal parameters in the spatial context at a moderate computational cost.

3.1 Composite likelihoods

For a parametric statistical model $\mathcal{F}$ with density function family $\mathcal{F} = \{f(\mathbf{y}; \boldsymbol{\psi}), \mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^K, \boldsymbol{\psi} \in \Psi \subseteq \mathbb{R}^d\}$, and a set of marginal or conditional events $\{\mathcal{I}_k : k \in \mathcal{K}\}$ (for some $\mathcal{K} \subseteq \mathbb{N}$) subset of some sigma algebra on $\mathcal{Y}$, the composite log-likelihood is defined by

$$\ell_{\mathcal{C}}(\boldsymbol{\psi}; \mathbf{y}) = \sum_{k \in \mathcal{K}} \log f(\mathbf{y} \in \mathcal{I}_k; \boldsymbol{\psi}),$$

where $\log f(\mathbf{y} \in \mathcal{I}_k; \boldsymbol{\psi})$ is the log-likelihood associated with event $\mathcal{I}_k$. First-order partial derivatives of $\ell_{\mathcal{C}}(\boldsymbol{\psi}; \mathbf{y})$ with respect to $\boldsymbol{\psi}$ yield the *composite score function* $\mathsf{D}_{\boldsymbol{\psi}} \ell_{\mathcal{C}}(\boldsymbol{\psi}; \mathbf{y})$, from which maximum composite likelihood estimator of $\boldsymbol{\psi}$, if unique, is obtained by solving $\mathsf{D}_{\boldsymbol{\psi}} \ell_{\mathcal{C}}(\widehat{\boldsymbol{\psi}}_{\mathrm{MCLE}}; \mathbf{y}) = 0$. Similarly, second-order partial derivatives of $\mathsf{D}_{\boldsymbol{\psi}} \ell_{\mathcal{C}}(\boldsymbol{\psi}; \mathbf{y})$ yield the Hessian matrix $\mathsf{H}_{\boldsymbol{\psi}} \ell_{\mathcal{C}}(\boldsymbol{\psi}; \mathbf{y})$ (Appendix A.1).

The key utility of the composite log-likelihood is it's ability, under the usual regularity conditions, to provide consistent and unbiased parameter estimates when standard likelihood estimators are not available. Under appropriate conditions (Lindsay, 1988; Cox and Reid, 2004) the maximum composite likelihood estimator is consistent and asymptotically distributed as

$$\widehat{\boldsymbol{\psi}}_{\mathrm{MCLE}} \,\dot\sim\, \mathrm{N}(\boldsymbol{\psi}, \tilde{\mathsf{I}}(\boldsymbol{\psi})^{-1}) \quad \text{with} \quad \tilde{\mathsf{I}}(\boldsymbol{\psi}) = \mathsf{H}(\boldsymbol{\psi}) \, \mathsf{J}(\boldsymbol{\psi})^{-1} \, \mathsf{H}(\boldsymbol{\psi}),$$

where $\mathsf{H}(\boldsymbol{\psi}) = \mathbb{E}\{-\mathsf{H}_{\boldsymbol{\psi}} \ell_{\mathcal{C}}(\boldsymbol{\psi}; \mathbf{Y})\}$ and $\mathsf{J}(\boldsymbol{\psi}) = \mathbb{V}\{\mathsf{D}_{\boldsymbol{\psi}} \ell_{\mathcal{C}}(\boldsymbol{\psi}; \mathbf{Y})\}$ are analogues of the expected information matrix and the variance of the score vector. Although the maximum composite likelihood estimator can be unbiased, it may not be asymptotically efficient in that $\tilde{\mathsf{I}}(\boldsymbol{\psi})^{-1}$, the inverse of the Godambe information matrix Godambe (1960), may not attain the Cramér–Rao bound Cox and Reid (2004).

3.2 The pairwise setting for spatial extremes

Recall, for the spatial setting we have observations $\{y_{n,k}\}$, each denoting the maximum of m samples over $n = 1, \ldots, N$ blocks and $k = 1, \ldots K$ locations in a continuous

region. E.g. for daily observations, $m = 366$ implies the $y_{n,k}$ describe annual maxima. Accordingly, the K univariate marginals are approximately GEV distributed. Despite the intractability of the multivariate max-stable process, availability of the bivariate form (3) implies a pairwise composite log-likelihood may be constructed as

$$\ell_{\mathcal{P}}(\boldsymbol{\psi}; \mathbf{y}) = \sum_{n=1}^{N} \sum_{i=1}^{\mathcal{K}} \sum_{j=i+1}^{\mathcal{K}-1} \log f(y_{n,i}, y_{n,j}; \boldsymbol{\psi}), \tag{5}$$

where each $f(y_{n,i}, y_{n,j}; \boldsymbol{\psi})$ is a bivariate marginal density based on data at locations i and j, taken over all distinct location pairs.

In order to characterized limiting behavior by a max-stable process (1) we require unit Fréchet marginal distributions. Accordingly, we consider the bijection $(Y_i, Y_j) = g(Z_i, Z_j)$ with inverse function given by

$$Z_i = \left(1 + \frac{\xi_i(Y_i - \mu_i)}{\lambda_i}\right)_{+}^{1/\xi_i} \qquad Z_j = \left(1 + \frac{\xi_j(Y_j - \mu_j)}{\lambda_j}\right)_{+}^{1/\xi_j} \tag{6}$$

where $Z_i \equiv Z(t_i)$ and $Y_j \equiv Y(t_j)$, and for each marginal Y, the constants μ, ξ and $\lambda > 0$ ensure that Z is unit Fréchet distributed. The resulting bivariate density is

$$f_{Y_i,Y_j}(y_i, y_j) = f_{Z_i,Z_j}\left[g^{-1}(y_i, y_j)\right] |J(y_i, y_j)|,$$

where $f_{Z_i,Z_j}(z_i, z_j)$ denotes the density of the Gaussian extreme value model (4), and

$$|J(y_i, y_j)| = \frac{1}{\lambda_i \lambda_j} \left(1 + \frac{\xi_i(y_i - \mu_i)}{\lambda_i}\right)_{+}^{1/\xi_i - 1} \left(1 + \frac{\xi_j(y_j - \mu_j)}{\lambda_j}\right)_{+}^{1/\xi_j - 1}.$$

This change of variable permits the use of GEV marginals (over unit Fréchet) without reforming the problem definition. Hence, the pairwise log-likelihood (5) allows simultaneous assessment of the tail dependence parameters (3) between pairs of sites and also the location, scale and shape parameters of the marginal distribution at each location. The parameters can not be estimated as an analytical solution of the composite score equation. Nonetheless quasi-Newton numerical maximization routines (e.g. Broyden, 1967) can be applied in order to obtain maximum likelihood estimates.

Variances of parameter estimates are provided through inverse of the Godambe information matrix, with estimates of the matrices $H(\boldsymbol{\psi})$ and $J(\boldsymbol{\psi})$ given by

$$\widehat{H}(\widehat{\boldsymbol{\psi}}_{\mathrm{MCLE}}) = -\sum_{n=1}^{N} \sum_{i=1}^{\mathcal{K}} \sum_{j=i+1}^{\mathcal{K}-1} \mathsf{H}_{\boldsymbol{\psi}} \log f(y_{n,i}, y_{n,j}; \widehat{\boldsymbol{\psi}}_{\mathrm{MCLE}})$$

and

$$\widehat{J}(\widehat{\boldsymbol{\psi}}_{\mathrm{MCLE}}) = \sum_{n=1}^{N} \sum_{i=1}^{\mathcal{K}} \sum_{j=i+1}^{\mathcal{K}-1} \mathsf{D}_{\boldsymbol{\psi}} \log f(y_{n,i}, y_{n,j}; \widehat{\boldsymbol{\psi}}_{\mathrm{MCLE}}) \, \mathsf{D}_{\boldsymbol{\psi}} \log f(y_{n,i}, y_{n,j}; \widehat{\boldsymbol{\psi}}_{\mathrm{MCLE}})^{\top},$$

each evaluated at the composite maximum likelihood value. In practice the matrix $\widehat{\mathrm{H}}$ is obtained straightforwardly with the numerical maximization routine employed for likelihood maximization. An explicit expression for $\widehat{\mathrm{J}}$ is derived in Appendix A.4.

In principle, estimating unique marginal parameters for each location ensures correct model application by respecting marginal constraints, though computational issues arise for large numbers of parameters. Alternatively, as is common in the modelling of univariate extremes, we may describe the GEV parameters through parsimonious regression models, which may be functions of space, environmental and other covariates and random effects. Specifically, we may express each parameter as

$$\eta(\mathbf{x}) \equiv h\{f(\mathbf{x})\} = \mathbf{X}\boldsymbol{\beta}, \tag{7}$$

where h is a link function, $\mathbf{x} = (x_1, \ldots, x_d)$ is a vector of predictors, $\mathbf{X}$ is a $N \times (d+1)$ design matrix, and $\boldsymbol{\beta}$ is a $(d+1) \times 1$ vector of unknown parameters.

For further flexibility, a non-parametric approach may provide a useful alternative. Non-parametric modelling of univariate extreme value responses has been recently proposed by Chavez-Demoulin and Davison (2005), Yee and Stephenson (2007) and Padoan and Wand (2008). The work of Kammann and Wand (2003), who developed a non-parametric spatial regression with Gaussian response, may be extended to the current spatial extremes setting.

3.3 Model selection

There are two model selection approaches under the composite likelihood framework. For nested models Varin (2008) the p-dimensional parameter, $\boldsymbol{\psi}$ is partitioned as $\boldsymbol{\psi} = (\boldsymbol{\psi}', \boldsymbol{\psi}'')$, where $\boldsymbol{\psi}'$ is q-dimensional, and testing $\boldsymbol{\psi}' = \boldsymbol{\psi}_0$ versus a two-sided hypothesis proceeds via the composite likelihood ratio test statistic

$$W(\boldsymbol{\psi}_0) = 2\{\ell_{\mathcal{C}}(\widehat{\boldsymbol{\psi}}) - \ell_{\mathcal{C}}(\boldsymbol{\psi}_0, \widehat{\boldsymbol{\psi}}''(\boldsymbol{\psi}_0))\}.$$

Under the null Kent (1982)

$$W \overset{.}{\sim} \sum_{j=1}^{q} \nu_i \chi_i^2, \tag{8}$$

where χ_i^2 are independent χ_1^2 random variables, and $\nu_1 \geq \ldots \geq \nu_q$ are the eigenvalues of $(\mathrm{H}^{\boldsymbol{\psi}'\boldsymbol{\psi}'})^{-1}\widetilde{\mathrm{I}}_{\boldsymbol{\psi}'\boldsymbol{\psi}'}$. Here, $\mathrm{H}^{\boldsymbol{\psi}'\boldsymbol{\psi}'}$ and and $\widetilde{\mathrm{I}}_{\boldsymbol{\psi}'\boldsymbol{\psi}'}$ respectively denote the information matrix and the Godambe information matrix, each restricted to those elements associated with parameter $\boldsymbol{\psi}'$. (Dependence on $\boldsymbol{\psi}$ under the null is omitted for brevity.) Hypothesis testing based on (8) either approximates null distribution using estimates of the eigenvalues ν_i Rotnitzky and Jewell (1990), or adjusts the composite likelihood such that the usual asymptotic χ_q^2 null is preserved Chandler and Bate (2007).

The *composite likelihood information criterion* (CLIC) Varin and Vidoni (2005), useful in the case of non-nested models, performs model selection on the basis of expected Kullback–Leibler divergence between the true unknown model and the adopted

8

model (Davison, 2003, p. 123). In the composite likelihood context, this is the AIC under model mis-specification (Takeuchi, 1976); (Davison, 2003, p. 150–152). Model selection is based on the model minimising

$$-2\{\ell_{\mathcal{C}}(\widehat{\boldsymbol{\psi}}_{\mathrm{MCLE}}; \mathbf{Y}) - \mathrm{tr}[\widehat{\mathrm{J}}(\widehat{\boldsymbol{\psi}}_{\mathrm{MCLE}})\,\widehat{\mathrm{H}}(\widehat{\boldsymbol{\psi}}_{\mathrm{MCLE}})^{-1}]\},$$

where the second term is the usual composite log-likelihood penalty term.

4   Simulation Study

We now evaluate the utility of the composite likelihood in the spatial extremes context. We examine various forms of extremal dependence with the Gaussian storm profile (4) characterised through the covariance, $\boldsymbol{\Sigma}$, including directional and strength of dependence variations (Table 1 and Figure 1). The covariance has direct meteorological interpretation and defines the extremal dependence directly. The K site locations are uniformly generated over a 40×40 region. Given the moderate computational demand for large site numbers, likelihood maximisation (and other) routines have been implemented in C and collected in the forthcoming R package SpatialExtremes.

Table 1: *Extremal dependence configurations.*

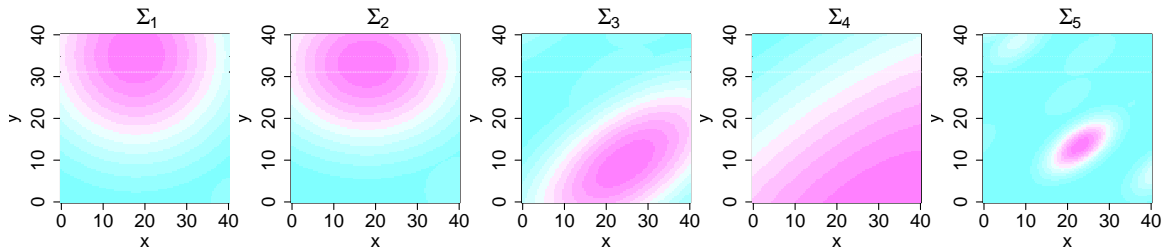|	Spatial dependence structure	σ_1^2	σ_2^2	σ_{12}
$\boldsymbol{\Sigma}_1$:	Same strength in both directions	300	300	0
$\boldsymbol{\Sigma}_2$:	Different strength in both directions	200	300	0
$\boldsymbol{\Sigma}_3$:	Spatial correlation	200	300	150
$\boldsymbol{\Sigma}_4$:	Strong dependence	2000	3000	1500
$\boldsymbol{\Sigma}_5$:	Weak dependence	20	30	15

Figure 1: *A Gaussian extreme value process realisation for* $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_5$. *Max-stable process simulation routines are available in the* RandomFields *package in* R *(Schlather, 2002).*

Table 2 summarizes estimator performance based on moderately sized datasets: $K = 50$ sites and $N = 100$ observations. Estimate means and standard errors over 500 data replications are reported, indicating good correspondence to the true values. There is no evidence of bias, even in cases where poorer performance may be expected

9

Table 2: *Composite MLE's based on 500 spatial extreme data simulations ($K = 50$ sites and $N = 100$) with the Gaussian extreme value model. True values are in [brackets]. Standard errors obtained through Godambe estimates and sample standard deviation (in parantheses).*

	$\widehat{\sigma}_1^2$ / s.e.	$\widehat{\sigma}_2^2$ / s.e.	$\widehat{\sigma}_{12}^2$ / s.e.
Σ_1:	306 [300] / 40.6 (44.7)	306 [300] / 39.8 (41.5)	1 [0] / 27.9 (27.7)
Σ_2:	204 [200] / 26.7 (28.5)	305 [300] / 39.6 (39.7)	1 [0] / 21.9 (21.2)
Σ_3:	202 [200] / 25.1 (26.1)	300 [300] / 37.3 (37.9)	150 [150] / 25.5 (26.1)
Σ_4:	2053 [2000] / 495.2 (300.1)	3065 [3000] / 664.8 (483.1)	1550 [1500] / 412.0 (322.4)
Σ_5:	20 [20] / 1.5 (1.6)	30 [30] / 2.3 (2.3)	15 [15] / 1.6 (1.6)

such as very strong or weak dependence. Overall, the Godambe standard errors and sample standard deviations are consistent, though there is some discrepancy in the case of strong dependence. Here, the eigenvalues of Σ_4^{-1} are $5 \pm \sqrt{10}/7500$ and so we are near the parameter space boundary. Consequently, for some of the 500 replications, the asymptotic normality of $\widehat{\psi}_{\mathrm{MCLE}}$ fails, and so the Godambe standard errors are not relevant.

Table 3: *Normalised mean squared error of extremal coefficient estimates based on 500 data simulations with $K = 50$ sites and $N = 100$ observations. Estimators are the composite MLE and those proposed by Smith (1990) and Schlather and Tawn (2003). Standard deviations are reported in parantheses. The values' order of magnitude is 10^{-4}.*

	Composite MLE	Smith	Schlather & Tawn
Σ_1:	3.1 (5.4)	18.4 (26.7)	17.3 (29.9)
Σ_2:	2.9 (4.8)	19.1 (30.3)	18.9 (28.6)
Σ_3:	2.5 (4.7)	21.9 (35.7)	21.1 (32.4)
Σ_4:	3.0 (35.6)	10.9 (18.9)	6.7 (13.3)
Σ_5:	0.3 (0.8)	30.3 (44.6)	25.5 (40.7)

Table 3 depicts normalised mean squared errors for three different estimators of the extremal coefficient functions: the composite MLE derived from the Gaussian extreme value model, and those proposed by Smith (1990) and Schlather and Tawn (2003). Normalized mean square errors are used to prevent the largest extremal coefficients from dominating. From Table 3, in general it is clear that the composite likelihood estimator is the most accurate. The standard deviation in the strong dependence case again suffers from the failure of the asymptotic normality of $\widehat{\psi}_{\mathrm{MCLE}}$.

Estimator performance for a range of dataset sizes ($N = 10, 50, 100, 500$) and site numbers ($K = 10, 50, 100$) is listed in Table 4, under 500 data replications of spatial dependence model Σ_3. As expected, the simulations indicate that there is some bias and larger variance for small N, and negligible bias and small variance for large N. Observe that, for fixed sample size, the number of sites does not impact the estimation results. This behavior is illustrated in Figure 2, highlighting, in particular, good estimator performance with increasing N.

Table 4: *Composite MLEs for a varying number of sites (K) and observations (N), based on 500 simulations of spatial extreme data using the Gaussian extreme value model with covariance Σ_3 (Table 1). Standard deviations are reported in parantheses.*

	K	$N = 10$	$N = 50$	$N = 100$	$N = 500$	True
$\widehat{\sigma}_1^2$	10	245 (120.2)	207 (43.8)	205 (31.7)	199 (13.3)	200
	50	244 (90.4)	208 (37.5)	200 (28.3)	199 (11.4)	
	100	239 (94.3)	205 (37.8)	202 (30.4)	199 (11.5)	
$\widehat{\sigma}_2^2$	10	353 (159.1)	305 (63.9)	301 (44.8)	298 (19.5)	300
	50	353 (131.6)	309 (56.6)	303 (44.3)	298 (16.9)	
	100	361 (143.4)	307 (59.5)	301 (44.8)	299 (16.7)	
$\widehat{\sigma}_{12}$	10	174 (108.9)	153 (41.8)	151 (31.9)	149 (13.4)	150
	50	179 (91.2)	156 (38.8)	149 (28.6)	149 (11.4)	
	100	181 (100.9)	154 (39.2)	150 (29.7)	150 (11.2)	

Finally, we consider model selection under misspecification. Figure 3 illustrates power curves (likelihood ratio tests) and rejection rates (CLIC) for two hypotheses, each versus their complement. Namely, $H_0 : \sigma_1^2 = 200$ fixing $\sigma_2^2 = 300$ and $\sigma_{12} = 150$ (top plots), and $H_0 : \sigma_1^2 = \sigma_2^2 = 200$ fixing $\sigma_{12} = 0$ (bottom plots). All resulting curves are near-quadratic and, for the pairwise likelihood ratio based tests, rejection rates are close to the confidence level $\alpha = 0.05$ when the null hypothesis is true. In this study, contrary to the results derived by Chandler and Bate (2007), the adjustment of the W statistic (8) Rotnitzky and Jewell (1990) appears to have slightly more power – even when testing mutiple parameters. In contrast the CLIC statistic demonstrates poor performance, with the rejection rate under the true model reaching only 20%.

5 Application to U.S. precipitation data

We illustrate the developed methodology in an analysis of U.S. precipitation data. These data consist of 46 gauging stations with daily rainfall records over a period of 91 years (Figure 4). We express the model GEV parameters as simple linear functions of longitude, lattitude and altitude. Further model variations, including additional environmental covariates, and regressions of the covariance Σ on these covariates to examine spatial variations in extremal dependence were not considered.

Exploratory analyses of the functional complexity of the GEV parameters was performed by fitting independent GEV models to the data from each station and evaluating appropriate surface responses using standard techniques (e.g. ANOVA, Fisher tests). After identifying an upper limit to model complexity, we perform model selection on the pairwise composite likelihood of the Gaussian extreme value process using the methodologies outlined in Section 3.3.

Table 5 summarizes some of the different models investigated. According to the CLIC criterion, model M_5 (in bold) is the preferred choice, although models M_2 and

Table 5: *Some Gaussian extreme value processes and their corresponding maximised negative composite log-likelihood, degrees of freedom and the CLIC score.*

| Model		$-\ell_{\mathcal{P}}(\widehat{\boldsymbol{\psi}}_{\mathrm{MCLE}}; \mathbf{y})$	d.f.	CLIC
M_0 :	$\mu(x) = \alpha_0 + \alpha_1(\mathrm{lat}) + \alpha_2(\mathrm{alt}) + \alpha_3(\mathrm{lon})$ $\sigma(x) = \beta_0 + \beta_1(\mathrm{lat}) + \beta_2(\mathrm{alt}) + \beta_2(\mathrm{lon})$ $\xi(x) = \gamma_0$	412,110.2	12	12,848,229
M_1 :	$\mu(x) = \alpha_0 + \alpha_1(\mathrm{lat}) + \alpha_2(\mathrm{alt})$ $\sigma(x) = \beta_0 + \beta_1(\mathrm{lat}) + \beta_2(\mathrm{alt}) + \beta_3(\mathrm{lon})$ $\xi(x) = \gamma_0$	412,110.9	11	16,096,068
M_2 :	$\mu(x) = \alpha_0 + \alpha_1(\mathrm{lat}) + \alpha_2(\mathrm{alt}) + \alpha_3(\mathrm{lon})$ $\sigma(x) = \beta_0 + \beta_1(\mathrm{lat}) + \beta_2(\mathrm{alt})$ $\xi(x) = \gamma_0$	412,113.3	11	1,008,997
M_3 :	$\mu(x) = \alpha_0 + \alpha_1(\mathrm{lat}) + \alpha_3(\mathrm{lon})$ $\sigma(x) = \beta_0 + \beta_1(\mathrm{lat}) + \beta_2(\mathrm{alt}) + \beta_3(\mathrm{lon})$ $\xi(x) = \gamma_0$	412,234.1	11	1,926,389,242
M_4 :	$\mu(x) = \alpha_0 + \alpha_1(\mathrm{lat}) + \alpha_2(\mathrm{alt}) + \alpha_3(\mathrm{lon})$ $\sigma(x) = \beta_0 + \beta_1(\mathrm{lat}) + \beta_3(\mathrm{lon})$ $\xi(x) = \gamma_0$	412,380.5	11	33,209,042
$\mathbf{M_5}$:	$\mu(\mathbf{x}) = \alpha_{\mathbf{0}} + \alpha_{\mathbf{1}}(\mathrm{lat}) + \alpha_{\mathbf{2}}(\mathrm{alt})$ $\sigma(\mathbf{x}) = \beta_{\mathbf{0}} + \beta_{\mathbf{1}}(\mathrm{lat}) + \beta_{\mathbf{2}}(\mathrm{alt})$ $\xi(\mathbf{x}) = \gamma_{\mathbf{0}}$	**412,113.3**	**10**	**1,008,261**
M_6 :	$\mu(x) = \alpha_0 + \alpha_1(\mathrm{lat})$ $\sigma(x) = \beta_0 + \beta_1(\mathrm{lat}) + \beta_2(\mathrm{alt})$ $\xi(x) = \gamma_0$	412,237.3	9	1,086,347
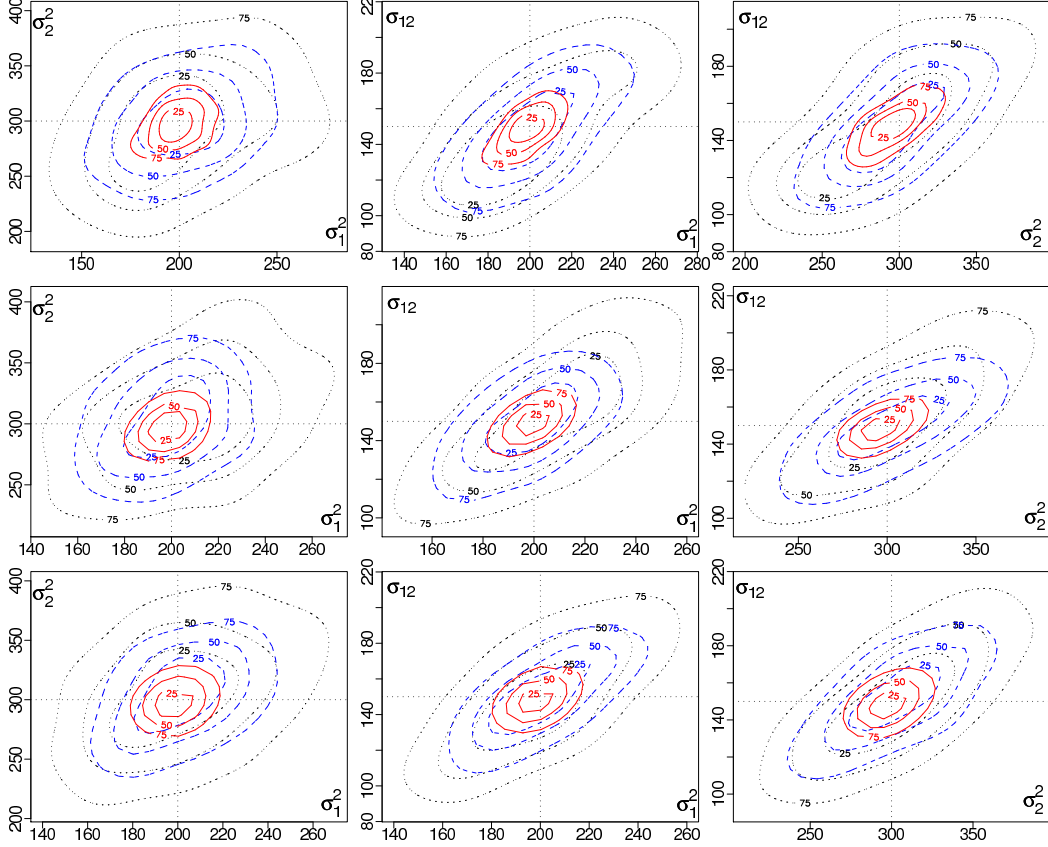
Figure 2: *Composite MLE distribution estimates for $N = 50$ (dotted line), $N = 100$ (broken line) and $N = 1000$ (solid line) observations over $K = 10$ (top row), $K = 50$ (centre row) and $K = 100$ (bottom row) sites. Contours correspond to 0.25, 0.5 and 0.75 percentiles.*

M_6 also appear competitive. Despite the relatively poor performance of the CLIC criterion for model selection (see Section 4), the same conclusions are obtained from the deviance based tests, with p-values around 0.9. For model M_5, the covariance $\boldsymbol{\Sigma}$ is estimated as $\sigma_1^2 = 0.06323$ (0.06463), $\sigma_{12} = 0.01334$ (0.03661) and $\sigma_2^2 = 0.02581$ (0.02538), where the parantheses indicate standard errors.

Figure 5 (right plot) illustrates the spatial variation of pointwise 50-year return level estimates. Comparison to the regional elevation map (left plot) indicates that the most extreme precipitation events occur in mountainous regions. Figure 6 (left plot) also depicts the strength of spatial dependence through the extremal coefficient function. There is clear evidence of anisotropy, with stronger dependence in the north-east/south-west direction. Interestingly, this axis corresponds to the shape of the Appalachian Mountains as well as the coastline. Accordingly, this directional extremal dependence may be the consequence of storms following either the coastline or the massif. Finally, conditioning of a fixed site (and for a given threshold) by using the pairwise conditional distribution, the conditional q-year return level estimates can
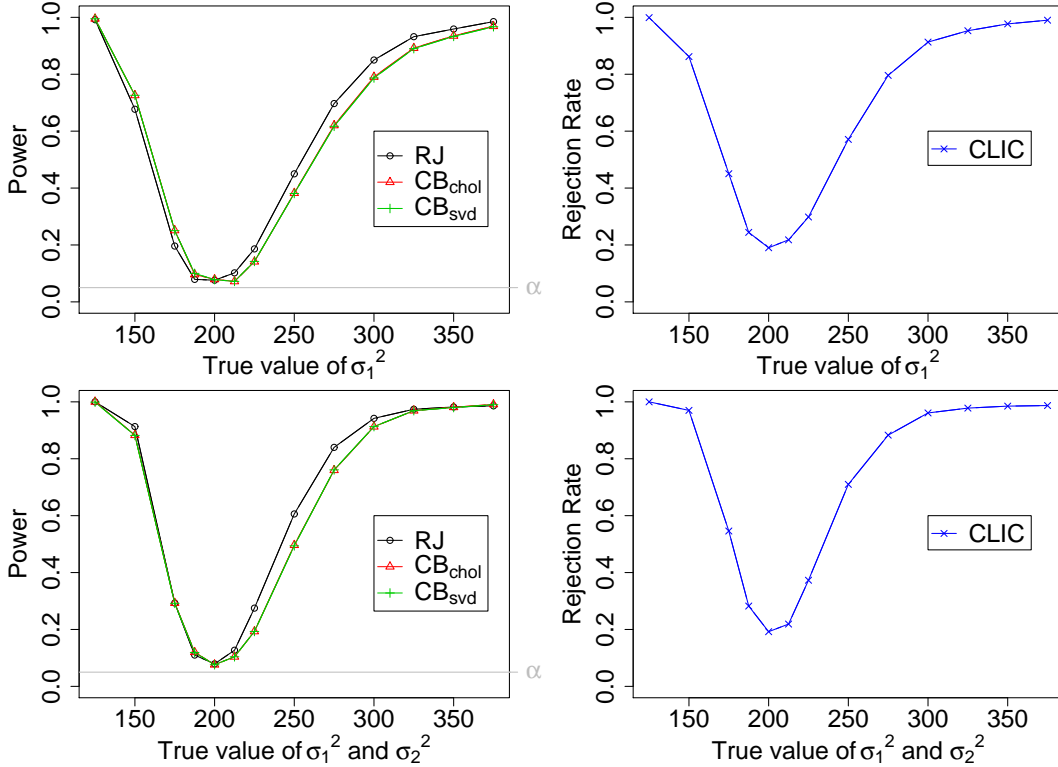
Figure 3: *Left panels: Power curves for the composite likelihood ratio tests; RJ (Rotnitzki and Jewell, 1990), and CB_{chol} and CB_{svd} (Chandler and Bate, 2007) using Cholesky and singular value decompositions. Right panels: CLIC rejection rates. Test levels are $\alpha = 0.05$. Point estimates are based on 1000 data replications.*

easily be provided. Figure 6 (right plot) illustrates the spatial variation of conditional pointwise 50-year return level estimates, given the fixed site indicated by the star.

6   Conclusion

As a natural generalisation of extremal dependence structures, max-stable processes are a powerful tool for the modelling of multivariate extremes. Unfortunately, the intractability of the multivariate density function precludes inference except in trivial cases (e.g. bivariate), or requires additional approximations and immense computational overheads (Jiang and Turnbull, 2004; Sisson et al., 2007; Peters et al., 2008).

This article has developed composite likelihood-based inferential methods for general max-stable processes. Our results demonstrate good applicability in the spatial context. The benefits of this likelihood-based approach are the flexible joint modelling of marginal and dependence parameters, coupled with good estimator behaviour with finite samples, all at moderate computational cost.

Modifications of the model formulation would draw alternative representations of
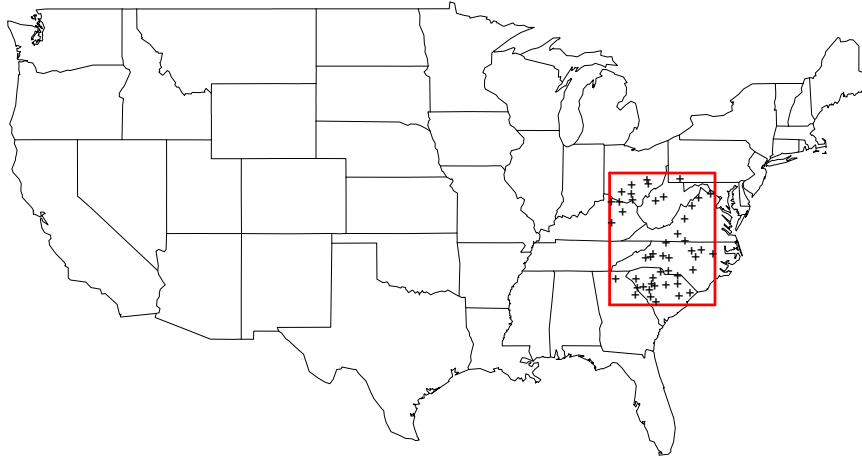
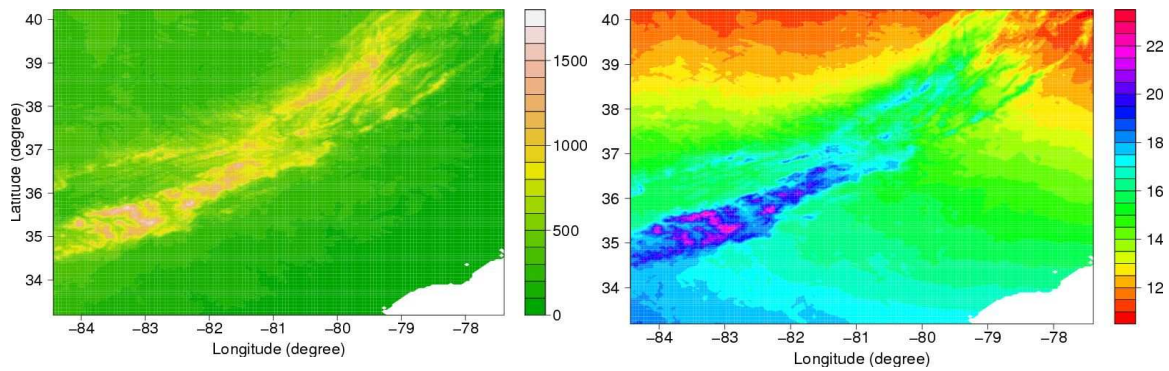Figure 4: *Locations of the 46 gauging stations.*

Figure 5: *Left: Elevation map (metres) of the region; Right: Pointwise 50-year return level map (cm) estimated from the fitted Gaussian extreme value process.*
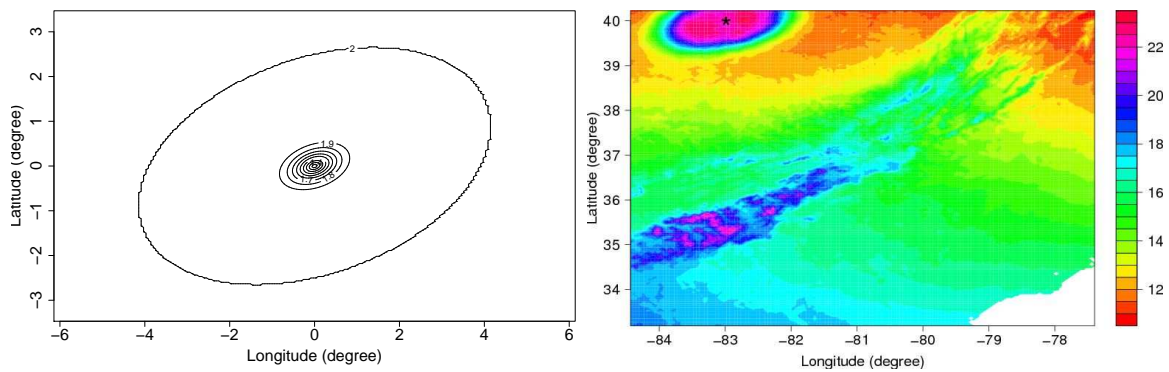

Figure 6: *Left: Contour plot of the extremal coefficient; Right: Pointwise 50-year, conditional return level map (cm) estimated from the fitted Gaussian extreme value process. The star indicates the fixed site used in the computation of the conditional return levels.*

15

extremal modelling into the composite-likelihood based framework, given the known links between these and block maxima (GEV) approaches (e.g. Coles (2001)). These include threshold excess models for marginals (Davison and Smith, 1990), and the limiting Poisson characterisation of extremes. The obvious practical benefit from these extensions would be the incorporation of more data into the modelling process.

Acknowledgements

References

Barndorff-Nielsen, O. E., Gupta, V. K., Pérez-Abreu, V., and Weymire, E., editors (1998). *Stochastic Methods in Hyrdology*. World Scienfitic, Singapore, New Jersey, London, Hong Kong.

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

Chavez-Demoulin, V. and Davison, A. C. (2005). Generalised additive modelling of sample extremes. *Applied Statistics*, 54:207–222.

Coles, S. G. (1993). Regional modelling of extreme storms via max-stable processes. *J. Roy. Statist. Soc. B*, 55:797–816.

Coles, S. G. (2001). *An Introduction to Statistical Modelling of Extreme Values*. Springer, London.

Coles, S. G. and Tawn, J. A. (1996). Modelling extremes of the areal rainfall process. *J. Roy. Statist. Soc. B*, 58:329–347.

Coles, S. G. and Walshaw, D. (1994). Directional modelling of extreme wind speeds. *Journal of Applied Statistics*, 33:139–158.

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

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York.

Davison, A. C. (2003). *Statistical Models*. Cambridge University Press.

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

De Haan, L. and Pereira, T. T. (2006). Spatial extremes: Models for the stationary case. *The Annals of Statistics*, 34:146–168.

de Haan, L. and Pickands, J. (1986). Stationary min-stable stochastic processes. *Probability Theory and Related Fields*, 74:477–492.

De Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 40:317–337.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24:180–190.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, 31:1208–1211.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 87:158–171.

Jiang, W. and Turnbull, B. (2004). The indirect method: Inference based on intermediate statistics – A synthesis and examples. *Statistical Science*, 19:239–263.

Kammann, E. E. and Wand, M. P. (2003). Geoadditive models. *Applied Statistics*, 52:1–18.

Kent, J. T. (1982). Information gain and a measure of correlation. *Biometrika*, 70:163–173.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics*, 80:221–239.

Padoan, S. A. and Wand, M. P. (2008). Mixed model-based additive models for sample extremes. *Statistics and Probability Letters*, page In press.

Peters, G. W., Fan, Y., and Sisson, S. A. (2008). On sequential Monte Carlo, partial rejection control and approximate Bayesian computation. Technical report, University of New South Wales.

Resnick, S. (1987). *Extreme values, point processes and regular variation*. Springer Verlag, New York.

Ripley, B. D. (2004). *Spatial statistics*. Wiley.

Rotnitzky, A. and Jewell, N. P. (1990). Hypothesis-testing of regression parameters in semiparametric generalised linear models for cluster correlated data. *Biometrika*, 77:485–497.

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

Schlather, M. and Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90:139–154.

Sisson, S. A., Fan, Y., and Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci.*, 104:1760–1765.

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

Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of models (in japenese). *Mathematical Science*, 153:12–18.

Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis*, 92:1–28.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92:519–528.

von Mises, R. (1954). La distribution de la plus grande de n valeurs. In *Selected Papers, Volume II*, pages 271–294. American Mathematical Society, Providence, Rhode Island, USA.

Wand, M. P. (2002). Vector differential calculus in statistics. *The American Statistician*, 56:55–62.

Yee, T. W. and Stephenson, A. G. (2007). Vector generalised linear and additive extreme value models. *Extremes*, 10:1–19.

Appendix

We present explicit expressions for the distribution function (3), density function (4) and the derivatives required for the estimated covariance matrix in Section 3.2

A.1: Vector notation

Let f be a real-valued function in the $d \times 1$ vector $\mathbf{x} = (x_1, \ldots, x_d)$. Then the $1 \times d$ derivative vector, $\mathsf{D}_{\mathbf{x}} f(\mathbf{x})$, has i-th element $\partial f(\mathbf{x})/\partial x_i$. The corresponding Hessian matrix is given by $\mathsf{H}_{\mathbf{x}} f(\mathbf{x}) = \mathsf{D}_{\mathbf{x}} \{\mathsf{D}_{\mathbf{x}} f(\mathbf{x})\}^{\top}$.

If $\mathbf{a} = (a_1, \ldots, a_d)$ and $\mathbf{b} = (b_1, \ldots, b_d)$ are two $d \times 1$ vectors, then element-wise multiplication is denoted by $\mathbf{a} \odot \mathbf{b} = (a_1 b_1, \ldots, a_d b_d)$. The expression $\mathbf{a}/\mathbf{b}$ denotes element-wise division $(a_1/b_1, \ldots, a_d/b_d)$. Scalar functions applied to vectors are also evaluated element-wise. For example, $\mathbf{a}^{-1/\xi} = (a_1^{-1/\xi}, \ldots, a_d^{-1/\xi})$.

A.2: Derivation of the bivariate distribution function

In order to derive the cumulative distribution function (3), considering formula (2) and the assumptions of Section 2.3, we need to solve:

$$
\begin{aligned}
F(z_i, z_j) &= \exp\left\{ -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \max\left(\frac{f(x_1, x_2)}{z_i}, \frac{f(x_1 - t_1, x_2 - t_2)}{z_j} \right) \mathrm{d}x_1 \, \mathrm{d}x_2 \right\} \\
&= \exp\left\{ -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{f(x_1, x_2)}{z_i} \mathbb{I}\left(\frac{f(x_1, x_2)}{z_i} \geq \frac{f(x_1 - t_1, x_2 - t_2)}{z_j} \right) \mathrm{d}x_1 \, \mathrm{d}x_2 \right. \\
&\quad \left. - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{f(x_1 - t_1, x_2 - t_2)}{z_j} \mathbb{I}\left(\frac{f(x_1 - t_1, x_2 - t_2)}{z_j} \geq \frac{f(x_1, x_2)}{z_i} \right) \mathrm{d}x_1 \, \mathrm{d}x_2 \right\},
\end{aligned}
$$

where $f(x_1, x_2)$ is the bivariate normal density of $(X_1, X_2) \sim N((0,0)^{\top}, \boldsymbol{\Sigma})$, and for brevity we set $\mathbf{h} = (\mathbf{t}_j - \mathbf{t}_i)^{\top} \equiv (t_1, t_2)^{\top}$. Recall from (3) that

$$
a(\mathbf{h}) = (\mathbf{h}^T \boldsymbol{\Sigma}^{-1} \mathbf{h})^{1/2} = \frac{1}{\sqrt{(1 - \rho^2)}} \left(\frac{t_1^2}{\sigma_1^2} - \frac{2\rho t_1 t_2}{\sigma_1 \sigma_2} + \frac{t_2^2}{\sigma_2^2} \right)^{1/2}
$$

where $\rho = \sigma_{12}/\sigma_1 \sigma_2$. Consider first the case $(t_1 \sigma_2 - \rho t_2 \sigma_1) > 0$. Note that $f(x_1, x_2)/z_i \geq f(x_1 - t_1, x_2 - t_2)/z_j$ implies that

$$
\begin{aligned}
&\exp\left\{ -\frac{1}{2(1 - \rho^2)} \left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1 \sigma_2} + \frac{x_2^2}{\sigma_2^2} \right) \right\} / z_i \\
&\geq \exp\left\{ -\frac{1}{2(1 - \rho^2)} \left(\frac{(x_1 - t_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - t_1)(x_2 - t_2)}{\sigma_1 \sigma_2} + \frac{(x_2 - t_2)^2}{\sigma_2^2} \right) \right\} / z_i \\
&\Leftrightarrow x_1 \leq \frac{\sigma_1^2 \sigma_2}{2(t_1 \sigma_2 - \rho t_2 \sigma_1)} \left(\frac{t_1^2}{\sigma_1^2} - \frac{2\rho t_1 t_2}{\sigma_1 \sigma_2} + \frac{t_2^2}{\sigma_2^2} \right) + \frac{\sigma_1^2 \sigma_2 (1 - \rho^2)}{t_1 \sigma_2 - \rho t_2 \sigma_1} \log \frac{z_j}{z_i} - \frac{x_2(t_2 \sigma_1^2 - \rho t_1 \sigma_1 \sigma_2)}{t_1 \sigma_2 - \rho t_2 \sigma_1} \\
&\Leftrightarrow x_1 \leq c.
\end{aligned}
$$

From this, it follows that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{f(x_1, x_2)}{z_i} \mathbb{I}\left(\frac{f(x_1, x_2)}{z_i} \geq \frac{f(x_1 - t_1, x_2 - t_2)}{z_j} \right) \mathrm{d}x_1 \, \mathrm{d}x_2$$

$$= \frac{1}{z_i} \int_{-\infty}^{\infty} \int_{-\infty}^{c} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1 \sigma_2} + \frac{x_2^2}{\sigma_2^2} \right) \right\} \mathrm{d}x_1 \, \mathrm{d}x_2$$

$$= \frac{1}{z_i} \int_{-\infty}^{\infty} \varphi(x_2) \int_{-\infty}^{c} \varphi\left(\frac{x_1 - \rho\sigma_1 x_2 / \sigma_2}{\sigma_1 \sqrt{1-\rho^2}} \right) \mathrm{d}x_1 \, \mathrm{d}x_2$$

$$= \frac{1}{z_i} \int_{-\infty}^{\infty} \varphi(x_2) \Phi\left(\frac{c - \rho\sigma_1 x_2 / \sigma_2}{\sigma_1 \sqrt{1-\rho^2}} \right) \mathrm{d}x_2$$

$$= \frac{1}{z_i} \Phi\left(\frac{1}{2\sqrt{(1-\rho^2)}} \left(\frac{t_1^2}{\sigma_1^2} - \frac{2\rho t_1 t_2}{\sigma_1 \sigma_2} + \frac{t_2^2}{\sigma_2^2} \right)^{1/2} + \left(\frac{t_1^2}{\sigma_1^2} - \frac{2\rho t_1 t_2}{\sigma_1 \sigma_2} + \frac{t_2^2}{\sigma_2^2} \right)^{-1/2} \frac{\log z_j / z_i}{(1-\rho^2)^{-1/2}} \right)$$

$$= \frac{1}{z_i} \Phi\left(\frac{a(\mathbf{h})}{2} + \frac{\log z_j / z_i}{a(\mathbf{h})} \right).$$

Similarly,

$$\frac{f(x_1 - t_1, x_2 - t_2)}{z_j} \geq \frac{f(x_1, x_2)}{z_i} \Leftrightarrow x_1 \geq c.$$

It then follows that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{f(x_1 - t_1, x_2 - t_2)}{z_j} \mathbb{I}\left(\frac{f(x_1, x_2)}{z_i} \geq \frac{f(x_1 - t_1, x_2 - t_2)}{z_j} \right) \mathrm{d}x_1 \, \mathrm{d}x_2$$

$$= \frac{1}{z_j} \int_{-\infty}^{\infty} \int_{c}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left(\frac{(x_1 - t_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - t_1)(x_2 - t_2)}{\sigma_1 \sigma_2} + \frac{(x_2 - t_2)^2}{\sigma_2^2} \right) \right\} \mathrm{d}x_1 \, \mathrm{d}x_2$$

$$= \frac{1}{z_j} \int_{-\infty}^{\infty} \varphi(x_2 - t_2) \int_{c}^{\infty} \varphi\left(\frac{(x_1 - t_1) - \rho\sigma_1(x_2 - t_2)/\sigma_2}{\sigma_1 \sqrt{1-\rho^2}} \right) \mathrm{d}x_1 \, \mathrm{d}x_2$$

$$= \frac{1}{z_j} \int_{-\infty}^{\infty} \varphi(x_2 - t_2) \left\{ 1 - \Phi\left(\frac{c - \rho\sigma_1(x_2 - t_2)/\sigma_2}{\sigma_1 \sqrt{1-\rho^2}} \right) \right\} \mathrm{d}x_2$$

$$= \frac{1}{z_j} \Phi\left(\frac{1}{2\sqrt{(1-\rho^2)}} \left(\frac{t_1^2}{\sigma_1^2} - \frac{2\rho t_1 t_2}{\sigma_1 \sigma_2} + \frac{t_2^2}{\sigma_2^2} \right)^{1/2} + \left(\frac{t_1^2}{\sigma_1^2} - \frac{2\rho t_1 t_2}{\sigma_1 \sigma_2} + \frac{t_2^2}{\sigma_2^2} \right)^{-1/2} \frac{\log z_i / z_j}{(1-\rho^2)^{-1/2}} \right)$$

$$= \frac{1}{z_j} \Phi\left(\frac{a(\mathbf{h})}{2} + \frac{\log z_i / z_j}{a(\mathbf{h})} \right),$$

and the form of the distribution (3) is confirmed. Observe, that the same result is obtained for the case $(t_1\sigma_2 - \rho t_2\sigma_1) < 0$. See also Smith (1990) and De Haan and Pereira (2006).

20

A.3: Derivation of the bivariate density function

In order to derive the bivariate density function (4) we require the second-order derivative of

$$F(z_i, z_j) = \exp\left(-\frac{\Phi(w)}{z_i} - \frac{\Phi(v)}{z_j}\right)$$

with respect to z_i and z_j, where for brevity we set $a \equiv a(h)$, $w \equiv w(h)$ and $v \equiv v(h)$ and write $w = a/2 + \log(z_j/z_i)/a$ and $v = a - w$. The differentiation gives

$$f(z_i, z_j) \equiv \frac{\partial^2}{\partial z_i \partial z_j} F(z_i, z_j) =$$

$$\exp\left(-\frac{\Phi(w)}{z_i} - \frac{\Phi(v)}{z_j}\right) \left\{ \frac{\partial}{\partial z_i}\left(-\frac{\Phi(w)}{z_i} - \frac{\Phi(v)}{z_j}\right) \frac{\partial}{\partial z_j}\left(-\frac{\Phi(w)}{z_i} - \frac{\Phi(v)}{z_j}\right) + \frac{\partial^2}{\partial z_i \partial z_j}\left(-\frac{\Phi(w)}{z_i} - \frac{\Phi(v)}{z_j}\right) \right\}.$$

First-order differentation gives

$$\frac{\partial}{\partial z_i}\left(-\frac{\Phi(w)}{z_i} - \frac{\Phi(v)}{z_j}\right) = \frac{\Phi(w)}{z_i^2} + \frac{\varphi(w)}{az_i^2} - \frac{\varphi(v)}{az_i z_j}, \qquad \frac{\partial}{\partial z_j}\left(-\frac{\Phi(w)}{z_i} - \frac{\Phi(v)}{z_j}\right) = \frac{\Phi(v)}{z_j^2} + \frac{\varphi(v)}{az_j^2} - \frac{\varphi(w)}{az_i z_j},$$

using the results

$$\frac{\partial \Phi(w)}{\partial z_i} = -\frac{\varphi(w)}{az_i}, \quad \frac{\partial \Phi(v)}{\partial z_i} = \frac{\varphi(v)}{az_i} \quad \text{and} \quad \frac{\partial w}{\partial z_i} = -\frac{1}{az_i}, \quad \frac{\partial v}{\partial z_i} = \frac{1}{az_i}.$$

Second-order differentation yields

$$\frac{\partial^2}{\partial z_i \partial z_j}\left(-\frac{\Phi(w)}{z_i} - \frac{\Phi(v)}{z_j}\right) = \frac{v\,\varphi(w)}{a^2 z_i^2 z_j} + \frac{w\,\varphi(v)}{a^2 z_i z_j^2},$$

using

$$\frac{\partial \varphi(w)}{\partial z_i} = \frac{w\varphi(w)}{az_i} \quad \text{and} \quad \frac{\partial \varphi(v)}{\partial z_i} = -\frac{v\varphi(v)}{az_i}.$$

Substituting, we obtain the probability density function

$$f(z_i, z_j) = \exp\left\{ -\frac{\Phi(w)}{z_i} - \frac{\Phi(v)}{z_j}\right\} \left\{ \left(\frac{\Phi(w)}{z_i^2} + \frac{\varphi(w)}{az_i^2} - \frac{\varphi(v)}{az_i z_j}\right) \right.$$

$$\left. \left(\frac{\Phi(v)}{z_j^2} + \frac{\varphi(v)}{az_j^2} - \frac{\varphi(w)}{az_i z_j}\right) + \left(\frac{v\,\varphi(w)}{a^2 z_i^2 z_j} + \frac{w\,\varphi(v)}{a^2 z_i z_j^2}\right) \right\}.$$

A.4: An expression for the squared score statistic

From Section 3.2 the term $J(\boldsymbol{\psi})$ of the Godambe information matrix can be estimated from

$$\sum_{i=1}^{\mathcal{K}} \sum_{j=i+1}^{\mathcal{K}-1} \mathsf{D}_{\boldsymbol{\psi}} \log f(\mathbf{y}_i, \mathbf{y}_j; \boldsymbol{\psi})^\top \mathsf{D}_{\boldsymbol{\psi}} \log f(\mathbf{y}_i, \mathbf{y}_j; \boldsymbol{\psi}),$$

where $\boldsymbol{\psi}^\top = (\boldsymbol{\sigma}, \boldsymbol{\beta}_\mu, \boldsymbol{\beta}_\lambda, \boldsymbol{\beta}_\xi)$, $\boldsymbol{\sigma}^\top = (\sigma_{11}^2, \sigma_{12}, \sigma_{22}^2)$ and where each parameter $\boldsymbol{\beta}$ is p-dimensional vector of coefficients. The bivariate log-density has the form

$$\log f(\mathbf{y}_i, \mathbf{y}_j; \boldsymbol{\psi}) = A + \log(B \odot C + D) + E,$$

where

$$A = -\frac{\Phi(\mathbf{w})}{\mathbf{z}_i} - \frac{\Phi(\mathbf{v})}{\mathbf{z}_j}, \qquad B = \frac{\Phi(\mathbf{w})}{\mathbf{z}_i^2} + \frac{\varphi(\mathbf{w})}{\mathbf{a} \odot \mathbf{z}_i^2} - \frac{\varphi(\mathbf{v})}{\mathbf{a} \odot \mathbf{z}_i \odot \mathbf{z}_j},$$

$$C = \frac{\Phi(\mathbf{v})}{\mathbf{z}_j^2} + \frac{\varphi(\mathbf{v})}{\mathbf{a} \odot \mathbf{z}_j^2} - \frac{\varphi(\mathbf{w})}{\mathbf{a} \odot \mathbf{z}_i \odot \mathbf{z}_j}, \qquad D = \frac{\mathbf{v} \odot \varphi(\mathbf{w})}{\mathbf{a}^2 \odot \mathbf{z}_i^2 \odot \mathbf{z}_j} + \frac{\mathbf{w} \odot \varphi(\mathbf{v})}{\mathbf{a}^2 \odot \mathbf{z}_i \odot \mathbf{z}_j^2}$$

$$\text{and} \quad E = \log\left\{ \frac{1}{\lambda_i \lambda_j} \left(1 + \xi_i \frac{\mathbf{y}_i - \mu_i \mathbf{1}}{\lambda_i} \right)_+^{\frac{1}{\xi_i}-1} \left(1 + \xi_j \frac{\mathbf{y}_j - \mu_j \mathbf{1}}{\lambda_j} \right)_+^{\frac{1}{\xi_j}-1} \right\}$$

and where $\mathbf{1}^\top = (1, \ldots, 1)$, $\mu_i = (\mathbf{X}_{\boldsymbol{\beta}_\mu} \boldsymbol{\beta}_\mu)_i$, $\xi_i = (\mathbf{X}_{\boldsymbol{\beta}_\xi} \boldsymbol{\beta}_\xi)_i$ and $\log(\psi_i) = (\mathbf{X}_{\boldsymbol{\beta}_\psi} \boldsymbol{\beta}_\psi)_i$. The GEV parameters are related to the predictors by the form (7). We assume identity link functions for the location and shape parameters and exponential for the scale. The term E corresponds to the log of the determinat of the Jacobian matrix associated with the transformation (6), see Section 3.2.

The first-order derivative term of the square score statistic is defined by

$$\mathsf{D}_{\boldsymbol{\psi}} \log f(\mathbf{y}_i, \mathbf{y}_j; \boldsymbol{\psi}) = (\mathsf{D}_{\boldsymbol{\sigma}} \log f(\boldsymbol{\psi}), \mathsf{D}_{\boldsymbol{\beta}_\mu} \log f(\boldsymbol{\psi}), \mathsf{D}_{\boldsymbol{\beta}_\lambda} \log f(\boldsymbol{\psi}), \mathsf{D}_{\boldsymbol{\beta}_\xi} \log f(\boldsymbol{\psi}))$$

where for brevity we write $\log f(\boldsymbol{\psi}) \equiv \log f(\mathbf{y}_i, \mathbf{y}_j; \boldsymbol{\psi})$. Vector differential calculus (e.g. Wand (2002)) leads to

$$
\begin{aligned}
\mathsf{D}_{\boldsymbol{\sigma}} \log f(\boldsymbol{\psi}) \equiv & \left[-\frac{\mathbf{v} \odot \varphi(\mathbf{w})}{\mathbf{a} \odot \mathbf{z}_i} - \frac{\mathbf{w} \odot \varphi(\mathbf{v})}{\mathbf{a} \odot \mathbf{z}_j} + \left\{ C \odot \left(\frac{(\mathbf{w}^2 - 1) \odot \varphi(\mathbf{v})}{\mathbf{a}^2 \odot \mathbf{z}_j^2} + \frac{(\mathbf{1} + \mathbf{w} \odot \mathbf{v}) \odot \varphi(\mathbf{w})}{\mathbf{a}^2 \odot \mathbf{z}_i \odot \mathbf{z}_j} \right) \right. \right. \\
& + B \odot \left(\frac{(\mathbf{v}^2 - 1) \odot \varphi(\mathbf{w})}{\mathbf{a}^2 \odot \mathbf{z}_i^2} + \frac{(\mathbf{1} + \mathbf{w} \odot \mathbf{v}) \odot \varphi(\mathbf{v})}{\mathbf{a}^2 \odot \mathbf{z}_i \odot \mathbf{z}_j} \right) + \frac{(\mathbf{w} - 2\mathbf{v} - \mathbf{w} \odot \mathbf{v}^2) \odot \varphi(\mathbf{w})}{\mathbf{a}^3 \odot \mathbf{z}_i^2 \odot \mathbf{z}_j} \\
& \left. \left. + \frac{(\mathbf{v} - 2\mathbf{w} - \mathbf{v} \odot \mathbf{w}^2) \odot \varphi(\mathbf{v})}{\mathbf{a}^3 \odot \mathbf{z}_j^2 \odot \mathbf{z}_i} \right\} / (B \odot C + D) \right] \mathbf{s}^\top,
\end{aligned}
$$

where $\mathbf{s}^\top \equiv (t_1^2, \ 2\, t_1\, t_2, \ t_2^2)$, using the results

$$\mathsf{D}_{\boldsymbol{\sigma}} \mathbf{w} = \frac{\mathbf{v}}{\mathbf{a}} \mathbf{s}^\top, \qquad \mathsf{D}_{\boldsymbol{\sigma}} \Phi(\mathbf{w}) = \frac{\mathbf{v} \odot \varphi(\mathbf{w})}{\mathbf{a}} \mathbf{s}^\top, \qquad \mathsf{D}_{\boldsymbol{\sigma}} \varphi(\mathbf{w}) = -\frac{\mathbf{w} \odot \mathbf{v} \varphi(\mathbf{w})}{\mathbf{a}} \mathbf{s}^\top$$

$$\text{and} \quad \mathsf{D}_{\boldsymbol{\sigma}} \mathbf{v} \odot \varphi(\mathbf{w}) = \frac{\mathbf{w} \odot (\mathbf{1} - \mathbf{v}^2) \varphi(\mathbf{w})}{\mathbf{a}} \mathbf{s}^\top.$$

The first-order derivatives of $\mathbf{v}$, $\Phi(\mathbf{v})$, $\varphi(\mathbf{v})$ and $\mathbf{w} \odot \varphi(\mathbf{v})$ are the same as the above, substituting $\mathbf{v}$ for $\mathbf{w}$. Similarly for the second term we have

$$
\begin{aligned}
\mathsf{D}_{\boldsymbol{\beta}_\mu} \log f(\boldsymbol{\psi}) \equiv &\left\{ \left(\frac{\varphi(\mathbf{w}) + \mathbf{a} \odot \Phi(\mathbf{w})}{\mathbf{a} \odot \mathbf{z}_i^2} - \frac{\varphi(\mathbf{v})}{\mathbf{a} \odot \mathbf{z}_i \odot \mathbf{z}_j} \right) \odot \frac{\mathbf{z}_i^{1-\xi_i}}{\lambda_i} \right\} (\mathbf{X}_{\boldsymbol{\beta}_\mu})_i \\
&+ \left\{ \left(\frac{\varphi(\mathbf{v}) + \mathbf{a} \odot \Phi(\mathbf{v})}{\mathbf{a} \odot \mathbf{z}_j^2} - \frac{\varphi(\mathbf{w})}{\mathbf{a} \odot \mathbf{z}_i \odot \mathbf{z}_j} \right) \odot \frac{\mathbf{z}_j^{1-\xi_j}}{\lambda_j} \right\} (\mathbf{X}_{\boldsymbol{\beta}_\mu})_j \\
&+ \left[\frac{\left\{ \mathrm{C} \odot \left(\frac{\mathbf{w} \odot \Phi(\mathbf{v})}{\mathbf{a}^2 \odot \mathbf{z}_i \odot \mathbf{z}_j^2} + \frac{\mathbf{v} \odot \Phi(\mathbf{w})}{\mathbf{a}^2 \odot \mathbf{z}_j \odot \mathbf{z}_i^2} \right) \odot \frac{\mathbf{z}_i^{1-\xi_i}}{\lambda_i} \right\}}{(\mathrm{B} \odot \mathrm{C} + \mathrm{D})} \right] (\mathbf{X}_{\boldsymbol{\beta}_\mu})_i \\
&+ \left[\frac{\left\{ \mathrm{C} \odot \left(\frac{(\mathbf{a}+\mathbf{w}) \odot \varphi(\mathbf{w})}{\mathbf{a}^2 \odot \mathbf{z}_i \odot \mathbf{z}_j^2} - \frac{(2\mathbf{a}+\mathbf{w}) \odot \varphi(\mathbf{v})}{\mathbf{a}^2 \odot \mathbf{z}_j^3} - \frac{2\Phi(\mathbf{v})}{\mathbf{z}_i^3} \right) \odot \frac{\mathbf{z}_j^{1-\xi_j}}{\lambda_j} \right\}}{(\mathrm{B} \odot \mathrm{C} + \mathrm{D})} \right] (\mathbf{X}_{\boldsymbol{\beta}_\mu})_j \\
&+ \left[\frac{\left\{ \mathrm{B} \odot \left(\frac{(\mathbf{a}+\mathbf{v}) \odot \varphi(\mathbf{v})}{\mathbf{a}^2 \odot \mathbf{z}_j \odot \mathbf{z}_i^2} - \frac{(2\mathbf{a}+\mathbf{v}) \odot \varphi(\mathbf{w})}{\mathbf{a}^2 \odot \mathbf{z}_i^3} - \frac{2\Phi(\mathbf{w})}{\mathbf{z}_j^3} \right) \odot \frac{\mathbf{z}_i^{1-\xi_i}}{\lambda_i} \right\}}{(\mathrm{B} \odot \mathrm{C} + \mathrm{D})} \right] (\mathbf{X}_{\boldsymbol{\beta}_\mu})_i \\
&+ \left[\frac{\left\{ \mathrm{B} \odot \left(\frac{\mathbf{v} \odot \Phi(\mathbf{w})}{\mathbf{a}^2 \odot \mathbf{z}_j \odot \mathbf{z}_i^2} + \frac{\mathbf{w} \odot \Phi(\mathbf{v})}{\mathbf{a}^2 \odot \mathbf{z}_i \odot \mathbf{z}_j^2} \right) \odot \frac{\mathbf{z}_j^{1-\xi_j}}{\lambda_j} \right\}}{(\mathrm{B} \odot \mathrm{C} + \mathrm{D})} \right] (\mathbf{X}_{\boldsymbol{\beta}_\mu})_j \\
&+ \left[\frac{\left\{ \left(\frac{(1-\mathbf{a} \odot \mathbf{v}-\mathbf{v}^2) \odot \varphi(\mathbf{w})}{\mathbf{a}^2 \odot \mathbf{z}_j \odot \mathbf{z}_i^3} - \frac{(1-\mathbf{a} \odot \mathbf{w}-\mathbf{w}^2) \odot \varphi(\mathbf{v})}{\mathbf{a}^2 \odot \mathbf{z}_i \odot \mathbf{z}_j^3} \right) \odot \frac{\mathbf{z}_i^{1-\xi_i}}{\lambda_i} \right\}}{(\mathrm{B} \odot \mathrm{C} + \mathrm{D})} \right] (\mathbf{X}_{\boldsymbol{\beta}_\mu})_i \\
&+ \left[\frac{\left\{ \left(\frac{(1-\mathbf{a} \odot \mathbf{w}-\mathbf{w}^2) \odot \varphi(\mathbf{v})}{\mathbf{a}^2 \odot \mathbf{z}_i \odot \mathbf{z}_j^3} - \frac{(1-\mathbf{a} \odot \mathbf{v}-\mathbf{v}^2) \odot \varphi(\mathbf{w})}{\mathbf{a}^2 \odot \mathbf{z}_j \odot \mathbf{z}_i^3} \right) \odot \frac{\mathbf{z}_j^{1-\xi_j}}{\lambda_j} \right\}}{(\mathrm{B} \odot \mathrm{C} + \mathrm{D})} \right] (\mathbf{X}_{\boldsymbol{\beta}_\mu})_j \\
&+ \frac{(\xi_i - 1)}{\lambda_i \mathbf{z}^{\xi_i}} (\mathbf{X}_{\boldsymbol{\beta}_\mu})_i + \frac{(\xi_j - 1)}{\lambda_j \mathbf{z}^{\xi_j}} (\mathbf{X}_{\boldsymbol{\beta}_\mu})_j.
\end{aligned}
$$

Observe that the above expression is obtained by deriving in order the components: $\mathsf{D}_{\boldsymbol{\beta}_\mu} \mathrm{A}$, $\mathsf{D}_{\boldsymbol{\beta}_\mu} \log(\mathrm{B} \odot \mathrm{C} + \mathrm{D})$ and $\mathsf{D}_{\boldsymbol{\beta}_\mu} \mathrm{E}$. These three components have the form $\mathsf{D}_{\boldsymbol{\beta}_\mu} f(\mathbf{x}) = \mathsf{D}_{\mathbf{z}_i} f(\mathbf{x}) \mathsf{D}_{\boldsymbol{\beta}_\mu} \mathbf{z}_i + \mathsf{D}_{\mathbf{z}_j} f(\mathbf{x}) \mathsf{D}_{\boldsymbol{\beta}_\mu} \mathbf{z}_j$, where $\mathsf{D}_{\boldsymbol{\beta}_\mu} \mathbf{z}_i = \mathbf{z}_i^{1-\xi_i} / \lambda_i \mathbf{X}_{\boldsymbol{\beta}_\mu}$. For this reason the derived expressions of $\mathsf{D}_{\boldsymbol{\beta}_\lambda} \log f(\boldsymbol{\psi})$ and $\mathsf{D}_{\boldsymbol{\beta}_\xi} \log f(\boldsymbol{\psi})$ are essentially the same but substituting $\mathsf{D}_{\boldsymbol{\beta}_\mu} \mathbf{z}_i$ with $\mathsf{D}_{\boldsymbol{\beta}_\lambda} \mathbf{z}_i$ and $\mathsf{D}_{\boldsymbol{\beta}_\xi} \mathbf{z}_i$, and $\mathsf{D}_{\boldsymbol{\beta}_\mu} \mathrm{E}$ with $\mathsf{D}_{\boldsymbol{\beta}_\lambda} \mathrm{E}$ and $\mathsf{D}_{\boldsymbol{\beta}_\xi} \mathrm{E}$.

We have

$$\mathsf{D}_{\boldsymbol{\beta}_\lambda}\mathbf{z}_i \equiv -\frac{\mathbf{z}_i(\mathbf{y}_i - \mu_i\mathbf{1})}{\lambda_i}(\mathbf{X}_{\boldsymbol{\beta}_\lambda})_i \quad \text{and} \quad \mathsf{D}_{\boldsymbol{\beta}_\xi}\mathbf{z}_i \equiv \left\{ \frac{1}{\xi_i}\left(\frac{\mathbf{z}_i^{1-\xi_i}(\mathbf{y}_i - \mu_i\mathbf{1})}{\lambda_i} - \mathbf{z}_i\log(\mathbf{z}_i) \right) \right\}(\mathbf{X}_{\boldsymbol{\beta}_\xi})_i.$$

Finally,

$$\mathsf{D}_{\boldsymbol{\beta}_\lambda}\mathrm{E} \equiv \left(\frac{(\xi_i - 1)(\mathbf{y}_i - \mu_i\mathbf{1})}{\lambda_i\mathbf{z}_i} - \mathbf{1} \right)(\mathbf{X}_{\boldsymbol{\beta}_\lambda})_i + \left(\frac{(\xi_j - 1)(\mathbf{y}_j - \mu_j\mathbf{1})}{\lambda_j\mathbf{z}_j} - \mathbf{1} \right)(\mathbf{X}_{\boldsymbol{\beta}_\lambda})_j \quad \text{and}$$

$$\begin{aligned}
\mathsf{D}_{\boldsymbol{\beta}_\xi}\mathrm{E} \equiv{}& \left[\frac{1}{\xi_i}\left\{ \frac{(1-\xi_i)(\mathbf{y}_i - \mu_i\mathbf{1})}{\mathbf{z}_i^{\xi_i}\lambda_i} - \log(\mathbf{z}_i) \right\} \right](\mathbf{X}_{\boldsymbol{\beta}_\xi})_i \\
&+ \left[\frac{1}{\xi_j}\left\{ \frac{(1-\xi_j)(\mathbf{y}_j - \mu_j\mathbf{1})}{\mathbf{z}_j^{\xi_j}\lambda_j} - \log(\mathbf{z}_j) \right\} \right](\mathbf{X}_{\boldsymbol{\beta}_\xi})_j.
\end{aligned}$$

In combination we obtain an expression for $\mathsf{D}_{\boldsymbol{\psi}}\log f(\mathbf{y}_i, \mathbf{y}_j; \boldsymbol{\psi})$, and from this the squared score statistic.