*Biometrika* (2011), **98**, 2, *pp.* 481–488 doi: 10.1093/biomet/asr020

*Printed in Great Britain*

# On the likelihood function of Gaussian max-stable processes

By MARC G. GENTON, YANYUAN MA and HUIYAN SANG

*Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.*

genton@stat.tamu.edu ma@stat.tamu.edu huiyan@stat.tamu.edu

## Summary

We derive a closed form expression for the likelihood function of a Gaussian max-stable process indexed by $\mathbb{R}^d$ at $p \leqslant d+1$ sites, $d \geqslant 1$. We demonstrate the gain in efficiency in the maximum composite likelihood estimators of the covariance matrix from $p=2$ to $p=3$ sites in $\mathbb{R}^2$ by means of a Monte Carlo simulation study.

*Some key words*: Composite likelihood; Extreme event; Multivariate index; Pairwise and triplewise inference; Spatial statistics.

## 1. Introduction

Max-stable processes (de Haan, 1984) have received sustained attention in recent years because of their relevance for studying extreme events in financial, environmental and climate sciences. In a seminal unpublished University of Surrey 1990 technical report, R. L. Smith defined Gaussian max-stable processes, where all margins follow a unit Fréchet distribution, in view of modelling spatial extremes. However, a closed form expression for the joint cumulative distribution function of the process $Z$ was provided only for two spatial sites $x_1, x_2 \in \mathbb{R}^2$,

$$\text{pr}\{Z(x_1) \leqslant z_1, Z(x_2) \leqslant z_2\} = \exp\left[-\frac{1}{z_1}\Phi\left\{\frac{\alpha}{2} + \frac{1}{\alpha}\log\left(\frac{z_2}{z_1}\right)\right\} - \frac{1}{z_2}\Phi\left\{\frac{\alpha}{2} + \frac{1}{\alpha}\log\left(\frac{z_1}{z_2}\right)\right\}\right], \quad (1)$$

where $\Phi$ denotes the univariate standard normal cumulative distribution function, $\alpha^2 = (x_1 - x_2)^{\mathrm{T}}\Sigma^{-1}(x_1 - x_2)$, and $\Sigma \in \mathbb{R}^{2\times 2}$ is the covariance matrix with variances $\sigma_{11}^2$ and $\sigma_{22}^2$, and correlation $\rho$. The square roots of the eigenvalues of $\Sigma$ control the range of the spatial dependence.

de Haan & Pereira (2006) proposed a consistent and asymptotically normal estimator for the parameters in $\Sigma$ based on a simple relationship between $\Sigma$ and a well-known pairwise extremal dependence coefficient; see the definition at the end of §2·1. Their approach starts from a sequence of independent replications of a stochastic process $U$ which is in the domain of attraction of a max-stable process $Z$. Then they estimate the extremal coefficient nonparametrically from the tails of the empirical two-dimensional marginal distributions of $U$ at each pair of locations. Recently, Padoan et al. (2010) introduced the pairwise composite likelihood approach (Lindsay, 1988; Cox & Reid, 2004; Varin & Vidoni, 2005; Varin et al., 2011) for inference in Gaussian max-stable processes. Unlike de Haan & Pereira (2006), their approach considers the max-stable process $Z$ directly instead of $U$.

In this note, we derive a closed form expression for the likelihood function of a Gaussian max-stable process at sites $x_j \in \mathbb{R}^d$ $(j = 1, \ldots, p)$, $p \leqslant d+1$, and $d \geqslant 1$. This allows for inference based on triples in spatial $\mathbb{R}^2$, on quadruples in spatial or space-time $\mathbb{R}^3$ and on quintuples in space-time $\mathbb{R}^4$. As a by-product, we obtain a simpler expression than Padoan et al. (2010) for the pairwise probability density function of a Gaussian max-stable process indexed by $\mathbb{R}^2$. We demonstrate the gain in efficiency in the maximum composite likelihood estimators of $\Sigma$ from $p=2$ to $p=3$ sites in $\mathbb{R}^2$ by means of a Monte Carlo simulation study. For $p > d+1$ sites, we show that a representation of type (1) does not exist.

## 2. MAIN RESULTS

### 2·1. *Joint cumulative distribution function*

Let $z=(z_1,\ldots,z_p)^{\mathrm{T}}\in\mathbb{R}^p$ and $c^{(j)}(z)=\{c_1^{(j)}(z),\ldots,c_{j-1}^{(j)}(z),c_{j+1}^{(j)}(z),\ldots,c_p^{(j)}(z)\}^{\mathrm{T}}\in\mathbb{R}^{p-1}$ ($j=1,\ldots,p$), where $c_k^{(j)}(z)=(x_j-x_k)^{\mathrm{T}}\Sigma^{-1}(x_j-x_k)/2-\log(z_j/z_k)$ ($j,k=1,\ldots,p$), and the covariance matrix $\Sigma\in\mathbb{R}^{d\times d}$ is positive definite.

Based on the Poisson process approach of constructing a max-stable process in R. L. Smith's unpublished University of Surrey 1990 technical report, the joint cumulative distribution function of a Gaussian max-stable process is

$$F(z)=\mathrm{pr}\{Z(x_1)\leqslant z_1,\ldots,Z(x_p)\leqslant z_p\}=\exp\left[-\int_{\mathbb{R}^d}\max_{i=1,\ldots,p}\left\{\frac{\phi_d(x-x_i;\Sigma)}{z_i}\right\}dx\right],$$

where $\phi_d(\cdot;\Sigma)$ denotes the probability density function of a $d$-dimensional Gaussian distribution with zero mean and covariance matrix $\Sigma$. Let $\Phi_d(\cdot;\Sigma)$ denote its cumulative distribution function. Consider the $d$-dimensional Gaussian random vectors $Y_j\sim N_d(0,\Sigma)$ ($j=1,\ldots,p$), assumed to be independent of each other. Then we have

$$\begin{aligned}F(z)&=\exp\left(-\sum_{j=1}^{p}E\left[\frac{1}{z_j}I\left\{\frac{\phi_d(Y_j;\Sigma)}{z_j}>\max_{k\neq j}\frac{\phi_d(Y_j+x_j-x_k;\Sigma)}{z_k}\right\}\right]\right)\\&=\exp\left[-\sum_{j=1}^{p}\frac{1}{z_j}\mathrm{pr}\left\{\frac{\phi_d(Y_j;\Sigma)}{z_j}>\max_{k\neq j}\frac{\phi_d(Y_j+x_j-x_k;\Sigma)}{z_k}\right\}\right]\\&=\exp\left[-\sum_{j=1}^{p}\frac{1}{z_j}\mathrm{pr}\left\{Y_j^{\mathrm{T}}\Sigma^{-1}(x_j-x_k)>-c_k^{(j)}(z),\ k=1,\ldots,p,\ k\neq j\right\}\right],\end{aligned}\tag{2}$$

where $I(\cdot)$ denotes the indicator function.

Define the matrices $X=(x_1,\ldots,x_p)\in\mathbb{R}^{d\times p}$ and $X_{-j}=X\backslash x_j\in\mathbb{R}^{d\times(p-1)}$, the matrix $X$ without the column $x_j$. Consider the matrix $\Sigma^{(j)}=(x_j1_{p-1}^{\mathrm{T}}-X_{-j})^{\mathrm{T}}\Sigma^{-1}(x_j1_{p-1}^{\mathrm{T}}-X_{-j})\in\mathbb{R}^{(p-1)\times(p-1)}$ ($j=1,\ldots,p$), where $1_{p-1}=(1,\ldots,1)^{\mathrm{T}}\in\mathbb{R}^{p-1}$. The matrix $\Sigma^{(j)}$ is invertible provided the $d\times(p-1)$ matrix $x_j1_{p-1}^{\mathrm{T}}-X_{-j}$ has full rank. Geometrically, this holds if and only if the simplex defined by the sites $x_1,\ldots,x_p$ does not degenerate, that is, it cannot be contained in any $(p-2)$-dimensional space. For $p>d+1$ sites, $\Sigma^{(j)}$ is not invertible. For $p\leqslant d+1$ sites, certain configurations of the sites $x_1,\ldots,x_p$ may also yield non-invertible $\Sigma^{(j)}$ matrices.

When $p\leqslant d+1$ and $\Sigma^{(j)}$ is invertible, we have from (2) that

$$\begin{aligned}F(z)&=\exp\left\{-\sum_{j=1}^{p}\frac{1}{z_j}\int_{-c^{(j)}(z)}^{+\infty}\phi_{p-1}(v^{(j)};\Sigma^{(j)})dv^{(j)}\right\}\\&=\exp\left[-\sum_{j=1}^{p}\frac{1}{z_j}\Phi_{p-1}\{c^{(j)}(z);\Sigma^{(j)}\}\right]\\&=\exp\{-V(z)\},\end{aligned}\tag{3}$$

where $v^{(j)}=(v_1,\ldots,v_{j-1},v_{j+1},\ldots,v_p)^{\mathrm{T}}\in\mathbb{R}^{p-1}$. The novelty is the representation (3) for $p>2$. For $p=d=2$, expression (3) reduces to (1). The exponent measure function $V(z)$ defined by (3) describes the dependence among the different sites. The extremal coefficient $V(1_p) \in [1, p]$ summarizes the degree of dependence between the maxima, ranging from total dependence, $V(1_p) = 1$, to independence, $V(1_p) = p$.

### 2·2. *Joint probability density function*

In order to derive the joint probability density function of a Gaussian max-stable process, we need the expressions for $V_{[i_1,i_2,\ldots,i_k]}(z) = \partial^k V(z)/(\partial z_{i_1} \partial z_{i_2} \cdots \partial z_{i_k})$. To derive these partial derivatives, we use the following formula for $x = (x_1^{\mathrm{T}}, x_2^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^d$, $x_1 \in \mathbb{R}^{d_1}$, $x_2 \in \mathbb{R}^{d_2}$, $d_1 + d_2 = d$, with corresponding block decomposition of $\Sigma$:

$$\frac{\partial}{\partial x_1} \Phi_d(x; \Sigma) = \phi_{d_1}(x_1; \Sigma_{11}) \Phi_{d_2}(x_2 - \Sigma_{21} \Sigma_{11}^{-1} x_1; \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}). \tag{4}$$

The explicit expressions for $V_{[i_1,i_2,\ldots,i_k]}(z)$ are derived in the Appendix.

From $F(z) = \exp\{-V(z)\}$, we then have the joint probability density function for $p = d = 2$:

$$\begin{aligned} f(z) &= F(z)\Big(V_{[1]}V_{[2]} - V_{[1,2]}\Big)(z) \\ &= \exp\left[-\frac{1}{z_1}\Phi\{c^{(1)}(z); \Sigma^{(1)}\} - \frac{1}{z_2}\Phi\{c^{(2)}(z); \Sigma^{(2)}\}\right] \\ &\quad \times \left[\frac{1}{z_1^2 z_2^2}\Phi\left\{c^{(1)}(z); \Sigma^{(1)}\right\}\Phi\left\{c^{(2)}(z); \Sigma^{(2)}\right\} + \frac{1}{z_1^2 z_2}\phi\left\{c^{(1)}(z); \Sigma^{(1)}\right\}\right]. \end{aligned} \tag{5}$$

The expression for the pairwise probability density function given by Padoan et al. (2010) can be further reduced to the simpler form (5).

For $p = 3$ and $d = 2$:

$$f(z) = F(z)\Big(-V_{[1]}V_{[2]}V_{[3]} + V_{[1]}V_{[2,3]} + V_{[2]}V_{[1,3]} + V_{[3]}V_{[1,2]} - V_{[1,2,3]}\Big)(z), \tag{6}$$

where, for $i \neq j \neq k \in \{1, 2, 3\}$ and $\Sigma_{a,b}^{(i)} = (x_i - x_a)^{\mathrm{T}}\Sigma^{-1}(x_i - x_b)$ for $a, b \in \{j, k\}$:

$$\begin{aligned} V_{[i]}(z) &= -\frac{1}{z_i^2}\Phi_2\big\{c^{(i)}(z); \Sigma^{(i)}\big\}, \\ V_{[i,j]}(z) &= -\frac{1}{z_i^2 z_j}\phi\big\{c_j^{(i)}(z); \Sigma_{j,j}^{(i)}\big\}\Phi\big\{c_k^{(i)}(z) - \Sigma_{j,k}^{(i)} c_j^{(i)}(z)/\Sigma_{j,j}^{(i)}; \Sigma_{k,k}^{(i)} - (\Sigma_{j,k}^{(i)})^2/\Sigma_{j,j}^{(i)}\big\}, \\ V_{[i,j,k]}(z) &= -\frac{1}{z_i^2 z_j z_k}\phi_2\big\{c^{(i)}(z); \Sigma^{(i)}\big\}. \end{aligned}$$

For $p \geqslant 4$ and $d \geqslant p - 1$, the likelihood can be computed by further differentiation of $F(z) = \exp\{-V(z)\}$. The resulting expressions become formidable yet they can be obtained symbolically with a computer. The cases when $d > 4$ are not physically realistic for applications.

## 3. Composite likelihoods

Consider a parametric statistical model with probability density function $\{f(z; \theta), z \in \mathcal{Z} \subseteq \mathbb{R}^K, \theta \in \Theta \subseteq \mathbb{R}^q\}$, and a set of marginal or conditional events $\{\mathcal{A}_i : \mathcal{A}_i \subseteq \mathcal{F}, i \in I \subseteq \mathbb{N}\}$, where $\mathcal{F}$ is some sigma algebra on $\mathcal{Z}$. The log composite likelihood (Lindsay, 1988) is defined as

$$\ell_c(\theta) = \sum_{i \in I} w_i \log f(z \in \mathcal{A}_i; \theta) \tag{7}$$

where $f(z \in \mathcal{A}_i; \theta)$ is the likelihood associated with the event $\mathcal{A}_i$, and $\{w_i, i \in I \subseteq \mathbb{N}\}$ is a set of weights. For example, assuming equal weights, we may define the pairwise log composite likelihood as

$$\ell_c(\theta) = \sum_{n=1}^{N} \sum_{i \neq j} \log f(z_i^{(n)}, z_j^{(n)}; \theta)$$

where $z_i^{(n)}$ is the sample of the $n$th replicate at the $i$th site. Analogously, we may define the triplewise log composite likelihood as

$$\ell_c(\theta) = \sum_{n=1}^{N} \sum_{i \neq j \neq k} \log f(z_i^{(n)}, z_j^{(n)}, z_k^{(n)}; \theta).$$

Then $\hat{\theta}$ is called the maximum composite likelihood estimator if it is the global maximum of $\ell_c(\theta)$. In the case of $N$ independent and identically distributed observations $z^{(1)}, \ldots, z^{(N)}$ from the model $f(z; \theta)$ on $\mathbb{R}^K$ with $K$ fixed, under the usual regularity conditions, $\hat{\theta}$ is a consistent parameter estimator (Lindsay, 1988; Varin & Vidoni, 2005):

$$N^{1/2}(\hat{\theta} - \theta) \to N_q\{0, H(\theta)^{-1} J(\theta) H(\theta)^{-1}\} \tag{8}$$

in distribution as $N \to +\infty$, where $H(\theta) = E\{-\nabla^2 \ell_c(\theta)\}$ and $J(\theta) = \text{var}\{\nabla \ell_c(\theta)\}$. The pairwise composite likelihood method has been introduced for inference in Gaussian max-stable processes by Padoan et al. (2010). However, to our knowledge, an efficiency study of the pairwise maximum composite likelihood estimators is still lacking for the Gaussian max-stable process model. In §2, we have derived a closed form expression for the likelihood function of a Gaussian max-stable process at sites $x_j \in \mathbb{R}^d$ $(j = 1, \ldots, p)$, $p \leqslant d + 1$, and $d \geqslant 1$. This allows a natural extension to triplewise composite likelihood inference and so on. We study the gain in efficiency of the maximum composite likelihood estimates from $p = 2$ to $p = 3$ sites in $\mathbb{R}^2$ in the next section.

## 4. EFFICIENCY GAIN

We present a simulation study to investigate the use of composite likelihood methods for inference in Gaussian max-stable processes. We compare pairwise with triplewise composite likelihood inference in the spatial domain $\mathbb{R}^2$ based on (5) and (6).

We randomly generate $K$ site locations uniformly in the square $[0, 100] \times [0, 100]$. We then simulate $N$ Gaussian max-stable process realizations at the sampled $K$ locations using the SpatialExtremes package in R (R Development Core Team, 2011). The multivariate Gaussian cumulative distribution function $\Phi_{p-1}$ in (3) can be evaluated numerically by means of the R or Matlab commands based on the algorithm by Genz (1992); see also Genz & Bretz (2002, 2009).

We adopt five parameter settings for $\Sigma$ to investigate the estimators' performances under varying spatial dependence structures. The true parameter values of $\Sigma$ are given in Tables 1 and 2. The square roots of the eigenvalues of $\Sigma$ for the first three settings are, respectively: (10, 10), (20, 20) and (30, 30), for which the correlation is 0. They represent an isotropic short-, mid- and long-range dependence structure on the $[0, 100] \times [0, 100]$ square. The square roots of the eigenvalues of $\Sigma$ for the last two settings are, respectively: (42·2, 26·8) and (13·3, 8·5), for which the correlation is 0·33. They represent an anisotropic short- and long-range dependence structure on the $[0, 100] \times [0, 100]$ square.

For each setting, we calculate the sample means and the sample standard deviations of the maximum composite likelihood estimates of $\Sigma$ based on 100 simulations. We also calculate the sample relative efficiency between the triplewise and pairwise maximum composite likelihood estimates with the same $K$ and $N$.

Table 1 summarizes the mean of the pairwise and the triplewise maximum composite likelihood estimates along with their sample standard deviations based on $K = 20$ sites and a varying number $N$ of observations. Given the same number of sites and observations, the triplewise maximum composite likelihood estimates have more accurate parameter estimates with smaller biases and standard deviations than

Table 1. *Maximum composite likelihood estimation of $\Sigma$ for a Gaussian max-stable process indexed by $\mathbb{R}^2$: mean and standard deviations over 100 simulations. $K = 20$ is the number of locations. $N$ is the number of observations. $r$ is the relative efficiency between the triplewise and pairwise maximum composite likelihood estimates with the same $K$ and $N$*

| $N$ | Pair | Triple | $r$ | Pair | Triple | $r$ | Pair | Triple | $r$ |
|---|---|---|---|---|---|---|---|---|---|
| | | $\sigma_{11} = 10$ | | | $\sigma_{22} = 10$ | | | $\rho = 0$ | |
| 5 | 10·9(2·46) | 11·1(2·30) | 0·95 | 11·0(2·75) | 10·6(1·89) | 0·46 | 0·10(0·25) | 0·10(0·23) | 0·88 |
| 10 | 10·4(1·58) | 10·5(1·28) | 0·70 | 10·4(1·56) | 10·2(0·84) | 0·29 | 0·05(0·17) | 0·05(0·13) | 0·65 |
| 20 | 10·1(1·14) | 10·2(0·74) | 0·44 | 10·2(1·11) | 10·2(0·51) | 0·22 | 0·02(0·13) | 0·02(0·10) | 0·65 |
| 50 | 10·1(0·77) | 10·1(0·43) | 0·34 | 10·1(0·73) | 10·0(0·20) | 0·08 | 0·01(0·10) | 0·01(0·05) | 0·22 |
| | | $\sigma_{11} = 20$ | | | $\sigma_{22} = 20$ | | | $\rho = 0$ | |
| 5 | 21·7(5·26) | 20·7(3·65) | 0·45 | 21·8(5·11) | 19·7(1·73) | 0·10 | 0·12(0·25) | 0·06(0·16) | 0·39 |
| 10 | 20·8(3·99) | 20·1(1·85) | 0·21 | 21·5(3·98) | 19·9(0·88) | 0·04 | 0·08(0·21) | 0·03(0·12) | 0·29 |
| 20 | 20·4(2·83) | 20·1(1·01) | 0·13 | 20·8(2·90) | 19·9(0·57) | 0·04 | 0·04(0·16) | 0·01(0·06) | 0·13 |
| 50 | 20·2(1·72) | 20·0(0·16) | 0·01 | 20·1(1·71) | 20·0(0·23) | 0·02 | 0·02(0·10) | 0·00(0·01) | 0·02 |
| | | $\sigma_{11} = 30$ | | | $\sigma_{22} = 30$ | | | $\rho = 0$ | |
| 5 | 33·6(9·48) | 30·3(4·84) | 0·23 | 33·1(9·38) | 29·1(2·86) | 0·09 | 0·08(0·22) | 0·04(0·14) | 0·40 |
| 10 | 31·8(7·28) | 29·8(3·33) | 0·20 | 31·6(5·95) | 29·4(1·52) | 0·07 | 0·03(0·18) | 0·01(0·10) | 0·31 |
| 20 | 31·1(5·16) | 29·7(1·50) | 0·08 | 30·8(4·21) | 29·6(1·02) | 0·06 | 0·05(0·15) | 0·00(0·06) | 0·14 |
| 50 | 30·2(2·82) | 29·9(0·62) | 0·05 | 30·3(2·84) | 29·9(0·53) | 0·04 | 0·02(0·11) | 0·00(0·02) | 0·03 |
| | | $\sigma_{11} = 40$ | | | $\sigma_{22} = 30$ | | | $\rho = 0{\cdot}33$ | |
| 5 | 46·0(13·6) | 43·3(9·85) | 0·49 | 35·0(8·39) | 30·3(3·46) | 0·13 | 0·36(0·27) | 0·37(0·18) | 0·47 |
| 10 | 41·7(7·91) | 40·5(4·93) | 0·38 | 32·5(6·17) | 29·7(1·92) | 0·09 | 0·31(0·21) | 0·33(0·12) | 0·31 |
| 20 | 41·6(6·10) | 40·4(2·82) | 0·20 | 31·0(3·97) | 29·8(0·89) | 0·05 | 0·30(0·15) | 0·34(0·06) | 0·14 |
| 50 | 41·0(3·92) | 40·1(0·41) | 0·01 | 30·5(2·89) | 30·0(0·22) | 0·01 | 0·33(0·11) | 0·34(0·02) | 0·03 |
| | | $\sigma_{11} = 12{\cdot}6$ | | | $\sigma_{22} = 9{\cdot}5$ | | | $\rho = 0{\cdot}33$ | |
| 5 | 13·8(3·26) | 14·0(3·13) | 0·97 | 11·5(3·27) | 10·9(2·47) | 0·54 | 0·36(0·29) | 0·39(0·24) | 0·70 |
| 10 | 12·9(2·25) | 13·2(2·08) | 0·91 | 10·1(1·81) | 10·0(1·33) | 0·55 | 0·31(0·25) | 0·37(0·18) | 0·55 |
| 20 | 12·7(1·38) | 13·0(1·22) | 0·83 | 9·8(1·15) | 9·7(0·77) | 0·47 | 0·31(0·17) | 0·35(0·12) | 0·47 |
| 50 | 12·7(0·92) | 12·9(0·71) | 0·65 | 9·6(0·70) | 9·5(0·27) | 0·16 | 0·31(0·11) | 0·34(0·05) | 0·20 |

the pairwise counterparts. For example, the triplewise maximum composite likelihood estimates based on $N = 10$ observations have comparable or even smaller biases and standard deviations than the pairwise maximum composite likelihood estimates based on $N = 20$. Overall, the efficiency gains from pairwise to triplewise maximum composite likelihood estimates are large; the relative efficiency ranges from 1 to 97% in this simulation study. In particular, we observe higher efficiency gains in the case of stronger spatial dependence or larger sample size $N$. The reason is that triplewise composite likelihood is able to borrow more spatial information to infer spatial extreme dependence. The parameter estimates under both methods are significantly improved as $N$ increases, which is expected considering the asymptotic theory in (8) as $N \to +\infty$.

It is also of interest to consider the case where $N$ is fixed and $K$ increases. However, the asymptotic theory for the composite likelihood inference in this case is much more challenging, especially under the infill asymptotic framework, and we are unaware of the existence of a rigorous theoretical investigation of this. Below, we examine the performance of the maximum composite likelihood estimators using a simulation study with an increasing number $K$ of sites.

Table 2 displays the mean of the pairwise and the triplewise maximum composite likelihood estimates along with their sample standard deviations based on $N = 2$ and an increasing number $K$ of sites. We again observe substantial efficiency gains from pairwise to triplewise maximum composite likelihood estimates. As $K$ increases, we observe smaller biases and standard deviations for both approaches. However, the number $K$ of sites does not impact the pairwise estimation results as much as the triplewise estimations, especially in the case of strong spatial dependence. Again, this suggests that triplewise composite

Table 2. *Maximum composite likelihood estimation of* $\Sigma$ *for a Gaussian max-stable process indexed by* $\mathbb{R}^2$*: mean and standard deviations over* 100 *simulations.* $K$ *is the number of locations.* $N = 2$ *is the number of observations.* $r$ *is the relative efficiency between the triplewise and pairwise maximum composite likelihood estimates with the same* $K$ *and* $N$

| $K$ | Pair | Triple | $r$ | Pair | Triple | $r$ | Pair | Triple | $r$ |
|---|---|---|---|---|---|---|---|---|---|
| | | $\sigma_{11} = 10$ | | | $\sigma_{22} = 10$ | | | $\rho = 0$ | |
| 10 | 28·9(29·6) | 22·6(20·34) | 0·47 | 27·5(26·09) | 26·3(25·71) | 0·94 | 0·32(0·48) | 0·28(0·46) | 0·89 |
| 20 | 13·0(5·77) | 12·9(4·65) | 0·69 | 13·1(6·05) | 12·0(4·44) | 0·51 | 0·20(0·36) | 0·19(0·34) | 0·86 |
| 30 | 11·7(3·88) | 11·1(2·95) | 0·55 | 11·8(4·07) | 11·0(3·35) | 0·62 | 0·16(0·31) | 0·13(0·24) | 0·64 |
| 50 | 11·2(3·48) | 10·0(1·52) | 0·17 | 11·3(3·11) | 10·3(1·38) | 0·18 | 0·12(0·26) | 0·05(0·17) | 0·37 |
| | | $\sigma_{11} = 20$ | | | $\sigma_{22} = 20$ | | | $\rho = 0$ | |
| 10 | 33·2(22·30) | 29·5(18·22) | 0·63 | 34·1(21·46) | 29·6(17·34) | 0·60 | 0·25(0·40) | 0·23(0·36) | 0·83 |
| 20 | 25·0(10·10) | 22·9(8·69) | 0·67 | 26·9(11·16) | 21·9(5·88) | 0·22 | 0·14(0·29) | 0·13(0·24) | 0·74 |
| 30 | 23·3(8·22) | 20·6(4·88) | 0·31 | 24·9(8·93) | 20·5(4·31) | 0·18 | 0·13(0·26) | 0·04(0·16) | 0·32 |
| 50 | 23·2(7·14) | 19·6(2·62) | 0·12 | 23·7(7·76) | 19·4(2·23) | 0·07 | 0·13(0·25) | 0·03(0·12) | 0·18 |
| | | $\sigma_{11} = 30$ | | | $\sigma_{22} = 30$ | | | $\rho = 0$ | |
| 10 | 48·2(24·77) | 39·9(17·62) | 0·43 | 43·6(22·01) | 37·0(15·33) | 0·42 | 0·24(0·38) | 0·20(0·35) | 0·80 |
| 20 | 40·0(18·34) | 34·4(13·55) | 0·46 | 38·2(15·61) | 30·5(6·90) | 0·16 | 0·19(0·33) | 0·15(0·27) | 0·65 |
| 30 | 38·9(15·97) | 30·3(7·98) | 0·19 | 36·7(13·11) | 30·1(5·89) | 0·16 | 0·19(0·32) | 0·07(0·20) | 0·33 |
| 50 | 38·5(14·71) | 28·4(4·12) | 0·07 | 36·8(12·53) | 29·2(4·45) | 0·10 | 0·19(0·33) | 0·04(0·17) | 0·22 |
| | | $\sigma_{11} = 40$ | | | $\sigma_{22} = 30$ | | | $\rho = 0{\cdot}33$ | |
| 10 | 57·2(25·15) | 52·5(21·77) | 0·68 | 48·2(22·2) | 38·2(14·32) | 0·33 | 0·43(0·38) | 0·44(0·32) | 0·71 |
| 20 | 54·8(22·61) | 50·2(20·60) | 0·72 | 41·5(16·1) | 33·6(10·21) | 0·30 | 0·41(0·35) | 0·46(0·29) | 0·76 |
| 30 | 53·4(21·98) | 40·2(9·61) | 0·14 | 41·0(15·1) | 30·2(6·14) | 0·11 | 0·43(0·35) | 0·36(0·23) | 0·41 |
| 50 | 53·2(20·83) | 38·6(8·38) | 0·12 | 41·3(14·4) | 29·1(5·11) | 0·08 | 0·43(0·35) | 0·35(0·20) | 0·31 |
| | | $\sigma_{11} = 12{\cdot}6$ | | | $\sigma_{22} = 9{\cdot}5$ | | | $\rho = 0{\cdot}33$ | |
| 10 | 17·8(7·20) | 17·6(7·22) | 0·98 | 15·0(6·71) | 14·8(6·69) | 0·97 | 0·35(0·42) | 0·31(0·42) | 0·96 |
| 20 | 16·1(5·56) | 16·0(5·47) | 0·97 | 13·3(5·22) | 12·7(4·72) | 0·78 | 0·38(0·36) | 0·41(0·35) | 0·99 |
| 30 | 15·3(5·00) | 14·5(4·97) | 0·88 | 12·6(4·62) | 11·0(3·70) | 0·52 | 0·36(0·33) | 0·39(0·27) | 0·67 |
| 50 | 15·0(4·28) | 12·7(2·21) | 0·20 | 11·6(3·75) | 10·0(2·16) | 0·27 | 0·35(0·30) | 0·35(0·19) | 0·43 |

likelihood inference can take more spatial information into account. Although both approaches appear to produce nonnegligible biases and large standard deviations when $K$ is small, triplewise inference is still better than pairwise.

## 5. DISCUSSION

For $p > d + 1$, the representation (3) is not valid. Nevertheless, the cumulative distribution function $F(z)$, and therefore also the exponent measure function $V(z)$, can be evaluated through Monte Carlo simulations based on:

$$\begin{aligned}\mathrm{pr}\left\{\frac{\phi_d(Y_j;\Sigma)}{z_j} > \max_{k\neq j}\frac{\phi_d(Y_j+x_j-x_k;\Sigma)}{z_k}\right\} &= \mathrm{pr}\left\{Y_j^{\mathrm{T}}\Sigma^{-1}(x_j-x_k) > -c_k^{(j)}(z),\ k=1,\ldots,p,\ k\neq j\right\}\\ &= \int_{\mathbb{R}^d}\phi_d(y;\Sigma)\prod_{k=1,\,k\neq j}^{p} I\left\{y^{\mathrm{T}}\Sigma^{-1}(x_j-x_k)+c_k^{(j)}(z)>0\right\}dy\\ &= E\left[\prod_{k=1,\,k\neq j}^{p} I\left\{Y^{\mathrm{T}}\Sigma^{-1}(x_j-x_k)+c_k^{(j)}(z)>0\right\}\right],\end{aligned}$$

where $Y \sim N_d(0, \Sigma)$. Unfortunately, the evaluation of the probability density function $f(z)$ by numerical partial differentiation of order $p$ of $F(z)$ seems infeasible for $p > d + 1$. Moreover, the maximization of the resulting likelihood function would add yet another level of difficulty.

When the margins do not have a unit Fréchet distribution, the bijective transformation defining the generalized extreme value distribution can be used easily to modify the likelihood function.

Further efficiency gains can be obtained by selecting appropriate unequal weights in (7). For instance, setting dummy weights to exclude distant pairs for spatially/temporally correlated observations may improve the efficiency of the pairwise composite likelihood (Davis & Yau, 2011; Varin et al., 2011).

## Acknowledgement

This research was sponsored by the U.S. National Science Foundation and by an award made by King Abdullah University of Science and Technology. The authors thank the editor and two referees for very useful comments.

## Appendix

### *Derivation of the joint probability density function*

Using (3), (4), and the convention $\Phi_0 \equiv 1$, we have

$$V_{[i]}(z) = \sum_{j \neq i} \left[ \frac{1}{z_i z_j} \frac{\partial \Phi_{p-1}\left\{c^{(j)}(z);\, \Sigma^{(j)}\right\}}{\partial c_i^{(j)}(z)} - \frac{1}{z_i^2} \frac{\partial \Phi_{p-1}\left\{c^{(i)}(z);\, \Sigma^{(i)}\right\}}{\partial c_j^{(i)}(z)} \right] - \frac{1}{z_i^2} \Phi_{p-1}\left\{c^{(i)}(z);\, \Sigma^{(i)}\right\}$$

$$= \sum_{j \neq i} \left[ \frac{1}{z_i z_j} \phi\left\{c_i^{(j)}(z);\, \Sigma_{i,i}^{(j)}\right\} \Phi_{p-2}\left\{c_{-i}^{(j)}(z) - (\Sigma_{i,-i}^{(j)})^{\mathrm{T}} c_i^{(j)}(z)/\Sigma_{i,i}^{(j)};\, \Sigma_{-i,-i}^{(j)} - (\Sigma_{i,-i}^{(j)})^{\mathrm{T}} \Sigma_{i,-i}^{(j)}/\Sigma_{i,i}^{(j)}\right\} \right.$$

$$\left. - \frac{1}{z_i^2} \phi\left\{c_j^{(i)}(z);\, \Sigma_{j,j}^{(i)}\right\} \Phi_{p-2}\left\{c_{-j}^{(i)}(z) - (\Sigma_{j,-j}^{(i)})^{\mathrm{T}} c_j^{(i)}(z)/\Sigma_{j,j}^{(i)};\, \Sigma_{-j,-j}^{(i)} - (\Sigma_{j,-j}^{(i)})^{\mathrm{T}} \Sigma_{j,-j}^{(i)}/\Sigma_{j,j}^{(i)}\right\} \right]$$

$$- \frac{1}{z_i^2} \Phi_{p-1}\left\{c^{(i)}(z);\, \Sigma^{(i)}\right\},$$

where $c_{-j}^{(i)}(z) = c^{(i)}(z) \backslash c_j^{(i)}(z)$, $\Sigma_{j,-j}^{(i)} = (x_i - x_j)^{\mathrm{T}} \Sigma^{-1} (x_i 1_{p-2}^{\mathrm{T}} - X_{-(ij)})$, and $\Sigma_{-j,-j}^{(i)} = (x_i 1_{p-2}^{\mathrm{T}} - X_{-(ij)})^{\mathrm{T}} \Sigma^{-1} (x_i 1_{p-2}^{\mathrm{T}} - X_{-(ij)})$ in which $X_{-(ij)} = X \backslash (x_i, x_j)$. Using the fact that $\Sigma_{j,j}^{(i)} = \Sigma_{i,i}^{(j)}$, $\Sigma_{-j,-j}^{(i)} - \Sigma_{-i,-i}^{(j)} = \Sigma_{j,-j}^{(i)} - \Sigma_{i,-j}^{(j)}$, $\Sigma_{j,-j}^{(i)} + \Sigma_{i,-i}^{(j)} = \Sigma_{j,j}^{(i)} 1_{p-2} = \Sigma_{i,i}^{(j)} 1_{p-2}$, it follows that

$$\phi\left\{c_j^{(i)}(z);\, \Sigma_{j,j}^{(i)}\right\} = (z_i/z_j)\phi\left\{c_i^{(j)}(z);\, \Sigma_{i,i}^{(j)}\right\},$$

$$\Sigma_{-j,-j}^{(i)} - (\Sigma_{j,-j}^{(i)})^{\mathrm{T}} \Sigma_{j,-j}^{(i)}/\Sigma_{j,j}^{(i)} = \Sigma_{-i,-i}^{(j)} - (\Sigma_{i,-i}^{(j)})^{\mathrm{T}} \Sigma_{i,-i}^{(j)}/\Sigma_{i,i}^{(j)},$$

$$c_{-j}^{(i)}(z) - (\Sigma_{j,-j}^{(i)})^{\mathrm{T}} c_j^{(i)}(z)/\Sigma_{j,j}^{(i)} = c_{-i}^{(j)}(z) - (\Sigma_{i,-i}^{(j)})^{\mathrm{T}} c_i^{(j)}(z)/\Sigma_{i,i}^{(j)}.$$

Therefore, the expression for $V_{[i]}(z)$ simplifies to

$$V_{[i]}(z) = -\frac{1}{z_i^2} \Phi_{p-1}\left\{c^{(i)}(z);\, \Sigma^{(i)}\right\}.$$

Denote $(i_2, \ldots, i_k)$, $k < p$, by $\mathcal{A}$. It follows that

$$V_{[i_1,i_2,\ldots,i_k]}(z) = -\frac{1}{z_{i_1} \prod_{j=1}^{k} z_{i_j}} \phi_{k-1}\left\{c_{\mathcal{A}}^{(i_1)}(z);\, \Sigma_{\mathcal{A}}^{(i_1)}\right\}$$

$$\times\, \Phi_{p-k}\left\{c_{-\mathcal{A}}^{(i_1)}(z) - (\Sigma_{\mathcal{A},-\mathcal{A}}^{(i_1)})^{\mathrm{T}} (\Sigma_{\mathcal{A},\mathcal{A}}^{(i_1)})^{-1} c_{\mathcal{A}}^{(i_1)}(z);\, \Sigma_{-\mathcal{A},-\mathcal{A}}^{(i_1)} - (\Sigma_{\mathcal{A},-\mathcal{A}}^{(i_1)})^{\mathrm{T}} (\Sigma_{\mathcal{A},\mathcal{A}}^{(i_1)})^{-1} \Sigma_{\mathcal{A},-\mathcal{A}}^{(i_1)}\right\}$$

where $c_{\mathcal{A}}^{(i_1)}(z) = \{c_{i_2}^{(i_1)}(z), \ldots, c_{i_k}^{(i_1)}(z)\}^{\mathrm{T}}$, $c_{-\mathcal{A}}^{(i_1)}(z) = c^{(i_1)}(z) \backslash c_{\mathcal{A}}^{(i_1)}(z)$, $\Sigma_{\mathcal{A},-\mathcal{A}}^{(i_1)} = (x_i 1_{k-1}^{\mathrm{T}} - X_{\mathcal{A}})^{\mathrm{T}} \Sigma^{-1} (x_i 1_{p-k}^{\mathrm{T}} - X_{-\mathcal{A}})$ and $\Sigma_{-\mathcal{A},-\mathcal{A}}^{(i_1)} = (x_i 1_{p-k}^{\mathrm{T}} - X_{-\mathcal{A}})^{\mathrm{T}} \Sigma^{-1} (x_i 1_{p-k}^{\mathrm{T}} - X_{-\mathcal{A}})$ in which $X_{\mathcal{A}} = (x_{i_2}, \ldots, x_{i_k})$ and $X_{-\mathcal{A}} = X \backslash (x_{i_1}, X_{\mathcal{A}})$. In particular, the $p$th order cross partial derivative of $V(z)$ is

$$V_{[1,2,\ldots,p]}(z) = -\frac{1}{z_i \prod_{j=1}^{p} z_j} \phi_{p-1} \left\{ c^{(i)}(z); \Sigma^{(i)} \right\}.$$

## REFERENCES


COX, D. R. & REID, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* **91**, 729–37.

DAVIS, R. A. & YAU, C. Y. (2011). Comments on pairwise likelihood in time series models. *Statist. Sinica* **21**, 255–77.

DE HAAN, L. (1984). A spectral representation for max-stable processes. *Ann. Prob.* **12**, 1194–204.

DE HAAN, L. & PEREIRA, T. T. (2006). Spatial extremes: Models for the stationary case. *Ann. Statist.* **34**, 146–68.

GENZ, A. (1992). Numerical computation of multivariate normal probabilities. *J. Comp. Graph. Statist.* **1**, 141–49.

GENZ, A. & BRETZ, F. (2002). Methods for the computation of multivariate t-probabilities. *J. Comp. Graph. Statist.* **11**, 950–71.

GENZ, A. & BRETZ, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Berlin: Springer.

LINDSAY, B. G. (1988). *Composite Likelihood Methods*. In: *Statistical Inference from Stochastic Processes*, Ed. N. U. Prabhu, pp. 221–39. Providence, RI: American Mathematical Society.

PADOAN, S. A., RIBATET, M. & SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Am. Statist. Assoc.* **105**, 263–77.

R DEVELOPMENT CORE TEAM (2011). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, ISBN 3-900051-07-0.

VARIN, C., REID, N. & FIRTH, D. (2011). An overview on composite likelihood methods. *Statist. Sinica* **21**, 5–42.

VARIN, C. & VIDONI, P. (2005). A note on composite likelihood inference and model selection. *Biometrika* **92**, 519–28.


[*Received May* 2010. *Revised February* 2011]