Biometrika (2013), **100**, 1, *pp.* 1–9

Advance Access publication on 30 June 2013
Printed in Great Britain

Composite likelihood estimation for the Brown–Resnick process

BY R. HUSER AND A. C. DAVISON

Ecole Polytechnique Fédérale de Lausanne, EPFL-FSB-MATHAA-STAT, Station 8, 1015 Lausanne, Switzerland

Raphael.Huser@epfl.ch Anthony.Davison@epfl.ch

SUMMARY

Genton *et al.* (2011) investigated the gain in efficiency when triplewise, rather than pairwise, likelihood is used to fit the popular Smith max-stable model for spatial extremes. We generalize their results to the Brown–Resnick model and show that the efficiency gain is substantial only for very smooth processes, which are generally unrealistic in applications.

Some key words: Brown–Resnick process; Composite likelihood; Max-stable process; Pairwise likelihood; Smith process; Triplewise likelihood.

1. INTRODUCTION

Max-stable processes are useful for the statistical modelling of spatial extreme events. No finite parametrization of such processes exists, but a spectral representation (de Haan, 1984) aids in constructing models. In a 1990 University of Surrey technical report, R. L. Smith proposed a max-stable model based on deterministic storm profiles which has become popular because it is simple, readily interpreted and easily simulated, but unfortunately it is too inflexible for realism in practice. Another popular model, the Brown–Resnick process, is based on intrinsically stationary log-Gaussian processes, can handle a wide range of dependence structures, and often provides a better fit to data; see, for example, Davison *et al.* (2012) or a 2012 University of North Carolina at Chapel Hill PhD thesis by Soyoung Jeon. Kabluchko *et al.* (2009) provided further underpinning for this process by showing that that under mild conditions, the Brown–Resnick process with variogram $2\gamma(h) = (\|h\|/\rho)^\alpha$ $(\rho > 0, 0 < \alpha \leq 2)$, where h is the spatial lag, is essentially the only isotropic limit of properly rescaled maxima of Gaussian processes. The Smith model is obtained by taking a Brown–Resnick process with variogram $2\gamma(h) = h^{\mathrm{T}}\Sigma^{-1}h$ for some covariance matrix Σ, corresponding after an affine transformation to taking $\alpha = 2$, whereas Davison *et al.* (2012) found that $1/2 < \alpha < 1$ for the rainfall data they examined.

Likelihood inference for max-stable models is difficult, since only the bivariate marginal density functions are known in most cases, and pairwise marginal likelihood is typically used (Padoan *et al.*, 2010; Davison and Gholamrezaee, 2012; Huser and Davison, 2012). This raises the question whether some other approach to inference would be preferable. Genton *et al.* (2011) derived the general form of the likelihood function for the Smith model and showed that large efficiency gains can arise when fitting it using triplewise, rather than pairwise, likelihood. In this paper we extend their investigation to the Brown–Resnick process and show that for rougher models, more realistic than those considered by Genton *et al.* (2011), the efficiency gains are much less striking. Thus pairwise likelihood inference provides a good compromise between statistical and computational efficiency in many applications.

2. BROWN–RESNICK PROCESS

2·1. *Definition and properties*

The Brown–Resnick process (Brown and Resnick, 1977; Kabluchko *et al.*, 2009) is a stationary max-stable process that may be represented as $Z(x) = \sup_{i\in\mathbb{N}} W_i(x)/T_i$ ($x \in \mathcal{X} \subset \mathbb{R}^d$), where $0 < T_1 < T_2 < \cdots$ are the points of a unit rate Poisson process on $\mathbb{R}_+$ and the $W_i(x)$ are independent replicates of the random process $W(x) = \exp\{\varepsilon(x) - \gamma(x)\}$. Here $\varepsilon(x)$ is an intrinsically stationary Gaussian random field with semi-variogram $\gamma(h)$ with $\varepsilon(0) = 0$ almost surely. One interpretation of $Z(x)$ is as the pointwise maximum of an infinite number of independent random storms $W_i(x)$, each rescaled by a corresponding storm size T_i^{-1}. The full distribution of $Z(x)$ at the set of sites $\mathcal{D} \subset \mathcal{X}$ is (Davison and Gholamrezaee, 2012)

$$\mathrm{pr}\{Z(x) \le z(x), x \in \mathcal{D}\} = \exp\left(-E\left[\sup_{x\in\mathcal{D}}\left\{\frac{W(x)}{z(x)}\right\}\right]\right),$$

where the exponent measure function $V_{\mathcal{D}}\{z(x)\} = E\left[\sup_{x\in\mathcal{D}}\{W(x)/z(x)\}\right]$ must satisfy certain constraints (see, e.g., Davison *et al.*, 2012). The full distribution is intractable when $\mathcal{D}$ is arbitrary, but explicit formulae for the marginal distributions are available when its size $|\mathcal{D}| = 1, 2$, and in certain cases more; see below. The univariate margins of $Z(x)$ equal $\exp(-1/z)$, for $z > 0$, and for $\mathcal{D} = \{x_1, x_2\}$ the exponent measure of the Brown–Resnick process is

$$V(z_1, z_2) = \frac{1}{z_1}\Phi\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{z_1}{z_2}\right)\right\} + \frac{1}{z_2}\Phi\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{z_2}{z_1}\right)\right\}, \tag{1}$$

where $z_i = z(x_i)$, $i = 1, 2$, $a = \{2\gamma(x_2 - x_1)\}^{1/2}$ and $\Phi(\cdot)$ denotes the standard normal distribution function. In this case expression (1) boils down to the Hüsler–Reiss (1989) model for bivariate extremes. The bivariate marginal density functions $f(z_1, z_2)$ are easily expressed using derivatives of (1).

Figure 1 shows how the variogram influences the smoothness of the max-stable process. In particular, when the smoothness parameter α equals 2, i.e., $2\gamma(h) = h^{\mathrm{T}}\Sigma^{-1}h$ for some covariance matrix Σ, the bivariate exponent measure of the Smith model is recovered (Kabluchko *et al.*, 2009; Padoan *et al.*, 2010), and the storm shapes are deterministic, taking the form of Gaussian densities.

2·2. *Triplewise margins*

Let $\mathcal{D} = \{x_1, x_2, x_3\} \subset \mathcal{X}$ and for simplicity write $z_1 = z(x_1)$, $\gamma_{1;2} = \gamma(x_1 - x_2)$, etc. Computations in Appendix A show that provided $R_1, R_2, R_3 \neq \pm 1$, the triplewise exponent measure may be expressed as

$$\begin{aligned} V(z_1, z_2, z_3) = {} & \frac{1}{z_1}\Phi_2\{\eta(z_1, z_2), \eta(z_1, z_3); R_1\} + \frac{1}{z_2}\Phi_2\{\eta(z_2, z_1), \eta(z_2, z_3); R_2\} \\ & + \frac{1}{z_3}\Phi_2\{\eta(z_3, z_1), \eta(z_3, z_2); R_3\}, \end{aligned} \tag{2}$$

where $\Phi_2(\cdot, \cdot; R)$ denotes the bivariate normal distribution function with zero mean, correlation R and unit variance, $\eta(z_i, z_j) = (\gamma_{i;j}/2)^{1/2} - \log(z_i/z_j)/(2\gamma_{i;j})^{1/2}$, and

$$R_1 = \frac{\gamma_{1;2} + \gamma_{1;3} - \gamma_{2;3}}{2(\gamma_{1;2}\gamma_{1;3})^{1/2}}, \quad R_2 = \frac{\gamma_{1;2} + \gamma_{2;3} - \gamma_{1;3}}{2(\gamma_{1;2}\gamma_{2;3})^{1/2}}, \quad R_3 = \frac{\gamma_{1;3} + \gamma_{2;3} - \gamma_{1;2}}{2(\gamma_{1;3}\gamma_{2;3})^{1/2}}.$$

The function $\Phi_2(\cdot, \cdot; R)$ is rapidly computed (Genz, 1992; Genz and Bretz, 2000, 2002), and the triplewise density $f(z_1, z_2, z_3)$ is easily found by differentiating $\exp\{-V(z_1, z_2, z_3)\}$. The resulting expressions are given in the Supplementary Material.

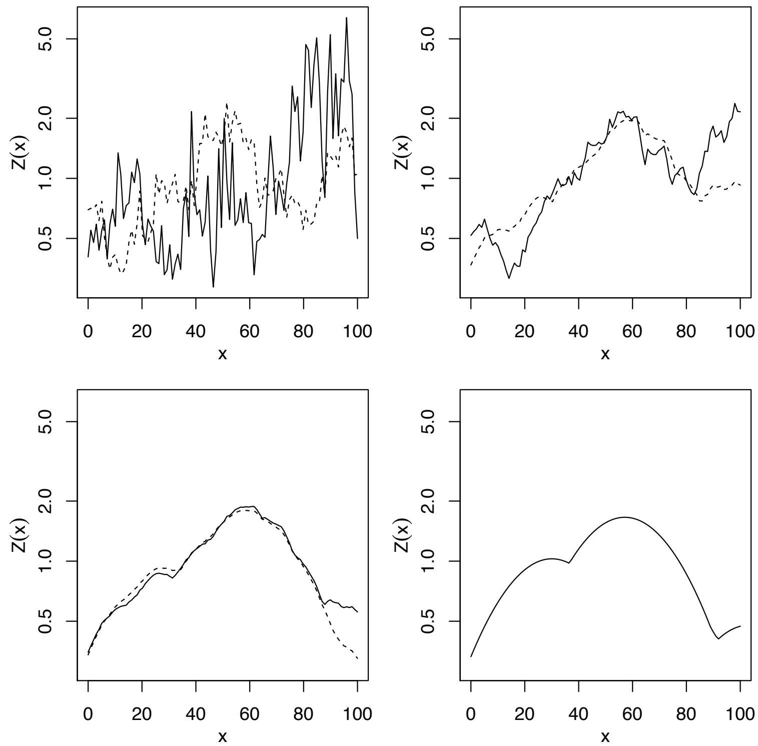

Fig. 1. Seven simulated Brown–Resnick processes in one dimension, i.e., $d = 1$, with variogram $2\gamma(h) = (\|h\|/28)^{\alpha}$, and different smoothness parameters. *Top left*: $\alpha = 0.5$ (solid), 1 (dashed); *top right*: $\alpha = 1.5$ (solid), 1.9 (dashed); *bottom left*: $\alpha = 1.95$ (solid), 1.98 (dashed); *bottom right*: $\alpha = 2$, which corresponds to the isotropic Smith model. The same random seed was used in all seven cases.

2·3. *Higher order margins*

Computations in Appendix B show that when $|\mathcal{D}| = p$, and $p \leq d + 1$ if $\alpha = 2$, the exponent measure for the Brown–Resnick process may be written as

$$V(z_1, \ldots, z_p) = \sum_{k=1}^{p} \frac{1}{z_k} \Phi_{p-1}(\eta_k; R_k), \tag{3}$$

where η_k is the $(p-1)$-dimensional vector with sth component $\eta(z_k, z_s)$ $(s = 1, \ldots, p; s \neq k)$, $\Phi_p(\cdot; R)$ denotes the cumulative distribution function of the p-variate normal distribution function with zero mean, unit variance and correlation matrix R, and R_k is the $(p-1) \times (p-1)$ correlation matrix whose (s,t)th entry is $(\gamma_{k;s} + \gamma_{k;t} - \gamma_{s;t})/\{2(\gamma_{k;s}\gamma_{k;t})^{1/2}\}$ $(s, t = 1, \ldots, p; s, t \neq k)$. We recover the results of Section 2·2 when $p = 3$ and those of Genton *et al.* (2011) when the variogram is $2\gamma(h) = h^{\mathrm{T}}\Sigma^{-1}h$ for some covariance matrix Σ. In principle the full likelihood can then be obtained by differentiation of the cumulative distribution, but the number of terms grows very fast as p increases, so direct likelihood inference seems infeasible except for small p. Moreover, when $\alpha \approx 2$ and large p, the matrices R_k may be numerically singular, causing computational problems in the evaluation of the likelihood; see the Supplementary Material for more details.

3. COMPOSITE LIKELIHOODS

Suppose that n independent replicates of a Brown–Resnick process with variogram $2\gamma(h)$ depending on parameters θ are observed at S sites in $\mathbb{R}^d$, and let $z_{i,j}$ denote the value of the ith process at site j. We consider only the pairwise and triplewise marginal likelihoods,

$$\ell_2(\theta) = \sum_{i=1}^{n} \sum_{j_1<j_2} \log f(z_{i,j_1}, z_{i,j_2}; \theta), \quad \ell_3(\theta) = \sum_{i=1}^{n} \sum_{j_1<j_2<j_3} \log f(z_{i,j_1}, z_{i,j_2}, z_{i,j_3}; \theta),$$

and the corresponding maximum likelihood estimators $\hat\theta_2$ and $\hat\theta_3$, which are consistent and asymptotically Gaussian as n increases (Lindsay, 1988; Cox and Reid, 2004; Varin *et al.*, 2011).

Since $\hat\theta_3$ might be thought to perform better than $\hat\theta_2$, the question of their relative statistical efficiency arises. In order to study this for random fields with different smoothness properties, we consider the isotropic semi-variogram $\gamma(h) = (\|h\|/\rho)^\alpha$ $(\rho > 0, 0 < \alpha \le 2)$, which corresponds to Brown–Resnick processes built from fractional Brownian motions. We consider the seven smoothness scenarios $\alpha = 0.5, 1, 1.5, 1.9, 1.95, 1.98, 2$, the last being equivalent to the Smith model. For each scenario we consider three levels of spatial dependence, with the range parameter $\rho = 14$, 28, 42, broadly corresponding to the three cases $\sigma_{11} = \sigma_{22} = 10$, 20, 30 in Genton *et al.* (2011). The number of replicates of the process was set to $n = 5$, 10, 20 and 50. Using the R package *SpatialExtremes* (Ribatet, 2012), we simulated n independent copies of the Brown–Resnick process with variogram $2\gamma(h)$ at the same set of 20 random sites uniformly generated in $[0, 100]^2$, and computed the estimates $\hat\theta_2 = (\hat\rho_2, \hat\alpha_2)$ and $\hat\theta_3 = (\hat\rho_3, \hat\alpha_3)$, the latter based on the expressions given in Appendix A. Such simulated datasets and random locations were generated 300 times and the resulting estimates were used to compute empirical covariance matrices V_2 and V_3 for $\hat\theta_2$ and $\hat\theta_3$, the empirical marginal efficiencies $\mathrm{RE}_\rho = \widehat{\mathrm{var}}(\hat\rho_3)/\widehat{\mathrm{var}}(\hat\rho_2)$ and $\mathrm{RE}_\alpha = \widehat{\mathrm{var}}(\hat\alpha_3)/\widehat{\mathrm{var}}(\hat\alpha_2)$, and the empirical global efficiency $\mathrm{RE}_\theta = \{\det(V_3)/\det(V_2)\}^{1/2}$.

These efficiencies are reported in Table 1. For rough processes, with $\alpha = 0.5, 1, 1.5$, maximum pairwise likelihood estimation has efficiency at least 70%, and often closer to 90%, relative to the use of triples, and the efficiencies depend little on n. For smooth processes, with $\alpha = 1.9, 1.95, 1.98, 2$, the efficiency of pairwise likelihood estimation can be markedly lower, and decreases rapidly as n increases. In particular, when $\alpha = 2$, i.e., for the Smith model, observations on the same storm profile at three different sites completely determine the profile and thus the underlying variogram. Since this event has non-zero probability the triplewise estimator is super-efficient compared to the pairwise one, explaining the dramatic drop in relative efficiency observed when $\alpha \approx 2$. This behaviour is more striking either when the range parameter ρ is big or for large n, since in either case it is then more likely that a single storm profile will be observed at three sites.

Further simulations described in the Supplementary Material show that when $\alpha = 0.5, 1, 1.5$ the efficiencies depend little on the number of sites S, but that when $\alpha = 2$ they decrease rapidly as S increases. Again, when S is larger, more triples observed on the same storm profile are likely to occur, so the super-efficiency of the triplewise likelihood estimator when $\alpha = 2$ has more impact in finite samples.

Figure 2 shows that the relevance of the limiting Gaussian distribution of $\hat\theta_3$ is questionable when $\alpha = 2$: the log triplewise likelihood is very asymmetric even for $n = 50$, whereas it is much more nearly quadratic when α is smaller. Inference based on profile marginal likelihood might thus be advisable when α is thought to be close to 2, even though classical likelihood theory does not apply in this setting. Numerical issues may be encountered when $\alpha \approx 2$, due to

Table 1. *Efficiency (%) of maximum pairwise likelihood estimators relative to maximum triplewise likelihood estimators for* $n = 5, 10, 20, 50$*, based on* 300 *simulations of the Brown–Resnick process with semi-variogram* $(\|h\|/\rho)^{\alpha}$ *observed at* 20 *random sites in* $[0, 100]^2$*. The numbers are respectively* $\mathrm{RE}_{\rho}/\mathrm{RE}_{\alpha}/\mathrm{RE}_{\theta}$.

	$n = 5$			$n = 10$		
$\alpha \setminus \rho$	14	28	42	14	28	42
0.5	83/89/86	89/93/91	87/93/91	94/95/94	90/93/92	93/94/93
1.0	96/92/94	97/84/90	98/88/92	96/89/93	93/90/93	95/85/90
1.5	87/81/83	93/72/79	89/67/74	89/77/82	91/71/81	89/69/78
1.9	79/81/80	72/60/61	74/56/58	84/76/79	76/48/54	66/35/47
1.95	77/80/78	67/54/54	72/54/53	76/75/74	64/46/51	60/38/43
1.98	73/80/77	63/62/58	55/42/46	70/67/66	56/38/39	49/22/29
2.0	74/80/76	61/59/52	53/48/44	64/74/68	42/39/38	26/11/16

	$n = 20$			$n = 50$		
$\alpha \setminus \rho$	14	28	42	14	28	42
0.5	94/94/93	92/93/93	92/95/95	92/92/92	91/97/94	89/92/91
1.0	94/89/91	96/87/92	94/86/92	93/84/88	95/85/91	95/90/95
1.5	88/77/82	90/68/78	88/69/76	92/77/84	90/65/76	87/69/77
1.9	79/60/67	74/36/47	66/28/39	75/48/58	69/22/35	62/18/32
1.95	73/60/64	59/24/35	50/15/26	73/44/55	54/11/22	48/8/17
1.98	68/56/60	49/22/29	38/7/16	68/42/51	40/5/12	33/2/7
2.0	62/65/63	20/6/11	16/3/6	38/30/33	6/0/1	1/0/0

the sharp drop in the likelihood as the range parameter exceeds its true value, and in experiments we have found that the computation often breaks down.

4. Discussion

This paper provides explicit expressions (2) and (3) for the exponent measure of the Brown–Resnick process in arbitrary dimensions, on which likelihood inference may be based. Use of triplewise likelihood rather than pairwise likelihood to fit these models can lead to an efficiency gain of up to 30% for rough processes, and much more if the process is very smooth. This augments the results of Genton *et al.* (2011), which show huge efficiency gains associated to high order composite likelihoods for the Smith model. Our more general results confirm those of Genton *et al.* (2011) for the Smith model, but in the more realistic setting when the process is rough, the small improvement afforded by the triplewise approach is probably not worth the additional computational and coding effort, particularly as issues of numerical precision may then arise. In principle it is possible to compute the full likelihood for the Brown–Resnick process in high dimensions, but the number of terms in the likelihood and the need to compute high-dimensional multivariate normal distribution functions in numerically near-singular cases seem to preclude this in practice.

In applications and for some other models, considerations other than statistical efficiency may arise; for example, the use of triples in a likelihood may be essential for parameter identifiability, as in work to be reported elsewhere on dimension reduction in extremes.

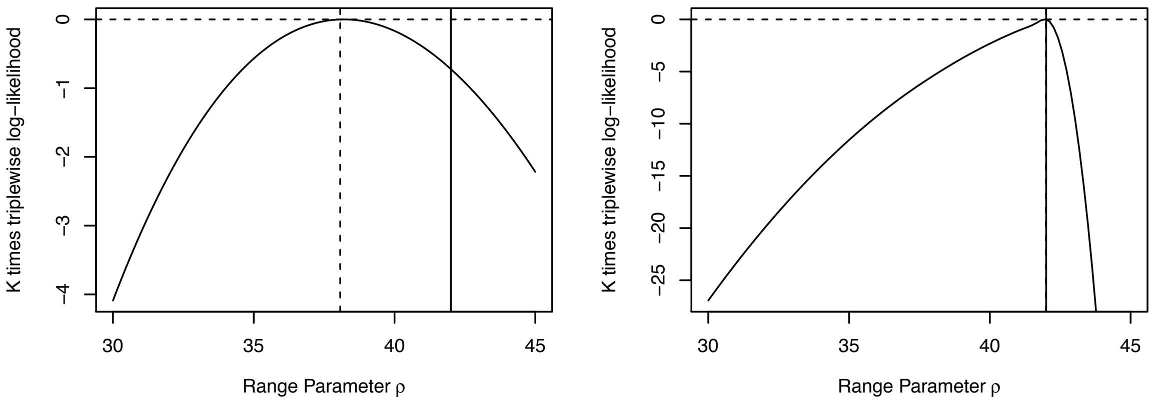

Fig. 2. Triplewise log-likelihoods for the range parameter ρ, with α fixed at its true value, shifted to have maximum at zero and scaled by the factor $K = \{(S-1)(S-2)/2\}^{-1}$, for two datasets generated from a Brown–Resnick process with variogram $2\gamma(h) = (\|h\|/\rho)^\alpha$, where $\alpha = 1$ (left) and $\alpha = 2$ (right). The true value $\rho = 42$ is represented by a solid vertical line. The vertical dashed line, which corresponds to the maximum triplewise likelihood estimator, coincides with the solid line in the right panel. The processes were simulated at the same 20 random sites in $[0, 100]^2$, with $n = 50$ replicates, using the same random seed.

It would be interesting to know whether the efficiency results given here generalize to weighted marginal composite likelihoods (Varin *et al.*, 2011). The best choice of subsets of sites is related to the separate topic of optimal design for likelihood estimation. Both topics are outside the scope of the present work.

ACKNOWLEDGEMENT

This research was funded by the Swiss National Science Foundation, and partly performed in the context of the ETH Competence Center Environment and Sustainability.

SUPPLEMENTARY MATERIAL

Supplementary material available online includes formulae for the computation of the trivariate density of the Brown–Resnick process, figures illustrating the performance of maximum pairwise and triplewise likelihood estimators, and a table summarizing further simulations showing how the efficiency of triplewise likelihood estimators changes with the number of locations.

APPENDIX A

Triplewise marginal distribution for the Brown–Resnick process

Recall the definition of the Brown–Resnick process in Section 2·1. For compactness we write $z_1 = z(x_1)$, $W_1 = W(x_1)$, $\varepsilon_1 = \varepsilon(x_1)$, $\gamma_1 = \gamma(x_1)$, $\gamma_{1;2} = \gamma(x_1 - x_2)$, etc. Since $\varepsilon(0) = 0$ almost surely, it is easy to see that $c_{i;i} = \mathrm{var}(\varepsilon_i) = 2\gamma_i$ and that $c_{i;j} = \mathrm{cov}(\varepsilon_i, \varepsilon_j) = \gamma_i + \gamma_j - \gamma_{i;j}$, for $i, j = 1, 2, 3$. Then $W_1/z_1 > W_2/z_2$ is equivalent to $\log W_1 - \log z_1 > \log W_2 - \log z_2$, and hence to $\varepsilon_1 - \gamma_1 - \log z_1 > \varepsilon_2 - \gamma_2 - \log z_2$ and thus to $\varepsilon_1 > \varepsilon_2 + a$, where $a = \gamma_1 - \gamma_2 + \log(z_1/z_2)$. Similarly, $W_1/z_1 > W_3/z_3$

if and only if $\varepsilon_1 > \varepsilon_3 + b$, where $b = \gamma_1 - \gamma_3 + \log(z_1/z_3)$. Let us write

$$V(z_1, z_2, z_3) = E\left\{\max\left(\frac{W_1}{z_1}, \frac{W_2}{z_2}, \frac{W_3}{z_3}\right)\right\} = I_1/z_1 + I_2/z_2 + I_3/z_3,$$

say, where

$$I_1 = E\left\{W_1 I\left(\frac{W_1}{z_1} > \frac{W_2}{z_2}, \frac{W_1}{z_1} > \frac{W_3}{z_3}\right)\right\}$$

and so forth. Now provided $x_1 \neq 0$ and with $w_i = \exp(\varepsilon_i - \gamma_i)$ and using ϕ to denote Gaussian densities, possibly multivariate, we have

$$\begin{aligned} I_1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(\varepsilon_1 - \gamma_1) I\left(\frac{w_1}{z_1} > \frac{w_2}{z_2}, \frac{w_1}{z_1} > \frac{w_3}{z_3}\right)\phi(\varepsilon_1, \varepsilon_2, \varepsilon_3) d\varepsilon_1 d\varepsilon_2 d\varepsilon_3 \\ &= \int_{-\infty}^{\infty} \exp(\varepsilon_1 - \gamma_1)\phi(\varepsilon_1) \int_{-\infty}^{\varepsilon_1 - a}\int_{-\infty}^{\varepsilon_1 - b} \phi(\varepsilon_2, \varepsilon_3 \mid \varepsilon_1) d\varepsilon_3 d\varepsilon_2 d\varepsilon_1 \qquad \text{(A1)} \\ &= \int_{-\infty}^{\infty} \frac{1}{(4\pi\gamma_1)^{1/2}} \exp\{-(\varepsilon_1 - 2\gamma_1)^2/(4\gamma_1)\} K(\varepsilon_1) d\varepsilon_1, \end{aligned}$$

say, where $K(\varepsilon_1)$ denotes the inner double integral in (A1), and thus

$$I_1 = \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}} \exp(-\xi^2/2) K\left\{(2\gamma_1)^{1/2}\xi + 2\gamma_1\right\} d\xi = E_\xi\left[K\left\{(2\gamma_1)^{1/2}\xi + 2\gamma_1\right\}\right],$$

where $\xi \sim \mathcal{N}(0,1)$. As the joint distribution of $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ is trivariate normal with zero mean and covariance matrix $C = (c_{i;j})$, the properties of the multivariate normal distribution imply that the joint density of $\varepsilon_2, \varepsilon_3$ conditional on ε_1 is $\mathcal{N}_2(\mu_{2,3|1}, \Sigma_{2,3|1})$, where

$$\mu_{2,3|1} = \begin{pmatrix} c_{1;2}\varepsilon_1/c_{1;1} \\ c_{1;3}\varepsilon_1/c_{1;1} \end{pmatrix}, \quad \Sigma_{2,3|1} = \begin{pmatrix} c_{2;2} - c_{1;2}^2/c_{1;1} & c_{2;3} - c_{1;2}c_{1;3}/c_{1;1} \\ c_{2;3} - c_{1;2}c_{1;3}/c_{1;1} & c_{3;3} - c_{1;3}^2/c_{1;1} \end{pmatrix}.$$

Therefore, conditional on ξ, we have

$$\begin{aligned} K\left\{(2\gamma_1)^{1/2}\xi + 2\gamma_1\right\} &= \int_{-\infty}^{(2\gamma_1)^{1/2}\xi + 2\gamma_1 - a}\int_{-\infty}^{(2\gamma_1)^{1/2}\xi + 2\gamma_1 - b} \phi\left\{\varepsilon_2, \varepsilon_3 \mid \varepsilon_1 = (2\gamma_1)^{1/2}\xi + 2\gamma_1\right\} d\varepsilon_3 d\varepsilon_2 \\ &= \mathrm{pr}\big[Z_1 \le (2\gamma_1)^{1/2}\xi + 2\gamma_1 - a - c_{1;2}\{(2\gamma_1)^{1/2}\xi + 2\gamma_1\}/c_{1;1}, \\ &\qquad Z_2 \le (2\gamma_1)^{1/2}\xi + 2\gamma_1 - b - c_{1;3}\{(2\gamma_1)^{1/2}\xi + 2\gamma_1\}/c_{1;1} \,\big|\, \xi\big], \end{aligned}$$

where Z_1 and Z_2 form a bivariate normal random variable with zero mean, and covariance matrix $\Sigma_{2,3|1}$. Integrating out over ξ, we get

$$\begin{aligned} E_\xi\left[K\left\{(2\gamma_1)^{1/2}\xi + 2\gamma_1\right\}\right] &= \mathrm{pr}\big\{Z_1 + \xi(-\gamma_1 + \gamma_2 - \gamma_{1;2})/(2\gamma_1)^{1/2} \le -a + \gamma_1 - \gamma_2 + \gamma_{1;2}, \\ &\qquad Z_2 + \xi(-\gamma_1 + \gamma_3 - \gamma_{1;3})/(2\gamma_1)^{1/2} \le -b + \gamma_1 - \gamma_3 + \gamma_{1;3}\big\} \\ &= \mathrm{pr}\left(Y_1 \le -a - \gamma_1 - \gamma_2 + \gamma_{1;2}, Y_2 \le -b - \gamma_1 - \gamma_3 + \gamma_{1;3}\right) \\ &= \mathrm{pr}\left\{Y_1 \le \gamma_{1;2} - \log(z_1/z_2), Y_2 \le \gamma_{1;3} - \log(z_1/z_3)\right\}, \qquad \text{(A2)} \end{aligned}$$

where (Y_1, Y_2) is a bivariate normal vector with zero mean and covariance matrix

$$\Omega_1 = \begin{pmatrix} 2\gamma_{1;2} & \gamma_{1;2} + \gamma_{1;3} - \gamma_{2;3} \\ \gamma_{1;2} + \gamma_{1;3} - \gamma_{2;3} & 2\gamma_{1;3} \end{pmatrix}.$$

The right-hand side of equation (A2) yields

$$I_1 = \Phi_2\{\eta(z_1, z_2), \eta(z_1, z_3); R_1\}, \qquad \text{(A3)}$$

where $\eta(z_i, z_j) = (2\gamma_{i;j})^{1/2}/2 - \log(z_i/z_j)/(2\gamma_{i;j})^{1/2}$, $R_1 = (\gamma_{1;2} + \gamma_{1;3} - \gamma_{2;3})/\{2(\gamma_{1;2}\gamma_{1;3})^{1/2}\}$. The case $x_1 = 0$ can be treated separately and turns out to give the same result. By interchanging the labels, I_2 and I_3 are derived similarly.

Expression (A3) and its counterparts hold if $|R_k| \neq 1$ $(k = 1, 2, 3)$, which is always true when $\alpha < 2$. However, if $\alpha = 2$ and the sites x_1, x_2 and x_3 form a degenerate simplex in $\mathbb{R}^d$, then $R_k = \pm 1$ $(k = 1, 2, 3)$. If $d = 1$, the simplex is always degenerate. In dimension $d \geq 2$, certain configurations of points may also be problematic, for example if the sites x_1, x_2, x_3 lie on a linear subset of $\mathbb{R}^2$. This will lead to problems when the sites of $\mathcal{D}$ form a grid.

APPENDIX B

Higher order margins of the Brown–Resnick process

For $p > 3$, the exponent measure may be written as $V(z_1, \ldots, z_p) = I_1/z_1 + \cdots + I_p/z_p$, where $I_k = E\left\{W_k I\left(W_k/z_k \geq W_s/z_s, s = 1, \ldots, p\right)\right\}$. Moreover, $I_k = E_\xi\left[K_k\left\{(2\gamma_k)^{1/2}\xi + 2\gamma_k\right\}\right]$, where

$$\xi \sim \mathcal{N}(0, 1), \quad K_k(x) = \int_{-\infty}^{x - a_{-k}} \phi(\varepsilon_{-k} \mid \varepsilon_k = x) d\varepsilon_{-k} \quad (k = 1, \ldots, p),$$

with ε_{-k} representing the $(p-1)$-dimensional vector $(\varepsilon_1, \ldots, \varepsilon_p)$ with the kth component removed, and where a_{-k} is the $(p-1)$-dimensional vector whose sth component equals $\gamma_k - \gamma_s + \log(z_k/z_s)$ $(s = 1, \ldots, p; s \neq k)$. The computations are the same as those above, and equation (A2) becomes

$$I_k = \mathrm{pr}\{Y_s \leq \gamma_{k;s} - \log(z_k/z_s); s = 1, \ldots, p, j \neq k\},$$

where the $(p-1)$-dimensional vector of Y_ss has a joint Gaussian distribution with $E(Y_s) = 0$, $\mathrm{var}(Y_s) = 2\gamma_{k;s}$ and $\mathrm{cov}(Y_s, Y_t) = \gamma_{k;s} + \gamma_{k;t} - \gamma_{s;t}$, from which we get $I_k = \Phi_{p-1}(\eta_k; R_k)$, where η_k and R_k are defined in Section 2·3. Thus $V(z_1, \ldots, z_p) = \sum_{k=1}^{p} z_k^{-1}\Phi_{p-1}(\eta_k; R_k)$.

This result holds if the correlation matrices R_k are invertible, which is always true when $\alpha < 2$. However, in the special case $\alpha = 2$, i.e., the Smith model, if the sites $x_1, \ldots, x_p$ form a degenerate simplex in $\mathbb{R}^d$, then the determinants of the correlation matrices equal zero and the result fails. If $p > d + 1$, the simplex is always degenerate (Genton *et al.*, 2011). Moreover, if $\alpha \approx 2$, so that the Brown–Resnick process is rather smooth, and especially for large p, the correlation matrices could be numerically singular.

REFERENCES

Brown, B. M. and Resnick, S. (1977) Extreme values of independent stochastic processes. *Journal of Applied Probability* **14**, 732–739.

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

Davison, A. C. and Gholamrezaee, M. M. (2012) Geostatistics of extremes. *Proceedings of the Royal Society of London Series A: Mathematical and Physical Sciences* **468**, 581–608.

Davison, A. C., Padoan, S. A. and Ribatet, M. (2012) Statistical modelling of spatial extremes (with Discussion). *Statistical Science* **27**, 161–186.

Genton, M. G., Ma, Y. and Sang, H. (2011) On the likelihood function of Gaussian max-stable processes. *Biometrika* **98**, 481–488.

Genz, A. (1992) Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics* **1**, 141–150.

Genz, A. and Bretz, F. (2000) Methods for the computation of multivariate t-probabilities. *Computing Sciences and Statistics* **25**, 400–405.

Genz, A. and Bretz, F. (2002) Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics* **11**(4), 950–971.

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

Huser, R. and Davison, A. C. (2012) Space-time modelling of extreme events. *arXiv:1201.3245* Submitted.

Hüsler, J. and Reiss, R. (1989) Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters* **7**(4), 283–286.

Kabluchko, Z., Schlather, M. and de Haan, L. (2009) Stationary max-stable fields associated to negative definite functions. *Annals of Probability* **37**, 2042–2065.
Lindsay, B. G. (1988) Composite likelihood methods. *Contemporary Mathematics* **80**, 221–239.
Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association* **105**, 263–277.
Ribatet, M. (2012) *SpatialExtremes: Modelling Spatial Extremes*. R package version 1.8-6.
Varin, C., Reid, N. and Firth, D. (2011) An overview of composite likelihood methods. *Statistica Sinica* **21**, 5–42.

[*Received July* 2012. *Revised December* 2012]