Space-time modelling of extreme events

R. Huser and A. C. Davison

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

Abstract. Max-stable processes are the natural analogues of the generalized extreme-value distribution when modelling extreme events in space and time. Under suitable conditions, these processes are asymptotically justified models for maxima of independent replications of random fields, and they are also suitable for the modelling of extreme measurements over high thresholds. This paper shows how a pairwise censored likelihood can be used for consistent estimation of the extremes of space-time data under mild mixing conditions, and illustrates this by fitting an extension of a model of Schlather (2002) to hourly rainfall data. A block bootstrap procedure is used for uncertainty assessment. Estimator efficiency is considered and the choice of pairs to be included in the pairwise likelihood is discussed. The proposed model fits the data better than some natural competitors.

Keywords: Composite likelihood; Extremal coefficient; Max-stable process; Rainfall data; Random set; Threshold-based inference.

1. Introduction

Under suitable conditions, max-stable processes are asymptotically justified models for maxima of independent replications of random fields. Since they extend the generalized extreme-value distribution of univariate extreme value theory to the spatial setting, they are natural models for spatial extremes. de Haan's (1984) spectral representation of such processes implies that there are infinitely many max-stable processes, and the practical challenge is to build flexible but parsimonious models that can capture a wide range of extremal dependencies. Parsimony is important since extremal data are often scarce, but flexibility is also crucial since a poor fit might lead to mis-estimation of risk. Several models for max-stable processes have been suggested; for example, Smith (1990) proposes a max-stable model with deterministic storm shapes, and Schlather (2002) proposes one based on a truncated Gaussian process. Other models include the Brown–Resnick processes (see Kabluchko and Schlather, 2010), and a Brownian motion model proposed by Buishand et al. (2008), which has the drawback of not being invariant with respect to coordinate axes. Wadsworth and Tawn (2012) generalize these models to hybrids able to capture both asymptotic dependence and asymptotic independence. Reich and Shaby (2012) propose a finite-dimensional construction of max-stable processes that can be fitted in the Bayesian framework through Markov chain Monte Carlo simulation. Other modelling approaches for spatial extremes, based on copulas or on latent processes, are presented by Davison et al. (2012).

The full likelihood cannot be obtained analytically for most max-stable processes (but see Genton et al., 2011; Huser and Davison, 2013). However, since the bivariate marginal densities can usually be derived, inference can be based on a pairwise composite likelihood. Much has been written on pseudo-, quasi- or composite-likelihood; see Varin (2008) and

Address for correspondence: A. C. Davison, Ecole Polytechnique Fédérale de Lausanne, EPFL-FSB-MATHAA-STAT, Station 8, 1015 Lausanne, Switzerland.
E-mail: Anthony.Davison@epfl.ch

Varin et al. (2011) for general overviews, and Padoan et al. (2010), Blanchet and Davison (2011), Davison et al. (2012) and Davison and Gholamrezaee (2012) for applications to spatial extremes. Such likelihoods are robust to misspecification of higher-order marginal distributions and have nice theoretical properties, but so far have been applied only to componentwise maxima. An important extension, which improves inference by incorporating more information, is to perform pairwise threshold-based inference for max-stable processes, analogous to the use of the generalized Pareto distribution for univariate peaks over threshold modelling. This will be addressed below.

In Section 2, we tie together geostatistics and statistics of extremes to construct asymptotically valid space-time models for extremes. The spatio-temporal aspect of this modelling is novel, though related work includes Davis and Mikosch (2008), Davis et al. (2013a) and Davis et al. (2013b). Section 3 is focused on inference and describes the methods based on pairwise likelihood, while Section 4 addresses the loss in efficiency of the estimation procedure and gives some suggestions about the choice of pairs to be included in the pairwise likelihood, based on a simulation study. Section 5 describes its application to space-time modelling of rainfall. Brief concluding discussion is given in Section 6.

2. Threshold modelling for extremes

2.1. Marginal modelling

The classical theory of extreme values addresses the large-sample fluctuations of the maximum M_n of a sequence of independent and identically distributed random variables $X_1, \ldots, X_n$ whose distribution F has upper terminal $x_F = \sup\{x : F(x) < 1\}$, possibly infinite. If sequences $\{a_n\} > 0$ and $\{b_n\} \subset \mathbb{R}$ exist such that the limiting distribution G of $(M_n - b_n)/a_n$ is non-degenerate, then it must necessarily be the generalized extreme-value (GEV) distribution, $G(x) = \exp[-\{1+\xi(x-\eta)/\tau\}^{-1/\xi}]$, defined on the set $\{y : 1+\xi(x-\eta)/\tau > 0\}$, with $\eta \in \mathbb{R}$, $\tau > 0$, $\xi \in \mathbb{R}$ and with the value for $\xi = 0$ interpreted as the limit when $\xi \to 0$.

A complementary result describes the stochastic behaviour of peaks over a high threshold u: if the limit distribution G holds for maxima, then as $u \to x_F$ the conditional distribution of $X - u$, suitably rescaled and conditional on $X > u$, converges to the generalized Pareto distribution, $\mathrm{GPD}(\sigma, \xi)$ (Davison and Smith, 1990). Its distribution function is

$$H(y) = 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-1/\xi}, \qquad y > 0,$$

where the scale parameter is linked to that of the GEV distribution by $\sigma = \tau + \xi(u - \eta)$, and the shape parameter ξ is the same as for G. A closely related characterization relies on point processes. If the limiting result holds for maxima, then as $n \to \infty$ the two-dimensional point process $\{i/(n+1), (X_i - b_n)/a_n\}_{i=1}^n$ converges to a non-homogeneous Poisson process on regions of the form $[t_1, t_2] \times [u, \infty)$, with $0 < t_1 < t_2 < 1$ (Leadbetter, 1991; Smith, 1989). In practice, the data often exhibit temporal dependence, and the aforementioned asymptotic results can be extended to stationary sequences with short-range dependence (Leadbetter et al., 1983). For more details about extreme-value statistics, see Coles (2001), Beirlant et al. (2004), Embrechts et al. (1997) or de Haan and Ferreira (2006).

As the upper tail may be well approximated by a GPD, the distribution F of X can be consistently estimated by

$$\tilde{F}(x) = \begin{cases} \hat{F}(x), & x \le u; \\ 1 - \hat{\zeta}_u \left\{1 + \hat{\xi}(x-u)/\hat{\sigma}\right\}^{-1/\hat{\xi}}, & x > u, \end{cases}$$

where $\hat{F}(x)$ is the empirical distribution function of the sample $X_1, \ldots, X_n$, $\hat{\zeta}_u$ is the estimated probability of exceeding the threshold u and $\hat{\xi}$ and $\hat{\sigma}$ are estimates of ξ and σ. The

transformation $t(x) = -1/\log \tilde{F}(x)$ therefore approximately standardizes the observations to have the unit Fréchet distribution $\exp(-1/x)$, for $x > 0$. The choice of threshold u is typically made informally using diagnostic plots, sometimes aided by theoretical arguments, though it can be vexing for nonstationary data; see Scarrott and MacDonald (2012) and Northrop and Jonathan (2011) and its discussion.

Joint modelling of extremes is crucial for a realistic assessment of risk, and the next section describes models for spatial or spatio-temporal extremes, with margins previously transformed to the unit Fréchet scale.

2.2. Max-stable processes

Suppose that $\{Y_i(x) : x \in \mathcal{X} \subset \mathbb{R}^d\}$, $i = 1, 2, \ldots$, are independent replicates of a random process, and that there exist sequences of continuous functions $\{a_n(x)\} > 0$ and $\{b_n(x)\}$ such that as $n \to \infty$, the rescaled process of maxima, $a_n^{-1}\{\max(Y_1, \ldots, Y_n) - b_n\}$, converges in distribution to a random process Z whose univariate margins are all non-degenerate. Then it can be shown that the class of possible limiting processes coincides with the class of max-stable processes with non-degenerate margins, i.e., those for which the maximum of n independent replicates of Z can be rescaled to have the same distribution as Z (de Haan and Ferreira, 2006, §9.2). The marginal distributions of Z are GEV and can be transformed to a standard form. Below we take $\Pr\{Z(x) \le z\} = \exp(-1/z)$, $z > 0$, so that these margins are unit Fréchet; such a process is called simple max-stable. Under mild technical conditions, one may represent a simple max-stable process Z as (de Haan, 1984; Schlather, 2002; de Haan and Ferreira, 2006, §9.4)

$$Z(x) = \sup_{i \ge 1} W_i(x)/P_i, \tag{1}$$

where the P_i's are the points of a unit rate Poisson process on $\mathbb{R}_+$ and the W_i's are independent replicates of a non-negative random process $W(x)$ with unit mean at each x. A common interpretation of Z is as a pointwise maximum of random storms W_i with corresponding intensities P_i^{-1}. Due to the characterization (1), no finite parametrization exists for such processes.

It follows from (1) that the joint distribution of the process Z at N distinct locations is

$$\Pr\{Z(x_1) \le z_1, \ldots, Z(x_N) \le z_N\} = \exp\left(-E\left[\max_{i=1,\ldots,N}\left\{\frac{W(x_i)}{z_i}\right\}\right]\right) = \exp\{-V_N(z_1, \ldots, z_N)\}, \tag{2}$$

where the exponent measure $V_N(\cdot)$, which summarises the extremal dependence structure, is homogeneous of order -1 and satisfies $V_N(\infty, \ldots, z, \ldots, \infty) = 1/z$ for any permutation of the N arguments. When $z_i = z$ for all $i = 1, \ldots, N$, we obtain $\Pr\{Z(x_1) \le z, \ldots, Z(x_N) \le z\} = \exp\{-V_N(1, \ldots, 1)/z\} = \{\exp(-1/z)\}^{\theta_N}$. The so-called extremal coefficient $\theta_N = V_N(1, \ldots, 1)$ can be seen as a summary of extremal dependence, and has two bounding cases: complete dependence, $\theta_N = 1$, and independence, $\theta_N = N$.

Different choices for $W(x)$ yield more or less flexible models for spatial maxima. For the space-time modelling of extreme rainfall (see §5), a model originally due to Schlather (2002) seems suitable. It comprises a truncated Gaussian random process for $W(x)$, so that storm shapes are stochastic, and includes a compact random set, which allows independent extremes. In the present context the points $x \in \mathcal{X}$ have coordinates in space $\mathcal{S} \subset \mathbb{R}^2$ and time $\mathcal{T} \subset \mathbb{R}$, that is, $x = (s, t) \in \mathcal{X} = \mathcal{S} \times \mathcal{T}$, where $\mathcal{S}$ and $\mathcal{T}$ are compact, and the model is defined by taking

$$W_i(s, t) \propto \max\{0, \varepsilon_i(s, t)\} I_{\mathcal{A}_i}\{(s, t) - X_i\}, \qquad (s, t) \in \mathcal{S} \times \mathcal{T}. \tag{3}$$

Here the $\varepsilon_i(s, t)$ are independent replicates of a Gaussian random field with space-time correlation function ρ, $I_{\mathcal{A}}$ is the indicator function of a compact random set $\mathcal{A} \subset \mathcal{S} \times \mathcal{T}$, the

$\mathcal{A}_i$ are independent replicates of $\mathcal{A}$, and the X_i are points of a unit rate Poisson process on $\mathcal{S} \times \mathcal{T}$, independent of the ε_i. The proportionality constant in (3) is chosen to satisfy $\mathrm{E}\{W_i(s,t)\} = 1$.

A common feature of the max-stable models thus far proposed is that the exponent measure V_N is known for $N = 2$. Genton et al. (2011) provide a closed-form expression of the likelihood function for the Smith max-stable model indexed by $\mathbb{R}^d$ when $N \leq d+1$, and Huser and Davison (2013) have generalized this to the Brown–Resnick process, but only the bivariate margins are known for other models. Moreover, the number of terms in the likelihood explodes as N increases, and this has led to the use of pairwise likelihood; see Section 3 below. The bivariate exponent measure for the model (3) can be expressed in the stationary case as (Schlather, 2002)

$$V_2(z_1, z_2) = \left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left\{1 - \frac{\alpha(h_s, h_t)}{2}\left(1 - \left[1 - 2\frac{\{\rho(h_s, h_t) + 1\}z_1 z_2}{(z_1 + z_2)^2}\right]^{1/2}\right)\right\}, \quad (4)$$

where $h_s = s_1 - s_2$ is the spatial lag, $h_t = t_1 - t_2$ is the temporal lag, $\alpha(h_s, h_t) = \mathrm{E}[|\mathcal{A} \cap \{(h_s, h_t) + \mathcal{A}\}|]/\mathrm{E}(|\mathcal{A}|)$ and $|\cdot|$ is used to denote the volume of a set. Hence the pairwise spatio-temporal extremal coefficients are

$$\theta_2(h_s, h_t) = V_2(1,1) = 2 - \alpha(h_s, h_t)\left\{1 - \sqrt{\frac{1 - \rho(h_s, h_t)}{2}}\right\}. \quad (5)$$

Usually in practice, $\rho(h_s, h_t) \to 0$ as $h_s \to \infty$ or $h_t \to \infty$. Therefore, for the Schlather model without a random set, which corresponds to setting $\mathcal{A} \equiv \mathcal{S} \times \mathcal{T}$ and $\alpha(h_s, h_t) \equiv 1$, $\theta_2(h_s, h_t)$ would be bounded above by 1.707 as $h_s \to \infty$ or $h_t \to \infty$: complete independence could not be captured by this model, even at very large distances. Since $\mathcal{A}$ is chosen to be compact, we can choose $\mathcal{A}$ so that $\alpha(h_s, h_t) \to 0$ and thus $\theta_2(h_s, h_t) \to 2$ as $h_s \to \infty$ or $h_t \to \infty$ for any spatio-temporal correlation function $\rho(h_s, h_t)$. The process (3) is built from random sets each with a truncated Gaussian process inside, so the short-range dependence is largely determined by the correlation function $\rho(h_s, h_t)$, while the longer-range dependence is regulated by the geometry of the random set $\mathcal{A}$. We describe several other models in §5.5, but (3) gives the best fit to our data.

In the geostatistics literature, stationary isotropic correlation functions have been used extensively. But in our general framework, the correlation function need be neither isotropic nor stationary, and could therefore depend on the space-time locations $x_1 = (s_1, t_1)$ and $x_2 = (s_2, t_2)$ rather than on their spatio-temporal distances $\|h_s\|$, $|h_t|$ and lag vectors h_s, h_t. We would then have non-stationary extremal coefficients.

In the next section, we show how to make the link from an asymptotic distribution for maxima to a model for multivariate threshold exceedances.

2.3. Censored threshold-based likelihood

The convergence of block maxima to a max-stable process implies that all finite-dimensional distributions converge to a max-stable distribution, i.e., to a multivariate extreme value distribution. As explained in Section 2.2, the joint distribution of properly scaled block maxima at N sites in $\mathcal{X}$ is well approximated by $\exp\{-V_N(z_1, \ldots, z_N)\}$, where the exponent measure V_N stems from the underlying spatial structure of the max-stable process. As shown by Beirlant et al. (2004, page 276), the joint tail of individual observations $Y(x_1), \ldots, Y(x_N)$ at the sites $x_1, \ldots, x_N$ is asymptotically equivalent to that of maxima, so when all of the $z_1, \ldots, z_N$ are large the joint distribution of $Y(x_1), \ldots, Y(x_N)$ can be approximated by $\exp\{-V_N(z_1, \ldots, z_N)\}$. Hence, the model for maxima in equation (2) also provides a model for extremes of individual observations. This argument has led several authors (see, e.g.,

Smith et al., 1997; Wadsworth and Tawn, 2012; Ledford and Tawn, 1996; Bortot et al., 2000; Coles, 2001, p.155) to consider censored modelling for bivariate extremes. Let u be a sufficiently high threshold, chosen so that $\exp\{-V_2(z_1, z_2)\}$ is a valid model for $\{Y(x_1), Y(x_2)\}$ when $z_1, z_2 > u$. The likelihood contribution $p_u(z_1, z_2)$ of a pair (z_1, z_2) is then taken to be

$$p_u(z_1, z_2) = \begin{cases} \frac{\partial^2}{\partial z_1 \partial z_2} \exp\{-V_2(z_1, z_2)\}, & \min(z_1, z_2) > u; \\ \frac{\partial}{\partial z_1} \exp\{-V_2(z_1, u)\}, & z_1 > u, z_2 \leq u; \\ \frac{\partial}{\partial z_2} \exp\{-V_2(u, z_2)\}, & z_1 \leq u, z_2 > u; \\ \exp\{-V_2(u, u)\}, & \max(z_1, z_2) \leq u, \end{cases} \tag{6}$$

where V_2 is the function appearing in (4), for which the density exists. Different marginal thresholds can be used (Bortot et al., 2000) and the approach generalizes to higher dimensions, though the probability that an observed N-uplet falls into the "upper right quadrant" decays geometrically with N, leading to potential inference problems in practice.

Alternative approaches recently proposed are based on generalized Pareto processes (Ferreira and de Haan, 2012; Aulbach and Falk, 2012) and multivariate generalized Pareto distributions (Rootzén and Tajvidi, 2006; Buishand et al., 2008; Beirlant et al., 2004, p.277). So far as we are aware, the former have not yet been used for inference, and unpublished simulations suggest that the approach to inference suggested for the latter may yield badly biased estimators; our approach has a larger variance but a smaller mean squared error.

In the next section, we will show that censored threshold-based pairwise likelihoods provide consistent inference.

3. Inference

3.1. Pairwise likelihood approach

As the full likelihood is not generally known for max-stable models, classical frequentist or Bayesian inference appears impossible, and we adopt an alternative approach based on composite likelihood. An analogous approach in the Bayesian framework using a pseudo-posterior distribution based on a pairwise likelihood has been developed by Ribatet et al. (2012), and Wadsworth and Tawn (2013) describe an approach to inference based on exceedances of a particular class of max-stable processes. Maximum composite likelihood estimators typically have similar asymptotic properties to the usual maximum likelihood estimator; often they are asymptotically normal and strongly consistent.

Assume that the spatio-temporal process $Z(s,t)$, $(s,t) \in \mathcal{X} = \mathcal{S} \times \mathcal{T}$, is observed at S monitoring stations and at times $1, \ldots, T$, that is at $N = ST$ locations in $\mathcal{X}$. For simplicity of notation we let $Z_{s,t}$ denote the value recorded at the sth station at time t, and consider the censored threshold-based pairwise log likelihood

$$l_{\mathcal{K}}(\psi) = \sum_{t=1}^{T} \sum_{h_t \in \mathcal{K}_t} \sum_{s_1=1}^{S} \sum_{s_2=1}^{S} (1 - I\{s_1 \geq s_2 \text{ and } h_t = 0\}) \log p_u\left(Z_{s_1,t}, Z_{s_2,t+h_t}; \psi\right), \tag{7}$$

with corresponding maximum pairwise likelihood estimator

$$\hat{\psi}_{p,\mathcal{K}} = \arg\max_{\psi \in \Psi} l_{\mathcal{K}}(\psi), \tag{8}$$

where $\mathcal{K}_t = \{h_t \in \mathcal{K} : h_t \leq T - t\}$ and $\mathcal{K} \subset \mathbb{N} \cup \{0\}$ is a finite collection of time lags, with p_u given by equation (6), the exponent measure V being given for example by (4) and where $I\{\cdot\}$ is the indicator function. If $\mathcal{K} = \{0, 1, \ldots, K\}$ for $K < \infty$, this pairwise log likelihood corresponds to summing all space-time pairwise contributions up to a maximum time lag K. If $K = T - 1$, it reduces to the full pairwise likelihood. However, the associated

computational burden could be reduced and statistical efficiency gained by taking a different subset $\mathcal{K}$. For example, we could take $\{\lfloor a^{k-1} \rfloor : k = 1, \ldots, K\} \cup \{0\}$, $a > 1$. In particular, when $a = 2$, we include the pairs at lag $0, 1, 2, 4, 8, \ldots$. Another choice could be based on the Fibonacci sequence: $0, 1, 2, 3, 5, 8, 13, \ldots$. In Section 4, we will see that the choice of pairs is closely linked to the efficiency of $\hat{\psi}_{p,\mathcal{K}}$, so careful selection of them is essential. Bevilacqua et al. (2012) discuss related issues in the context of Gaussian random fields.

3.2. Asymptotics

Davison and Gholamrezaee (2012) and Padoan et al. (2010) use pairwise likelihood for inference on max-stable processes, assuming independence between distinct annual maxima. In the case of spatio-temporal extremes, the asymptotic normality of $\hat{\psi}_{p,\mathcal{K}}$ stems from a central limit theorem for stationary time series applied to the score $U(\psi) = \nabla l(\psi) = \sum_{t=1}^{T} U_t(\psi)$, where $U_t(\psi)$ is the derivative of the rightmost triple sums in equation (7) with respect to ψ. However, as the elements $U_t(\psi)$ are generally correlated over time t, we need an additional mixing condition in order for classical asymptotics to hold. A mild sufficient condition is that the process $Z(s,t)$ be temporally α-mixing, along with a condition on the rate at which the mixing coefficients $\alpha(n)$ must decay, ensuring that the correlation vanishes sufficiently fast at infinity. With this condition, two events become more and more independent as their time lag increases. In particular, all m-dependent processes are α-mixing.

We call a space-time process $Z(s,t)$, $(s,t) \in \mathcal{X} = \mathcal{S} \times \mathcal{T}$, temporally α-mixing with coefficients $\alpha(n)$ if for all $s \in \mathcal{S}$, for all sequences $t_n \subset \mathcal{T}$, the time series $\{Z(s,t_n) : n \in \mathbb{N}\}$ is α-mixing with coefficients $\alpha_s(n)$ and where $\sup_{s\in\mathcal{S}} \alpha_s(n) \leq \alpha(n) \to 0$ as $n \to \infty$. For the definition of an α-mixing time series, see Bradley (2007, Definition 1.6). We can then obtain the following theorem, whose proof, which relies on the theory of estimating equations, is given in Appendix A.

THEOREM 1. *Let $Z(s,t)$ be a stationary spatio-temporal max-stable process that is temporally α-mixing with coefficients $\alpha(n)$. Moreover, suppose that for all $\psi \in \Psi$, $\mathrm{E}[\{U_1(\psi)\}^2] < \infty$ and that for some $\delta > 0$, one has $\mathrm{E}(|U_1(\psi)|^{2+\delta}) < \infty$ and $\sum_{n\geq 1} |\alpha(n)|^{\delta/(2+\delta)} < \infty$. If ψ is identifiable from the bivariate densities, then*

$$T^{1/2} K(\psi)^{-1/2} J_1(\psi)(\hat{\psi}_{p,\mathcal{K}} - \psi) \to \mathcal{N}(0, I_p)$$

in distribution as $T \to \infty$, where

$$\begin{aligned} J_1(\psi) &= \mathrm{E}\{-\nabla_\psi U_1(\psi)\}, && (9)\\ K(\psi) &= T^{-1}\mathrm{var}\left\{\sum_{t=1}^{T} U_t(\psi)\right\} \\ &= \mathrm{E}\{U_1(\psi)U_1(\psi)^T\} + \sum_{t=1}^{T-1}\left(1 - \frac{t}{T}\right)\left[\mathrm{E}\left\{U_1(\psi)U_{t+1}(\psi)^T\right\} + \mathrm{E}\left\{U_{t+1}(\psi)U_t(\psi)^T\right\}\right] && (10)\\ &\to \mathrm{E}\{U_1(\psi)U_1(\psi)^T\} + \sum_{t=1}^{\infty}\left[\mathrm{E}\left\{U_1(\psi)U_{t+1}(\psi)^T\right\} + \mathrm{E}\left\{U_{t+1}(\psi)U_t(\psi)^T\right\}\right] < \infty, \qquad T \to \infty. \end{aligned}$$

This result shows that the standard asymptotic normality result for composite likelihoods (Hjort and Varin, 2008; Lindsay, 1988; Godambe and Heyde, 1987; Varin, 2008; Varin and Vidoni, 2005; Cox and Reid, 2004; Padoan et al., 2010) still holds under mild conditions for moderately temporally dependent processes. The asymptotic variance is of sandwich form, as is standard for misspecified models.

If the process $Z(s,t)$ were instead assumed to be Gaussian, and hence not max-stable, and if the pairwise likelihood were defined in terms of the marginal bivariate normal densities, then the moment conditions of the theorem, that $\mathrm{E}[\{U_1(\psi)\}^2], \mathrm{E}[|U_1(\psi)|^{2+\delta}] < \infty$, would be automatically satisfied for all $\delta > 0$, and thus the mixing condition would reduce to $\sum_{n\geq 1} |\alpha(n)|^{1-\epsilon} < \infty$, for some $\epsilon > 0$. Similar results were obtained by Davis and Yau (2011), who establish the asymptotic normality and the strong consistency of the maximum consecutive pairwise likelihood estimator for ARMA models, under a condition on the autocorrelation function, and treat certain long-memory models.

3.3. Variance estimation

Variance estimation for $\hat{\psi}_{p,\mathcal{K}}$ is difficult owing to the complicated form of the sandwich matrices in equations (9) and (10). The pairwise log likelihood is formed by summing the pairwise contributions for the time lags in the set $\mathcal{K}$ and across all S stations, so a single evaluation of the pairwise log likelihood requires $\mathrm{O}(T|\mathcal{K}|S^2)$ operations, and the computation of (10) is yet more intensive.

The temporal dependence of the data suggests that block bootstrap or jackknife methods can be used. In our application we apply a block bootstrap, treating rainfall data from different summers as independent. We resample the summers with replacement and use replicates of $\hat{\psi}_{p,\mathcal{K}}$ to estimate its variability. Fortunately, the replicates can be computed in parallel. The bootstrap was originally developed for independent data, and its consistency is discussed in Davison and Hinkley (1997, §2.6), for example. Dependence may be expected between observations within the same block, but when the blocks (in our case the summers) may be treated as independent, the bootstrap may be applied to the blocks.

4. Efficiency considerations

In Section 3, we introduced our maximum pairwise likelihood estimator for spatio-temporal extremes. Although it inherits its asymptotic properties from the usual maximum likelihood estimator, the natural question of statistical efficiency remains to be addressed. It turns out that the loss in efficiency is closely related to the pairs that are included in the pairwise likelihood, i.e., to the choice of $\mathcal{K}$. Adding pairs might simultaneously increase the variability $K(\psi)$ of the score and the amount of information $J(\psi)$, so it is unclear how the selection of pairs acts on the variance $T^{-1}J(\psi)^{-1}K(\psi)J(\psi)^{-1}$; the amount of information contained in a single pair might be insufficient to counteract the increase of variability due to including it, so the choice of the optimal subset of pairs is not obvious. However, one might suspect that for short-range dependent processes, pairs that are far apart in $\mathcal{S} \times \mathcal{T}$ are not as relevant for the estimation of a dependence parameter as are nearby ones. Varin et al. (2011), Varin and Vidoni (2005) and Varin and Czado (2010) have suggested the elimination of non-neighbouring pairs.

In the online supplementary material, we give the asymptotic relative efficiency of the maximum pairwise likelihood estimator for AR(1) and MA(1) models, for which maximum likelihood estimators can be computed, in order to gain a qualitative understanding of how composite likelihoods behave in more complex settings. As mentioned by Davis and Yau (2011), the efficiency for these simple models is maximized when pairs at lag 1 only are included. However, in practice, more lags must often be included in order that the model be identifiable, and our results shed some light on how to choose them. Complementary results on the efficiency of pairwise likelihood may be found in Cox and Reid (2004), Varin and Vidoni (2009), Hjort and Varin (2008) and Joe and Lee (2009).

Since these theoretical results do not apply directly to max-stable processes, we conducted a simulation study in a one-dimensional framework, using the statistical software R (R Core

Table 1. Mean squared errors (MSE) ($\times 1000$) for estimation of $\log\lambda$, the logarithm of the correlation range parameter, based on 1000 replications of the Schlather model, for different sets of pairs included in the pairwise likelihood, and μ known. There are three estimation procedures: margins known (MK); margins unknown, two-step approach (MU-2); margins unknown, one-step approach (MU-1).

	Number of time lags K								
	1	3		6			9		
Set $\mathcal{K}$	$\mathcal{K}^K_{a/b/c}$	$\mathcal{K}^K_{a/b}$	$\mathcal{K}^K_c$	$\mathcal{K}^K_a$	$\mathcal{K}^K_b$	$\mathcal{K}^K_c$	$\mathcal{K}^K_a$	$\mathcal{K}^K_b$	$\mathcal{K}^K_c$
MK	19	21	21	26	24	22	29	24	23
MU-2	42	45	46	54	50	48	59	50	49
MU-1	37	41	42	48	45	44	52	46	44

Team, 2012). We simulated the Schlather model (3) on the time axis, taking $\mathcal{X} = [0, 10000]$, with random sets of the form $\mathcal{A} = [0, D]$, where $D = 24\delta$ and $\delta \sim \text{beta}(10, 240/\mu - 10)$. The parameter μ corresponds to the mean length of the random set, which lies in the range $(0, 24)$, and we set $\mu = \text{E}(D) = 40/3 \simeq 13.3$. We chose an exponential correlation for the underlying Gaussian random field ε, with range parameter $\lambda = 4$; the effective range is 12. We then transformed the simulated processes to the Student t_5 scale, so that the exceedances above some high threshold u are approximately $\text{GPD}(\sigma, \xi)$ with shape parameter $\xi = 0.2$ (Beirlant et al., 2004, p. 59). The parameters were chosen to mimic rainfall data. The threshold u was set to the empirical 95% quantile, so that we have 500 exceedances contributing to the pairwise likelihood; in our application in Section 5, about 3000 exceedances were available at each station. A realization from this model is shown in the supplementary material. To assess the influence of the marginal estimation on the overall fit, we consider three estimation procedures: (i) estimation of the dependence parameters, treating the margins as known; (ii) a two-step approach, first estimating the marginal parameters by fitting the approximate GPD model, and then using the data thereby transformed to the unit Fréchet scale to estimate the dependence parameters; and (iii) a one-step approach, estimating marginal and dependence parameters simultaneously.

We first fixed the random set parameter μ to its true value, and estimated the logarithm of the range parameter, $\log\lambda$, with the threshold-based pairwise likelihood estimator (8), using the empirical 95% quantile threshold. We tested estimators corresponding to three sets of time lags: (a) $\mathcal{K}^K_a = \{1, \ldots, K\}$, for which all time lags are used up to some maximum time lag K; (b) $\mathcal{K}^K_b = \{b_k : k = 1, \ldots, K\}$, where b_k is based on the Fibonacci sequence; and (c) $\mathcal{K}^K_c = \{2^{k-1} : k = 1, \ldots, K\}$, where the lags increase geometrically. We considered $K = 1, 3, 6, 9$. Table 4 reports the mean squared errors (MSE) of these estimates based on 1000 realizations of the Schlather model.

With the random set parameter μ known, the MSE is minimized for $\mathcal{K} = \{1\}$, whatever the estimation procedure, corroborating the findings of Davis and Yau (2011) for AR(1) or MA(1) processes. Moreover, the MSE is systematically lower when $\mathcal{K}^6_b$ is used instead of $\mathcal{K}^6_a$ or when $\mathcal{K}^9_b$ is used instead of $\mathcal{K}^9_a$, even though the observations separated by more than 24 time units were independent. The same is true for $\mathcal{K}^K_c$. Thus, the inclusion of some distant, less dependent, pairs can improve inference significantly for fixed K, and marginal estimation does not influence the conclusions about the pairs that should be included in the pairwise likelihood.

A plot in the supplementary material shows how the bias and variance of the dependence parameter estimator decrease as the number of observations T increases, confirming the theoretical results of Section 3.

We then estimated the correlation range parameter $\lambda > 0$ and the mean duration $\mu \in (0, 24)$ of the random set simultaneously, using the maximum composite likelihood estimator

Table 2. Mean squared errors (MSE) for the joint estimation of the mean duration μ of the random set and the logarithm of the range parameter, $\log \lambda$, when different sets of pairs are included in the pairwise likelihood. The percentages of time when $\hat{\mu}$ reaches its upper bound is also reported. We considered three estimation procedures: margins known (MK); margins unknown, two-step approach (MU-2); and margins unknown, one-step approach (MU-1). This simulation is based on 1000 replications of the Schlather model.

		Number of time lags K								
		1	3		6			9		
	Set $\mathcal{K}$	$\mathcal{K}^K_{a/b/c}$	$\mathcal{K}^K_{a/b}$	$\mathcal{K}^K_c$	$\mathcal{K}^K_a$	$\mathcal{K}^K_b$	$\mathcal{K}^K_c$	$\mathcal{K}^K_a$	$\mathcal{K}^K_b$	$\mathcal{K}^K_c$
MK	1000×MSE for $\log \hat{\lambda}$	28	28	24	24	23	22	24	23	22
	MSE for $\hat{\mu}$	22.5	16.0	10.9	6.9	2.1	2.8	3.3	2.2	2.9
	Bound reached (%)	21	10	6	1	0	0	0	0	0
MU-2	1000×MSE for $\log \hat{\lambda}$	67	70	66	62	48	47	52	49	48
	MSE for $\hat{\mu}$	27.3	20.9	17.5	9.4	2.1	2.7	3.5	2.3	2.7
	Bound reached (%)	33	22	10	2	0	0	0	0	0
MU-1	1000×MSE for $\log \hat{\lambda}$	58	60	56	53	45	45	48	47	46
	MSE for $\hat{\mu}$	24.7	18.4	15.2	8.5	2.1	2.6	3.3	2.3	2.7
	Bound reached (%)	30	19	8	1	0	0	0	0	0

(8) and the threshold used above. The estimation of μ is more difficult, especially when only pairs at lag 1 are included, and in some cases its estimate reached the upper bound used in the optimisation. Table 4 shows that with $\mathcal{K} = \{1\}$ this happens on 21%, 33% and 30% of occasions for known margins, the two-step estimator and the one-step estimator, respectively. This could be anticipated since the pairs at lag 1 are uninformative for the estimation of μ. When further lags are added, the difference between use of the set $\mathcal{K}^K_a$ and the other sets becomes striking, especially for $K = 6$. The upper bound for μ is not attained for $\mathcal{K}^6_b$ and $\mathcal{K}^6_c$, and the MSEs are two or three times lower than those for $\mathcal{K}^6_a$; λ is also better estimated. The estimators including distant pairs in the composite likelihood outperform those that do not or that use only the most dependent pairs. The same phenomenon is observed when $K = 9$, but the difference is less striking than for $K = 6$. In fact, the pairs at lags less than 6 are probably ineffective for estimation of the duration of sets that in this case last on average 13.3 time units, and so $\mathcal{K}^K_b$ or $\mathcal{K}^K_c$ are better choices than $\mathcal{K}^K_a$. To sum up, if K is fixed and not too large compared to the "true" independence range, then both estimators that include pairs at higher lags, $\mathcal{K}^K_b$ or $\mathcal{K}^K_c$, behave appreciably better than that based on $\mathcal{K}^K_a$, which does not.

As might be expected, the one-step estimation procedure outperforms the two-step procedure overall, though by an amount that depends on the choice of pairs; the differences are rather small for $\mathcal{K}^K_b$ and $\mathcal{K}^K_c$. The one-step approach performs relatively better for smaller samples or higher thresholds, but the procedures are essentially equivalent for samples of 3000 exceedances as used in our application.

5. Data analysis

5.1. Description of the dataset

The dataset used for our application is composed of hourly rainfall measurements (mm) recorded from 1981 to 2007 at ten monitoring stations in western Switzerland. Figure 1 illustrates the location and topography of the area of study. All stations are located between the Alps and the Jura mountains, and their altitudes are similar. Only the periods from midnight on June 21st to 11 pm on September 20th were considered, these summers being treated as mutually independent. The entire dataset comprises 503988 measurements, with up to 59616 data points per station. The rainfall time series, shown in Figure 2, were

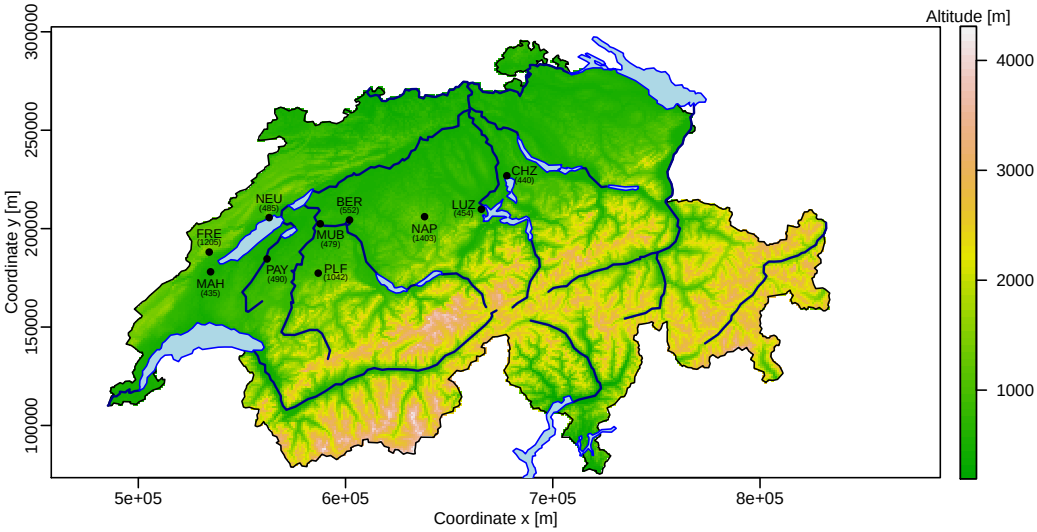

Figure 1. Topographic map of Switzerland, showing the location and altitude of the monitoring stations used. Their elevations are all close to 500 m above mean sea level (amsl), except for three stations (FRE, NAP, PLF) at about 1000 m amsl. The x and y axes use the Swiss coordinate system. The closest stations (FRE, MAH) are 10 km apart and the most distant ones (CHZ, MAH) are 151 km apart.

independently transformed to the unit Fréchet scale, following Section 2.1, with quantile-quantile plots showing satisfactory agreement between the empirical and fitted quantiles. The thresholds were the empirical 95% quantiles of each series. Due to the size of the dataset at each station, the margins were fitted with negligible variability. Below we focus on the modelling of extremal dependence, rather than on the marginal behaviour.

Figure 3 shows empirical space-time pairwise extremal coefficients for a subset of 5 stations at different time lags, based on a censored version of the naive Schlather–Tawn (2003) estimator. There is evidence of significant spatial and temporal dependence between the different series. Panel $(1,1)$ shows the temporal extremal coefficients at Bern-Zollikofen; it starts with the value 1 (complete dependence at lag 0), and tends smoothly to the value 2 (independence) as the time lag increases. This pattern repeats itself for the other stations. The off-diagonal panels represent extremal coefficients for the different pairs of stations, and hence display space-time interactions. For example, Panel $(1,4)$, in the 1st row and 4th column, displays the extremal coefficients between the rainfall time series at Luzern at time t and the rainfall time series at Bern-Zollikofen at time $t+h$, for $h = 0, 1, \ldots, 24$. Panel $(4,1)$ reverses the roles of the stations. The extremal coefficient functions differ for the panels, showing that the orientation of the stations matters. The extremal coefficient decreases at lags 1 or 2 when the stations are west-east oriented: during the summer months, western Switzerland is governed by dominant winds from the west or north-west, so that the clouds tend to discharge their rain first in the west. The same rainfall event could therefore be recorded by two distant monitoring stations at a lag of 1 or 2 hours, depending on their location and on the wind velocity. Consequently, extremal dependence might be higher at lag 1 or 2 than at lag 0. A model for the data should be able to capture such features.

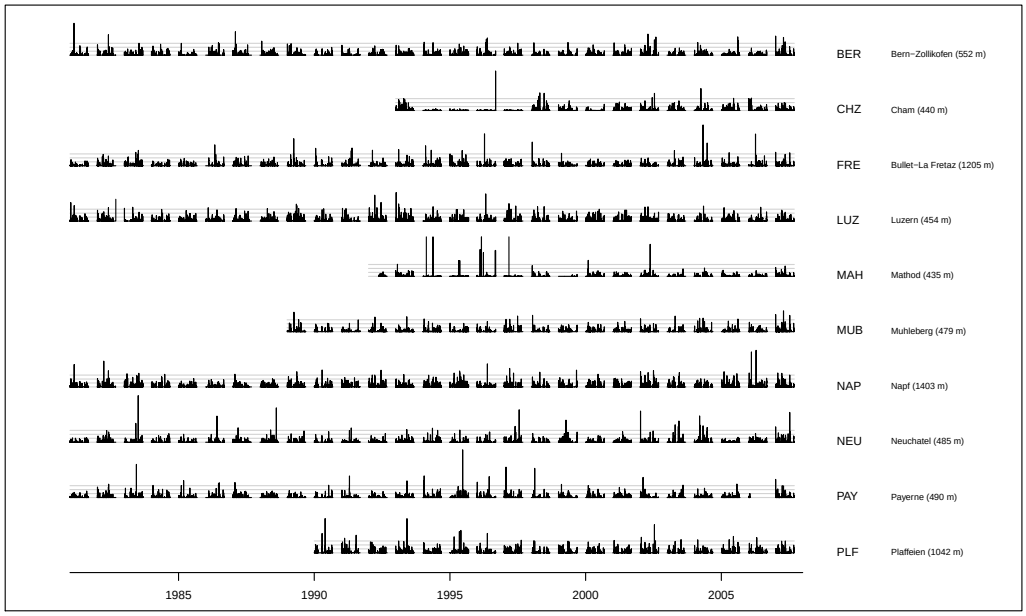

Figure 2. Summer hourly rainfall data (mm) at ten monitoring stations. The light grey lines show $0, 5, 10, 15$ mm. 75% of the measurements equal zero. The univariate thresholds used for transformation to the unit Fréchet scale are the 0.95-quantiles, ranging from 0.7 – 1.9 mm depending on the station. The gaps indicate that summers were treated as independent from one year to the next.

5.2. Model construction

We now discuss the construction of a model based on (3) for the rainfall data described in Section 5.1. This space-time model comprises a standard Gaussian random field $\varepsilon(s,t)$ with correlation function $\rho(h_s, h_t)$ and a random set element $\mathcal{A}$, both defined on a space $\mathcal{X} = \mathcal{S} \times \mathcal{T}$, where $\mathcal{S} = \mathbb{R}^2$ denotes space and $\mathcal{T} = \mathbb{R}_+$ denotes time.

The Gaussian random field is supposed to model the short-range behaviour of the process within single storms, so it is important to have a correlation function that can flexibly capture space-time interactions. For a good review of space-time correlation functions and a discussion of properties such as stationarity, separability and full symmetry, see Gneiting et al. (2007) and the references therein. Cressie and Huang (1999) propose classes of nonseparable spatio-temporal stationary covariance functions based on Bochner's theorem, and Gneiting (2002) extends their work by providing other very general flexible space-time covariance models. Davis et al. (2013a) show that this class of covariance functions satisfies a natural smoothness property at the origin, directly linked to the smoothness of the random field, and is therefore suitable for the modelling of physical processes such as rainfall. As a simple but fairly flexible possibility we used the isotropic nonseparable space-time correlation function (Gneiting, 2002)

$$\rho(h_s, h_t) = \frac{1}{\left\{\left(\frac{|h_t|}{\alpha_t}\right)^{\beta_t} + 1\right\}^{1+d\gamma/2}} \exp\left[-\frac{\left(\frac{\|h_s\|}{\alpha_s}\right)^{\beta_s}}{\left\{\left(\frac{|h_t|}{\alpha_t}\right)^{\beta_t} + 1\right\}^{\beta_s\gamma/2}}\right], \tag{11}$$

where h_s and h_t are lags in space and time, $\alpha_s, \alpha_t > 0$ determine spatial and temporal scale parameters, $\beta_s, \beta_t \in (0, 2]$ are spatial and temporal shape parameters, $d = 2$ is the spatial

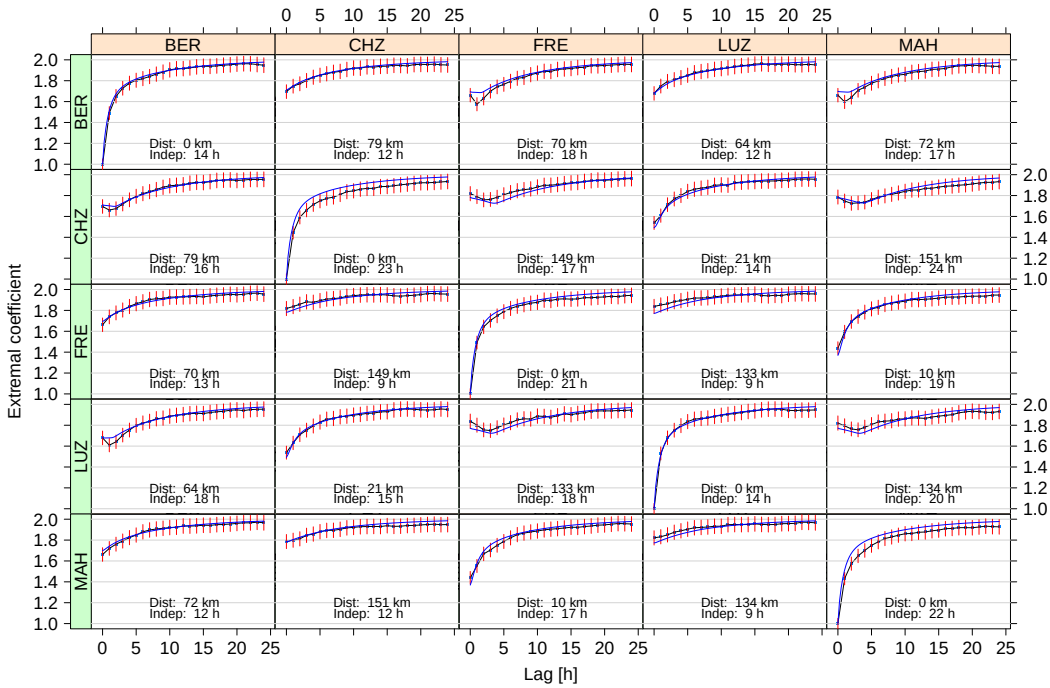

Figure 3. Empirical and model-based pairwise extremal coefficients θ_2 for five stations. The black lines join the empirical extremal coefficients found using the censored Schlather–Tawn estimator at the 0.95-quantile threshold, the vertical red segments being 95% confidence intervals. The blue lines correspond to the extremal coefficient curves derived from the fitted model in (3). The panel at the $r\text{th}$ row and $c\text{th}$ column shows the extremal coefficients between Z_t^{c} and Z_{t+h}^{r}, for $h = 0, 1, 2, \ldots, 24$. "Dist" stands for the distance between stations, and "Indep" is the time needed to reach independence (the first lag for which the value $\theta_2 = 2$ lies within the confidence interval).

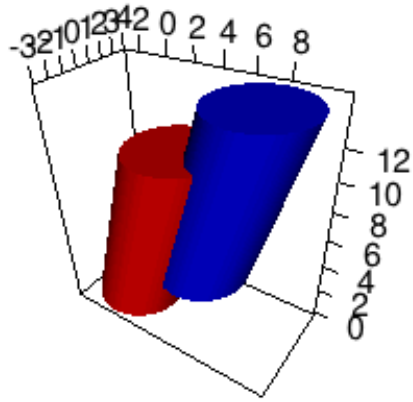

Figure 4. Illustration of the random set element $\mathcal{A}$ in space $\mathcal{S}$ (horizontal plane) and time $\mathcal{T}$ (vertical axis). The storms are conceptualized as random disks with a random radius moving at a random velocity for a random duration. The red tilted cylinder represents a realization $\mathcal{A}$ of such a storm in $\mathcal{S} \times \mathcal{T}$, and the blue one is $\mathcal{A} + (h_s, h_t)$, for a given spatio-temporal lag vector (h_s, h_t). The coefficient $\alpha(h_s, h_t)$ needed for the fitting is the expected volume of intersection between the two cylinders.

dimension, and $\gamma \in [0, 1]$ is a separability parameter quantifying the space-time interactions. When $\gamma = 0$, the space-time correlation function (11) is separable, i.e., it reduces to the product of a purely temporal correlation and a purely spatial correlation, whereas as γ approaches 1, the spatial and temporal components become increasingly entwined.

The random set $\mathcal{A}$ is interpreted as a random storm having a finite extent, which enables the model to capture complete independence. Conceptualizing storms as disks of random radius R moving at a random velocity V for a random duration D starting from a random position, the storm extent $\mathcal{A}$ in space and time becomes a tilted cylinder in $\mathcal{S} \times \mathcal{T}$, with a truncated Gaussian process inside; see Figure 4. For tractability we assume that $R \sim \text{Gamma}(m_R/k_R, k_R)$ (with mean m_R km), $V \sim \mathcal{N}_2(m_V, \Omega)$ (km/hour) and $D \sim \text{Gamma}(m_D/k_D, k_D)$ (with mean m_D hours). Furthermore, we parametrize the mean velocity as the vector $m_V = \{\|V\| \cos(\nu), \|V\| \sin(\nu)\}^T$, where ν is the angle of the main winds with respect to West-East direction, and the covariance of V as $\Omega = \|V\|^2 \omega^2 I_2$, where I_2 is the identity matrix. The factor ω is a dispersion parameter.

Other models were also considered, but were outperformed by the model described above; see Section 5.5.

5.3. Model fitting

The fitting of our model requires the computation of the coefficient $\alpha(h_s, h_t) = \text{E}[|\mathcal{A} \cap \{(h_s, h_t) + \mathcal{A}\}|]/\text{E}(|\mathcal{A}|)$ for $(h_s, h_t) \in \mathcal{X}$, i.e., the normalized expected volume of overlap between the random set $\mathcal{A}$ and itself shifted by the space-time lag (h_s, h_t). Several mild approximations, some analytical calculations and a single one-dimensional finite integration yield a good approximation to $\alpha(h_s, h_t)$, which is then used in computing the pairwise likelihood; see Appendix B.

After some exploratory analysis, we fixed $k_D = 9$ and $k_R = 0.4$, since these parameters are difficult to estimate; the model then has five parameters for the correlation function, and five

Table 3. Parameter estimates and 95%-confidence intervals from fitting our random set model to the rainfall data.

				Estimate	Conf. interval
Correlation	Scale	Space	α_s (km)	37.9	(31.7,43.1)
		Time	α_t (hr)	2.2	(1.6,2.4)
	Shape	Space	β_s	0.93	(0.89,1.09)
		Time	β_t	1.45	(1.33,1.65)
	Separability		γ	1.00	(0.43,1.00)
Random Set	Duration	Mean	m_D (hr)	111	(99,120)
		Shape	k_D	9	(—)
	Radius	Mean	m_R (km)	68	(61,80)
		Shape	k_R	0.4	(—)
	Velocity	Absolute speed	$\|V\|$ (km/hr)	39.6	(37.2,46.3)
		Angle	ν (rad)	0.24	(0.09,0.29)
		Dispersion	ω (km/hr)	0.12	(0.08,0.12)

for the random set. Due to the complexity of the problem, we split the estimation procedure into two steps: we first estimate $\alpha_t, \alpha_s, \gamma, m_R, \|V\|, m_D$, with the other parameters held fixed, and then all ten parameters together, with the former estimates as starting values. We always use the pairwise likelihood estimator (8). Confidence intervals are calculated by the block bootstrap described in Section 3.3 using yearly blocks. Based on the results in Section 4, we include the pairs at lags in $\mathcal{K} = \{0, 1, 2, 3, 5, 8, 13, 21\}$ in the pairwise log likelihood, a single evaluation of which involves contributions for about $T|\mathcal{K}|S^2 = 50000 \times 8 \times 10^2 = 40$ million pairs—the full pairwise likelihood would have 7 billion pairs, completely impractical for inference purposes! We coded the pairwise likelihood in C, parallelized the work on 8 CPUs, and fitted the model using the Nelder–Mead optimization routine in R. Even though there is a large amount of data and our model is very complex, a full fit took only about 10 minutes. Uncertainty assessment was based on 300 bootstrap replicates. The results are presented in Table 5.3.

The estimated mean speed of the dominant winds is 39.6 km/hr and the estimated angle is about 14° in the Argand diagram, which seem reasonable when compared to radar images of precipitation for the same region and time of year. This means that the clouds are likely to move in a rough east-northeasterly direction, in agreement with the summer climate in Western Switzerland. However, as the estimated angle coincides more or less with the main orientation of our monitoring stations and as the information along the perpendicular axis is likely to be small, one should interpret it with care.

The mean duration and mean radius of a storm are estimated as 111 hr and 68 km. Given the shape parameters, one half of the clouds have durations of over 107 hr and a radius of over 25 km. These estimates seem rather large. However, as shown in the supplementary material, the pairwise likelihood for the duration parameter is almost flat in its right tail, and that for the radius parameter is somewhat asymmetric. With a speed of about 40 km/hr on average, a cloud moves across the region of study in less than 4 hr, so it is hard to estimate these parameters based on the available observations. Data collected at a larger number of monitoring stations in a wider region of study would give more reliable conclusions. Hence the bootstrap confidence intervals for m_D, $(99, 120)$, and for m_R, $(61, 80)$, seem optimistic.

The correlation parameters appear to be better estimated. In particular, the separability parameter γ reaches its upper bound and its 95% confidence interval does not include zero, suggesting that the data are highly nonseparable and that γ tries to capture this.

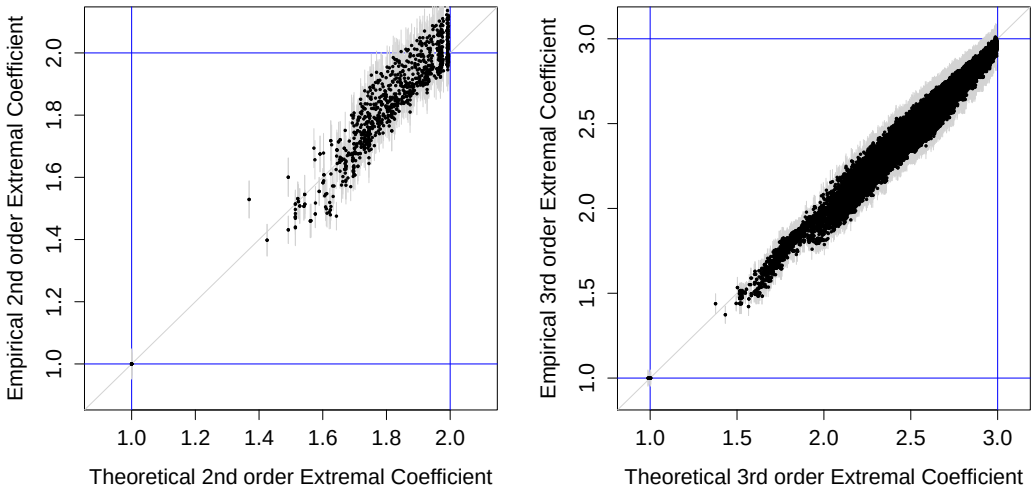

Figure 5. Comparison of empirical estimates of pairwise (left) and trivariate (right) extremal coefficients at the fitted lags for the rainfall data with their model-based counterparts. The light-grey vertical lines are 95% confidence intervals. A perfect agreement would place all points on the grey diagonal line.

5.4. Model checking

Figure 3 compares empirical estimates of the pairwise extremal coefficients with their model-based counterparts for a subset of 5 representative stations. There is a good agreement overall, but the fitted model often provides less extremal dependence at lag 1 than is present in the data. This lack of fit at short time lags might be explained either by a lack of flexibility due to the (conceptually) simplistic model that we used or by optimisation difficulties. The diagonal plots, showing the marginal temporal dependence of the extremes, show a good fit. The small differences at Cham (CHZ) or Mathod (MAH) may be due to nonstationarity or because data at those monitoring stations seem unreliable; see Figure 2. The left panel of Figure 5 shows pairwise extremal coefficients θ_2 in (5) for all pairs of stations.

As the model was fitted using pairs of observations, one might wonder whether it can capture higher-order interactions. We therefore computed the trivariate extremal coefficients (see Appendix C) and found good agreement between nonparametric estimates of trivariate extremal coefficients and their model-based counterparts; see the right panel of Figure 5. It seems that the trivariate interactions are fairly well modelled, though there is strong dependence among their estimates. The biggest discrepancies are from stations CHZ (Cham) and MAH (Mathod), but without these stations the points lie very close to the diagonal.

In order to assess the sensitivity of the results to initial conditions, we re-fitted the model with different starting values. The results were sometimes fairly different, but with similar bivariate properties and with almost the same pairwise likelihood. Consequently, we believe that some parameters are likely to play a similar role, giving rise to identifiability issues.

5.5. Alternative spatio-temporal models

We also fitted models based on the Brown–Resnick process (Brown and Resnick, 1977; Kabluchko et al., 2009), which, unlike the Schlather process, can capture full independence without a random set component. This is a stationary max-stable process that may

be represented in the space-time framework as $Z(s,t) = \sup_{i\geq 1} W_i(s,t)/P_i$, as in equation (1), where the $W_i(s,t)$ are independent replicates of the random process $W(s,t) = \exp\{\varepsilon(s,t) - \gamma(s,t)\}$ and $\varepsilon(s,t)$ is an intrinsically stationary Gaussian random field with space-time semi-variogram $\gamma(h_s, h_t)$, with $\varepsilon(0) = 0$ almost surely. The bivariate exponent measure for this model can be expressed as

$$V_2(z_1, z_2) = \frac{1}{z_1}\Phi\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{z_1}{z_2}\right)\right\} + \frac{1}{z_2}\Phi\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{z_2}{z_1}\right)\right\},$$

where $a = \{2\gamma(h_s, h_t)\}^{1/2}$, $h_s = s_2 - s_1$ is the spatial lag, $h_t = t_2 - t_1$ is the temporal lag, and where $\Phi(x)$ is the standard normal cumulative distribution function. Inference can be made similarly to the model (3), using the threshold-based censored pairwise likelihood (7).

Four different space-time semi-variograms were considered: (i) Model 1: $\gamma(h_s, h_t) = 1 - \rho(h_s, h_t)$, where the correlation function $\rho(h_s, h_t)$ is defined in (11); (ii) Model 2: $\gamma(h_s, h_t) = (\|h_s - h_t V\|/\alpha)^{2\beta}$, where $V = (V_1, V_2)^T$ is the wind velocity, $\alpha > 0$ is the range parameter and $\beta \in (0, 1]$ is the smoothness parameter; (iii) Model 3: $\gamma(h_s, h_t) = (h^T\Sigma^{-1}h)^\beta$, where $h = h_s - h_t V$, and Σ is a 2×2 covariance matrix; (iv) Model 4: $\gamma(h_s, h_t) = \{\|h_s - h_t V\|^2/\alpha_1^2 + |h_t|^2/\alpha_2^2\}^\beta$, where $\alpha_1, \alpha_2 > 0$ are range parameters capturing spatial and temporal dependence decays.

Model selection was performed by minimizing the composite likelihood information criterion, $\text{CLIC} = -2l_{\mathcal{K}}(\hat{\psi}_{p,\mathcal{K}}) + \text{tr}\{J_1(\hat{\psi}_{p,\mathcal{K}})^{-1}K(\hat{\psi}_{p,\mathcal{K}})\}$, an analogue of the Akaike information criterion (AIC) in a composite likelihood framework (see Varin and Vidoni, 2005). We considered a variant, CLIC*, which is scaled to be comparable with AIC for independent data (Davison and Gholamrezaee, 2012). The matrix $K(\hat{\psi}_{p,\mathcal{K}})$ has a complicated form, so we estimated the product $J_1(\hat{\psi}_{p,\mathcal{K}})^{-1}K(\hat{\psi}_{p,\mathcal{K}})$ by right-multiplication of the covariance matrix $V \approx J_1(\hat{\psi}_{p,\mathcal{K}})^{-1}K(\hat{\psi}_{p,\mathcal{K}})J_1(\hat{\psi}_{p,\mathcal{K}})^{-1}/T$ found using the block bootstrap by the Hessian matrix $J_1(\hat{\psi}_{p,\mathcal{K}})^{-1}/T$ estimated by finite differences. The estimated CLIC*s were 515370 for (3), 516170 (Model 1), 515956 (Model 2), 516191 (Model 3) and 515969 (Model 4). Based on this criterion, model (3) is greatly preferable to the others even though it has more parameters. More informal model-checking, using the pairwise extremal coefficients, confirmed that model (3) is the best of those considered here. Radar images of summer rain fields show areas where there is heavy rain and others where there is none. The random set model captures this, but the space-time Brown–Resnick processes do not.

6. Discussion

The work described above proposes inference for space-time extremes using a censored pairwise likelihood, and illustrates this by fitting a model for extreme rainfall; there are clear possibilities for extension to other phenomena. 'Dynamic' space-time modelling of extremes thus seems to be feasible; complex models can be consistently fitted using composite censored likelihood based on threshold exceedances. However, the large amount of data involved and the consequent usefulness of parallel computation underline the advantages of access to substantial computing resources when tackling such problems.

Although highly idealized, our extreme rainfall model is fairly complex, and estimation and simulation are demanding. Moreover, the assessment of fit is tricky, due to the computational burden that it requires. After a major effort we were able to check the trivariate interactions by means of the third-order extremal coefficients, and although it would be feasible to use simulation to investigate higher order interactions, it would be awkward.

An important modelling issue is that asympototic independence cannot be captured by our model, which is based solely on max-stable processes. However, it is common in practice to observe two distinct events becoming less and less dependent as their rarity increases, in

particular when rainfall data are considered (Davison et al., 2013). Wadsworth and Tawn (2012) have proposed models that can handle both asymptotic independence and asymptotic dependence, and it seems entirely feasible to extend our approach to them.

Acknowledgement

The authors are grateful to the reviewers for their constructive comments. This research was funded by the Swiss National Science Foundation, and partly performed in the context of the Competence Center Environment and Sustainability (CCES).

Supplementary material

Supplementary material available online includes analytical results on the asymptotic relative efficiencies of our pairwise threshold-based estimator for AR(1) and MA(1) models, discussion of the choice of time lags to be included in the pairwise likelihood in these cases, and additional plots for the simulation study of §4 and for the data analysis of §5.

Appendix

A. Proof of Theorem 1

Proof. For notational simplicity, we give the proof in the case where the parameter ψ is scalar, but the argument can be extended to the vector case.

By definition of the pairwise likelihood in equation (7), and as the observations $Z_{s,t}$ are realizations of a censored max-stable process, we have

$$\mathrm{E}\{U_t(\psi)\} = \sum_{h_t \in \mathcal{K}_t} \sum_{s_2=1}^{S} \sum_{s_1=1}^{S} \underbrace{\mathrm{E}\left\{ \frac{d}{d\psi} \log p_u \left(Z_{s_1,t}, Z_{s_2,t+h_t}; \psi\right)\right\}}_{=0} (1 - I\{s_1 \geq s_2 \text{ and } h_t = 0\}) = 0.$$

Therefore, we also have that $\mathrm{E}\{U(\psi)\} = \mathrm{E}\{\sum_{t=1}^{T} U_t(\psi)\} = 0$.

The variance of $U(\psi)$ renormalised by T is (Shumway and Stoffer, 2011, p.510)

$$\begin{aligned} T^{-1}\mathrm{var}\{U(\psi)\} &= \mathrm{E}\{U_1(\psi)^2\} + 2\sum_{t=1}^{T-1}\left(1 - \frac{t}{T}\right)\mathrm{E}\{U_1(\psi)U_{t+1}(\psi)\} \\ &\to \mathrm{E}\{U_1(\psi)^2\} + 2\sum_{t=1}^{\infty}\mathrm{E}\{U_1(\psi)U_{t+1}(\psi)\}, \quad T \to \infty, \end{aligned}$$

if the sum converges absolutely. Now, as $\hat{\psi}_{p,\mathcal{K}}$ is the maximum pairwise likelihood estimator, second-order Taylor expansion of $U_t(\hat{\psi}_{p,\mathcal{K}})$ around the true parameter ψ gives

$$0 = \sum_{t=1}^{T} U_t(\hat{\psi}_{p,\mathcal{K}}) \doteq \sum_{t=1}^{T}\left\{U_t(\psi) + \frac{d}{d\psi}U_t(\psi)(\hat{\psi}_{p,\mathcal{K}} - \psi)\right\},$$

which gives, up to a term of the order $O\{(\hat{\psi}_{p,\mathcal{K}} - \psi)^2\}$, that

$$\hat{\psi}_{p,\mathcal{K}} \doteq \psi + \left\{\sum_{t=1}^{T} H_t(\psi)\right\}^{-1} \sum_{t=1}^{T} U_t(\psi) = \psi + H(\psi)^{-1}U(\psi), \tag{12}$$

where $H_t(\psi) = -dU_t(\psi)/d\psi$ and $H(\psi) = \sum_{t=1}^{T} H_t(\psi)$ is the observed information. Moreover, since the process $Z(s,t)$ is assumed to be temporally α-mixing with coefficients $\alpha(n)$, the time series $U_t(\psi)$ is also α-mixing with coefficients $\alpha'(n) = \alpha(n - \max\mathcal{K})$. Hence,

$$\alpha'(n) \to 0, \qquad \sum_{n \geq 1} |\alpha'(n)|^{\delta/(2+\delta)} < \infty,$$

with the same $\delta > 0$. These results, along with the assumptions $\mathrm{E}(U_1^2) < \infty$ and $\mathrm{E}(|U_1|^{2+\delta}) < \infty$, ensure that the Central Limit Theorem 10.7 of Bradley (2007) applies, and thus

$$T^{-1/2}U(\psi) \xrightarrow{D} \mathcal{N}\{0, K(\psi)\}, \quad T \to \infty,$$

where $K(\psi) = \mathrm{E}\{U_1(\psi)^2\} + 2\sum_{t=1}^{\infty} \mathrm{E}\{U_1(\psi)U_{t+1}(\psi)\} < \infty$ and $\xrightarrow{D}$ denotes convergence in distribution. Therefore, returning to equation (12), and by definition of $J_1(\psi)$, by the law of large numbers, and by Slutsky's theorem, we get

$$\begin{aligned} T^{1/2}(\hat{\psi}_{p,\mathcal{K}} - \psi) &\doteq T^{1/2}H(\psi)^{-1}U(\psi) \\ &= \{T^{-1}H(\psi)\}^{-1}\{T^{-1/2}U(\psi)\} \\ &\xrightarrow{D} J_1(\psi)^{-1}\mathcal{N}\{0, K(\psi)\} \qquad \text{as } T \to \infty \\ &\stackrel{D}{=} \mathcal{N}(0, J_1(\psi)^{-1}K(\psi)J_1(\psi)^{-1}), \end{aligned}$$

where $\stackrel{D}{=}$ denotes equality in distribution. But $K(\psi)$ is the asymptotic variance of the score, renormalized by T. Hence, the result is proved.

B. Computation of the volume of overlap $\alpha(h_s, h_t)$

The coefficient $\alpha(h_s, h_t)$ is defined as $\mathrm{E}[|\mathcal{A} \cap \{(h_s, h_t) + \mathcal{A}\}|]/\mathrm{E}(|\mathcal{A}|)$, where $\mathcal{A}$ is a tilted cylinder in $\mathcal{X} = \mathcal{S} \times \mathcal{T} = \mathbb{R}^2 \times \mathbb{R}_+$ (see Figure 4), and $(h_s, h_t) \in \mathcal{X}$. If the cylinder were vertical (zero wind velocity), the volume of overlap would simply be the product of the area of overlap between two discs distant by $\|h_s\|$ and the corresponding height, the storm duration minus h_t.

Let R be the storm radius, $V = (V_1, V_2) \in \mathbb{R}^2$ be its velocity and D be its lifetime. A good linear approximation to the area of overlap of two discs of radius R distant by d is $\pi R^2 \max\{0, 1 - d/(2R)\}$ (Davison and Gholamrezaee, 2012). Therefore, for a vertical cylinder $\mathcal{A}$, $|\mathcal{A} \cap \{(h_s, h_t) + \mathcal{A}\}|$ can be approximated by

$$\pi R^2 \left(1 - \frac{\|h_s\|}{2R}\right)_+ (D - h_t)_+,$$

where $a_+ = \max\{0, a\}$. When the cloud is moving, giving a tilted cylinder, a geometric argument shows that in the general case, the volume of overlap is transformed to

$$|\mathcal{A} \cap \{(h_s, h_t) + \mathcal{A}\}| \doteq \pi R^2 \left(1 - \frac{d^*}{2R}\right)_+ (D - h_t)_+,$$

where $d^* = \left[\|h_s\|^2 + h_t^2(V_1^2 + V_2^2) - 2\|h_s\|h_t\{V_1\cos(\nu) + V_2\sin(\nu)\}\right]^{1/2}$, $\nu = \arctan(h_{s;1}/h_{s;2})$ being the angle between the stations with respect to a reference axis in the West-East direction. Careful checking suggests that this provides adequate approximations for the values of h_s and h_t and the parameter values used in the pairwise likelihood in our application.

In order to compute the coefficient $\alpha(h_s, h_t)$, which depends upon the spatial distance $\|h_s\|$, the temporal lag h_t and the orientation of the stations ν, we need to obtain the expected volume of overlap $\mathrm{E}[|\mathcal{A} \cap \{(h_s, h_t) + \mathcal{A}\}|]/\mathrm{E}(|\mathcal{A}|)$, by putting tractable distributions on R, D, and $V = (V_1, V_2)$. We choose to set

- $R \sim \mathrm{Gamma}(m_R/k_R, k_R)$ (with mean m_R km);
- $V \sim \mathcal{N}_2(m_V, \Omega)$ (km/hour), with $m_V = (m_1, m_2)^T$ and $\Omega = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 \end{pmatrix}$;
- $D \sim \mathrm{Gamma}(m_D/k_D, k_D)$ (with mean m_D hours);

and we assume that R, D and V are mutually independent. To compute this expectation, note first that $(D - t)_+$ can be integrated out analytically. Second, by conditioning on V, it is possible to integrate over R as well. We can then reduce the full computation to this single expectation with respect to $V = (V_1, V_2)$:

$$\alpha(h_s, h_t) = \mathrm{E}_V \left\{ \Pr(G_{m_R/k_R;k_R+2} > d^*/2) - \frac{d^* k_R}{2(k_R+1)m_R} \Pr(G_{m_R/k_R;k_R+1} > d^*/2) \right\}, \qquad (13)$$

where $G_{\theta;k}$ is a gamma random variable with scale parameter θ and shape parameter k; its mean equals $m = \theta k$. Expression (13) does not have a closed form, but it can be remarkably well approximated by a function of the form $\exp\left[-a\left\{(V_1-\mu_1)^2+(V_2-\mu_2)^2\right\}\right]$, where a is real number that does not depend upon $V = (V_1, V_2)$ and can be estimated with a few points by least squares, and where $\mu_1 = \|h_s\|\cos(\nu)/h_t$ and $\mu_2 = \|h_s\|\sin(\nu)/h_t$. Therefore, we have

$$\begin{aligned}
\alpha(h_s,h_t) &\approx \mathrm{E}_V\left\{\exp\left[-a\left\{(V_1-\mu_1)^2+(V_2-\mu_2)^2\right\}\right]\right\} \\
&= \int_{\mathbb{R}^2} e^{-a\{(v_1-\mu_1)^2+(v_2-\mu_2)^2\}^{1/2}} \frac{1}{2\pi\det(\Omega)^{1/2}} e^{-\frac{1}{2}(v_1-m_1;v_2-m_2)\Omega^{-1}(v_1-m_1;v_2-m_2)^T} dv_1 dv_2 \\
&= \frac{1}{2\pi\det(\Omega)^{1/2}} \int_{\mathbb{R}^2} e^{-a\{(v_1-\mu_1)^2+(v_2-\mu_2)^2\}^{1/2}-\frac{1}{2\det(\Omega)}\{(v_1-m_1)^2\sigma_2^2-2(v_1-m_1)(v_2-m_2)\sigma_1\sigma_2\rho_{12}+(v_2-m_2)^2} \\
&= \frac{1}{2\pi\det(\Omega)^{1/2}} \int_0^{2\pi} d\xi \int_{\mathbb{R}_+} r e^{-ar-\frac{1}{2\det(\Omega)}\{r^2a(\xi)+rb(\xi)+c(\xi)\}} dr \\
&= \frac{1}{(2\pi)^{1/2}} \int_0^{2\pi} \frac{1}{\sqrt{a(\xi)}} e^{-\frac{1}{2\sigma(\xi)^2}\left(\frac{c(\xi)}{a(\xi)}-\mu(\xi)^2\right)} d\xi \int_{\mathbb{R}_+} r \frac{1}{\sqrt{2\pi}\sigma(\xi)} e^{-\frac{1}{2}\left(\frac{r-\mu(\xi)}{\sigma(\xi)}\right)^2} dr \\
&= \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{\sqrt{a(\xi)}} e^{-\frac{1}{2\sigma(\xi)^2}\left(\frac{c(\xi)}{a(\xi)}-\mu(\xi)^2\right)} \left[\sigma(\xi)e^{-\frac{1}{2}\frac{\mu(\xi)^2}{\sigma(\xi)^2}} + \sqrt{2\pi}\mu(\xi)\left\{1-\Phi\left(-\frac{\mu(\xi)}{\sigma(\xi)}\right)\right\}\right] d\xi,
\end{aligned}$$

where $\Phi(\cdot)$ is the normal cumulative distribution function and

$$\begin{aligned}
a(\xi) &= \cos^2(\xi)\sigma_2^2 + \sin^2(\xi)\sigma_1^2 - 2\cos(\xi)\sin(\xi)\sigma_1\sigma_2\rho_{12}, \\
b(\xi) &= 2\cos(\xi)(\mu_1-m_1)\sigma_2^2 + 2\sin(\xi)(\mu_2-m_2)\sigma_1^2 - 2\cos(\xi)(\mu_2-m_2)\sigma_1\sigma_2\rho_{12} \\
&\qquad -2\sin(\xi)(\mu_1-m_1)\sigma_1\sigma_2\rho_{12}, \\
c(\xi) &= (\mu_1-m_1)^2\sigma_2^2 + (\mu_2-m_2)^2\sigma_1^2 - 2(\mu_1-m_1)(\mu_2-m_2)\sigma_1\sigma_2\rho_{12}, \\
\mu(\xi) &= -\frac{b(\xi)}{2a(\xi)} - \frac{a\det(\Omega)}{a(\xi)}, \qquad \sigma(\xi) = \sqrt{\det(\Omega)/|a(\xi)|}, \qquad \det(\Omega) = \sigma_1^2\sigma_2^2(1-\rho_{12}).
\end{aligned}$$

Expression (14) was computed with a straightforward change of variables $v_1 = r\cos(\xi) + \mu_1$, $v_2 = r\sin(\xi) + \mu_2$, and expression (15) stems from the properties of the normal cumulative distribution function. Since the integral (15) is impossible to handle analytically, we can use a finite approximation to estimate $\alpha(h)$, based on 100 points equi-spaced in the interval $[0, 2\pi]$. The approximation seems to be adequate when $\sigma_1^2, \sigma_2^2 > 5$, which we impose in the R optimization routine.

C. Trivariate extremal coefficients for model (3)

From equation (2), we know that the multivariate extremal coefficient in dimension N is

$$\theta_N = V_N(1,\ldots,1) = \mathrm{E}\left[\max_{i=1,\ldots,N}\{W(x_i)\}\right].$$

This takes values between 1 and N, ranging from complete dependence to asymptotic independence. Therefore, the extremal coefficient of order $N = 3$ is

$$\theta_3 = \mathrm{E}\left[\max\{W(x_1), W(x_2), W(x_3)\}\right],$$

where, for model (3), $W(x) \propto \max\{0, \varepsilon(x)\}I_{\mathcal{A}}(x-X)$, $x \in \mathcal{X}$, $\varepsilon(x)$ being an isotropic Gaussian random field with zero mean, unit variance and correlation function $\rho(\cdot)$ and $I_{\mathcal{A}}$ being the indicator that the point $x - X$ belongs to a random set $\mathcal{A}$ (where X is a Poisson process of unit rate in $\mathcal{X}$). The proportionality constant is such that $W(x)$ has mean 1, so it must be

$$\frac{1}{\mathrm{E}\{\max(0,\varepsilon)I_{\mathcal{A}}\}} = \frac{1}{\mathrm{E}\{\max(0,\varepsilon)\}\mathrm{E}(I_{\mathcal{A}})} = \frac{\sqrt{2\pi}}{\Pr(x\in\mathcal{A})} = \frac{\sqrt{2\pi}|\mathcal{X}|}{\mathrm{E}(|\mathcal{A}|)}.$$

Below we write $W_1 = W(x_1)$, $\varepsilon_1 = \varepsilon(x_1)$, $I_1 = I_{\mathcal{A}}(x_1)$, $I_{1;2;-} = I\{x_1 \in \mathcal{A} \text{ and } x_2 \in \mathcal{A} \text{ and } x_3 \notin \mathcal{A}\}$ and so forth. Then the required extremal coefficient is

$$\begin{aligned}
\theta_3 &= \mathrm{E}\left\{\max(W_1, W_2, W_3)\right\} \\
&= \frac{\sqrt{2\pi}}{\Pr(x_1 \in \mathcal{A})}\mathrm{E}\left\{\max\left(0, \varepsilon_1 I_1, \varepsilon_2 I_2, \varepsilon_3 I_3\right)\right\} \\
&= \frac{\sqrt{2\pi}}{\Pr(x_1 \in \mathcal{A})}\Big[\mathrm{E}\left\{\max\left(0, \varepsilon_1, \varepsilon_2, \varepsilon_3\right) I_{1;2;3}\right\} + \mathrm{E}\left\{\max\left(0, \varepsilon_1, \varepsilon_2\right) I_{1;2;-}\right\} \\
&\quad + \mathrm{E}\left\{\max\left(0, \varepsilon_1, \varepsilon_3\right) I_{1;-;3}\right\} + \mathrm{E}\left\{\max\left(0, \varepsilon_2, \varepsilon_3\right) I_{-;2;3}\right\} \\
&\quad + \mathrm{E}\left\{\max\left(0, \varepsilon_1\right) I_{1;-;-}\right\} + \mathrm{E}\left\{\max\left(0, \varepsilon_2\right) I_{-;2;-}\right\} + \mathrm{E}\left\{\max\left(0, \varepsilon_3\right) I_{-;-;3}\right\}\Big] \\
&= \Pr(x_2 \in \mathcal{A}, x_3 \in \mathcal{A} \mid x_1 \in \mathcal{A})\sqrt{2\pi}\mathrm{E}\left\{\max\left(0, \varepsilon_1, \varepsilon_2, \varepsilon_3\right)\right\} \\
&\quad + \Pr(x_2 \in \mathcal{A}, x_3 \notin \mathcal{A} \mid x_1 \in \mathcal{A})\sqrt{2\pi}\mathrm{E}\left\{\max\left(0, \varepsilon_1, \varepsilon_2\right)\right\} \\
&\quad + \Pr(x_2 \notin \mathcal{A}, x_3 \in \mathcal{A} \mid x_1 \in \mathcal{A})\sqrt{2\pi}\mathrm{E}\left\{\max\left(0, \varepsilon_1, \varepsilon_3\right)\right\} \\
&\quad + \Pr(x_1 \notin \mathcal{A}, x_3 \in \mathcal{A} \mid x_2 \in \mathcal{A})\sqrt{2\pi}\mathrm{E}\left\{\max\left(0, \varepsilon_2, \varepsilon_3\right)\right\} \\
&\quad + \Pr(x_2 \notin \mathcal{A}, x_3 \notin \mathcal{A} \mid x_1 \in \mathcal{A}) + \Pr(x_1 \notin \mathcal{A}, x_3 \notin \mathcal{A} \mid x_2 \in \mathcal{A}) \\
&\quad + \Pr(x_1 \notin \mathcal{A}, x_2 \notin \mathcal{A} \mid x_3 \in \mathcal{A}).
\end{aligned}$$

The expression $\sqrt{2\pi}\mathrm{E}\left\{\max\left(0, \varepsilon_1, \varepsilon_2, \varepsilon_3\right)\right\}$ above is the trivariate extremal coefficient for the Schlather model without random sets, and can be evaluated quickly and accurately by simulation, whereas $\sqrt{2\pi}\mathrm{E}\left\{\max\left(0, \varepsilon_i, \varepsilon_j\right)\right\}$ is the bivariate extremal coefficient between station i and station j, and can be computed analytically with the exponent measure $V_{i;j}(1,1)$.

The probabilities above correspond to the normalized expected volumes of overlap of three sets $\mathcal{A}$ centered at x_1, x_2 and x_3. For example,

$$\Pr(x_2 \in \mathcal{A}, x_3 \in \mathcal{A} \mid x_1 \in \mathcal{A}) = \mathrm{E}\{|\mathcal{A} \cap \{\mathcal{A} + (x_2 - x_1)\} \cap \{\mathcal{A} + (x_3 - x_1)\}|\}/\mathrm{E}(|\mathcal{A}|),$$

$$\Pr(x_2 \in \mathcal{A}, x_3 \notin \mathcal{A} \mid x_1 \in \mathcal{A}) = \mathrm{E}[|\mathcal{A} \cap \{\mathcal{A} + (x_2 - x_1)\} \cap \{\mathcal{A} + (x_3 - x_1)\}^c|]/\mathrm{E}(|\mathcal{A}|).$$

For given radius R, duration D and velocity V, the random set is fixed and the volume of overlap can be calculated analytically. Simulation can then be used to compute the expectation of such random quantities.

The same approach could be used to compute extremal coefficients at a higher order N, though it would be painful to compute all the areas of overlap between N discs.

References

Aulbach, S. and M. Falk (2012). Testing for a generalized Pareto process. *Electronic Journal of Statistics 6*, 1779–1802.

Beirlant, J., Y. Goegebeur, J. Teugels, and J. Segers (2004). *Statistics of Extremes: Theory and Applications*. New York: Wiley.

Bevilacqua, M., C. Gaetan, J. Mateu, and E. Porcu (2012). Estimating space and space-time covariance functions: A weighted composite likelihood approach. *Journal of the American Statistical Association 107*, 268–280.

Blanchet, J. and A. C. Davison (2011). Spatial modelling of extreme snow depth. *Annals of Applied Statistics 5*, 1699–1725.

Bortot, P., S. Coles, and J. A. Tawn (2000, January). The multivariate Gaussian tail model: An application to oceanographic data. *Applied Statistics 49*, 31–49.

Bradley, R. C. (2007). *Introduction to Strong Mixing Conditions*, Volume 1. Heber City, Utah: Kendrick Press.

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

Buishand, T. A., L. de Haan, and C. Zhou (2008). On spatial extremes: With application to a rainfall problem. *Annals of Applied Statistics 2*, 624–642.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* London, UK: Springer Verlag.

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

Cressie, N. and H.-C. Huang (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association 94*, 1330–1340.

Davis, R. A., C. Klüppelberg, and C. Steinkohl (2013a). Max-stable processes for modelling extremes observed in space and time. *Journal of the Korean Statistical Society ??*, to appear.

Davis, R. A., C. Klüppelberg, and C. Steinkohl (2013b). Statistical inference for max-stable processes in space and time. *Journal of the Royal Statistical Society, series B 75*, to appear.

Davis, R. A. and T. Mikosch (2008). Extreme value theory for space-time processes with heavy-tailed distributions. *Stochastic Processes and Their Applications 118*, 560–584.

Davis, R. A. and C. Y. Yau (2011). Comments on pairwise likelihood in time series models. *Statistica Sinica 21*, 255–277.

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

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods and their Application.* Cambridge University Press.

Davison, A. C., R. Huser, and E. Thibaud (2013). Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences ??*, to appear.

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

Davison, A. C. and R. L. Smith (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society, Series B 52*, 393–442.

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

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction.* New York: Springer Verlag.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance.* Berlin: Springer Verlag.

Ferreira, A. and L. de Haan (2012, November). The Generalized Pareto process; with a view towards application and simulation. *arXiv:1203.2551v2*.

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

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association 97*, 590–560.

Gneiting, T., M. G. Genton, and P. Guttorp (2007). Geostatistical Space-Time Models, Stationarity, Separability, and Full Symmetry. In B. Finkenstädt, L. Held, and V. Isham (Eds.), *Statistical Methods for Spatio-Temporal Systems*, pp. 151–175. Boca Raton: Chapman & Hall/CRC.

Godambe, V. P. and C. C. Heyde (1987). Quasi-likelihood and optimal estimation. *International Statistical Review 55*, 231–244.

Hjort, N. L. and C. Varin (2008). ML, PL, QL in Markov chain models. *Scandinavian Journal of Statistics 35*, 64–82.

Huser, R. and A. C. Davison (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika*. DOI: 10.1093/biomet/ass089.

Joe, H. and Y. Lee (2009, April). On weighting of bivariate margins in pairwise likelihood. *J. Multivar. Anal. 100*, 670–685.

Kabluchko, Z. and M. Schlather (2010). Ergodic properties of max-infinitely divisible processes. *Stochastic Processes and their Applications 120*, 281–295.

Kabluchko, Z., M. Schlather, and L. de Haan (2009). Stationary max-stable fields associated to negative definite functions. *Annals of Probability 37*, 2042–2065.

Leadbetter, M. R. (1991). On a basis for 'Peaks over Threshold' modeling. *Statistics & Probability Letters 12*, 357–362.

Leadbetter, M. R., G. Lindgren, and H. Rootzén (1983). *Extreme and Related Properties of Random Sequences and Processes*. New York: Springer Verlag.

Ledford, A. W. and J. A. Tawn (1996). Statistics for near independence in multivariate extreme values. *Biometrika 83*, 169–187.

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

Northrop, P. J. and P. Jonathan (2011). Threshold modelling of spatially-dependent non-stationary extremes with application to hurricane-induced wave heights (with discussion). *Environmetrics 22*, 799–809.

Padoan, S. A., M. Ribatet, and S. A. Sisson (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association 105*, 263–277.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Reich, B. J. and B. A. Shaby (2012). A hierarchical max-stable spatial model for extreme precipitation. *Annals of Applied Statistics 6*, 1430–1451.

Ribatet, M., D. Cooley, and A. C. Davison (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica 22*, 813–845.

Rootzén, H. and N. Tajvidi (2006). Multivariate generalized Pareto distributions. *Bernoulli 12*, 917–930.

Scarrott, C. and A. MacDonald (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT—Statistical Journal 10*, 33–60.

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

Schlather, M. and J. Tawn (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika 90*, 139–156.

Shumway, R. H. and D. S. Stoffer (2011). *Time Series Analysis and its Applications, with R Examples* (Third ed.). Springer.

Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science 4*, 367–393.

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

Smith, R. L., J. A. Tawn, and S. G. Coles (1997, January). Markov chain models for threshold exceedances. *Biometrika 84*, 249–268.

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

Varin, C. and C. Czado (2010). A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics 11*, 127–138.

Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica 21*, 5–42.

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

Varin, C. and P. Vidoni (2009). Pairwise likelihood inference for general state space models. *Econometric Reviews 28*, 170–185.

Wadsworth, J. and J. Tawn (2012). Dependence modelling for spatial extremes. *Biometrika 99*, 253–272.

Wadsworth, J. L. and J. A. Tawn (2013). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika 100*, submitted.