

# Efficient inference for spatial extreme value processes associated to log-Gaussian random functions

BY JENNIFER L. WADSWORTH

*Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland*

jenny.wadsworth@epfl.ch

AND JONATHAN A. TAWN

*Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, U.K.*

j.tawn@lancaster.ac.uk

## SUMMARY

Max-stable processes arise as the only possible nontrivial limits for maxima of affinely normalized identically distributed stochastic processes, and thus form an important class of models for the extreme values of spatial processes. Until recently, inference for max-stable processes has been restricted to the use of pairwise composite likelihoods, due to intractability of higher-dimensional distributions. In this work we consider random fields that are in the domain of attraction of a widely used class of max-stable processes, namely those constructed via manipulation of log-Gaussian random functions. For this class, we exploit limiting  $d$ -dimensional multivariate Poisson process intensities of the underlying process for inference on all  $d$ -vectors exceeding a high marginal threshold in at least one component, employing a censoring scheme to incorporate information below the marginal threshold. We also consider the  $d$ -dimensional distributions for the equivalent max-stable process, and perform full likelihood inference by exploiting the methods of Stephenson & Tawn (2005), where information on the occurrence times of extreme events is shown to dramatically simplify the likelihood. The Stephenson–Tawn likelihood is in fact simply a special case of the censored Poisson process likelihood. We assess the improvements in inference from both methods over pairwise likelihood methodology by simulation.

*Some key words:* Extreme value theory; Likelihood inference; Max-stable process; Poisson process; Spatial extreme.

## 1. INTRODUCTION

Max-stable processes form an important class of models for extreme values of spatial processes (e.g., Padoan et al., 2010; Davison et al., 2012; Davison & Gholamrezaee, 2012). Max-stable processes arise as the only possible nondegenerate laws of normalized identically distributed stochastic processes in the following manner. Let  $\mathcal{S} \subset \mathbb{R}^2$  be a compact index space, and let  $\{Y_i(s) : s \in \mathcal{S}\}$  be independent and identically distributed copies of a stochastic process on  $\mathcal{S}$ . If there exist functions  $a_n(s) > 0$  and  $b_n(s) \in \mathbb{R}$  such that for a pointwise max operation, the distributional convergence

$$M_{\text{GEV}}(s) = \lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \{Y_i(s) - b_n(s)\} / a_n(s) \quad (1)$$

yields a process  $M_{\text{GEV}}(s)$  with nondegenerate margins for all  $s$ , then  $M_{\text{GEV}}(s)$  is a max-stable process. The univariate margins of  $M_{\text{GEV}}(s)$  are necessarily generalized extreme value distributed, that is, for  $a_+ = \max(a, 0)$ , they take the form

$$\text{pr}\{M_{\text{GEV}}(s) \leq x\} = \exp\left(-[1 + \xi(s)\{x - \mu(s)\}/\sigma(s)]_+^{-1/\xi(s)}\right), \quad (2)$$

where  $\mu(s) \in \mathbb{R}$ ,  $\sigma(s) > 0$  and  $\xi(s) \in \mathbb{R}$  are location, scale and shape parameters, respectively, with a limit interpretation for  $\xi = 0$ . We denote distribution (2) by  $\text{GEV}\{\mu(s), \sigma(s), \xi(s)\}$ .

Spatial data are often available as point-referenced  $d$ -dimensional multivariate data, with replications over time, and this is the situation we assume. To appeal to the limit theory (1), one takes pointwise maxima over a sequence of blocks of a large number of replications of the field  $Y(s)$ , to yield a set of approximate realizations of a max-stable field  $M_{\text{GEV}}(s)$ , in the form of a  $d$ -variate vector  $M_{\text{GEV}} = \{M_{\text{GEV}}(s_1), \dots, M_{\text{GEV}}(s_d)\}$ . Both marginal and dependence parameters must be estimated, and to simplify matters it is common to proceed in two stages. The marginal parameters are estimated initially, and then used with the probability integral transform to convert the field  $M_{\text{GEV}}(s)$  to a common marginal scale, frequently the unit Fréchet scale,  $\text{GEV}(1, 1, 1)$ . We will denote a max-stable field with unit Fréchet margins by  $M_{\text{F}}(s)$ .

Models for spatial dependence usually follow from a constructive approach to max-stable fields (de Haan, 1984; Schlather, 2002), expressed in terms of spectral functions. Let  $\{r_i\}$ ,  $i \geq 1$ , be points of a Poisson process on  $(0, \infty)$ , with intensity  $dr$ , and let  $\{W_i(s)\}$  be independent and identically distributed copies of a nonnegative-valued continuous random function on  $\mathcal{S}$  satisfying  $E\{W(s)\} = 1$  for all  $s \in \mathcal{S}$ , termed the spectral function. Then  $X_i(s) = W_i(s)/r_i$  are points of a Poisson process on the space of continuous nonnegative functions on  $\mathcal{S}$ , and

$$M_{\text{F}}(s) = \max_i X_i(s) \quad (3)$$

is a max-stable random field with unit Fréchet margins and  $d$ -dimensional distribution function

$$\text{pr}\{M_{\text{F}}(s_1) \leq x_1, \dots, M_{\text{F}}(s_d) \leq x_d\} = \exp\left(-E\left[\max\left\{\frac{W(s_1)}{x_1}, \dots, \frac{W(s_d)}{x_d}\right\}\right]\right) \quad (4)$$

for  $x_i > 0$ ,  $i = 1, \dots, d$ . Further theoretical aspects of extreme value theory for functions are treated in, e.g., de Haan & Ferreira (2006, Ch. 9). The dependence structure is thus prescribed through  $W$  via the exponent function,  $V: \mathbb{R}_+^d \rightarrow \mathbb{R}_+$ , defined by

$$V(x) = E\left[\max\left\{\frac{W(s_1)}{x_1}, \dots, \frac{W(s_d)}{x_d}\right\}\right]. \quad (5)$$

The function  $V(x)$  is also the integrated intensity measure of the Poisson process of points  $X_i = \{X_i(s_1), \dots, X_i(s_d)\}$  on the region  $\mathbb{R}_+^d \setminus \{[0, x_1] \times \dots \times [0, x_d]\} = [0, x]^c$ . Typically (5) is difficult to calculate, and this is the limiting factor in finding tractable models for which inference is feasible. There exists a slowly growing set of models for which the bivariate distributions are available; see, e.g., Schlather (2002), Kabluchko et al. (2009), Davison et al. (2012) and Wadsworth & Tawn (2012). For such models, pairwise composite likelihood methods (Padoan et al., 2010) have become the standard for inference.

The focus of this paper is inference for a class of processes, commonly known as Brown–Resnick processes (Brown & Resnick, 1977; Kabluchko et al., 2009), whose spectral functions are log-Gaussian random fields. The  $d$ -dimensional distribution functions for all such

processes are known (Genton et al., 2011; Huser & Davison, 2013; Engelke et al., 2014), but owing to the exponential form of distribution function (4), even when the expectation is calculable, differentiation to yield high-dimensional densities produces an explosion of terms in the density: the  $d$ -variate density consists of  $B_d$  summands, where  $B_d$  is the  $d$ th Bell number. Furthermore, the known representations of the exponent function admit apparently awkward derivatives, which are themselves sums of several terms. Thus in practice tripletwise composite likelihood inference is possible, though Huser & Davison (2013) suggest that, except for the Gaussian extreme value process model, a special boundary case of the Brown–Resnick class, the relative gain in parameter estimation efficiency is not overwhelming.

When taking pointwise maxima, the resulting dependence structure is rather complex, since the max-stable random field may be formed from a composition of several different underlying fields. In this work we explore how this dependence can be substantially simplified by modelling the dependence of individual extreme events. In a multivariate setting, Coles & Tawn (1991) introduced methodology for modelling original events via a Poisson process representation, where previous work had focused only on componentwise maxima (Tawn, 1990). However, assuming the Poisson process model to be accurate in regions where not all variables are extreme can induce bias in parameter estimation (e.g., Ledford & Tawn, 1996). Smith et al. (1997) introduced a related approach, but using a censored likelihood to reduce this bias. Rootzén & Tajvidi (2006) also introduced a model for multivariate threshold exceedances in the form of the multivariate generalized Pareto distribution.

In this work, we adapt these principles to the spatial setting, reducing the number of summands in the likelihood from  $B_d$  to one. Our premise is the Poisson process convergence of the extreme events to the elements  $X_i(s)$  in representation (3). For  $W_i(s)$  log-Gaussian, simple forms arise for the intensity of the limiting Poisson process. At a limiting level, any random field which is extreme in at least one coordinate would follow the Poisson process intensity. At a practical level, to use the full limit model would be likely to induce bias due to lack of applicability of limiting arguments to non extreme components. In order to eliminate any bias that may thus arise, we consider a censored Poisson process likelihood, which also admits relatively simple forms. A main contribution of the work is therefore to adapt the inferential ideas of Coles & Tawn (1991) and Smith et al. (1997) to the spatial setting, and demonstrate that this alternative to max-stable process modelling permits more efficient inference. Threshold-based inference for extremal events has also been considered by Huser & Davison (2014), who adapt the pairwise composite likelihood approach for maxima to a threshold setting.

Engelke et al. (2014) offer a different perspective, deriving the multivariate conditional distribution of extremal increments of distributions in the domain of attraction of the process with respect to a single extreme component; in this regard there are similarities to the conditional approach of Heffernan & Tawn (2004) and Heffernan & Resnick (2007). For such distributions, this limiting conditional distribution is multivariate Gaussian, therefore offering the possibility to perform high-dimensional inference on the process, subject to applicability of the limiting arguments. Engelke et al. (2014) also provide the associated multivariate spectral density. For a  $d$ -vector  $X^*$  in the domain of attraction of  $M_F = (M_{F,1}, \dots, M_{F,d})$ , the spectral density is the limiting density of  $X^*/\|X^*\|$ , conditional upon  $\|X^*\| > t$  as  $t \rightarrow \infty$ , with  $\|\cdot\|$  a norm. Aside from an alternative formulation and construction, the key difference between our approach and that of Engelke et al. (2014) is the censoring at high marginal thresholds to eliminate bias that may arise due to non-applicability of limiting arguments.

The focus of our work is on tools for improved inference in existing classes of models, rather than the development of new types of models. For this reason we keep practical implementation to simulation results that demonstrate efficiency gains, rather than applications to data; it is not

that we can now model data that could not be modelled previously, but that we can now estimate the parameters of certain models for spatial extremes much more efficiently.

## 2. POISSON PROCESS MODELLING

### 2.1. Introduction

Consider  $X^*(s)$  in the domain of attraction of  $M_F(s)$ , and the associated  $d$ -dimensional vector  $X^*$ . Then it is a standard result (e.g., [Resnick, 2007](#), Ch. 6) that  $\{X_i^*/n : i = 1, \dots, n\}$  converges as  $n \rightarrow \infty$  to a Poisson process on  $\mathbb{R}_+^d \setminus \{0\}$ , with mean measure  $\mu$  satisfying  $\mu([0, x]^c) = V(x)$ , where  $V$  is the exponent function defined in (5). Consequently, modelling the extremes of  $X^*(s)$  can be achieved through modelling the Poisson process of points  $\{X_i\}$  with measure defined by  $V$ , restricted to appropriate subspaces.

In this section we consider the likelihood for points of the process  $X_i(s) = W_i(s)/r_i$ , as defined for construction (3), when the  $W(s)$  are log-Gaussian random fields. We shall focus upon the finite-dimensional distributions of points  $X_i = \{X_i(s_1), \dots, X_i(s_d)\}$  on the region  $[0, u1_d]^c$ , with  $1_d = (1, \dots, 1)^T$  a  $d$ -vector, i.e., distributions conditional upon the exceedance of some threshold  $u > 0$  in one or more coordinates. Here and throughout  $W_i = \{W_i(s_1), \dots, W_i(s_d)\}$ , or  $W$ , represents a  $d$ -vector, whilst  $r_i$ , or  $r$ , is a scalar quantity.

Aspects of the distribution of the  $X_i$  are most simply considered through  $X_i = W_i/r_i$ . We begin with the form of the exponent function or integrated intensity,  $V$ , and then calculate partial derivatives of all orders. The approach yields simple forms for these in terms of standard functions. These forms, combined with the Poisson process being considerably simpler than the corresponding max-stable process, lead to inference procedures that are practical in  $d$  dimensions. The main restriction on  $d$  comes from speed of calculation of (5), and is discussed below.

### 2.2. Integrated intensity

We assume the max-stable process construction (3), and begin by considering the  $d$ -dimensional exponent function of the process  $M_F(s)$ , with

$$W(s) = \exp\{Z(s) - \sigma^2(s)/2\}, \quad (6)$$

where  $Z(s)$  in general is a nonstationary Gaussian process with mean zero and variance  $\text{var}\{Z(s)\} = \sigma^2(s)$ . Then the exponent function takes the form

$$V(x) = \int_{\mathbb{R}^d} \max_{1 \leq i \leq d} \left\{ \frac{\exp(z_i - \sigma_i^2/2)}{x_i} \right\} \phi_d(z; \Sigma) dz, \quad (7)$$

where  $\phi_d(\cdot; \Sigma)$  denotes the  $d$ -dimensional multivariate normal probability density function with covariance matrix  $\Sigma$ , composed of diagonal elements  $\sigma_i^2 = \sigma^2(s_i)$  and off-diagonal elements  $\sigma_{ij} = \text{cov}\{Z(s_i), Z(s_j)\}$ . [Huser & Davison \(2013\)](#) use (7) to derive

$$V(x) = \sum_{i=1}^d \frac{1}{x_i} \Phi_{d-1} \left[ \{\log(x_j/x_i) + \sigma_j^2/2 + \sigma_i^2/2 - \sigma_{ij}\}_{j \neq i}; \Sigma|_i \right], \quad (8)$$

where  $\Phi_j(\cdot; \Sigma|_i)$  represents the  $j$ -dimensional multivariate normal distribution function with mean zero and covariance matrix  $\Sigma|_i = T_i \Sigma T_i^T$ , for  $T_i$  a  $(d-1) \times d$  transformation matrix with the  $i$ th column having  $-1$  for each entry and the other columns being the  $d-1$  standard basis

vectors of  $\mathbb{R}^{d-1}$ , i.e.,

$$T_i = \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & -1 & 0 & \cdots & 1 \end{pmatrix}.$$

Derivatives of  $V$  are required for likelihood inference. Direct differentiation of (8) is awkward, but considering  $X_i = W_i/r_i$  directly produces a simple form for the intensity function. Conditioning on lower-dimensional intensities and integrating with respect to the other coordinates yield elegant forms for the partial derivatives of  $V$ , making likelihood inference feasible.

For the remainder of this section, we will restrict ourselves to  $W$  constructed from stationary Gaussian random fields  $Z$ , i.e.,  $\sigma_i^2 \equiv \sigma^2$ ,  $\sigma_{ij} = \sigma^2 \rho(s_i - s_j)$ , for  $\rho$  a correlation function. We often consider the original events to be stationary, and consequently for modelling original events this seems a sensible approach. However, the theory extends readily to nonstationary fields  $Z$  that produce a stationary limiting max-stable process; we return to these in § 3.3.

### 2.3. Point process intensity

The point process integrated intensity on  $[0, x]^c$  can be written (Dombry et al., 2013) as

$$V(x) = \int_0^\infty \text{pr}(W/r \in [0, x]^c) dr = \int_0^\infty \text{pr}(W/r \not\leq x) dr. \quad (9)$$

Recalling (5), an alternative way to obtain (9) is to observe that since  $\max\{W(s_1)/x_1, \dots, W(s_d)/x_d\}$  is a nonnegative random variable,

$$\begin{aligned} E \left[ \max \left\{ \frac{W(s_1)}{x_1}, \dots, \frac{W(s_d)}{x_d} \right\} \right] &= \int_0^\infty \text{pr} \left[ \max \left\{ \frac{W(s_1)}{x_1}, \dots, \frac{W(s_d)}{x_d} \right\} > r \right] dr \\ &= \int_0^\infty \text{pr}\{W/r \not\leq x\} dr. \end{aligned}$$

Use of the inclusion-exclusion formula followed by differentiation with respect to each of  $x_1, \dots, x_d$  yields

$$-V_{1:d}(x) = \int_0^\infty r^d f_W(xr) dr, \quad x \in \mathbb{R}_+^d \setminus \{0\}, \quad (10)$$

where  $f_W$  denotes the  $d$ -dimensional log-Gaussian density associated to  $W$ . We use subscripts of  $V$  to denote partial differentiation with respect to all elements in the subscript; in particular,  $i : j$  ( $i < j$ ) is shorthand notation for  $\{i, \dots, j\}$ , so that  $1 : d = \{1, \dots, d\}$ . Straightforward calculations in the Appendix show integral (10) to be

$$\frac{|\Sigma|^{-1/2} (1_d^\top q)^{-1/2}}{(2\pi)^{(d-1)/2} \prod_{i=1}^d x_i} \exp \left[ -\frac{1}{2} \left\{ \log x^\top A \log x + \log x^\top \left( \frac{2q}{1_d^\top q} \right) + \sigma^2 - \frac{1}{1_d^\top q} \right\} \right], \quad (11)$$

where  $\log x^\top = (\log x_1, \dots, \log x_d)$ ,  $q = \Sigma^{-1} 1_d$ , and  $A = (\Sigma^{-1} - qq^\top / 1_d^\top q)$  is a  $d \times d$  matrix of rank  $d - 1$ . In order to consider (11) as a density, we simply need to restrict the intensity to some region on which the Poisson process is defined, and normalize by the integrated intensity

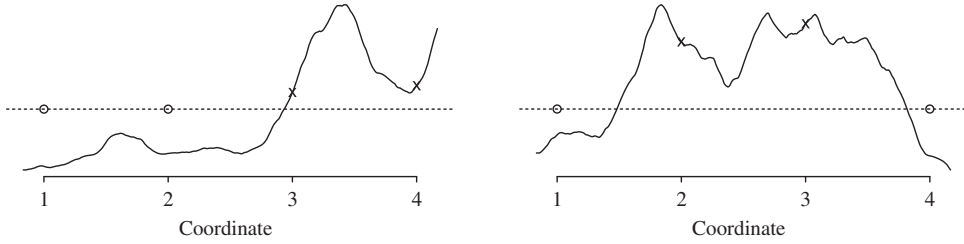


Fig. 1. Illustration of the censoring scheme. The dashed line denotes the threshold; circles denote values censored at the threshold; crosses denote uncensored values.

on this region. Taking the region  $[0, u1_d]^c$ ,  $u > 0$ , provides the density

$$-\frac{V_{1:d}(x)}{V(u1_d)}, \quad x \in [0, u1_d]^c.$$

For  $N_u$  vectors  $X_i$  observed on this region, the Poisson process likelihood is

$$\exp\{-V(u1_d)\} \prod_{i=1}^{N_u} \{-V_{1:d}(x_i)\}, \quad (12)$$

where  $x_1, \dots, x_{N_u}$  is an enumeration of the vectors in  $[0, u1_d]^c$ .

In practice we modify the set-up in two ways to provide an inferential procedure for observed extreme fields. We assume that we have realizations from the process  $X_i^*(s)$ , in the domain of attraction of  $M_F(s)$ , with normalization constants  $a_n(s) \equiv n$ ,  $b_n(s) \equiv 0$ . Specifically we assume that the  $X_i^*(s)$  have standard Pareto tails above some point  $t \geq 1$ ,  $\text{pr}\{X^*(s_j) > x\} = t/x$  for  $x > t$ ; this can be achieved through marginal transformations of the data. If we have observed  $n$  such  $X_i^*(s)$ , then it would be reasonable to take  $X_i(s) \approx X_i^*(s)/n$ , and thus a Poisson process likelihood for  $X_i^*$  is attained through that for  $nX_i$  on  $[0, nu1_d]^c$ . We relabel the threshold  $u^* = nu$ , such that it represents a high quantile of the marginal distribution of the  $X_i^*$ . Convergence of the dependence structure in  $d$  dimensions is a strong assumption when not all coordinates are extreme. Thus we modify likelihood (12) to censor components that do not exceed the threshold  $u^*$ . Our likelihood on the  $X_i^*$  scale for  $N_{u^*}$  fields each having at least one exceedance of  $u^*$  is

$$\exp\{-nV(u^*1_d)\} \prod_{i=1}^{N_{u^*}} \{-V_{a_i}(y_i)\}, \quad (13)$$

where in repetition  $i$ , coordinates in set  $a_i \subseteq \{1, \dots, d\}$  are exceedances of  $u^*$ , and  $y_i = (y_{i1}, \dots, y_{id})$  with

$$y_{ij} = \begin{cases} x_{ij}^*, & x_{ij}^* > u^* \\ u^*, & x_{ij}^* \leq u^* \end{cases} \quad (j = 1, \dots, d).$$

The form of likelihood (13) follows from consideration of the Poisson process likelihood as the product of the Poisson component and the probability density component. We assume that the number of fields with at least one exceedance of  $u^*$  is still Poisson-distributed, but conditional upon this number, we censor the locations of the points in any coordinates lying below  $u^*$ . Figure 1 offers a conceptual illustration in one dimension: if our coordinates were at  $\{1, 2, 3, 4\}$ , we would have  $a_i = \{3, 4\}$  in the left-hand panel, and  $a_i = \{2, 3\}$  in the right-hand panel.

2.4. Partial derivatives of  $V$ 

Likelihood (13) requires calculation of partial derivatives of all orders of  $V$ . For convenience of notation we now assume that the exceedances of  $u^*$  occur in coordinates  $1:k$ , for  $1 \leq k \leq d$ . We define vectors subscripted with a set of indices to be the subvector corresponding only to those indices, e.g.,  $x_{1:k} = (x_1, \dots, x_k)$ . The appropriate partial derivatives can be calculated by considering the conditional intensity of  $X_{k+1:d}$  given  $X_{1:k} = x_{1:k}$ , followed by integration with respect to  $x_{k+1:d}$ . The key idea is that since this conditional intensity is a probability density function, its integral is a cumulative distribution function. The conditional intensity of  $X_{k+1:d}$  given  $X_{1:k} = x_{1:k}$  is

$$\frac{-V_{1:d}(x)}{-V_{1:k}(x_{1:k}, \infty 1_{d-k})}, \quad (14)$$

with the denominator denoting the  $k$ -dimensional marginal intensity. The partial derivative is therefore obtained by integrating (14) with respect to  $x_{k+1:d}$  and multiplying by  $-V_{1:k}(x_{1:k}, \infty 1_{d-k})$ . The conditional intensity, presented in the Appendix, is a log-Gaussian probability density function (Dombry et al., 2013), which leads to partial derivatives of the form  $-V_{1:k}(x_{1:k}, u^* 1_{d-k})$ , equal to

$$\frac{\Phi_{d-k}(\log u^* 1_{d-k} - \mu; \Gamma)}{(2\pi)^{(k-1)/2} |\Sigma_{1:k}|^{1/2} (1_k^T q_k)^{1/2} \prod_{i=1}^k x_i} \times \exp \left[ -\frac{1}{2} \left\{ (\log x_{1:k})^T A_{1:k} \log x_{1:k} + (\log x_{1:k})^T \left( \frac{2q_k}{1_k^T q_k} \right) + \sigma^2 - \frac{1}{1_k^T q_k} \right\} \right], \quad (15)$$

where we define  $\Phi_0$  to be unity,  $\Sigma_{1:k}$  to be the  $k \times k$  matrix derived from  $\Sigma$  corresponding to the indices  $\{1, \dots, k\}$  and, analogously to § 2.3,  $q_k = \Sigma_{1:k}^{-1} 1_k$  and  $A_{1:k} = (\Sigma_{1:k}^{-1} - q_k q_k^T / 1_k^T q_k)$ . The quantities  $\mu$  and  $\Gamma$  are defined in the Appendix.

From expressions (8) and (15), we observe that the principal restrictions on  $d$  stem from the requirement to calculate the multivariate normal distribution function up to dimension  $d - 1$ . Efficient calculation of such quantities is discussed for example in Genz (1992) and Genz & Bretz (2002), who suggest reasonably quick computational times for dimensions up to  $d = 20$ .

## 3. EXTENSION TO INFERENCE FOR THE MAX-STABLE PROCESS

## 3.1. Introduction

The simplified likelihood (13) for extreme events arises because the dependence in original events is less complicated than the dependence amongst componentwise maxima. In particular, one source of the complexity of the latter is the fact that the resulting max-stable field is formed from an unknown composition of original events. Stephenson & Tawn (2005) demonstrated how the density of the multivariate extreme value distribution is greatly simplified by incorporating knowledge of which maxima occurred simultaneously. Davison & Gholamrezaee (2012) have exploited these ideas for efficiency gains in the pairwise modelling of spatial extremes, but the benefit is even clearer in higher dimensions. The  $B_d$  summands in the usual  $d$ -variate max-stable process density cover the  $B_d$  possible partitions of the set  $\{1, \dots, d\}$ , representing the possible partitions of the occurrence of the maxima. By incorporating knowledge of this configuration, the number of summands is reduced to one.

### 3.2. Stephenson–Tawn max-stable process likelihood

Let  $D = \{1, \dots, d\}$ , and let  $\Omega_d$  be the set of all possible partitions of  $D$ . Let  $\Pi \in \Omega_d$  denote a partition that indicates which elements of the componentwise maxima occurred simultaneously, i.e., in the same event. For example, if  $d = 4$ , and if the maxima at sites 1 and 2 occur together but separately from those at 3 and 4, which are also separate from each other, then  $\Pi = \{\{1, 2\}, \{3\}, \{4\}\}$ . Denote the  $|\Pi|$  elements of  $\Pi$  by  $\pi_j$  ( $j = 1, \dots, |\Pi|$ ) and denote partial differentiation of  $V$  with respect to the indices in  $\pi_j$  by  $V_{\pi_j}$ . Then [Stephenson & Tawn \(2005\)](#) showed that the joint probability contribution from the values of the maximized process, and their occurrence times, is

$$\exp\{-V(x)\} \prod_{j=1}^{|\Pi|} \{-V_{\pi_j}(x)\}. \quad (16)$$

Comparing with likelihood (13), observe that this is also the likelihood of the censored Poisson process, but with the censoring threshold set at the observed componentwise maxima,  $x$ . Likelihood (16) offers substantially simplified and more efficient inference for max-stable processes than previously available. However, if the original events are available to us, inference via likelihood (13) should be preferable due to its capacity to incorporate more extreme data into the model. All data that exceed a high threshold may be modelled, with the probability of being nonextreme also accounted for. By contrast, with the max-stable process, only the most extreme observations in a block can be used for inference.

### 3.3. Nonstationary Gaussian random fields

Up to now we have focused on the case where the random field  $Z(s)$  is assumed stationary. From the perspective of considering the domain of attraction process, this makes sense if we believe the underlying fields  $X^*(s)$  to be stationary. Stationary max-stable Brown–Resnick processes can be constructed from suitable nonstationary constituent Gaussian processes. As models for maxima, such a construction has the advantage over the stationary Gaussian process construction of being able to capture independence between sites as the distance between sites tends to infinity. Intrinsically stationary Gaussian processes  $Z(s)$  are characterized by the semivariogram,

$$\gamma(h) = \frac{1}{2} E[\{Z(s) - Z(s+h)\}^2].$$

We take  $Z(s)$  to be fractional Brownian motion, with  $Z(0) = 0$  almost surely and semivariogram  $\gamma(h) = \|h/\lambda\|_2^\alpha$  for  $\lambda > 0$  and  $\alpha \in (0, 2]$ , with  $\|\cdot\|_2$  being the  $L_2$ , or Euclidean, norm. The distribution of an arbitrary  $X_i(s) = W_i(s)/r_i$  depends heavily on the underlying coordinates and does not in general offer a suitable model for original processes in the domain of attraction of a max-stable process. However, the limiting model for maxima, being stationary ([Kablichko et al., 2009](#)), can provide a convenient and flexible model for componentwise maxima data. The principles that were applied in §§ 2.3 and 2.4 to obtain full likelihoods for the stationary Gaussian process case also apply here, though the resulting expressions are slightly more complicated, as some cancellations occurring in the stationary case do not occur under nonstationarity. Following derivations similar to those in the Appendix, we find the partial derivatives of the exponent measure to be

$$-V_{1:k}(x) = \frac{\Phi_{d-k}(\log x_{k+1:d} - \bar{\mu}; \bar{\Gamma})}{(2\pi)^{(k-1)/2} |\Sigma_{1:k}|^{1/2} (1_k^T q_k)^{1/2} \prod_{i=1}^k x_i}$$



$$\begin{aligned} & \times \exp \left\{ -\frac{1}{2} \left( \frac{1}{4} \sigma_k^\top \Sigma_{1:k}^{-1} \sigma_k - \frac{1}{4} \frac{\sigma_k^\top q_k q_k^\top \sigma_k}{1_k^\top q_k} + \frac{\sigma_k^\top q_k}{1_k^\top q_k} - \frac{1}{1_k^\top q_k} \right) \right\} \\ & \times \exp \left[ -\frac{1}{2} \left\{ \log x_{1:k}^\top A_{1:k} \log x_{1:k} + \log x_{1:k}^\top \left( \frac{2q_k}{1_k^\top q_k} + \Sigma_{1:k}^{-1} \sigma_k - \frac{q_k q_k^\top \sigma_k}{1_k^\top q_k} \right) \right\} \right], \end{aligned}$$

where  $\sigma_k = \{\sigma^2(s_1), \dots, \sigma^2(s_k)\}$ ,  $\sigma_d = \{\sigma^2(s_1), \dots, \sigma^2(s_d)\}$ , and

$$\bar{\mu} = -\bar{\Gamma} \left\{ K_{01}^\top A K_{10} \log x_{1:k} + K_{01}^\top \left( \frac{q}{1_d^\top q} - \frac{1}{2} \frac{q q^\top \sigma_d}{1_d^\top q} + \frac{1}{2} \Sigma^{-1} \sigma_d \right) \right\}.$$

The matrices  $K_{01}$ ,  $K_{10}$  are defined in the Appendix; the matrix  $\bar{\Gamma}$  has the same expression as  $\Gamma$ , also defined in the Appendix, but with  $\Sigma$  representing a nonstationary covariance matrix. Such expressions are then used in likelihood (16) for inference.

#### 4. COMPARISON OF INFERENCE

##### 4.1. Stationary $Z$

We compare the inference from the censored Poisson process of § 2, for threshold exceedances of stationary processes, with inference for the corresponding max-stable process. For the latter, we use pairwise composite likelihood methods and the full likelihood given in § 3. In this subsection we focus on stationary  $Z$  in construction (6); the nonstationary case is treated in § 4.2.

We consider inference for 16 sites, on a regular grid  $\{1, 2, 3, 4\}^2$ . Rather than simulating exactly from the limiting max-stable or Poisson process, we simulate events from the process

$$X^*(s) = \exp\{Z(s) - \sigma^2/2\}/R, \quad (17)$$

with  $R \sim \text{Un}(0, 1)$  and  $Z(s)$  a zero-mean stationary Gaussian random field with variance  $\sigma^2$ , independent of  $R$ . The process  $X^*$  is in the domain of attraction of a max-stable Brown–Resnick process with standard Fréchet margins. We use the isotropic Whittle–Matérn correlation function,  $\rho(h; \phi)$  (e.g., Diggle & Ribeiro, 2007, p. 51), with fixed shape parameter  $\kappa = 1$  and variable scale parameter  $\phi$ ; thus  $\text{cov}\{Z(s_1), Z(s_2)\} = \sigma^2 \rho(\|s_1 - s_2\|; \phi)$ .

For data simulated from model (17), we consider thresholded Poisson process inference with  $u$  the 95% quantile of all  $16 \times m$  variates, where  $m$  is the number of replicates of process (17) that were simulated. We find that this is a sufficiently high threshold not to induce bias, but that still incorporates more extreme data points than is possible with the pointwise maximum process.

Since  $M_F(s) = \lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} X_i^*(s)/n$ , we can approximate  $M_F(s)$  by taking rescaled pointwise maxima over the original fields for some sufficiently large  $n$ . It is necessary to determine the number of fields over which to take maxima, such that the model is not biased but data are not wasted by maximizing over too many fields. For blocks of size  $n = 10, 20, \dots, 100$  we assess the marginal and dependence structure convergence through simulation. Figure 2 presents convergence diagnostics, displaying estimated quantities for different  $n$ , with true limiting quantities overlaid. Figures 2 (c) and (d) display estimates of the dependence parameters  $(\sigma, \phi)$  via pairwise composite likelihood. These suggest that the dependence parameters can be estimated without much bias with fairly low  $n$ . Figure 2(a) displays estimation of the marginal scale parameter,  $\gamma$ , assuming a Fréchet model,  $\text{pr}(\max_{1 \leq i \leq n} X_i^*/n \leq x) = \exp(-\gamma/x)$ . Marginal convergence is suggested by estimated values of  $\gamma$  close to one. Convergence of the margins is much slower than that of the dependence structure, principally due to the behaviour in the lower tail: the marginal distribution of  $X^*$  does not admit as many small values as the standard Fréchet

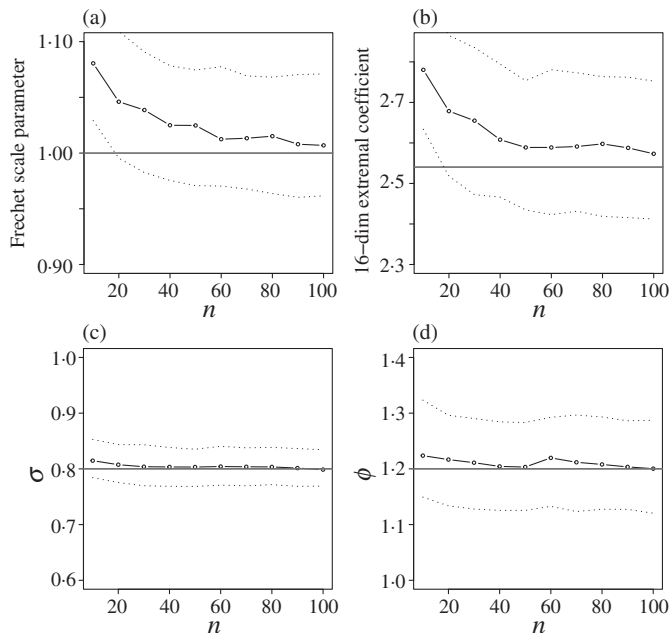


Fig. 2. Convergence diagnostics for the max-stable random fields: (a) maximum likelihood estimates of the marginal Fréchet scale parameter; (b) maximum likelihood estimates of the 16-dimensional extremal coefficient; (c), (d) maximum composite likelihood estimates of the parameters  $(\sigma, \phi)$ . Solid line, median; dotted lines, 50% confidence regions, each on the basis of 500 repetitions; horizontal lines, the true limiting quantities in each case.

distribution, which thus inflates the maximum likelihood estimate from a sample  $(x_1, \dots, x_m)$  assumed to be Fréchet( $\gamma$ ),  $\hat{\gamma} = m / \sum_{i=1}^m x_i^{-1}$ . Figure 2(b), which offers a joint marginal and dependence convergence diagnostic, displays estimates of the 16-dimensional extremal coefficient function, defined for a max-stable field with standard Fréchet margins as the scale parameter  $V(1_d)$  of the distribution of  $\max\{M_F(s_1), \dots, M_F(s_d)\}$  via

$$\text{pr}[\max\{M_F(s_1), \dots, M_F(s_d)\} \leq x] = \exp\{-V(1_d)/x\}, \quad x > 0.$$

The true values of  $V(1_d)$  are calculated using equation (8). Clearly the accuracy of estimation of  $V(1_d)$  depends on both the margins and the dependence structure; Figs. 2 (a), (c) and (d) suggest that discrepancies for smaller  $n$  are due to lack of marginal convergence. Based upon Fig. 2, we take  $n = 50$ ; for larger values of  $\sigma$ , larger  $n$  will be required; see also the comments in § 4.2. We simulate a total of  $m = 1000$  replicates of process (17) per repetition, so that there are 20 approximately max-stable fields used for inference in the max-stable process likelihoods. In each likelihood, the margins are assumed to be those from the limiting distribution.

Table 1 details bias of parameter estimates, normalized by their true values, from 200 repetitions of each of three different parameter configurations; results for the pairwise likelihood were obtained using the R package *SpatialExtremes* (Ribatet, 2013). All methods exhibit low bias. Table 2 displays the corresponding variances of the normalized parameters, as a proportion of the variance attained from the pairwise likelihood for maxima. In all cases the censored Poisson process likelihood is more efficient than the Stephenson–Tawn likelihood, which in turn is more efficient than the pairwise likelihood. This is entirely in line with expectations: the Stephenson–Tawn likelihood offers a  $d$ -dimensional inference procedure for maxima, which should be more efficient than the corresponding pairwise composite procedure; the censored Poisson process

Table 1. *Sample bias, multiplied by 10 and displayed to two significant figures, of normalized parameter estimates over 200 repetitions for each configuration*

	$\sigma = 0.5, \phi = 1$			$\sigma = 1, \phi = 0.5$			$\sigma = 1, \phi = 1$		
	PW	ST	PP	PW	ST	PP	PW	ST	PP
$\hat{\sigma}/\sigma$	0.12	0.046	0.054	0.076	0.090	0.0062	0.14	0.034	-0.067
$\hat{\phi}/\phi$	0.24	0.18	0.075	-0.029	0.21	0.0047	0.36	0.055	-0.042

PW, pairwise likelihood; ST, Stephenson–Tawn likelihood; PP, censored Poisson process likelihood.

Table 2. *Relative variance of normalized parameter estimates over 200 repetitions for each configuration, as a proportion of the variance of the pairwise likelihood. Notation as per Table 1. From left to right, parameter values as per Table 1*

	ST	PP	ST	PP	ST	PP
$\hat{\sigma}/\sigma$	0.50	0.15	0.32	0.10	0.18	0.076
$\hat{\phi}/\phi$	0.42	0.11	0.34	0.14	0.15	0.067

Table 3. *Sample bias and variance, multiplied by 10, for 150 repetitions on the regular 4 by 4 grid, with true values  $\lambda = 3, \alpha = 1$ . Notation as in Table 1*

Parameter	Bias		Variance	
	PW	ST	PW	ST
$\lambda/\hat{\lambda}$	-0.078	-0.0034	0.080	0.030
$\alpha/\hat{\alpha}$	0.12	-0.016	0.059	0.015

likelihood offers a  $d$ -dimensional procedure and the ability to incorporate more extreme events, as well as more information per event, by recording the configuration of exceedances and non-exceedances, as well as the size distributions of all exceedances.

#### 4.2. Nonstationary $Z$

For this class of models, we compare pairwise composite likelihood inference for component-wise maxima to that from the Stephenson–Tawn likelihood incorporating occurrence times. For the results of § 4.1, the max-stable process was simulated using a sufficient number of normalized maxima of the process (17). This idea does not extend easily to nonstationary underlying random fields. We simulate these max-stable processes using representation (3) and Method 2 of Oesting et al. (2012), which involves random shifting of the centres of the Brownian motion. We then record which maxima occur simultaneously.

Table 3 provides variances and biases for the normalized parameters ( $\hat{\lambda}/\lambda, \hat{\alpha}/\alpha$ ) from 150 repetitions on the same regular 4 by 4 grid as in § 4.1, with true values  $\lambda = 3, \alpha = 1$ . Both methods again produce approximately unbiased inference. The variances from the Stephenson–Tawn inference are well under half the size of those from the corresponding pairwise composite likelihood, which is broadly similar to the results in Table 2.

We comment on some difficulties with the simulation of such Brown–Resnick processes that can seemingly affect the inference. These difficulties pertain to the variability of the Brownian

Table 4. Means and maxima of the maximum indices that contribute to the Brown–Resnick max-stable process approximation, divided by  $K = 20\,000$ , for different values of the scale parameter  $\lambda$ , with fixed  $\alpha = 1$ . Results are recorded to two significant figures

$\lambda$	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2.0
Mean	0.77	0.49	0.37	0.18	0.16	0.058	0.040	0.024	0.011	0.0060
Max	1.00	0.95	0.90	0.84	0.93	0.51	0.32	0.27	0.068	0.03

motion,  $Z$ , for certain parameter values, and the number of points involved in the finite approximation of (3). Simulation based on (3) is carried out by using a finite approximation of  $K$  points and ordering the points of the Poisson process  $\{r_i\}$  so that  $r_1 < \dots < r_K$ . We then set

$$M_F(s) = \max_{1 \leq i \leq K} W_i(s)/r_i.$$

The idea is that for  $i$  sufficiently large,  $r_i$  should be sufficiently large that the associated  $W_i/r_i$  has a negligible probability of contributing to the maximum. Here, since  $\text{var}\{W(s)\} = \exp(2\|s/\lambda\|^\alpha) - 1$ , small values of the scale parameter  $\lambda$  lead to highly variable  $W(s)$ . Consequently, processes  $W$  with extremely large values are more likely to occur. Thus even for a large value of the index  $i$ , the probability of occurrence of a  $W_i$  large enough for  $W_i/r_i$  to contribute to the maximum is still nonnegligible. Therefore if sufficient numbers of fields are not simulated, then the distribution of the configuration of maxima will be wrong, and this can lead to upward bias in the estimation of  $\lambda$ . For relatively large values of  $\lambda$ , such as in the example provided, this does not seem to be an issue. To reinforce these ideas, we summarize the indices at which maxima occur for fixed  $\alpha = 1$  and varying  $\lambda \in \{0.2, 0.4, \dots, 2\}$ , on the same regular 4 by 4 grid, using a finite approximation of  $K = 20\,000$  points. For each  $\lambda$ , we simulated 50 fields and recorded the maximum index that contributed to the field. Table 4 provides the means of these maximum indices, as well as the maxima over all 50 repetitions.

For small  $\lambda$  values, even using 20 000 replicates does not appear to suffice for simulating accurately the configuration of events in the max-stable process. By contrast, for larger  $\lambda$  values,  $K$  could be taken much smaller and accurate representations attained. The best way to simulate Brown–Resnick processes is open; Oesting et al. (2012) examine different methods, which perform best on different regions of the parameter space. Their Method 4 appears to perform better on larger spatial domains, analogous to smaller scale parameters in our case.

## 5. DISCUSSION

We have developed an event-based threshold approach for efficient inference in spatial extremes. By exploiting the spectral decomposition  $X_i = W_i/r_i$  for calculation of censored intensities, along with a censored Poisson process likelihood, we demonstrated that  $d$ -dimensional inference is feasible for classes of models where  $W$  is log-Gaussian. For max-stable processes, we showed that the likelihood of Stephenson & Tawn (2005) is a special case of the censored Poisson process, and thus extended the likelihood to inference for the max-stable process.

There are two further notable advantages to the availability of full likelihoods. The first is the ability to exploit Bayesian inferential mechanisms. The adjustment of composite likelihoods to full posterior distributions has been explored by Ribatet et al. (2012), though the results are approximations. Original data are often available, but if only aggregated maxima are available, then the full max-stable process likelihood is still required for inference. The explosion of terms

from differentiating  $\exp\{-V(x)\}$  would typically still make this prohibitive in more than a few dimensions. However, recent work by [Dombry et al. \(2013\)](#) has employed a Gibbs sampler for sampling over partitions; such ideas could potentially be adapted and incorporated into Bayesian inference when the partition of maxima is unknown. The second advantage to having full likelihoods would be the capacity to simultaneously model marginal and dependence parameters, and obtain uncertainty measures without the need for additional procedures. Since we have a full likelihood, standard asymptotic normality arguments could be applied to parameter estimators, under mild conditions.

The models discussed in this work are useful only if the data exhibit asymptotic dependence between all sites. Such data have a nonzero limiting probability of being simultaneously extreme everywhere, given that one component is extreme. What may often be more realistic in practice, particularly for environmental applications, is that the level of dependence decays with the level of the process, and the most extreme events ultimately occur in isolation. This is consistent with the notion of asymptotic independence. [Wadsworth & Tawn \(2012\)](#) considered modelling asymptotic independence in spatial data. They demonstrated that a widely applicable class of models for such data is the so-called inverted max-stable process, which translates the lower tail of a max-stable copula to its upper tail. Max-stable processes thus have modelling potential beyond asymptotically dependent data, but the inferential improvements in this paper cannot easily be adapted to asymptotically independent data.

It is natural to wonder whether simplified inference is possible for other classes of max-stable process models. There is realistically only one other such class in use, which consists of models constructed with zero-truncated Gaussian random fields ([Schlather, 2002](#); [Wadsworth & Tawn, 2012](#)). The truncation makes calculation more awkward, though numerical approximations could in theory allow this. Since the Gaussian process forms the usual basis for spatial statistics, there are not, as yet, other types of model available for consideration.

#### ACKNOWLEDGEMENT

This work was undertaken whilst J. L. W. was based at both Lancaster University, U.K. and EPFL, Switzerland. Funding from the U.K. Engineering and Physical Sciences Research Council, Shell Research, and the Swiss National Science Foundation is gratefully acknowledged. We would also like to thank Raphaël Huser and Anthony Davison for helpful discussions.

#### APPENDIX

##### *Poisson process intensity*

The required intensity is integral (10), with  $f_W$  the  $d$ -dimensional log-Gaussian density function. Changing variables to  $v = -\log r$ , this integral is

$$\int_{-\infty}^{\infty} \frac{|\Sigma|^{-1/2}}{(2\pi)^{d/2} \prod_{i=1}^d x_i} \exp \left[ -\frac{1}{2} \left\{ \left( \log x + \frac{\sigma^2}{2} \mathbf{1}_d - v \mathbf{1}_d \right)^\top \Sigma^{-1} \left( \log x + \frac{\sigma^2}{2} \mathbf{1}_d - v \mathbf{1}_d \right) + 2v \right\} \right] dv.$$

Consider  $(\log x + \frac{\sigma^2}{2} \mathbf{1}_d - v \mathbf{1}_d)^\top \Sigma^{-1} (\log x + \frac{\sigma^2}{2} \mathbf{1}_d - v \mathbf{1}_d) + 2v$ , and complete the square in  $v$ . With some rearrangement, the integrand is equal to

$$\frac{(1_d^\top \Sigma^{-1} \mathbf{1}_d)^{1/2}}{(2\pi)^{1/2}} \exp \left( -\frac{1_d^\top \Sigma^{-1} \mathbf{1}_d}{2} \left[ v - \{ \log x^\top \Sigma^{-1} \mathbf{1}_d + (\sigma^2/2) 1_d^\top \Sigma^{-1} \mathbf{1}_d - 1 \} / 1_d^\top \Sigma^{-1} \mathbf{1}_d \right]^2 \right)$$

$$\begin{aligned} & \times \frac{|\Sigma|^{-1/2} (1_d^\top \Sigma^{-1} 1_d)^{-1/2}}{(2\pi)^{(d-1)/2} \prod_{i=1}^d x_i} \exp \left[ -\frac{1}{2} \left\{ \log x^\top \left( \Sigma^{-1} - \frac{\Sigma^{-1} 1_d 1_d^\top \Sigma^{-1}}{1_d^\top \Sigma^{-1} 1_d} \right) \log x + \log x^\top \left( \frac{2\Sigma^{-1} 1_d}{1_d^\top \Sigma^{-1} 1_d} \right) \right\} \right] \\ & \times \exp \left\{ -\frac{1}{2} \left( \sigma^2 - \frac{1}{1_d^\top \Sigma^{-1} 1_d} \right) \right\}. \end{aligned}$$

The expression on the first line integrates to unity, leaving the second and third lines. Defining  $q$  and  $A$  as in § 2.3 leaves the expression in (11).

#### Partial derivatives of $V$

The calculations follow Dombry et al. (2013). The ratio of intensities of equation (14) in § 2.4,  $V_{1:d}(x)/V_{1:k}(x_{1:k}, \infty 1_{d-k})$ , is

$$\begin{aligned} & \frac{(2\pi)^{(k-1)/2} |\Sigma_{1:k}|^{1/2} (1_k^\top \Sigma_{1:k}^{-1} 1_k)^{1/2} \prod_{i=1}^k x_i}{(2\pi)^{(d-1)/2} |\Sigma|^{1/2} (1_d^\top \Sigma^{-1} 1_d)^{1/2} \prod_{i=1}^d x_i} \exp \left\{ -\frac{1}{2} \left( \frac{1}{1_k^\top \Sigma_{1:k}^{-1} 1_k} - \frac{1}{1_d^\top \Sigma^{-1} 1_d} \right) \right\} \\ & \times \exp \left[ -\frac{1}{2} \left\{ \log x^\top A \log x + \log x^\top \left( \frac{2q}{1_d^\top q} \right) - \log x_{1:k}^\top A_{1:k} \log x_{1:k} - \log x_{1:k}^\top \left( \frac{2q_k}{1_k^\top q_k} \right) \right\} \right], \end{aligned}$$

with notation as defined in § 2.3 and 2.4. Define the  $d \times k$  matrix  $K_{10}$  and the  $d \times (d-k)$  matrix  $K_{01}$ :

$$K_{10} = \begin{pmatrix} I_k \\ 0_{d-k,k} \end{pmatrix}, \quad K_{01} = \begin{pmatrix} 0_{k,d-k} \\ I_{d-k} \end{pmatrix},$$

where  $I_k$  is the  $k \times k$  identity matrix and  $0_{j,k}$  is a  $j \times k$  matrix of zeroes. By writing  $\log x = K_{10} \log x_{1:k} + K_{01} \log x_{k+1:d}$  and completing the square in  $\log x_{k+1:d}$ , one can verify that the conditional intensity is

$$\frac{1}{(2\pi)^{(d-k)/2} |\Gamma|^{1/2} \prod_{i=k+1}^d x_i} \exp \left\{ -\frac{1}{2} (\log x_{k+1:d} - \mu)^\top \Gamma^{-1} (\log x_{k+1:d} - \mu) \right\},$$

with  $\Gamma^{-1} = K_{01}^\top A K_{01}$  and  $\mu = -\Gamma (K_{01}^\top A K_{10} \log x_{1:k} + K_{01}^\top \Sigma^{-1} 1_d / 1_d^\top \Sigma^{-1} 1_d)$ , i.e., a log-Gaussian density. In order to calculate the required likelihood contribution, integration with respect to  $x_{k+1:d}$  yields

$$\Phi_{d-k}(\log u^* 1_{d-k} - \mu; \Gamma).$$

This cumulative normal probability can be efficiently computed in a likelihood using standard routines for multivariate normal probability calculation. Postmultiplication by the marginal intensity  $-V_{1:k}(x_{1:k}, \infty 1_{d-k})$  gives the expression  $-V_{1:k}(x_{1:k}, u^* 1_{d-k})$  in equation (15).

#### REFERENCES

- BROWN, B. M. & RESNICK, S. I. (1977). Extreme values of independent stochastic processes. *J. Appl. Prob.* **14**, 732–9.
- COLES, S. G. & TAWN, J. A. (1991). Modelling extreme multivariate events. *J. R. Statist. Soc. B* **53**, 377–92.
- DAVISON, A. C. & GHOLAMREZAEI, M. M. (2012). Geostatistics of extremes. *Proc. R. Soc. A* **468**, 581–608.
- DAVISON, A. C., PADOAN, S. & RIBATET, M. (2012). Statistical modelling of spatial extremes (with Discussion). *Statist. Sci.* **27**, 161–86.
- DE HAAN, L. (1984). A spectral representation for max-stable processes. *Ann. Prob.* **12**, 1194–204.
- DE HAAN, L. & FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. New York: Springer.
- DIGGLE, P. J. & RIBEIRO, P. J. (2007). *Model-based Geostatistics*. New York: Springer.
- DOMBRY, C., EYI-MINKO, F. & RIBATET, M. (2013). Conditional simulation of max-stable processes. *Biometrika* **100**, 111–24.
- ENGELKE, S., MALINOWSKI, A., KABLUCHKO, Z. & SCHLATHER, M. (2014). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *J. R. Statist. Soc. B* To appear.
- GENTON, M. G., MA, Y. & SANG, H. (2011). On the likelihood function of Gaussian max-stable processes. *Biometrika* **98**, 481–8.
- GENZ, A. (1992). Numerical computation of multivariate normal probabilities. *J. Comp. Graph. Statist.* **1**, 141–9.

- GENZ, A. & BRETZ, F. (2002). Methods for the computation of multivariate t-probabilities. *J. Comp. Graph. Statist.* **11**, 950–71.
- HEFFERNAN, J. E. & RESNICK, S. I. (2007). Limit laws for random vectors with an extreme component. *Ann. Appl. Prob.* **17**, 537–71.
- HEFFERNAN, J. E. & TAWN, J. A. (2004). A conditional approach for multivariate extreme values (with Discussion). *J. R. Statist. Soc. B* **66**, 497–546.
- HUSER, R. & DAVISON, A. C. (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika* **100**, 511–8.
- HUSER, R. & DAVISON, A. C. (2014). Space-time modeling of extreme events. *J. R. Statist. Soc. B* **76**, 439–61.
- KABLUCHKO, Z., SCHLATHER, M. & DE HAAN, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Prob.* **37**, 2042–65.
- LEDFOORD, A. W. & TAWN, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169–87.
- OESTING, M., KABLUCHKO, Z. & SCHLATHER, M. (2012). Simulation of Brown–Resnick processes. *Extremes* **15**, 89–107.
- PADOAN, S. A., RIBATET, M. & SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Am. Statist. Assoc.* **105**, 263–77.
- RESNICK, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer.
- RIBATET, M. (2013). SpatialExtremes: R package.
- RIBATET, M., COOLEY, D. & DAVISON, A. C. (2012). Bayesian inference for composite likelihoods with an application to spatial extremes. *Statist. Sinica* **22**, 813–45.
- ROOTZÉN, H. & TAJVIDI, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli* **12**, 917–30.
- SCHLATHER, M. (2002). Models for stationary max-stable random fields. *Extremes* **5**, 33–44.
- SMITH, R. L., TAWN, J. A. & COLES, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika* **84**, 249–68.
- STEPHENSON, A. G. & TAWN, J. A. (2005). Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika* **92**, 213–27.
- TAWN, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika* **77**, 245–53.
- WADSWORTH, J. L. & TAWN, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika* **99**, 253–72.

[Received November 2012. Revised July 2013]