Likelihood-Based Inference for Max-Stable Processes

S. A. PADOAN, M. RIBATET, and S. A. SISSON

The last decade has seen max-stable processes emerge as a common tool for the statistical modeling 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 modeling of marginal and dependence 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 United States precipitation extremes.

KEY WORDS: Extreme value theory; Multivariate extreme analysis; Pseudo-likelihood; Rainfall; Spatial dependence; Spatial extremes.

1. INTRODUCTION

A common objective of spatial analysis is to quantify and characterize the behavior of environmental phenomena such as precipitation levels, wind speed, or daily temperatures. A number of generic approaches to spatial modeling have been developed (e.g., Cressie 1993; Barndorff-Nielsen et al. 1998; 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 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 modeling 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 continuous region. Observations $\{y_{m,k}\}$ each denote the maximum of s samples over $m = 1, \ldots, M$ temporal blocks. For example, for daily observations, $s = 366$ implies the $\{y_{m,k}\}$ 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 Generalized Extreme Value (GEV) distribution (von Mises 1954; Jenkinson 1955), providing a general approach to modeling process extremes incorporating temporal or spatial dependence. Statistical

methods for max-stable processes and data analyses 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 nonparametric estimators). In this article we develop flexible and inferentially practical methods for the fitting of max-stable processes to spatial data based on nonstandard, *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 at a moderate computational cost.

The article is organized 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 studies. We conclude with an illustration of a real extremal data analysis of United States precipitation levels.

2. MAX–STABLE PROCESSES AND SPATIAL EXTREMES

2.1 Max-Stable Processes

Max-stable processes provide a natural generalization 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

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

S. A. Padoan is Postdoctoral Researcher, Laboratory of Environmental Fluid Mechanics and Hydrology, EPFL-ENAC-EFLUM, Bâtiment GR, Station 2, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (E-mail: *Simone.Padoan@epfl.ch*). M. Ribatet is Postdoctoral Researcher, Institute of Mathematics, EPFL-FSB-IMA-STAT, Station 8, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (E-mail: *Mathieu.Ribatet@epfl.ch*). S. A. Sisson is Senior Lecture, School of Mathematics and Statistics, University of New South Wales, Sydney 2052, Australia (E-mail: *Scott.Sisson@unsw.edu.au*). This work was commenced while SAP was visiting the School of Mathematics and Statistics, University of New South Wales, Sydney, Australia. Their hospitality is gratefully acknowledged. The authors are grateful to Anthony Davison and Stuart Coles for their helpful suggestions, and to Richard Smith for providing the U.S. precipitation data. The authors also wish to thank the associate editor and two anonymous referees for their helpful comments. The work was supported by the CCES Extremes project, *http://www.cces.ethz.ch/projects/hazri/EXTREMES*. SAS is supported by the Australian Research Council through the Discovery Project scheme (DP0877432).

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). First, the one-dimensional marginal distributions belong to the class of generalized extreme value distributions, $Y \sim \text{GEV}(\mu, \lambda, \xi)$ with distribution function

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

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.

Without loss of generality, if $a_n(t) = n$, $b_n(t) = 0$, 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 standardization 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 its *spectral representation* (de Haan and Pickands 1986).

In detail, let $\{X_j, U_j\}_{j \geq 1}$ be a Poisson process, Π, on $\mathbb{R}^d \times \mathbb{R}_+$, with counting measure $\Pi(\cdot) := \sum_j \mathbb{I}_{(X_j, U_j)}(\cdot)$ 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 $A \subset \mathbb{R}^d \times \mathbb{R}_+$ and ν is a positive measure. For a nonnegative measurable function $f(x - t)$ (for fixed $t \in T$) such that $\int_{\mathbb{R}^d} f(x - t)\nu(dx) = 1, \forall t \in T$ the stochastic process

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

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, center, and 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)

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

It then follows that the marginal distributions are unit Fréchet:

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

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

2.2 Extremal Coefficients

Given n independent replications $Y^{(1)}, \ldots, Y^{(n)}$, of the random vector $Y = (Y_1, \ldots, Y_K) \in \mathbb{R}^K$ with unit Fréchet margins for each Y_k, then the joint distribution of componentwise maxima satisfies (de Haan and Resnick 1977; Resnick 1987)

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

for $k = 1, 2, \ldots, K$ and common threshold z, where the rightmost term, obtained in the limit as $n \to \infty$, 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 $\theta = K$ corresponds to full independence.

The extremal dependence of stochastic processes can be extended in a similar way. The stationary max-stable process $Z(t)$ given by (1) is the limiting process of n iid copies of the standardized maxima of a stationary stochastic process $\tilde{Y}(t)$, where $a_n(t) = n$ and $b_n(t) = 0$. From (2), for all $z > 0$, the asymptotic distribution of $Z(t)$ at K sites can be written as

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

under weak conditions (de Haan 1984), and so

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

where $\theta \in \mathbb{R}$ again represents the effective number of independent variables. Schlather and Tawn (2003) discuss the extremal coefficient within the context of max-stable processes.

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

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

where $\mathbf{h} = (\mathbf{t}_j - \mathbf{t}_i)^\top$, $\mathbf{0}$ is the origin, Φ is the standard Gaussian distribution function, $a(\mathbf{h}) = (\mathbf{h}^\top \mathbf{\Sigma}^{-1} \mathbf{h})^{1/2}$ and $\mathbf{\Sigma}$ is the covariance matrix of f, with covariance σ_{12} and standard deviations $\sigma_1, \sigma_2 > 0$. A derivation of (3) is provided in Appendix A.3. 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 two-dimensional density function

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

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.4.

Observe that $a(\mathbf{h})$ measures the strength of extremal dependence: $a(\mathbf{h}) \to 0$ represents complete dependence, and (in the limit) $a(\mathbf{h}) \to \infty$ 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, characterization of extremal dependence is determined by the covariance matrix, $\mathbf{\Sigma}$, 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

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

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.3. Alternative models result by considering, for example, 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 modeling 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 modeling 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 $\mathcal{F} = \{f(\mathbf{y}; \boldsymbol{\psi}), \mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^K, \boldsymbol{\psi} \in \Psi \subseteq \mathbb{R}^q\}$ satisfying the usual regularity conditions (e.g., Davison 2003, p. 118), and assuming a set of measurable marginal or conditional events

$\{\mathcal{I}_k : k \in \mathcal{K}\}$ (for some $\mathcal{K} \subseteq \mathbb{N}$) as a subset of some sigma algebra on $\mathcal{Y}$, then the composite log-likelihood contribution for the random vector $\mathbf{Y}$ may be defined by

$$\ell_{\mathcal{C}}(\boldsymbol{\psi}; \mathbf{y}) = \sum_{k \in \mathcal{K}} w_k \log f(\mathbf{y} \in \mathcal{I}_k; \boldsymbol{\psi}), \qquad (5)$$

where $f(\mathbf{y} \in \mathcal{I}_k; \boldsymbol{\psi}) = f(\{y_i : y_i \in \mathcal{I}_k\}; \boldsymbol{\psi})$, with $\mathbf{y} = (y_1, \ldots, y_K)$, is the likelihood associated with event $\mathcal{I}_k$ and $\{w_k\}_{k \in \mathcal{K}}$ are appropriate nonnegative weights. 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 the maximum composite likelihood estimator (MCLE) of $\boldsymbol{\psi}$, if unique, is obtained by solving the *composite score equation*: $\mathsf{D}_{\boldsymbol{\psi}}\ell_{\mathcal{C}}(\widehat{\boldsymbol{\psi}}_{\mathrm{MCLE}}; \mathbf{y}) = 0$. Similarly, second-order partial derivatives of $\ell_{\mathcal{C}}(\boldsymbol{\psi}; \mathbf{y})$ 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 its ability, under the usual regularity conditions, to provide asymptotically unbiased parameter estimates when standard likelihood estimators are not available. Specifically, the composite score equation is an unbiased estimating equation, as it is a linear combination of individual likelihood components, each of which is an unbiased estimating equation. Accordingly, if the regulatory conditions in Appendix A.2 hold, similarly to the derivation of the ordinary MLE asymptotics (Huber 1967; Davison 2003, p. 147), the MCLE is consistent and asymptotically distributed as

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

where $\tilde{\mathrm{I}}(\boldsymbol{\psi})$ is the *sandwich information matrix* stemming from unbiased estimating function arguments (Godambe 1960), and $\mathrm{H}(\boldsymbol{\psi}) = \mathbb{E}\{-\mathsf{H}_{\boldsymbol{\psi}}\ell_{\mathcal{C}}(\boldsymbol{\psi}; \mathbf{Y})\}$ and $\mathrm{J}(\boldsymbol{\psi}) = \mathbb{V}\{\mathsf{D}_{\boldsymbol{\psi}}\ell_{\mathcal{C}}(\boldsymbol{\psi}; \mathbf{Y})\}$ are analogues of the expected information matrix and the covariance matrix of the score vector. Under ordinary maximum likelihood, $\mathrm{H}(\boldsymbol{\psi}) = \mathrm{J}(\boldsymbol{\psi})$ are equal. Note that although the MCLE is asymptotically unbiased, it is not asymptotically efficient in that $\tilde{\mathrm{I}}(\boldsymbol{\psi})^{-1}$, the inverse of the sandwich information matrix, does not attain the Cramér–Rao lower bound. Hence estimation using the composite likelihood results in a loss of efficiency over ordinary maximum likelihood. In addition, consistency and asymptotic normality of the MCLE may not always be satisfied (Cox and Reid 2004).

3.2 The Pairwise Setting for Spatial Extremes

Within the spatial setting we have observations $\{y_{m,k}\}$, that are the maximum of s samples over $m = 1, \ldots, M$ blocks and for $k = 1, \ldots, K$ locations in a continuous region. For example, for daily observations, $s = 366$ implies the $y_{m,k}$ describe annual maxima. Accordingly, the K univariate marginals are approximately GEV distributed. Assume that $\{\mathbf{y}_m\}_{m=1}^M$, are M iid realizations from the Gaussian extreme value process with GEV margins denoted by $Y_1, \ldots, Y_K$. Despite the intractability of the multivariate max-stable process, availability of the bivariate marginal form (4) implies the definition of a pairwise composite log-likelihood as a sum of log-likelihoods corresponding to each potential bivariate contribution. Specifically from (5), for each m of the M iid data replications, by defining the events $\mathcal{I}_k = \{(y_{m,i}, y_{m,j})\}$ as the sets of bivariate subvectors of $\mathbf{y}$ taken

over all $K(K-1)/2$ distinct location pairs i and j, the pairwise composite log-likelihood function is given by

$$\ell_{\mathcal{P}}(\boldsymbol{\psi}; \mathbf{y}) = \sum_{m=1}^{M} \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} w_{ij} \ell_{ij}(\boldsymbol{\psi}), \qquad (6)$$

where each $\ell_{ij}(\boldsymbol{\psi}) = \log f(y_{m,i}, y_{m,j}; \boldsymbol{\psi})$ is the bivariate marginal log-likelihood based on data at locations i and j, and w_{ij} is its respective weight. Setting all weights $w_{ij} > 0$, $\sum_{ij} w_{ij} = 1$ induces a pairwise log-likelihood constructed from all observational pairs. However, computational requirements and statistical efficiency may be improved by alternative specifications. For instance, as in the case of spatial–temporal processes (Bevilacqua et al. 2007), one may set $w_{ij} = 1$ if $\|\mathbf{t}_j - \mathbf{t}_i\| \le \delta$ and $w_{ij} = 0$ if $\|\mathbf{t}_j - \mathbf{t}_i\| > \delta$, where δ is a parameter that can be used to improve the performance of the procedure. For instance, it might be chosen as the value that minimizes the total amount of variation of the asymptotic variance, $\text{tr}(\tilde{\text{I}}(\boldsymbol{\psi})^{-1})$ (Heyde 1997, pp. 19–21), when this minimum exists. In Section 4, we show empirically that including only closely located pairs in the construction of the composite likelihood improves the statistical efficiency of the maximum composite likelihood estimator for the max-stable process.

In order to characterize 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

$$Z_i = \left(1 + \frac{\xi_i(Y_i - \mu_i)}{\lambda_i}\right)_+^{1/\xi_i},$$
$$\qquad\qquad\qquad\qquad\qquad\qquad (7)$$
$$Z_j = \left(1 + \frac{\xi_j(Y_j - \mu_j)}{\lambda_j}\right)_+^{1/\xi_j},$$

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

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

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

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

This change of variable permits the use of GEV marginals (over unit Fréchet) without reforming the problem definition. Hence, the pairwise log-likelihood (6) 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 cannot 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 the inverse of the sandwich information matrix, with estimates of the matrices $\text{H}(\boldsymbol{\psi})$ and $\text{J}(\boldsymbol{\psi})$ given by

$$\widehat{\text{H}}(\widehat{\boldsymbol{\psi}}_{\text{MCLE}}) = -\sum_{m=1}^{M} \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} w_{ij} \text{H}_{\boldsymbol{\psi}} \ell_{ij}(\widehat{\boldsymbol{\psi}}_{\text{MCLE}})$$

and

$$\widehat{\text{J}}(\widehat{\boldsymbol{\psi}}_{\text{MCLE}}) = \sum_{m=1}^{M} \left\{ \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} w_{ij} \text{D}_{\boldsymbol{\psi}} \ell_{ij}(\widehat{\boldsymbol{\psi}}_{\text{MCLE}}) \right\}$$
$$\times \left\{ \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} w_{ij} \text{D}_{\boldsymbol{\psi}} \ell_{ij}(\widehat{\boldsymbol{\psi}}_{\text{MCLE}}) \right\}^{\top},$$

each evaluated at the composite maximum likelihood value. In practice the matrix $\widehat{\text{H}}$ is obtained straightforwardly with the numerical maximization routine employed for likelihood maximization. An explicit expression for $\widehat{\text{J}}$ is derived in Appendix A.5.

In principle, estimating unique marginal parameters for each location ensures correct model application by respecting marginal constraints, although computational issues arise for large numbers of parameters. Alternatively, as is common in the modeling 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

$$\eta(\mathbf{x}_k) \equiv h\{p(\mathbf{x}_k)\} = [\mathbf{X}\boldsymbol{\beta}]_k \qquad (8)$$

for $k = 1, \ldots, K$, where $p(\cdot)$ represents the parameters $\mu(\cdot)$, $\lambda(\cdot)$, or $\xi(\cdot)$, h is a link function, $\mathbf{x}_k = (x_{k1}, \ldots, x_{kI})$ is a vector of predictors, $\mathbf{X}$ is a $K \times I$ design matrix, and $\boldsymbol{\beta}$ is an $I \times 1$ vector of unknown parameters. The link function may differ according to the parameter being modeled.

For further flexibility, a nonparametric approach may provide a useful alternative. Nonparametric modeling of univariate extreme value responses has been recently proposed by Chavez-Demoulin and Davison (2005), Yee and Stephenson (2007), and Padoan and Wand (2008). For example, a nonparametric model extending (8) is the spline model

$$\eta(\mathbf{x}_k) \equiv h\{p(\mathbf{x}_k)\} = \beta_0 + \sum_{i=2}^{I} \beta_{1i} x_{ki} + \sum_{i=2}^{I} \sum_{j=1}^{J} u_{ij} b_{ij}(x_{ki}),$$

with an intercept given by $x_{k1} = 1$, and where for the covariate x_{ki}, for $i \ge 2$, $b_{i1}, \ldots, b_{iJ}$ is a set of spline basis functions [e.g., $b_{ij}(x_{ki}) = (x_{ki} - \kappa_{ij})_+$ for a set of knots, $\kappa_{i1}, \ldots, \kappa_{iJ}$, within the range of the x_{ki}'s] and u_{ij} are the associated coefficients. The work of Kammann and Wand (2003), who developed a nonparametric spatial regression with Gaussian response, may be similarly extended to the current spatial extremes setting. Spline-based, nonparametric regression, such as penalized splines, employs penalized likelihoods for parameter estimation to avoid over-fitting (e.g., Green and Silverman 1994). Model fitting therefore requires the definition of a penalized composite likelihood for parameter estimation.

3.3 Model Selection

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

$$W \overset{.}{\sim} \sum_{j=1}^{r} \nu_i \chi_i^2, \tag{9}$$

where χ_i^2 are independent χ_1^2 random variables, and $\nu_1 \geq \cdots \geq \nu_r$ are the eigenvalues of $(H^{\boldsymbol{\psi}'\boldsymbol{\psi}'})^{-1}\tilde{I}_{\boldsymbol{\psi}'\boldsymbol{\psi}'}$. Here, $H^{\boldsymbol{\psi}'\boldsymbol{\psi}'}$ and $\tilde{I}_{\boldsymbol{\psi}'\boldsymbol{\psi}'}$ respectively denote the information matrix and the sandwich 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 via (9) either approximates the null distribution using estimates of the eigenvalues ν_i (Rotnitzky and Jewell 1990), or adjusts the composite likelihood such that the usual asymptotic χ_r^2 null is preserved (Chandler and Bate 2007).

The *composite likelihood information criterion* (CLIC) (Varin and Vidoni 2005), useful in the case of nonnested models, performs model selection on the basis of expected Kullback–Leibler divergence between the true unknown model and the adopted model (e.g., Davison 2003, p. 123). In the composite likelihood context, this is the AIC under model misspecification (Takeuchi 1976; Davison 2003, pp. 150–152). Model selection is based on the model minimizing

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

where the second term is the 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) characterized through the covariance matrix, $\boldsymbol{\Sigma}$, 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 moderately large computational demand for large site numbers, likelihood maximization (and other) routines have been implemented in C and collected in the R package SpatialExtremes.

Table 2 summarizes covariance ($\boldsymbol{\Sigma}$) estimator performance based on moderately sized datasets: $K = 50$ sites and $N = 100$ observations. Estimated means and standard errors of the MCLE based on 500 data replications are reported for each spatial dependence structure. Good correspondence with true values is achieved. There is no evidence of bias, even in cases where poorer performance may be expected, such as very strong or weak dependence, and the sandwich standard errors and sample standard deviations are consistent.

We note that in situations where the conditions of Appendix A.2 are not satisfied, the asymptotic distribution of the MCLE can be far from normal. Specifically, as the spatial extremes become increasingly dependent, the true parameter $\boldsymbol{\psi}_0$ approaches the boundary of the parameter space, Ψ, threatening the viability of condition C.1 ($\boldsymbol{\psi}_0$ is an interior point of Ψ). As a consequence, estimation of the Hessian matrix can become problematic and so the Gaussian asymptotics of the MCLE may fail.

For the Gaussian storm profile (4) we require that the covariance $\boldsymbol{\Sigma}$ (or equivalently its inverse) is positive-semidefinite, with nonnegative eigenvalues. In the case of strong dependence in Table 2, the eigenvalues of $\boldsymbol{\Sigma}_4^{-1}$ are $(5 \pm \sqrt{10})/7500$ and so $\boldsymbol{\psi}_0$ is near the parameter space boundary. For a few of the 500 replications, numerical estimates of the Hessian matrix were unstable. In this case, stable estimates were obtained by using a slightly larger than the default relative step-size parameter when estimating the Hessian using finite-difference methods. Performance of the MCLE in Table 2 under strong dependence is otherwise good.

The performance of the estimation of the extremal coefficient function, $\theta(\mathbf{h})$, can be assessed by computing the mean integrated squared error (MISE) and the mean integrated absolute error (MIAE), between the true and estimated extremal coefficients. Table 3 depicts MISE and MIAE for three different estimators of the extremal coefficient functions: the MCLE derived from the Gaussian extreme value model, and those proposed by Smith (1990) and Schlather and Tawn (2003). From Table 3, in general it is clear that the composite likelihood estimator is the most accurate.

Covariance ($\boldsymbol{\Sigma}$) 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 $\boldsymbol{\Sigma}_3$. As expected, the simulations indicate that there is some bias and large 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. Estimator performance for the marginal parameters $\mu(\cdot)$, $\lambda(\cdot)$, and $\xi(\cdot)$ is also examined, under assumed quadratic spatial surfaces for the GEV parameters, and the spatial dependence model $\boldsymbol{\Sigma}_3$. Figure 3 (top panels) compares the true quadratic surfaces (black lines) with the mean of the trend surface estimates (grey lines). Under this model, the integrated absolute error is computed for each of 500 data replications (bottom panels). The correspondence of the marginal parameter estimates appears to be reasonable.
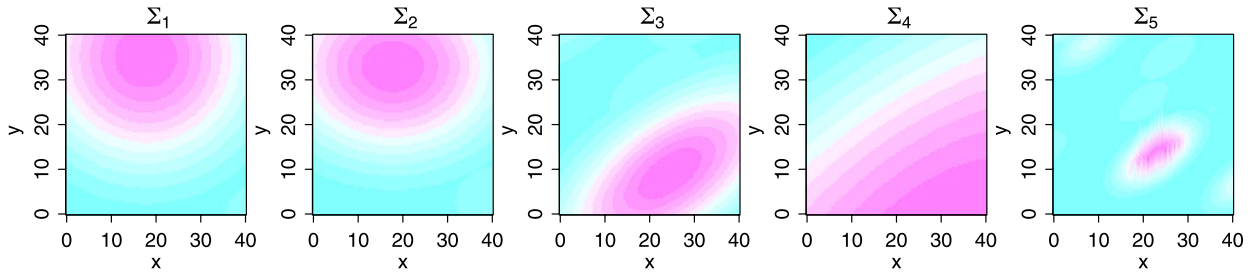
Table 1. Extremal dependence configurations

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 realization for $\Sigma_1, \ldots, \Sigma_5$. Max-stable process simulation routines are available in the `RandomFields` package in `R` (Schlather 2002). A color version of this figure is available in the electronic version of this article.

We consider now model selection under misspecification. Figure 4 illustrates power curves (CLRTs) 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 significance 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 (9) (Rotnitzky and Jewell 1990) appears to have slightly more power—even when testing multiple parameters. In contrast the CLIC statistic demonstrates poor performance, with the rejection rate under the true model reaching only 20%. However, despite this underperformance, the CLIC statistic may be implemented in the case of nonnested models, where the CLRT is unavailable. Accordingly, it may have utility as a model selection tool when searching through large numbers of models.

Finally, we consider MCLE efficiency under composite likelihoods constructed only using neighboring sites. We follow the weighting scheme of Bevilacqua et al. (2007) as described in Section 3.2, for a range of distances, δ. Data are generated under spatial dependence (using Σ_3) for both uniformly and regularly distributed grid site locations over a $[0, 40]^2$ region (Figure 5, top). For the uniform sites (left panels), the trace of the asymptotic covariance, $\mathrm{tr}(\tilde{I}(\psi)^{-1})$, decreases with decreasing distance, δ, and then increases for small distances. The minimum, obtained at $\delta = 12$, corresponds to a composite log-likelihood formed by only 290 (out of 1225) pairs of observations (Figure 5, bottom left). For the regular grid site locations (right panels), the trace decreases monotonically with decreasing δ, with a minimum obtained at $\delta = 6$ (using 112 out of 2016 pairs of observations). In conclusion, utilizing only

neighboring rather than all distinct observational pairs produces gains in computation efficiency, and statistical efficiency of the MCLE.

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 6). We express the model GEV parameters as simple linear functions of longitude, latitude 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 analysis of the model 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). No explanatory variables were found relevant for the shape parameter, whereas first-order terms for location and scale parameters were sufficient. Consequently, based on this linear 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 M_1 also appear competitive. Given the relatively poor performance of the CLIC criterion for model selection (see Section 4), a composite likelihood ratio test is performed to check if M_5 is indeed our best model against the alternatives M_1 and M_2. The same conclusions are obtained from the deviance based tests, with p-values around 0.9. For model M_5, the covariance Σ is estimated as $\hat{\sigma}_1^2 = 0.06026\ (0.0096)$, $\hat{\sigma}_{12} = 0.01420\ (0.0058)$, and

Table 2. Means and standard errors of the MCLE based on 500 spatial extreme data simulations ($K = 50$ sites and $N = 100$ observations) with the Gaussian extreme value model. True values are in [brackets]. Standard errors are obtained through sandwich estimates and sample standard deviation (in parentheses). Computations using Σ_4 were performed using a larger than default relative step-size parameter when estimating the Hessian matrix using finite-difference methods

	$\hat{\sigma}_1^2$ / s.e.	$\hat{\sigma}_2^2$ / s.e.	$\hat{\sigma}_{12}^2$ / 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:	2043 [2000] / 286.2 (314.3)	3033 [3000] / 417.6 (455.4)	1522 [1500] / 287.5 (308.7)
Σ_5:	20 [20] / 1.5 (1.6)	30 [30] / 2.3 (2.3)	15 [15] / 1.6 (1.6)

Table 3. MISE and MIAE of extremal coefficient estimates based on 500 data simulations with $K = 50$ sites and $N = 100$ observations. Estimators are the MCLE and those proposed by Smith (1990) and Schlather and Tawn (2003). Standard deviations are reported in parentheses

	MCLE		Smith		Schlather and Tawn	
	MISE	MIAE	MISE	MIAE	MISE	MIAE
Σ_1:	0.86 (0.92)	24.09 (12.66)	5.14 (3.18)	60.09 (17.45)	4.96 (2.70)	58.83 (16.25)
Σ_2:	0.85 (0.83)	24.50 (11.92)	5.63 (3.02)	63.15 (15.78)	5.74 (2.96)	63.57 (16.28)
Σ_3:	0.76 (0.87)	22.94 (12.25)	7.17 (3.99)	70.64 (18.07)	7.04 (3.78)	70.22 (18.15)
Σ_4:	0.36 (0.53)	14.53 (9.09)	2.15 (2.04)	40.04 (18.21)	1.45 (1.24)	30.42 (12.30)
Σ_5:	0.10 (0.10)	6.20 (3.04)	13.77 (1.83)	102.13 (6.73)	11.46 (2.42)	85.29 (10.97)

Table 4. Means and standard deviations (in parantheses) of the MCLE 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

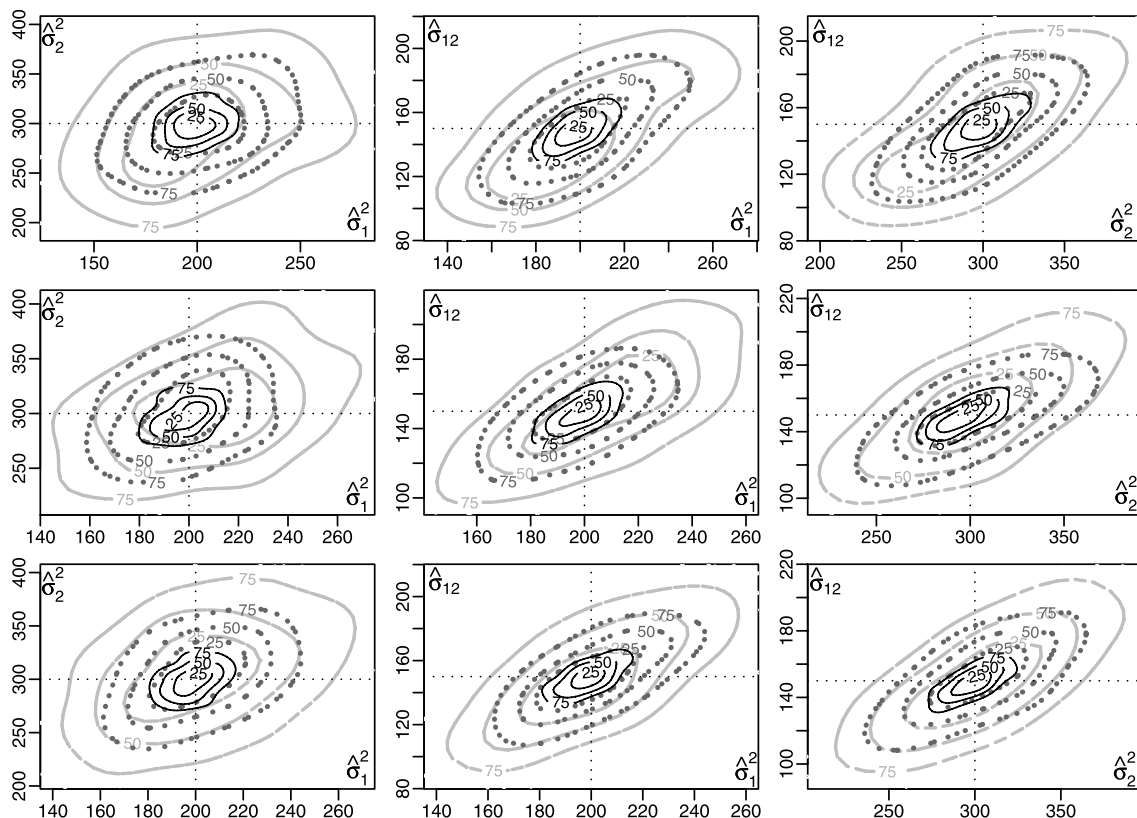|	K	$N = 10$	$N = 50$	$N = 100$	$N = 500$	True
$\hat{\sigma}_1^2$	10	245 (120.2)	207 (43.8)	205 (31.7)	199 (13.3)	200
	50	244 (90.4)	208 (37.5)	200 (28.3)	199 (11.4)	
	100	239 (94.3)	205 (37.8)	202 (30.4)	199 (11.5)	
$\hat{\sigma}_2^2$	10	353 (159.1)	305 (63.9)	301 (44.8)	298 (19.5)	300
	50	353 (131.6)	309 (56.6)	303 (44.3)	298 (16.9)	
	100	361 (143.4)	307 (59.5)	301 (44.8)	299 (16.7)	
$\hat{\sigma}_{12}$	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)	

Figure 2. MCLE distribution estimates for $N = 50$ (grey line), $N = 100$ (dotted line), and $N = 1000$ (solid line) observations over $K = 10$ (top row), $K = 50$ (center row), and $K = 100$ (bottom row) sites. Contours correspond to 0.25, 0.5, and 0.75 percentiles.
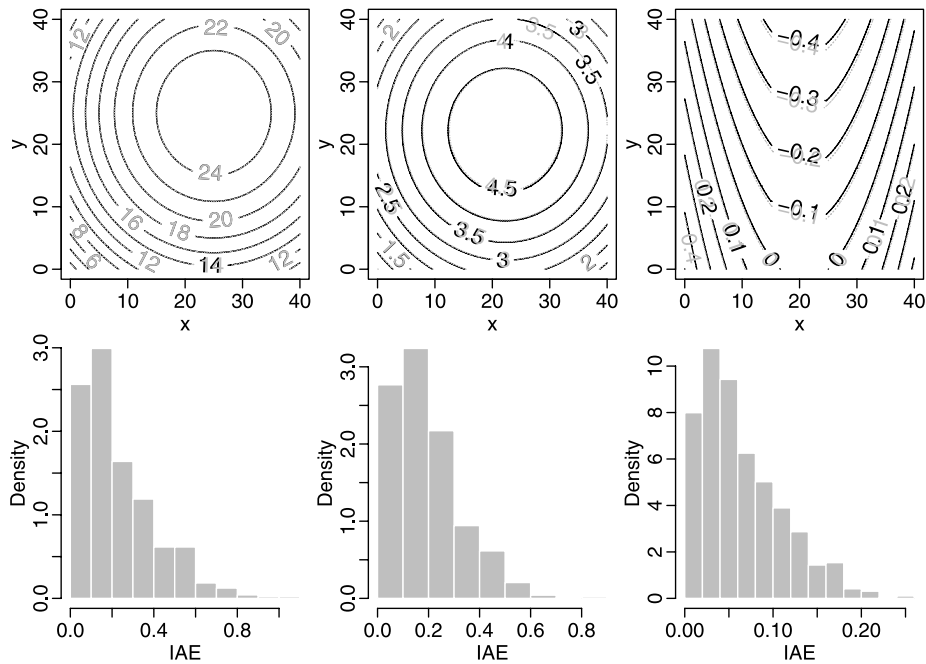
Figure 3. Quadratic spatial surface for location, μ (top left panel), scale, λ (top center), and shape, ξ (top right). Black lines denote true quadratic surfaces. Grey lines display the estimated surfaces corresponding to the mean of the replication-wise simulation. Bottom panels show the respective integrated absolute errors.

$\widehat{\sigma}_2^2 = 0.02849\ (0.0072)$, where the numbers in parentheses denote standard errors.

Figure 7 illustrates the spatial variation of pointwise 50-year return level estimates. Comparison to the regional elevation map (Figure 6) indicates that the most extreme precipitation events occur in mountainous regions.

Figure 8 (left plot) depicts the strength of spatial dependence through the extremal coefficient function $\theta(\mathbf{h})$. There is clear evidence of anisotropy, with stronger dependence in the northeast/southwest direction. Interestingly, this axis corresponds to the shape of the Appalachian Mountains as well as the coastline. Accordingly, this directional extremal dependence may be
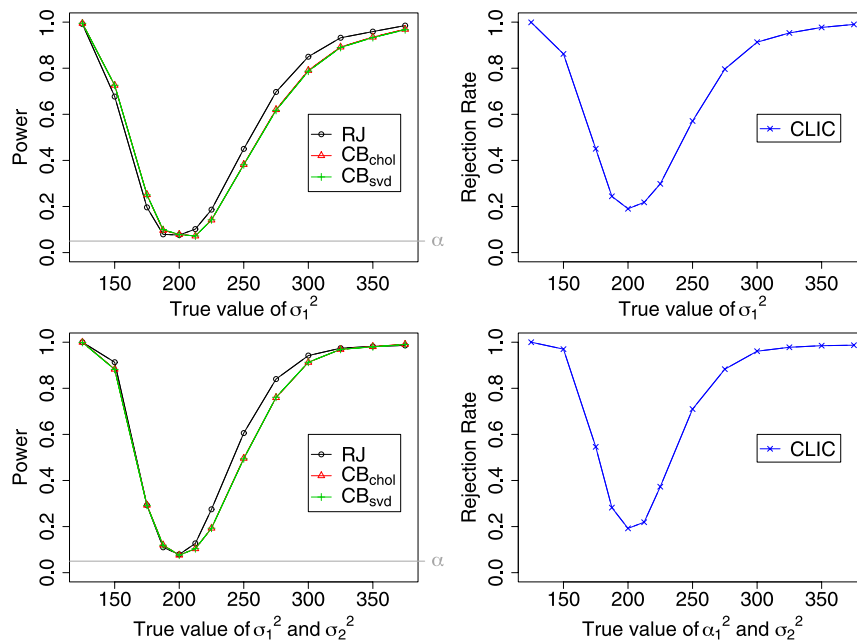

Figure 4. Left panels: Power curves for the CLRTs; RJ (Rotnitzki and Jewell 1990), and CB$_{\text{chol}}$ and CB$_{\text{svd}}$ (Chandler and Bate 2007) using Cholesky and singular value decompositions. Right panels: CLIC rejection rates. Hypotheses are (top panels): $H_0 : \sigma_1^2 = 200$ fixing $\sigma_2^2 = 300$ and $\sigma_{12} = 150$; (bottom panels) $H_0 : \sigma_1^2 = \sigma_2^2 = 200$ fixing $\sigma_{12} = 0$. Test levels are $\alpha = 0.05$. Point estimates are based on 1000 data replications. A color version of this figure is available in the electronic version of this article.
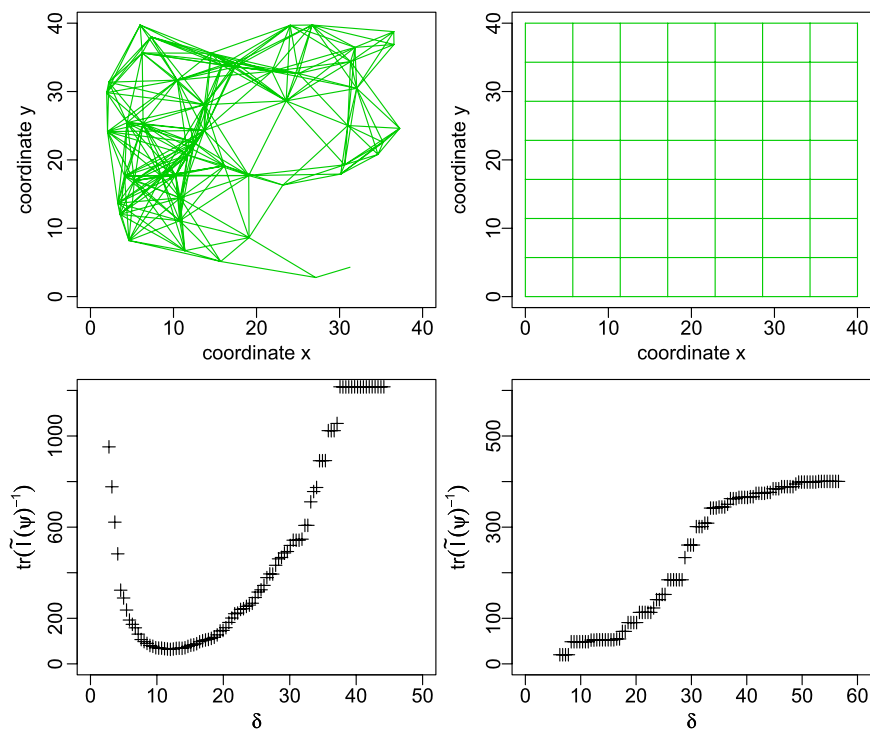
Figure 5. Top panels: Simulated site locations (open circles) over a $[0, 40]^2$ region using both random uniformly (right panel) and regularly (left panel) distributed sites, for which artificial data are generated under spatial dependence. Lines connect site-pairs used to construct the composite log-likelihood with minimum asymptotic covariance trace. Bottom panels: Trace of the asymptotic variance, $\text{tr}(\tilde{I}(\psi)^{-1})$, obtained by using neighboring sites only in computation of the MCLE. Neighboring sites are site pairs which are located within distance δ of each other. A color version of this figure is available in the electronic version of this article.

the consequence of storms following either the coastline or the massif. Moreover, we note that conditional predictive inference is available via the pairwise conditional distribution. Figure 8 (right plot) illustrates the spatial variation of pointwise 50-year return level estimates, conditional on observing an event of magnitude 13 cm at the site indicated by the star.

Finally, we consider model goodness of fit based on repeated simulation under the fitted model. From the observed
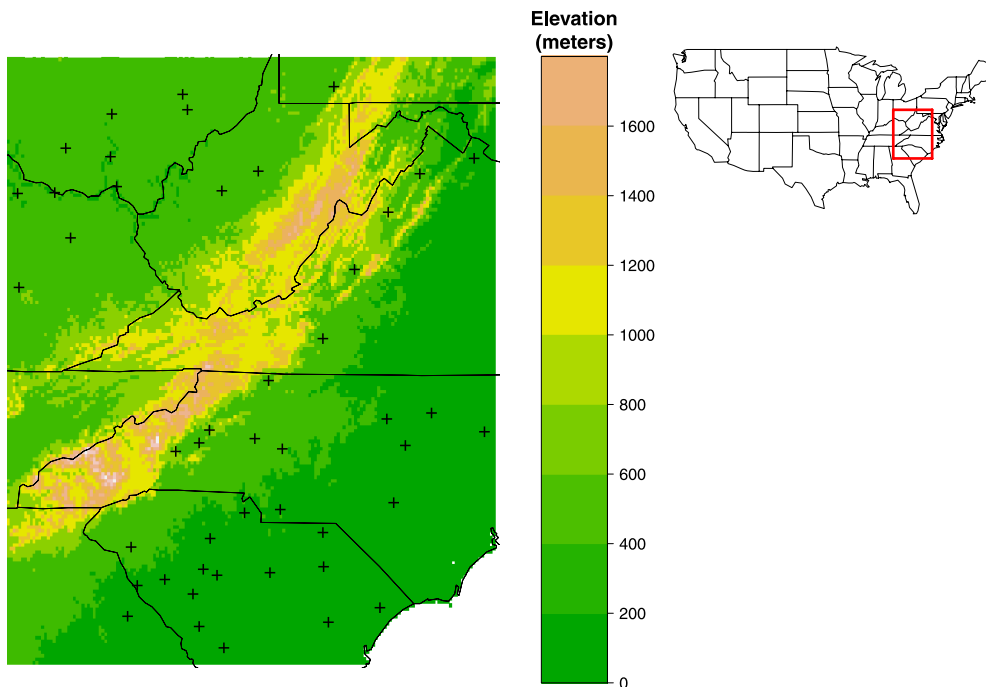

Figure 6. Locations of the 46 gauging stations (crosses) on the east coast of the U.S., with superimposed elevation map (meters) highlighting the shape of the Appalachian Mountains.

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

Model		$-\ell_{\mathcal{P}}(\widehat{\boldsymbol{\psi}}_{\mathrm{MCLE}}; \mathbf{y})$	df	CLIC
M_0:	$\mu(x) = \alpha_0 + \alpha_1(\mathrm{lat}) + \alpha_2(\mathrm{alt}) + \alpha_3(\mathrm{lon})$ $\lambda(x) = \beta_0 + \beta_1(\mathrm{lat}) + \beta_2(\mathrm{alt}) + \beta_2(\mathrm{lon})$ $\xi(x) = \gamma_0$	412,110.5	12	825,679
M_1:	$\mu(x) = \alpha_0 + \alpha_1(\mathrm{lat}) + \alpha_2(\mathrm{alt})$ $\lambda(x) = \beta_0 + \beta_1(\mathrm{lat}) + \beta_2(\mathrm{alt}) + \beta_3(\mathrm{lon})$ $\xi(x) = \gamma_0$	412,111.7	11	825,526
M_2:	$\mu(x) = \alpha_0 + \alpha_1(\mathrm{lat}) + \alpha_2(\mathrm{alt}) + \alpha_3(\mathrm{lon})$ $\lambda(x) = \beta_0 + \beta_1(\mathrm{lat}) + \beta_2(\mathrm{alt})$ $\xi(x) = \gamma_0$	412,113.6	11	825,459
M_3:	$\mu(x) = \alpha_0 + \alpha_1(\mathrm{lat}) + \alpha_3(\mathrm{lon})$ $\lambda(x) = \beta_0 + \beta_1(\mathrm{lat}) + \beta_2(\mathrm{alt}) + \beta_3(\mathrm{lon})$ $\xi(x) = \gamma_0$	412,234.4	11	825,840
M_4:	$\mu(x) = \alpha_0 + \alpha_1(\mathrm{lat}) + \alpha_2(\mathrm{alt}) + \alpha_3(\mathrm{lon})$ $\lambda(x) = \beta_0 + \beta_1(\mathrm{lat}) + \beta_3(\mathrm{lon})$ $\xi(x) = \gamma_0$	412,380.9	11	826,177
M_5:	$\mu(x) = \alpha_0 + \alpha_1(\mathrm{lat}) + \alpha_2(\mathrm{alt})$ $\lambda(x) = \beta_0 + \beta_1(\mathrm{lat}) + \beta_2(\mathrm{alt})$ $\xi(x) = \gamma_0$	**412,113.9**	**10**	**825,327**
M_6:	$\mu(x) = \alpha_0 + \alpha_1(\mathrm{lat})$ $\lambda(x) = \beta_0 + \beta_1(\mathrm{lat}) + \beta_2(\mathrm{alt})$ $\xi(x) = \gamma_0$	412,314.4	9	825,684

data $\{y_{m,k}\}$, we compute $\tilde{y}_m = \max_{k \in K'}\{y_{m,k}\}$ for each block $m = 1, \ldots, M$, where the set $K' \subseteq \{1, \ldots, K\}$ is formed by some subset of sites. The model-based distribution of each $\tilde{y}_m$ is estimated based on 10,000 simulated datasets generated under the fitted model. Figure 9 illustrates in quantile-plot style the ith order-statistic $\tilde{y}_m(i)$ versus the mean of its distribution under the fitted model, in close analogy with diagnostic plots for univariate extremes (e.g., Coles 2001). Dashed lines correspond to 95% confidence envelopes (e.g., Davison and Hinkley 1997, p. 153).

Specifically, Figure 9 considers [top plots, (a)–(c)] three different pairs of locations, representing sites separated by short, medium, and long distances, and [bottom plots, (d)–(f)] the 10 lowest, the 10 highest, and all $K = 46$ sites. The displayed plots are typical of the many examined for these data and

model. Overall, the diagnostics suggest that the observed extreme precipitation data for this region is consistent with the fitted Gaussian extreme value model.

6. CONCLUSION

As a natural generalization of extremal dependence structures, max-stable processes are a powerful tool for the modeling 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 (e.g., Jiang and Turnbull 2004; Bortot, Coles, and Sisson 2007; Sisson, Fan, and Tanaka 2007).

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 modeling of marginal and dependence parameters, coupled with good estimator behavior with finite samples, all at moderate computational cost.

Modifications of the model formulation would draw alternative representations of extremal modeling 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 characterization of extremes. The obvious practical benefit from these extensions would be the incorporation of more data into the modeling process.

APPENDIX

We present conditions for the asymptotic normality and consistency of the MCLE, explicit expressions for the distribution function (3), the density function (4), and the derivatives required for the estimated covariance matrix in Section 3.2.
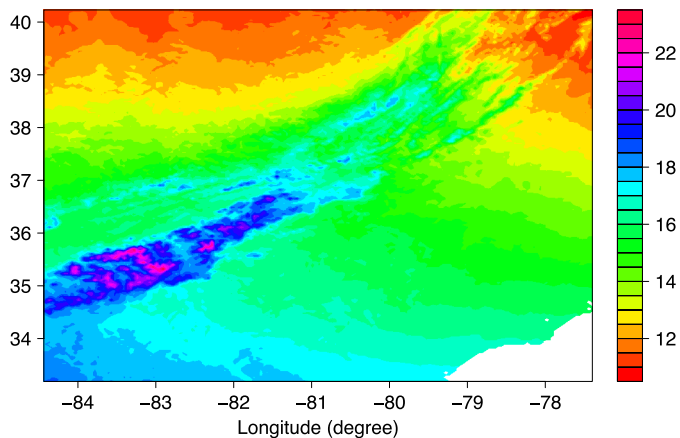

Figure 7. Pointwise 50-year return level map (cm) estimated from the fitted Gaussian extreme value process, model M_5.
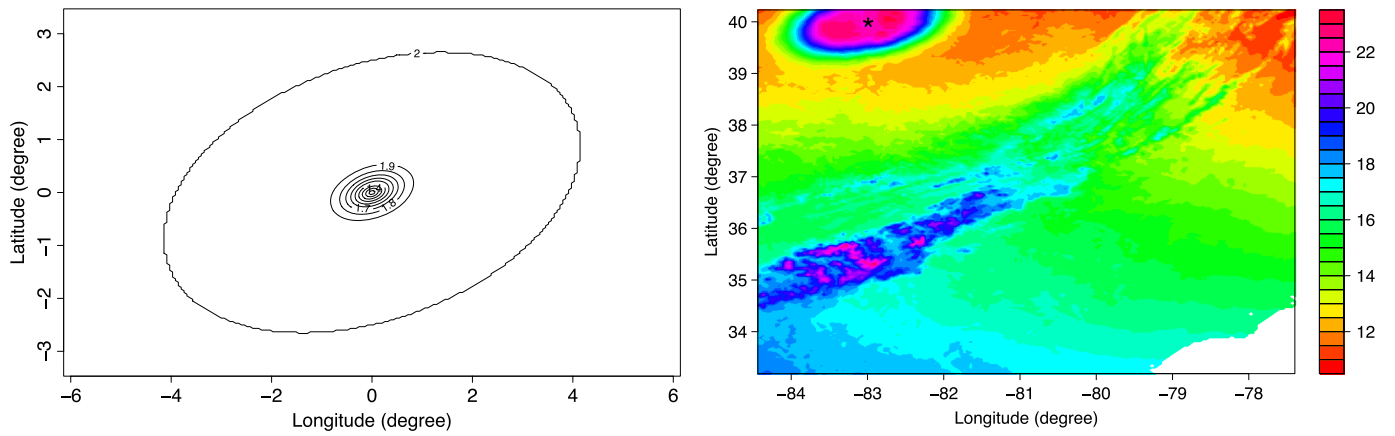
Figure 8. Left: Contour plot of the fitted extremal coefficient $\theta(\mathbf{h}) = 2\Phi(a(\mathbf{h})/2)$; Right: Pointwise 50-year, conditional return level map (cm) estimated from the fitted Gaussian extreme value process. Estimates are conditioned on observing an event of magnitude 13 cm at the site indicated by a star.

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 ith element $\partial f(\mathbf{x})/\partial x_i$. The corresponding Hessian matrix is given by $\mathsf{H}_{\mathbf{x}}f(\mathbf{x}) = \mathsf{D}_{\mathbf{x}}\{\mathsf{D}_{\mathbf{x}}f(\mathbf{x})\}^{\top}$.

A.2 Conditions for the Consistency and Asymptotic Normality of the MCLE

We provide conditions under which the MCLE (Section 3.1) is consistent and asymptotically normal. While in this article we consider the specific case of the marginal pairwise composite log-likelihood (6),
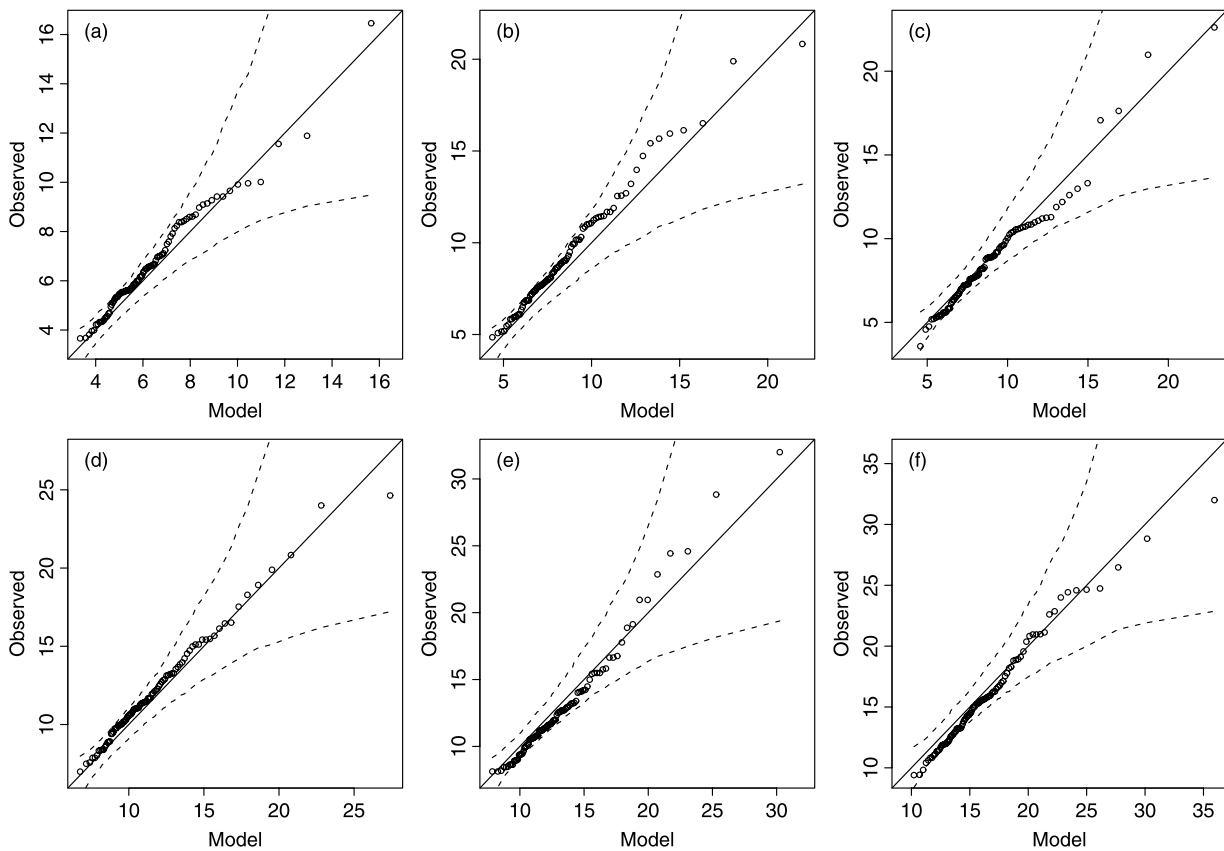

Figure 9. Goodness-of-fit diagnostics for the fitted model for various subsets of site locations, $K' \in \{1, \ldots, K\}$. Panels display quantile plots of observed block-maxima $\tilde{y}_m = \max_{k \in K'}\{y_{m,k}\}$ versus means of their respective distributions, obtained by simulation, under the fitted model (10,000 replications). Dashed lines indicate simulation-based 95% confidence envelopes. Top panels: K' consisting of three pairs of locations separated by (a) short (≈ 20 km), (b) medium (≈ 350 km), and (c) long (≈ 735 km) distances. Bottom panels: K' consisting of (d) the 10 lowest, (e) the 10 highest, and (f) all $K = 46$ sites.

these conditions are relevant for both marginal and conditional events $\{\mathcal{I}_k, k \in \mathcal{K}\}$.

Consider the max-stable process, $\{Y(t)\}$ (Section 2.1), and some parametric family of distributions for the dependence function (e.g., the storm profile model; cf. Smith 1990) derived from the specification of some function f in (1). Then the associated density family $\mathcal{F} = \{f(\mathbf{y}; \boldsymbol{\psi}), \mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^K, \boldsymbol{\psi} \in \Psi \subseteq \mathbb{R}^q\}$ of the process, for fixed K, is assumed to be the parametric statistical model for the data. Define $\widehat{\boldsymbol{\psi}}_{\text{MCLE}}$ as the root of the equation $\mathsf{D}_{\boldsymbol{\psi}} \ell_{\mathcal{C}}(\boldsymbol{\psi}; \mathbf{Y}) = 0$. Suppose that:

C.1 The support, $\mathcal{Y}$, does not depend on $\boldsymbol{\psi}$; the true parameter, $\boldsymbol{\psi}_0$ is an interior point of Ψ; and $F_{\boldsymbol{\psi}} \equiv F_{\boldsymbol{\psi}_0} \Leftrightarrow \boldsymbol{\psi} = \boldsymbol{\psi}_0$ ensures parameter identifiability through the distribution function F. (Alternatively, identifiability may be ensured through the conditional or marginal distributions.)

C.2 $f(\mathbf{y}; \boldsymbol{\psi})$ is continuous in $\boldsymbol{\psi}$ for all $\mathbf{y}$. Further, $\forall \mathbf{y}$ and $\forall \boldsymbol{\psi}$ such that $f(\mathbf{y}; \boldsymbol{\psi}) > 0$, $\mathsf{D}_{\boldsymbol{\psi}} \log f(\mathbf{y}; \boldsymbol{\psi}) < \infty$, and $\mathsf{H}_{\boldsymbol{\psi}} \log f(\boldsymbol{\psi}; \mathbf{y}) < \infty$ (these conditions satisfied componentwise).

C.3 Interchange of differentiation and integration of $f(\mathbf{y}; \boldsymbol{\psi})$ with respect to $\boldsymbol{\psi}$ is valid for $\int f(\mathbf{y}; \boldsymbol{\psi}) \, d\mathbf{y}$ and $\int \mathsf{D}_{\boldsymbol{\psi}} \log f(\mathbf{y}; \boldsymbol{\psi}) \, d\mathbf{y}$, $\forall \boldsymbol{\psi} \in \Psi$.

C.4 $\mathbb{E}_{\boldsymbol{\psi}}[\{\mathsf{D}_{\boldsymbol{\psi}} \log f(\mathbf{y}; \boldsymbol{\psi})\}^2]$ is a positive definite matrix $\forall \boldsymbol{\psi} \in \Psi$.

C.5 $\mathbb{E}_{\boldsymbol{\psi}}\{\mathsf{D}_{\boldsymbol{\psi}} \ell_{\mathcal{C}}(\mathbf{y} \in \mathcal{I}_k; \boldsymbol{\psi})\} = 0$ (for each component of $\boldsymbol{\psi}$), $\forall \boldsymbol{\psi} \in \Psi$. Note that the unbiasedness of the composite score equation stems from the assumption that C.3 also holds for each $f(\mathbf{y} \in \mathcal{I}_k; \boldsymbol{\psi})$.

C.6 $\mathsf{H}_{\boldsymbol{\psi}} \ell_{\mathcal{C}}(\mathbf{y}; \boldsymbol{\psi})$ exists (with finite components) $\forall \mathbf{y} \in \mathcal{I}_{k \in \mathcal{K}}$ and $\forall \boldsymbol{\psi} \in \Psi$.

C.7 Interchange of differentiation and integration of $\ell_{\mathcal{C}}(\mathbf{y}; \boldsymbol{\psi})$ with respect to $\boldsymbol{\psi}$ is valid for $\int \ell_{\mathcal{C}}(\mathbf{y}; \boldsymbol{\psi}) f(\mathbf{y}; \boldsymbol{\psi}) \, d\mathbf{y}$, $\forall \boldsymbol{\psi} \in \Psi$.

C.8 $[\mathbb{E}_{\boldsymbol{\psi}}\{\mathsf{H}_{\boldsymbol{\psi}} \ell_{\mathcal{C}}(\mathbf{y}; \boldsymbol{\psi})\}]^2$ is a positive definite matrix, $\forall \boldsymbol{\psi} \in \Psi$.

C.9 $\ell_{\mathcal{C}}(\mathbf{y}; \boldsymbol{\psi})$ admits a Taylor expansion up to the third order in an open neighborhood of $\boldsymbol{\psi}_0$, where $m^{-1} \mathbb{E}\{|\mathsf{D}_{\boldsymbol{\psi}} \mathsf{H}_{\boldsymbol{\psi}} \ell_{\mathcal{C}}(\mathbf{y}; \boldsymbol{\psi})|\}$ is uniformly bounded $\forall \boldsymbol{\psi}$ in this neighborhood.

Under these assumptions, if the root is unique, then $\widehat{\boldsymbol{\psi}}_{\text{MCLE}}$ is consistent and asymptotically normal as sample size $m \to \infty$. Consistency results of the MCLE are obtained by requiring MLE consistency assumptions on each component in the composite likelihood. The normality of the asymptotic distribution of the MCLE is derived from Taylor expansion arguments of the composite log-likelihood (e.g., similar to those of Davison 2003, p. 147, and Huber 1967).

Note that when modeling with GEV margins, condition C.1 is not strictly satisfied as the support of the parameter space does depend on the parameter vector. However despite this, standard MLE asymptotics are still available under most practical modeling situations (broadly speaking, when $\xi > -1/2$ as shown by Smith (1985).

Observe that the MCLE asymptotics examined here result from assuming an increasing sample size M (of $\mathbf{Y}$) which has a fixed dimension K. However equivalent results are available for the asymptotic behaviour of the MCLE for fixed M and increasing K (e.g., Cox and Reid 2004). Known in spatial statistics as fixed or increasing-domain (Cressie 1993), typically a single replication of a process is observed at many spatial points. MCLE asymptotics in this setting are discussed by, for example, Nott and Rydén (1999) and Bevilacqua et al. (2007).

A.3 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

$$F(z_i, z_j) = \exp\left\{-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \max\left(\frac{f(x_1, x_2)}{z_i}, \right.\right.$$
$$\left.\left. \frac{f(x_1 - t_1, x_2 - t_2)}{z_j}\right) dx_1\, dx_2\right\}$$

$$= \exp\left\{-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{f(x_1, x_2)}{z_i}\right.$$
$$\times \mathbb{I}\left(\frac{f(x_1, x_2)}{z_i} \geq \frac{f(x_1 - t_1, x_2 - t_2)}{z_j}\right) dx_1\, dx_2$$
$$- \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{f(x_1 - t_1, x_2 - t_2)}{z_j}$$
$$\left. \times \mathbb{I}\left(\frac{f(x_1 - t_1, x_2 - t_2)}{z_j} \geq \frac{f(x_1, x_2)}{z_i}\right) dx_1\, dx_2\right\},$$

where $f(x_1, x_2)$ is the bivariate normal density of $(X_1, X_2) \sim \text{N}((0, 0)^\top, \boldsymbol{\Sigma})$, 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

$$\exp\left\{-\frac{1}{2(1 - \rho^2)}\left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1 \sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right\}/z_i$$

$$\geq \exp\left\{-\frac{1}{2(1 - \rho^2)}\left(\frac{(x_1 - t_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - t_1)(x_2 - t_2)}{\sigma_1 \sigma_2}\right.\right.$$
$$\left.\left. + \frac{(x_2 - t_2)^2}{\sigma_2^2}\right)\right\}/z_j$$

$$\Leftrightarrow \quad x_1 \leq \frac{\sigma_1^2 \sigma_2}{2(t_1 \sigma_2 - \rho t_2 \sigma_1)}\left(\frac{t_1^2}{\sigma_1^2} - \frac{2\rho t_1 t_2}{\sigma_1 \sigma_2} + \frac{t_2^2}{\sigma_2^2}\right)$$
$$+ \frac{\sigma_1^2 \sigma_2 (1 - \rho^2)}{t_1 \sigma_2 - \rho t_2 \sigma_1} \log \frac{z_j}{z_i} - \frac{x_2(t_2 \sigma_1^2 - \rho t_1 \sigma_1 \sigma_2)}{t_1 \sigma_2 - \rho t_2 \sigma_1}$$

$$\Leftrightarrow \quad x_1 \leq c.$$

From this, it follows that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{f(x_1, x_2)}{z_i} \mathbb{I}\left(\frac{f(x_1, x_2)}{z_i} \geq \frac{f(x_1 - t_1, x_2 - t_2)}{z_j}\right) dx_1\, dx_2$$

$$= \frac{1}{z_i}\int_{-\infty}^{\infty}\int_{-\infty}^{c} \frac{1}{2\pi\sqrt{1 - \rho^2}}$$
$$\times \exp\left\{-\frac{1}{2(1 - \rho^2)}\left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1 \sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right\} dx_1\, dx_2$$

$$= \frac{1}{z_i}\int_{-\infty}^{\infty} \varphi(x_2)\int_{-\infty}^{c} \varphi\left(\frac{x_1 - \rho\sigma_1 x_2/\sigma_2}{\sigma_1\sqrt{1 - \rho^2}}\right) dx_1\, dx_2$$

$$= \frac{1}{z_i}\int_{-\infty}^{\infty} \varphi(x_2)\Phi\left(\frac{c - \rho\sigma_1 x_2/\sigma_2}{\sigma_1\sqrt{1 - \rho^2}}\right) dx_2$$

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

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

Similarly,

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

It then follows that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{f(x_1 - t_1, x_2 - t_2)}{z_j}$$

$$\times \mathbb{I}\left(\frac{f(x_1, x_2)}{z_i} \geq \frac{f(x_1 - t_1, x_2 - t_2)}{z_j}\right) dx_1\, dx_2$$

$$= \frac{1}{z_j}\int_{-\infty}^{\infty}\int_{c}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-t_1)^2}{\sigma_1^2}\right.\right.$$

$$\left.\left. -\frac{2\rho(x_1-t_1)(x_2-t_2)}{\sigma_1\sigma_2} + \frac{(x_2-t_2)^2}{\sigma_2^2}\right)\right\} dx_1\, dx_2$$

$$= \frac{1}{z_j}\int_{-\infty}^{\infty} \varphi(x_2 - t_2)$$

$$\times \int_{c}^{\infty} \varphi\left(\frac{(x_1-t_1) - \rho\sigma_1(x_2-t_2)/\sigma_2}{\sigma_1\sqrt{1-\rho^2}}\right) dx_1\, dx_2$$

$$= \frac{1}{z_j}\int_{-\infty}^{\infty} \varphi(x_2 - t_2)\left\{1 - \Phi\left(\frac{c - \rho\sigma_1(x_2-t_2)/\sigma_2}{\sigma_1\sqrt{1-\rho^2}}\right)\right\} dx_2$$

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

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

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

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).

A.4 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 $a \equiv a(\mathbf{h})$, $w \equiv w(\mathbf{h})$, and $v \equiv v(\mathbf{h})$ and write $w = a/2 + \log(z_j/z_i)/a$ and $v = a - w$. The differentiation gives

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

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

$$\times \left\{\frac{\partial}{\partial z_i}\left(-\frac{\Phi(w)}{z_i} - \frac{\Phi(v)}{z_j}\right)\frac{\partial}{\partial z_j}\left(-\frac{\Phi(w)}{z_i} - \frac{\Phi(v)}{z_j}\right)\right.$$

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

First-order differentiation gives

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

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

using the results

$$\frac{\partial\Phi(w)}{\partial z_i} = -\frac{\varphi(w)}{az_i}, \qquad \frac{\partial\Phi(v)}{\partial z_i} = \frac{\varphi(v)}{az_i}$$

and

$$\frac{\partial w}{\partial z_i} = -\frac{1}{az_i}, \qquad \frac{\partial v}{\partial z_i} = \frac{1}{az_i}.$$

Second-order differentiation yields

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

using

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

Substituting, we obtain the probability density function

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

$$\times \left\{\left(\frac{\Phi(w)}{z_i^2} + \frac{\varphi(w)}{az_i^2} - \frac{\varphi(v)}{az_iz_j}\right)\right.$$

$$\times \left(\frac{\Phi(v)}{z_j^2} + \frac{\varphi(v)}{az_j^2} - \frac{\varphi(w)}{az_iz_j}\right)$$

$$\left. + \left(\frac{v\varphi(w)}{a^2z_i^2z_j} + \frac{w\varphi(v)}{a^2z_iz_j^2}\right)\right\}.$$

A.5 An Expression for the Squared Score Statistic

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

$$\sum_{m=1}^{M}\left\{\sum_{i=1}^{K-1}\sum_{j=i+1}^{K} w_{ij}\mathsf{D}_{\boldsymbol{\psi}}\ell_{ij}(\boldsymbol{\psi})\right\}\left\{\sum_{i=1}^{K-1}\sum_{j=i+1}^{K} w_{ij}\mathsf{D}_{\boldsymbol{\psi}}\ell_{ij}(\boldsymbol{\psi})\right\}^{\top},$$

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

$$\ell_{ij}(\boldsymbol{\psi}) = \mathrm{A} + \log(\mathrm{BC} + \mathrm{D}) + \mathrm{E},$$

where

$$\mathrm{A} = -\frac{\Phi(w)}{z_i} - \frac{\Phi(v)}{z_j}, \qquad \mathrm{B} = \frac{\Phi(w)}{z_i^2} + \frac{\varphi(w)}{az_i^2} - \frac{\varphi(v)}{az_iz_j},$$

$$\mathrm{C} = \frac{\Phi(v)}{z_j^2} + \frac{\varphi(v)}{az_j^2} - \frac{\varphi(w)}{az_iz_j}, \qquad \mathrm{D} = \frac{v\varphi(w)}{a^2z_i^2z_j} + \frac{w\varphi(v)}{a^2z_iz_j^2},$$

and

$$\mathrm{E} = \log\left\{\frac{1}{\lambda_i\lambda_j}\left(1 + \xi_i\frac{y_i - \mu_i}{\lambda_i}\right)_+^{1/\xi_i - 1}\left(1 + \xi_j\frac{y_j - \mu_j}{\lambda_j}\right)_+^{1/\xi_j - 1}\right\}$$

and where $\mu_i = (\mathbf{X}_{\boldsymbol{\beta}_\mu}\boldsymbol{\beta}_\mu)_i$, $\xi_i = (\mathbf{X}_{\boldsymbol{\beta}_\xi}\boldsymbol{\beta}_\xi)_i$, and $\log(\psi_i) = (\mathbf{X}_{\boldsymbol{\beta}_\psi}\boldsymbol{\beta}_\psi)_i$. The GEV parameters are related to the predictors by the form (8). 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 determinant of the Jacobian matrix associated with the transformation (7); see Section 3.2.

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

$$\mathsf{D}_{\boldsymbol{\psi}}\ell_{ij}(\boldsymbol{\psi}) = \left(\mathsf{D}_{\boldsymbol{\sigma}}\ell_{ij}(\boldsymbol{\psi}), \mathsf{D}_{\boldsymbol{\beta}_\mu}\ell_{ij}(\boldsymbol{\psi}), \mathsf{D}_{\boldsymbol{\beta}_\lambda}\ell_{ij}(\boldsymbol{\psi}), \mathsf{D}_{\boldsymbol{\beta}_\xi}\ell_{ij}(\boldsymbol{\psi})\right).$$

Vector differential calculus (e.g., Wand 2002) leads to

$$\mathsf{D}_{\boldsymbol{\sigma}}\ell_{ij}(\boldsymbol{\psi}) \equiv \left[-\frac{v\varphi(w)}{az_i} - \frac{w\varphi(v)}{az_j}\right.$$

$$+ \left\{\mathrm{C}\left(\frac{(w^2 - 1)\varphi(v)}{a^2z_j^2} + \frac{(1 + wv)\varphi(w)}{a^2z_iz_j}\right)\right.$$

$$+ \mathrm{B}\left(\frac{(v^2-1)\varphi(w)}{a^2 z_i^2} + \frac{(1+wv)\varphi(v)}{a^2 z_i z_j}\right)$$

$$+ \frac{(w-2v-wv^2)\varphi(w)}{a^3 z_i^2 z_j}$$

$$+ \left.\frac{(v-2w-vw^2)\varphi(v)}{a^3 z_j^2 z_i}\right\} \Big/ (\mathrm{BC}+\mathrm{D})\right] \mathsf{D}_{\boldsymbol\sigma} a^\top,$$

where we have used the results

$$\mathsf{D}_{\boldsymbol\sigma} a^\top = \frac{1}{2a|\boldsymbol\Sigma|^2}\,(t_1^2 \quad t_1 t_2 \quad t_2^2)$$

$$\times \begin{bmatrix} -\sigma_2^4 & 2\sigma_2^2\sigma_{12} & -\sigma_{12}^2 \\ 2\sigma_2^2\sigma_{12} & -2(\sigma_1^2\sigma_2^2+\sigma_{12}^2) & 2\sigma_1^2\sigma_{12} \\ -\sigma_{12}^2 & 2\sigma_1^2\sigma_{12} & -\sigma_1^4 \end{bmatrix},$$

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

$$\mathsf{D}_{\boldsymbol\sigma}\varphi(w) = -\frac{wv\varphi(w)}{a}\mathsf{D}_{\boldsymbol\sigma} a^\top,$$

and

$$\mathsf{D}_{\boldsymbol\sigma} v\varphi(w) = \frac{w(1-v^2)\varphi(w)}{a}\mathsf{D}_{\boldsymbol\sigma} a^\top.$$

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

$$\mathsf{D}_{\boldsymbol\beta_\mu}\ell_{ij}(\boldsymbol\psi)$$

$$\equiv \left\{\left(\frac{\varphi(w)+a\Phi(w)}{a z_i^2} - \frac{\varphi(v)}{a z_i z_j}\right)\frac{z_i^{1-\xi_i}}{\lambda_i}\right\}(\mathbf{X}_{\boldsymbol\beta_\mu})_i$$

$$+ \left\{\left(\frac{\varphi(v)+a\Phi(v)}{a z_j^2} - \frac{\varphi(w)}{a z_i z_j}\right)\frac{z_j^{1-\xi_j}}{\lambda_j}\right\}(\mathbf{X}_{\boldsymbol\beta_\mu})_j$$

$$+ \left\{\mathrm{C}\left(\frac{w\varphi(v)}{a^2 z_i z_j^2} + \frac{v\varphi(w)}{a^2 z_j z_i^2}\right)\frac{z_i^{1-\xi_i}}{\lambda_i}\Big/(\mathrm{BC}+\mathrm{D})\right\}(\mathbf{X}_{\boldsymbol\beta_\mu})_i$$

$$+ \left\{\mathrm{B}\left(\frac{v\varphi(w)}{a^2 z_j z_i^2} + \frac{w\varphi(v)}{a^2 z_i z_j^2}\right)\frac{z_j^{1-\xi_j}}{\lambda_j}\Big/(\mathrm{BC}+\mathrm{D})\right\}(\mathbf{X}_{\boldsymbol\beta_\mu})_j$$

$$+ \left\{\mathrm{C}\left(\frac{(a+w)\varphi(w)}{a^2 z_i z_j^2} - \frac{(2a+w)\varphi(v)}{a^2 z_j^3} - \frac{2\Phi(v)}{z_i^3}\right)\frac{z_j^{1-\xi_j}}{\lambda_j}\right.$$

$$\left.\Big/(\mathrm{BC}+\mathrm{D})\right\}(\mathbf{X}_{\boldsymbol\beta_\mu})_j$$

$$+ \left\{\mathrm{B}\left(\frac{(a+v)\varphi(v)}{a^2 z_j z_i^2} - \frac{(2a+v)\varphi(w)}{a^2 z_i^3} - \frac{2\Phi(w)}{z_j^3}\right)\frac{z_i^{1-\xi_i}}{\lambda_i}\right.$$

$$\left.\Big/(\mathrm{BC}+\mathrm{D})\right\}(\mathbf{X}_{\boldsymbol\beta_\mu})_i$$

$$+ \left\{\left(\frac{(1-av-v^2)\varphi(w)}{a^3 z_j z_i^3} - \frac{(1+aw+vw)\varphi(v)}{a^3 z_i^2 z_j^2}\right)\frac{z_i^{1-\xi_i}}{\lambda_i}\right.$$

$$\left.\Big/(\mathrm{BC}+\mathrm{D})\right\}(\mathbf{X}_{\boldsymbol\beta_\mu})_i$$

$$+ \left\{\left(\frac{(1-aw-w^2)\varphi(v)}{a^3 z_i z_j^3} - \frac{(1+av+vw)\varphi(w)}{a^3 z_j^2 z_i^2}\right)\frac{z_j^{1-\xi_j}}{\lambda_j}\right.$$

$$\Big/(\mathrm{BC}+\mathrm{D})\bigg\}(\mathbf{X}_{\boldsymbol\beta_\mu})_j$$

$$+ \frac{(\xi_i-1)}{\lambda_i z_i^{\xi_i}}(\mathbf{X}_{\boldsymbol\beta_\mu})_i + \frac{(\xi_j-1)}{\lambda_j z_j^{\xi_j}}(\mathbf{X}_{\boldsymbol\beta_\mu})_j.$$

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

$$\mathsf{D}_{\boldsymbol\beta_\lambda}z_i \equiv -\frac{z_i(y_i-\mu_i)}{\lambda_i}(\mathbf{X}_{\boldsymbol\beta_\lambda})_i$$

and

$$\mathsf{D}_{\boldsymbol\beta_\xi}z_i \equiv \left\{\frac{1}{\xi_i}\left(\frac{z_i^{1-\xi_i}(y_i-\mu_i)}{\lambda_i} - z_i\log(z_i)\right)\right\}(\mathbf{X}_{\boldsymbol\beta_\xi})_i.$$

Finally,

$$\mathsf{D}_{\boldsymbol\beta_\lambda}\mathrm{E} \equiv \left(\frac{(\xi_i-1)(y_i-\mu_i)}{\lambda_i z_i} - \right)(\mathbf{X}_{\boldsymbol\beta_\lambda})_i$$

$$+ \left(\frac{(\xi_j-1)(y_j-\mu_j)}{\lambda_j \mathbf{z}_j} - \right)(\mathbf{X}_{\boldsymbol\beta_\lambda})_j$$

and

$$\mathsf{D}_{\boldsymbol\beta_\xi}\mathrm{E} \equiv \left[\frac{1}{\xi_i}\left\{\frac{(1-\xi_i)(y_i-\mu_i)}{\mathbf{z}_i^{\xi_i}\lambda_i} - \log(z_i)\right\}\right](\mathbf{X}_{\boldsymbol\beta_\xi})_i$$

$$+ \left[\frac{1}{\xi_j}\left\{\frac{(1-\xi_j)(y_j-\mu_j)}{\mathbf{z}_j^{\xi_j}\lambda_j} - \log(z_j)\right\}\right](\mathbf{X}_{\boldsymbol\beta_\xi})_j.$$

Combining all these results together we obtain an expression for the squared score statistic.

[Received October 2008. Revised September 2009.]

REFERENCES

Barndorff-Nielsen, O. E., Gupta, V. K., Pérez-Abreu, V., and Waymire, E. (eds.) (1998), *Stochastic Methods in Hyrdology*, Singapore, New Jersey, London, Hong Kong: World Scienfitic. [263]

Bevilacqua, M., Gaetan, C., Mateu, J., and Porcu, E. (2007), "Estimating Space and Space–Time Covariance Functions: A Weighted Composite Likelihood Approach," Technical Report 117–2007, Universitat Jaume I. [266,268,274]

Bortot, P., Coles, S. G., and Sisson, S. A. (2007), "Inference for Stereological Extremes," *Journal of the American Statistical Association*, 102, 84–92. [272]

Broyden, C. G. (1967), "Quasi-Newton Methods and Their Application to Function Minimization," *Mathematics of Computation*, 21, 368–381. [266]

Chandler, R. E., and Bate, S. (2007), "Inference for Clustered Data Using the Independence Loglikelihood," *Biometrika*, 94, 167–183. [267,268,270]

Chavez-Demoulin, V., and Davison, A. C. (2005), "Generalised Additive Modelling of Sample Extremes," *Applied Statistics*, 54, 207–222. [266]

Coles, S. G. (1993), "Regional Modelling of Extreme Storms via Max-Stable Processes," *Journal of the Royal Statistical Society, Ser. B*, 55, 797–816. [263,264]

———— (2001), *An Introduction to Statistical Modelling of Extreme Values*, London: Springer. [272]

Coles, S. G., and Tawn, J. A. (1996), "Modelling Extremes of the Areal Rainfall Process," *Journal of the Royal Statistical Society, Ser. B*, 58, 329–347. [263]

Coles, S. G., and Walshaw, D. (1994), "Directional Modelling of Extreme Wind Speeds," *Journal of Applied Statistics*, 33, 139–158. [263]

Cox, D. R., and Reid, N. (2004), "A Note on Pseudolikelihood Constructed From Marginal Densities," *Biometrika*, 91, 729–737. [265,274]

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York: Wiley. [263, 274]

Davison, A. C. (2003), *Statistical Models*, New York: Cambridge University Press. [265,267,274]

Davison, A. C., and Hinkley, D. V. (1997), *Bootstrap Methods and Their Application*, New York: Cambridge University Press. [272]

Davison, A. C., and Smith, R. L. (1990), "Models for Exceedances Over High Thresholds" (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 52, 393–402. [272]

de Haan, L. (1984), "A Spectral Representation for Max-Stable Processes," *The Annals of Probability*, 12, 1194–1204. [263,264]

de Haan, L., and Pereira, T. T. (2006), "Spatial Extremes: Models for the Stationary Case," *The Annals of Statistics*, 34, 146–168. [263-265,275]

de Haan, L., and Pickands, J. (1986), "Stationary Min-Stable Stochastic Processes," *Probability Theory and Related Fields*, 74, 477–492. [263,264]

de Haan, L., and Resnick, S. I. (1977), "Limit Theory for Multivariate Sample Extremes," *Zeitschrift Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40, 317–337. [264]

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. [264]

Godambe, V. P. (1960), "An Optimum Property of Regular Maximum Likelihood Estimation," *Annals of Mathematical Statistics*, 31, 1208–1211. [265]

Green, P. J., and Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, London: Chapman & Hall. [266]

Heyde, C. (1997), *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*, New York: Springer. [266]

Huber, P. J. (1967), "The Behaviour of Maximum Likelihood Estimates Under Nonstandard Conditions," in *Proceedings of the 5th Berkeley Symposium of Mathematic, Statistic and Probability*, Berkeley: University of California Press, pp. 221–233. [265,274]

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. [263]

Jiang, W., and Turnbull, B. (2004), "The Indirect Method: Inference Based on Intermediate Statistics—A Synthesis and Examples," *Statistical Science*, 19, 239–263. [272]

Kammann, E. E., and Wand, M. P. (2003), "Geoadditive Models," *Applied Statistics*, 52, 1–18. [266]

Kent, J. T. (1982), "Information Gain and a Measure of Correlation," *Biometrika*, 70, 163–173. [267]

Lindsay, B. G. (1988), "Composite Likelihood Methods," *Contemporary Mathematics*, 80, 221–239. [263,265]

Nott, D. J., and Rydén, T. (1999), "Pairwise Likelihood Methods for Inference in Image Models," *Biometrika*, 86, 661–676. [274]

Padoan, S. A., and Wand, M. P. (2008), "Mixed Model-Based Additive Models for Sample Extremes," *Statistics and Probability Letters*, 78, 2850–2858. [266]

Resnick, S. (1987), *Extreme Values, Point Processes and Regular Variation*, New York: Springer-Verlag. [263,264]

Ripley, B. D. (2004), *Spatial Statistics*, New Jersey: Wiley. [263]

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. [267,268,270]

Schlather, M. (2002), "Models for Stationary Max-Stable Random Fields," *Extremes*, 5, 33–44. [264,265,268]

Schlather, M., and Tawn, J. A. (2003), "A Dependence Measure for Multivariate and Spatial Extreme Values: Properties and Inference," *Biometrika*, 90, 139–154. [264,267,269]

Sisson, S. A., Fan, Y., and Tanaka, M. M. (2007), "Sequential Monte Carlo Without Likelihoods," *Proceedings of the National Academy of Sciences*, 104, 1760–1765. [272]

Smith, R. L. (1985), "Maximum Likelihood Estimation in a Class of Nonregular Cases," *Biometrika*, 72, 67–90. [274]

—— (1990), "Max-Stable Processes and Spatial Extremes," unpublished manuscript, University of North California. [263,264,267,269,274,275]

Takeuchi, K. (1976), "Distribution of Information Statistics and Criteria for Adequacy of Models," *Mathematical Science*, 153, 12–18 (in Japanese). [267]

Varin, C. (2008), "On Composite Marginal Likelihoods," *Advances in Statistical Analysis*, 92, 1–28. [265,267]

Varin, C., and Vidoni, P. (2005), "A Note on Composite Likelihood Inference and Model Selection," *Biometrika*, 92, 519–528. [267]

von Mises, R. (1954), "La Distribution de la Plus Grande de *n* Valeurs," in *Selected Papers*, Vol. II, Providence, RI: American Mathematical Society, pp. 271–294. [263]

Wand, M. P. (2002), "Vector Differential Calculus in Statistics," *The American Statistician*, 56, 55–62. [275]

Yee, T. W., and Stephenson, A. G. (2007), "Vector Generalised Linear and Additive Extreme Value Models," *Extremes*, 10, 1–19. [266]