Geostatistics of extremes

By A. C. Davison[1,*] and M. M. Gholamrezaee[2]

[1]*Ecole Polytechnique Fédérale de Lausanne, EPFL-FSB-MATHAA-STAT,
Station 8, 1015 Lausanne, Switzerland*
[2]*Centre Hospitalier Universitaire Vaudois, CHUV-CEPP, Hôpital de Cery,
Site de Cery, 1008 Prilly, Switzerland*

We describe a prototype approach to flexible modelling for maxima observed at sites in a spatial domain, based on fitting of max-stable processes derived from underlying Gaussian random fields. The models we propose have generalized extreme-value marginal distributions throughout the spatial domain, consistent with statistical theory for maxima in simpler cases, and can incorporate both geostatistical correlation functions and random set components. Parameter estimation and fitting are performed through composite likelihood inference applied to observations from pairs of sites, with occurrence times of maxima taken into account if desired, and competing models are compared using appropriate information criteria. Diagnostics for lack of model fit are based on maxima from groups of sites. The approach is illustrated using annual maximum temperatures in Switzerland, with risk analysis proposed using simulations from the fitted max-stable model. Drawbacks and possible developments of the approach are discussed.

1. Introduction

(a) Spatial extremes

The prospect of climatic change and its impacts have brought spatial statistics of extreme events into sharper focus. For example, the European summer heatwave of 2003 cost the lives of many vulnerable individuals, and affected crops and water supply. Heatwaves and heavy rainfall are predicted to become more frequent in some regions of a warming Europe (Beniston *et al.* 2007), and it is important to have a mathematically sound and statistically efficient basis for modelling them and assessing their possible consequences. This paper represents a step towards this goal, based on data from a number of sites. The key idea is to use a so-called pairwise likelihood to fit models for spatial extremal processes, for which standard likelihood or Bayesian inference seem to be out of reach. By incorporating elements of classical Gaussian geostatistics, we are able to find flexible formulations for spatial extremes that extend univariate extremal models in a natural way.

*Author for correspondence (anthony.davison@epfl.ch).

This journal is © 2011 The Royal Society

Statistics of extremes have over the past two decades developed remarkably, with books now available at various levels of sophistication and a journal, *Extremes*, devoted to recent research. The underlying mathematical basis is now thoroughly established (Leadbetter *et al.* 1983; Resnick 1987; Embrechts *et al.* 1997; de Haan & Ferreira 2006) and statistical tools and methods for use with a single time series of data, or with a few series, are well developed (Coles 2001; Beirlant *et al.* 2004) and widely used. The main approaches to univariate extremal analysis are the fitting of the generalized extreme-value distribution to maxima, and of the generalized Pareto distribution or a related point process model to exceedances over high thresholds, though many elaborations appear in applications.

Multi-variate extreme-value models, the springboard for modelling of spatial aspects of rare events, are surveyed by Beirlant *et al.* (2004). Key mathematical elements in this classical theory, which is based on the limiting behaviour of tails of distributions, are regular variation and point processes, and these also play a central role in the theory of spatial extremes, as summarized in ch. 9 of de Haan & Ferreira (2006). There are many published applications of multi-variate extremes, but the literature on applications of the spatial theory is limited owing to a lack of flexible statistical models and corresponding inferential tools.

Four main approaches to modelling spatial extremes have been proposed. The first of these employs underlying latent processes, conditional on which standard extremal models are applied (Coles & Casson 1998; Casson & Coles 1999; Cooley *et al.* 2006, 2007; Fawcett & Walshaw 2006; Sang & Gelfand 2009*a*). This approach has the advantages of fitting naturally into Bayesian computational set-ups using Markov chain Monte Carlo simulation, and being applicable to large datasets, but because most such models postulate independence of extremes conditional on the latent process, the spatial modelling may yield unrealistic estimates of certain risk measures because the spatial clustering of rare events is not properly accounted for. An exception to this is Sang & Gelfand (2009*b*), who use a Gaussian copula model to mitigate the independence assumption, but this does not admit a max-stable model for extremes (see the next paragraph). A further difficulty with such models is that the marginal distributions thus generated are mixtures of standard extremal distributions and so do not fall naturally into the usual extremal paradigm.

A second approach, sometimes called Gaussian anamorphosis and related to the use of copulas, involves the transformation of the marginal distributions of the data to a Gaussian scale, on which standard multi-variate normal models may be used. This seems unsatisfactory because for any fixed correlation, the extremes of a bivariate Gaussian distribution occur independently (Sibuya 1960), and this is unrealistic for many applications. Davison *et al.* (in press) discuss these and other types of copula models, and compare them with the approaches suggested here.

The third potential approach, owing to Heffernan & Tawn (2004) and further applied by Butler *et al.* (2007), relies upon a conditional formulation for rare events. This can be applied to large datasets, but, among other drawbacks, it is not obvious how to extend it to non-gridded observations.

The present paper builds on the pioneering work of Smith (R. L. Smith 1990, unpublished data) in discussing a fourth approach, based on representations of max-stable processes suggested by de Haan (1984) and Schlather (2002), and
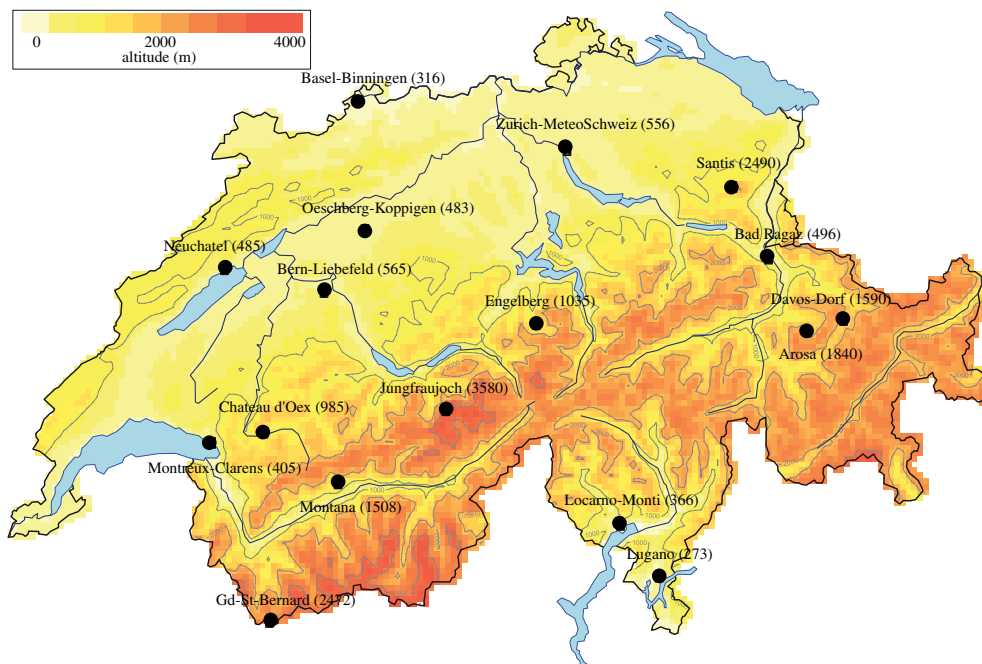
Figure 1. Topographical map of Switzerland showing the sites and altitudes in metres above sea level of 17 weather stations for which daily maximum temperature data are available.

applied to the modelling of rainfall data by Smith & Stephenson (2009), Padoan *et al.* (2010) and Davison *et al.* (in press); rainfall modelling with max-stable processes was also described by Buishand *et al.* (2008), and modelling of maxima of annual snow heights is described by Blanchet & Davison (in press). The approach we describe has the advantages of a full integration of classical extremal results, so that the marginal distributions for maxima are appropriately modelled, and a treatment of spatial dependence suitable for rare events.

Before describing the methodological content of the paper, we give an example of the type of data and application that our work is intended to address.

(b) Data

Below, we shall consider annual maximum temperatures from the 17 sites in Switzerland shown in figure 1. The data are abstracted from daily maximum temperature time series for the years 1961–2005, part of which are shown in figure 2 as temperature anomalies, obtained as residuals after subtraction of the site-wise average summer temperatures for the five years shown. There are evidently very strong correlations between the different series: many of the maxima occur simultaneously or as part of the same sequence of a few successive hot days. If we let $Y(x)$ denote the temperature at site x of the set $\mathcal{X}$, then the quantity of ultimate interest might be an integral such as

$$R_{\mathcal{X}} = \int_{\mathcal{X}} r(x)\{Y(x) \geq y_{\text{danger}}\}_+ \, dx, \qquad (1.1)$$
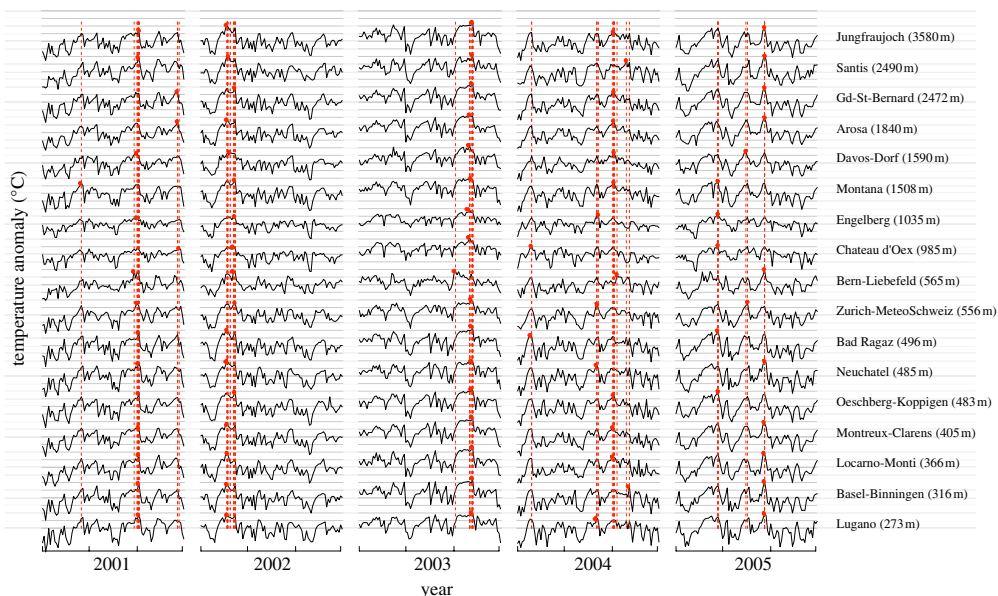
Figure 2. Daily temperature anomalies (°C; solid lines) for the months June, July and August 2001–2005 at 17 sites in Switzerland at differing heights above sea level. The heavy grey lines show the site-wise averages of the summer data for these 5 years, and the lighter grey lines show intervals of 5°C. The annual maxima are shown by the blobs, and the vertical dashed lines show the timing of the annual maxima. The rugs at the foot indicate the months.

where $r(x)$ might represent a population or a crop at risk at site x if $Y(x)$ exceeds some level y_{danger}, and $\{u\}_+ = \max(u, 0)$. Of course, many other functions might be of interest; the key point is that joint spatial modelling of $Y(x)$ is required for useful inference on $R_{\mathcal{X}}$. This entails extrapolation from the points at which data are available to the whole of $\mathcal{X}$, and estimation of the joint distribution of $\{Y(x)\}_{x \in \mathcal{X}}$.

At each individual site $x \in \mathcal{X}$, standard arguments imply that if a limiting law for the maximum exists, then it must be a generalized extreme-value distribution

$$
G(y) = \begin{cases} \exp\left[-\left\{ 1 + \dfrac{\xi(y - \eta)}{\tau} \right\}_+^{-1/\xi} \right], & \xi \neq 0, \\[2mm] \exp\left[-\exp\left\{ -\dfrac{(y - \eta)}{\tau} \right\} \right], & \xi = 0, \end{cases}
\tag{1.2}
$$

where η and τ are, respectively, real location and positive scale parameters and the shape parameter ξ determines the weight of the upper tail of the density. These parameters may depend on the site x or on covariates indexed by x. For example, temperature depends strongly on altitude $a(x)$.

In many applications, it would be necessary to take semi-parametric formulations for dependence on x, but as we have data at relatively few sites, we use parametric forms. Davison & Ramesh (2000), Hall & Tajvidi (2000),

Pauli & Coles (2001), Chavez-Demoulin & Davison (2005) and Padoan & Wand (2008) describe approaches to more flexible modelling, applied by Ramesh & Davison (2002) and Butler *et al.* (2007).

Whatever the chosen regression structure, we shall suppose for simplicity of exposition that it has been used to transform the maxima to have the unit Fréchet distribution, $\exp(-1/z)$, for $z > 0$. If $\hat{G}_d(y)$ denotes the distribution at site x_d, estimated from the annual maxima, then the variables $Z_{dj} = -1/\log \hat{G}_d(Y_{dj})$, for the sites $d = 1, \ldots, 17$ and the years $j = 1, \ldots, n$, should possess approximately the unit Fréchet distribution. The quality of the approximation may be verified by quantile–quantile plots.

(*c*) *Layout*

The following section brings together elements from the statistics of extremes and from Gaussian geostatistics, and uses them to construct flexible models for spatial extremes for which pairwise joint density functions are available. Section 3 outlines how these may be used to perform exploratory analysis, and how information on the times of the maxima may be incorporated. It appears impossible to perform full likelihood inference for such models, and in §4, we discuss how composite marginal likelihood inference may be used for parameter estimation and model comparison. Section 5 describes its application to the Swiss temperature data, and some discussion is given in §6.

2. Spatial maxima

(*a*) *Max-stable processes*

The generalization of the classical multi-variate extreme-value distributions to the spatial case is a max-stable process, discussion of which is simplified under the assumption that all of its univariate marginal distributions are unit Fréchet. There is no loss of generality in making this assumption, which may be achieved by the marginal transformation described in §1*b*.

A max-stable process $\{Z(x)\}$ for $x \in \mathcal{X} \subset \mathbb{R}^2$ with unit Fréchet margins, also called a simple max-stable process, is a random field, all of whose marginal distributions satisfy the property of max-stability. That is, $\mathrm{pr}\{Z(x) \leq z\} = \exp(-1/z)$ for $z > 0$ and every $x \in \mathcal{X}$, and if $\mathcal{D} = \{x_1, \ldots, x_D\}$ is a finite disjoint subset of $\mathcal{X}$, then for $k = 1, 2, \ldots$ (de Haan 1984),

$$\mathrm{pr}\{Z(x_1) \leq kz_1, \ldots, Z(x_D) \leq kz_D\}^k$$
$$= \mathrm{pr}\{Z(x_1) \leq z_1, \ldots, Z(x_D) \leq z_D\}, \quad z_1, \ldots, z_D > 0, \qquad (2.1)$$

implying that the exponent measure function (Resnick 1987, ch. 5)

$$-\log \mathrm{pr}\{Z(x_1) \leq z_1, \ldots, Z(x_D) \leq z_D\} = V_{\mathcal{D}}(z_1, \ldots, z_D)$$

say, is homogeneous of order -1 and that for each $d \in \{1, \ldots, D\}$,

$$V_{\mathcal{D}}(\infty, \ldots, \infty, z_d, \infty, \ldots, \infty) = \frac{1}{z_d}.$$

Max-stable processes provide the natural generalization of models for the limiting distributions of multi-variate extremes to the process setting, in that the key property of max-stability is satisfied for all finite-dimensional marginal distributions, and the appropriate univariate marginal distributions, in this case unit Fréchet, are obtained.

There are several canonical representations of such processes (Schlather 2002; de Haan & Ferreira 2006, ch. 9). For our purposes, the most useful has two elements, a Poisson process Π of intensity ds/s^2 on $\mathbb{R}_+$ and a set of independent random processes $\{W_s(x) : x \in \mathcal{X}, s \in \mathbb{R}_+\}$, the latter being replicates of a non-negative process $\{W(x) : x \in \mathcal{X}\}$ with measure ν on $\mathcal{W} = \mathbb{R}_+^{\mathcal{X}}$ that satisfies $\mathrm{E}\{W(x)\} = 1$ for all x. We then define

$$Z(x) = \max_{s \in \Pi} s\, W_s(x), \quad x \in \mathcal{X}. \tag{2.2}$$

Following Schlather (2002), it is straightforward to verify that $Z(x)$ is max-stable, with unit Fréchet marginal distributions, and that for suitable functions $z(x)$ on $\mathcal{X}$,

$$\mathrm{pr}\{Z(x) \le z(x), x \in \mathcal{X}\} = \exp\left(-\mathrm{E}\left[\sup_{x \in \mathcal{X}} \left\{\frac{W(x)}{z(x)}\right\}\right]\right). \tag{2.3}$$

The very general expression (2.2) suggests several specific types of model, and in particular 'random storm' and 'random process' formulations. For the first, suppose that $\mathcal{X} = \mathbb{R}^2$ and that $W_s(x) = f(x - X_s)$, where f is a density function on $\mathcal{X}$, and X_s is a point of a Poisson process of unit rate in $\mathcal{X}$. Then, if we regard $s\,W_s(x)$ as the impact at x of a storm of magnitude s, centred at X_s, and of shape f, we interpret $Z(x)$ as the impact of the largest storm experienced at x. This formulation has been exploited by R. L. Smith (1990, unpublished data), Coles (1993), Coles & Tawn (1996), Coles & Walshaw (1994), Smith & Stephenson (2009) and Padoan *et al.* (2010) for the modelling of rainfall and maximum wind speeds. R. L. Smith (1990, unpublished data) obtained the marginal distribution for pairs $Z(x_1), Z(x_2)$ with f taken to be bivariate normal, and de Haan & Pereira (2006) extended this to Student t and Laplace densities.

A second type of max-stable model arises on taking $W(x)$ to be a stationary random process on $\mathcal{X}$. When modelling heatwaves, for example, we might take $W_s(x)$ to represent the temperature over $\mathcal{X}$ on the sth independent occasion of severity s, so that $Z(x)$ represents the overall maximum at x. We discuss this in §2c.

Difficulties with parametric inference on max-stable processes are immediately apparent on contemplating equation (2.3). In practice, data will be available at $\mathcal{D} = \{x_1, \ldots, x_D\} \subset \mathcal{X}$, and the joint distribution function for maxima observed on this set will be given by

$$\mathrm{pr}\{Z(x_1) \le z_1, \ldots, Z(x_D) \le z_D\} = \exp\{-V_{\mathcal{D}}(z_1, \ldots, z_D)\} \tag{2.4}$$

and

$$V_{\mathcal{D}}(z_1, \ldots, z_D) = \mathrm{E}\left[\sup_{d=1,\ldots,D} \left\{\frac{W(x_d)}{z_d}\right\}\right], \tag{2.5}$$

obtained from equation (2.3) by taking $z(x) = +\infty$ for $x \notin \mathcal{D}$. Two difficulties are associated with these expressions. First, a combinatorial explosion results if a full likelihood is formed by differentiation of the joint distribution function (2.4) with respect to $z_1, \ldots, z_D$: the number of terms is the Dth Bell number, around 8.3×10^{10} for $D = 17$ (Cameron 1994). Direct computation thus seems infeasible. Second, even if a suitable algorithm for computation of the necessary derivatives was available, the complexity of analytical computation of equation (2.5) rapidly mounts with D: it is feasible for a variety of simple random processes $W(x)$ when $D = 2$, but is possible only for special processes for larger D (Genton *et al.* 2011). Thus, standard likelihood-based inference appears to be out of reach.

(b) Extremal coefficients

A natural way to measure dependence among spatial maxima stems from considering the distribution of the largest value that might be observed on $\mathcal{X}$. If this distribution is close to unit Fréchet, then the maxima observed on $\mathcal{X}$ will be close to fully dependent, whereas lower degrees of dependence would yield stochastically larger distributions. To see the implications of this, we set $z(x) \equiv z$ in equation (2.3) and thus obtain

$$\mathrm{pr}\{Z(x) \leq z, \text{for all } x \in \mathcal{X}\} = \exp\left(-\frac{\theta_{\mathcal{X}}}{z}\right), \tag{2.6}$$

where

$$\theta_{\mathcal{X}} = \mathrm{E}\left\{\sup_{x \in \mathcal{X}} W(x)\right\} \tag{2.7}$$

is called the extremal coefficient corresponding to the set $\mathcal{X}$. It is more useful to consider the extremal coefficient $\theta_{\mathcal{D}}$ for a finite subset $\mathcal{D} = \{x_1, \ldots, x_D\}$ of $\mathcal{X}$, obtained by taking $z(x)$ to equal z for $x \in \mathcal{D}$ and infinity elsewhere; thus,

$$\theta_{\mathcal{D}} = \mathrm{E}\left\{\max_{x \in \mathcal{D}} W(x)\right\} = V_{\mathcal{D}}(1, \ldots, 1).$$

If $\theta_{\mathcal{D}} = 1$, then the maxima at $\mathcal{D}$ are perfectly dependent, whereas if $\theta_{\mathcal{D}} = D$, they are independent. Thus, $\theta_{\mathcal{D}}$ may be interpreted as the effective number of independent Frechét variables in the set indexed by $\mathcal{D}$. Schlather & Tawn (2003) consider the constraints that must be satisfied by the extremal coefficients corresponding to different subsets of $\mathcal{D}$, and show how the coefficients may be estimated both with and without these constraints.

Extremal coefficients are useful summaries of the joint tail behaviour of multi-variate extreme-value distributions, though they do not fully characterize dependence. We let $\theta(h)$ denote the extremal coefficient of $Z(x), Z(x + h)$; if the field is isotropic, then this depends only on the length of h. If replicate data are available, then the fact that $\max\{Z(x), Z(x + h)\}$ has distribution $\exp\{-\theta(h)/z\}$ enables maximum-likelihood estimation of $\theta(h)$ for each distinct pair of spatial sites (Schlather & Tawn 2003, §4.2). We use this below to check the fit of our models.

There is a close relation between the extremal coefficient and the madogram proposed by Poncet *et al.* (2006) and further developed by Naveau *et al.* (2009), an extreme-value analogue of the variogram (Cressie 1993, p. 58). The construction of a madogram requires a relatively large number of sites, and we do not pursue it here.

(c) Gaussian process models

A special case of equation (2.2) is to suppose, following Schlather (2002), that $W_s(x) = \max\{0, (2\pi)^{1/2}\varepsilon_s(x)\}$ is proportional to the non-negative part of a stationary Gaussian process $\{\varepsilon_s(x)\}$ with zero mean, unit variance and correlation function $\rho(x)$. In this case, the process $\{Z(x)\}$ is called an extremal Gaussian process, and its marginal bivariate distributions are determined by the exponent measure

$$V(z_1, z_2) = \frac{1}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left(1 + \left[1 - 2\frac{\{\rho(h) + 1\}z_1 z_2}{(z_1 + z_2)^2}\right]^{1/2}\right), \tag{2.8}$$

where $h = \|x_1 - x_2\|$. If $\rho(h) > -1$, then $Z(x_1)$ and $Z(x_2)$ are positively related; the extremal coefficient $\theta(h)$ for $\mathcal{D} = \{x_1, x_2\}$ equals $1 + 2^{-1/2}\{1 - \rho(h)\}^{1/2}$, so if the underlying Gaussian random field has $\rho(h) \to 0$ as $h \to \infty$, then $\lim_{h \to \infty} \theta(h) = 1 + 2^{-1/2}$. Moreover, as a positive definite isotropic correlation function on $\mathbb{R}^2$ can given correlations no smaller than -0.403, it follows that $\theta(h) \le 1.838$ for any h in the case of a spatial process. Thus, it is impossible to produce independent extremes using such a process, irrespective of the distance of the sites. This drawback may be overcome by taking

$$W_s(x) = c \max\{\varepsilon_s(x), 0\}I_{\mathcal{B}_s}(x - X_s), \tag{2.9}$$

where $I_{\mathcal{B}}$ is the indicator function of a compact random set $\mathcal{B} \subset \mathcal{X}$, of which the $\{\mathcal{B}_s\}$ are independent replicates, and the $\{X_s\}$ are the points of a homogeneous Poisson process of unit rate on $\mathcal{X}$, independent of the ε_s. Then modulo minor edge effects, if $c^{-1} = \mathrm{E}[\max\{W(x), 0\}]\mathrm{E}(|\mathcal{B}|)$, then the process (2.2) is again max-stable on $\mathcal{X}$. If $W_s(x)$ is an extremal Gaussian process, as above, then equation (2.8) becomes

$$V(z_1, z_2) = \left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left\{1 - \frac{\alpha(h)}{2}\left(1 - \left[1 - 2\frac{\{\rho(h) + 1\}z_1 z_2}{(z_1 + z_2)^2}\right]^{1/2}\right)\right\}, \tag{2.10}$$

where $\alpha(h) = \mathrm{E}\{|\mathcal{B} \cap (h + \mathcal{B})|\}/\mathrm{E}(|\mathcal{B}|)$ lies in the unit interval. The extremal coefficient for this model,

$$\theta(h) = 2 - \alpha(h)[1 - 2^{-1/2}\{1 - \rho(h)\}^{1/2}], \tag{2.11}$$

can take any value in the interval $[1, 2]$. Independent extremes can be generated if the distribution of $\mathcal{B}$ is chosen so that $\mathcal{B} \cap (h + \mathcal{B}) = \emptyset$ with probability one for large enough h, as in that case $\alpha(h) = 0$. Thus, equation (2.9) has the appealing interpretation that the effects of the event corresponding to $sW_s(x)$ are felt only over the set $\mathcal{B} + X_s$.

In §2*d*,*e*, we discuss some possible choices of $\rho(h)$ and $\alpha(h)$. Below, we merely sketch some generic possibilities for modelling: more plausible forms will be prompted by consideration of particular applications.

(d) Geostatistics

Geostatistics is based largely on the theory of Gaussian random processes (see Cressie 1993; Stein 1999; Wackernagel 2003; Banerjee *et al.* 2004; Schabenberger & Gotway 2005; Diggle & Ribeiro 2007; Cressie & Wikle 2011). For a stationary Gaussian random field $\{S(x)\}$ defined on $x \in \mathbb{R}^2$ and having standard normal marginal distributions, the joint distribution of the field at any subset of points $\mathcal{D} = \{x_1, \ldots, x_D\}$ must be multi-variate normal. The key element is the correlation function $\rho(x_2 - x_1) = \mathrm{cov}\{S(x_1), S(x_2)\}$, which is constrained to be positive definite. If the field is also isotropic, then the correlation depends only on the distance $h = \|x_2 - x_1\|$, and for the rest of this subsection we assume this and write $\rho(x_2 - x_1) = \rho(h)$.

Standard families of correlation functions include the powered exponential family

$$\rho(h) = \exp\left\{-\left(\frac{h}{\lambda}\right)^\kappa\right\}, \quad h \geq 0, \ \lambda > 0, \ 0 < \kappa \leq 2, \tag{2.12}$$

the Whittle–Matérn family

$$\rho(h) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}\left(\frac{h}{\lambda}\right)^\kappa K_\kappa\left(\frac{h}{\lambda}\right), \quad h \geq 0, \ \lambda, \kappa > 0, \tag{2.13}$$

in which K_κ denotes the modified Bessel function of order κ, and the Cauchy family

$$\rho(h) = \frac{1}{1 + (h/\lambda)^2}, \quad h \geq 0, \ \lambda > 0. \tag{2.14}$$

In all three families, λ represents a scale parameter with the dimensions of distance, and in the first two families, κ is a shape parameter that controls the properties of the random field and in particular the roughness of its realizations. Visually smooth realizations arise only when $\rho(h)$ is sufficiently well behaved at the origin. Rather than trying to estimate the shape parameter based on limited data, it may be more practicable to choose among a small set of values of κ that represent qualitatively different behaviour of the Gaussian process. The Matérn correlation family is the more flexible and is strongly advocated by Stein (1999), but in practice equations (2.12)–(2.14) may be hard to distinguish.

The correlation functions above impose $\rho(h) \to 1$ as $h \to 0$, but this is unrealistic in many applications. One way to relax this condition is to include a so-called nugget effect by adding to $S(x)$ an independent Gaussian white noise random field. The resulting correlation function is

$$(1 - \nu)\delta(h) + \nu\rho(h), \quad 0 \leq \nu \leq 1, \tag{2.15}$$

where $\delta(h)$ is the Kronecker delta function, $\rho(h)$ is a correlation function and the nugget effect $1 - \nu$ is the proportion of the variance stemming from the local variation represented by $\{\epsilon(x)\}$.

As these correlation functions have $\rho(h) > 0$, the extremal coefficient based on equation (2.8) satisfies $1 \leq \theta(h) < 1.71$; as mentioned previously, independence of the extremal Gaussian processes resulting from using equations (2.12) and (2.13), either as such or in equation (2.15), cannot be achieved.

The simplest departure from isotropy, geometric anisotropy, presumes that there is an invertible matrix V such that the process $Z(Vx)$ is isotropic, and corresponds to an affine transformation of the plane containing the sites.

(e) The set $\mathcal{B}$

Detailed modelling of the random set $\mathcal{B}$ that appears in equation (2.9) will require data at many sites and appears to be difficult. In practice, it may be preferable to use flexible parametric shapes that generate simple forms for the parameter $\alpha(h)$ appearing in equation (2.10). An obvious possibility is to take $\mathcal{B}$ to be a disc of fixed radius r. Then, the area of intersection of $\mathcal{B}$ and $h + \mathcal{B}$ is

$$|\mathcal{B} \cap (h + \mathcal{B})| = \begin{cases} 2r^2 \left[\cos^{-1}\left\{ \dfrac{|h|}{2r} \right\} - \left(\dfrac{|h|}{2r} \right) \left\{ 1 - \dfrac{|h|^2}{2r^2} \right\}^{1/2} \right], & 0 \leq |h| \leq 2r, \\ 0, & |h| > 2r, \end{cases}$$

where $|h|$ is the length of the vector h. This function is close to linear in $|h|$ over much of its range, and a crude approximation is to replace it by $\pi r^2 \{1 - |h|/(2r)\}_+$. This yields $\alpha(h) = \mathrm{E}\{|\mathcal{B} \cap (h + \mathcal{B})|\}/\mathrm{E}(|\mathcal{B}|) \doteq \{1 - |h|/(2r)\}_+$, a model with one unknown parameter r that ensures the independence of extremes at sites for which $|h| > 2r$.

It may be thought to be more realistic to take sets of random size. A simple possibility is to take the above model, but to suppose that the radius of the disc is random with a generalized Pareto density function

$$g(r) = \left(1 + \frac{\gamma r}{\delta} \right)_+^{-1/\gamma - 1}, \quad r > 0. \tag{2.16}$$

Thus, g is positive only for $0 < r < r_{\max}$, where $r_{\max} = \infty$ if $\gamma \geq 0$ and $r_{\max} = -\delta/\gamma$ otherwise. For $\gamma > 0$, the density g is heavy-tailed, for $\gamma = 0$, it is exponential, and for $\gamma < 0$, it has a finite upper support point $r_{\max}$, with $\gamma = -1$ corresponding to the uniform distribution, and discs of constant radius $r_{\max}$ appearing as $\gamma \to -\infty$. With the linear approximation to the overlap area mentioned above and with this choice of g, after some further integration, we find that provided $\gamma < 1/2$,

$$\alpha(h) \approx \begin{cases} \left(1 + \dfrac{|h|}{4\delta} \right) \left(1 + \dfrac{\gamma |h|}{2\delta} \right)_+^{1 - 1/\gamma}, & \gamma \neq 0, \\ \left(1 + \dfrac{|h|}{4\delta} \right) \exp\left\{ -\dfrac{|h|}{2\delta} \right\}, & \gamma = 0. \end{cases} \tag{2.17}$$

When $\gamma \geq 1/2$, the expectation is infinite, corresponding to $\alpha(h) \equiv 1$.

A natural generalization is to suppose that the set $\mathcal{B}$ is the interior of an ellipse having area πr^2 and eccentricity ϵ, and whose major axis subtends an angle ω with the horizontal axis. If $|h|$ and φ are the polar coordinates of h, then the elliptical sets $\mathcal{B}$ are accounted for by replacing $|h|$ in equation (2.17) by

$$|h|(1 - \epsilon^2)^{-1/4}\{1 - \epsilon^2 \cos^2(\varphi - \omega)\}^{1/2}, \quad 0 \leq \omega < \pi, \ 0 \leq \epsilon < 1.$$

This provides a four-parameter family of sets, whose shapes are determined by ϵ and ω, and the distribution of whose sizes is determined by γ and δ.

3. Pairwise analysis

(*a*) *Direct analysis*

Before discussing the fitting of the models described above to data from a number of sites $x_1, \ldots, x_D$, we consider exploratory data analysis based on data from pairs of sites.

We suppose that the data have been transformed to have unit Fréchet marginal distributions. We suppose that data pairs (z_{1j}, z_{2j}) $(j = 1, \ldots, n)$ are available, and take them to be independent realizations of variables $(Z(x_1), Z(x_2))$ from a max-stable random field. In an environmental context, such data might be annual maxima recorded at the sites x_1 and x_2 in n successive years, and for ease of exposition, we shall use corresponding language.

The marginal pairwise density of variables Z_1, Z_2 is obtained from equation (2.4) as

$$f(z_1, z_2) = \left\{ \frac{\partial V(z_1, z_2)}{\partial z_1} \frac{\partial V(z_1, z_2)}{\partial z_2} - \frac{\partial^2 V(z_1, z_2)}{\partial z_1 \partial z_2} \right\} \exp\{-V(z_1, z_2)\}, \quad z_1, z_2 > 0. \tag{3.1}$$

Given a parametric form for V such as equations (2.8) or (2.10), the log likelihood is then readily computed, and standard methods may be used to obtain estimates and confidence intervals (CIs).

Figure 3*a*,*b* shows the results of such fits for the annual maximum temperature data described in §2*b*. The parameter estimates when model (2.10) is fitted to the standardized Fréchet observations from all 136 distinct pairs of sites are plotted against the inter-site distances, with pairs involving the two southern sites Locarno-Monti and Lugano in canton Ticino, and the highest, the Jungfraujoch, indicated differently. The approximate 95% CIs are obtained using profile likelihood. There is a clear general decline of ρ with distance, but the values of α mostly equal unity; the latter is not surprising, as hot days tend to result from large-scale climatic events. Aberrant pairs involving the Jungfraujoch and the sites in Ticino are visible in both graphs, though the uncertainty is large.

(*b*) *Inclusion of event times*

The approach to estimation of the parameters for each pair sketched in §3*a* and based on forming a log likelihood by adding contribution (3.1) may be improved if the times of maxima are available. In this case, arguments of Stephenson & Tawn (2005) show that if the maxima at the two sites are known to occur simultaneously, then the likelihood contribution should be

$$f_{\text{sim}}(z_1, z_2) = -\frac{\partial^2 V(z_1, z_2)}{\partial z_1 \partial z_2} \exp\{-V(z_1, z_2)\}, \quad z_1, z_2 > 0, \tag{3.2}$$

whereas if they are known to occur at different times, then equation (3.2) should be replaced by

$$f_{\text{sep}}(z_1, z_2) = \frac{\partial V(z_1, z_2)}{\partial z_1} \frac{\partial V(z_1, z_2)}{\partial z_2} \exp\{-V(z_1, z_2)\}, \quad z_1, z_2 > 0. \tag{3.3}$$
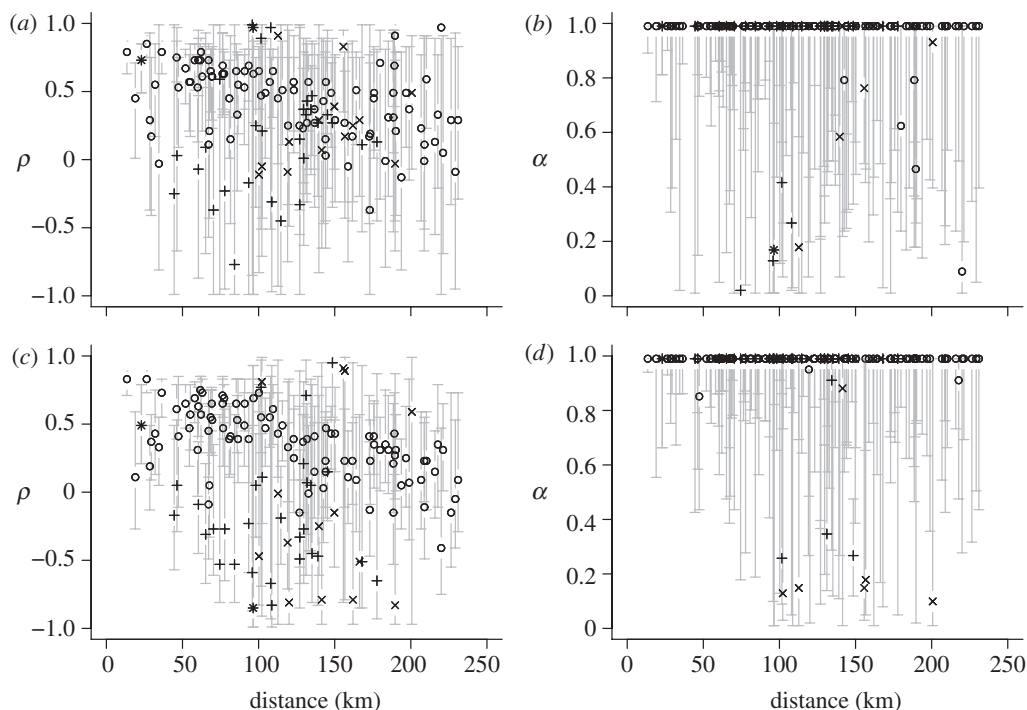
Figure 3. Pairwise analysis of annual maximum temperature data for the sites shown in figure 1. Maximum-likelihood estimates of parameters (*a*) ρ and (*b*) α plotted against inter-site distance, with approximate 95% CIs (grey) obtained by profile likelihood. The panels (*c,d*) are the same, but with estimation improved by inclusion of the occurrence times. (*a–d*) Pairs involving the sites Lugano and Locarno-Monti in canton Ticino, south of the Alps, are shown as plus symbols, and pairs involving the Jungfraujoch, the highest station, are shown as cross symbols.

Use of these expressions may give improved efficiency relative to use of equation (3.1). The potential gains are illustrated by comparing figure 3*a,b* and figure 3*c,d*, of which figure 3*c,d* are constructed using equations (3.2) and (3.3). When the times are used, there is a general shrinking of the CIs, and it becomes clearer that the Ticino sites and Jungfraujoch are somewhat disconnected from the others, though the CIs remain too wide to draw firm conclusions. However, some degree of disconnection is entirely plausible on general grounds. At an altitude of 3580 m, the Jungfraujoch is above the snow line all year round, relatively little influenced by human activities, and more than a kilometre higher than the next highest site, Santis. The Ticino is more strongly influenced by meterological conditions in northern Italy and the Mediterranean basin than by those within and north of the Alps.

(*c*) Model checking

A check on the quality of a parametric model such as equation (2.10) is to compare its fitted extremal coefficients, obtained by replacing α and ρ in equation (2.11) by their maximum-likelihood estimates, with the naive estimator of Schlather & Tawn (2003) or the madogram-based estimator of Poncet *et al.*
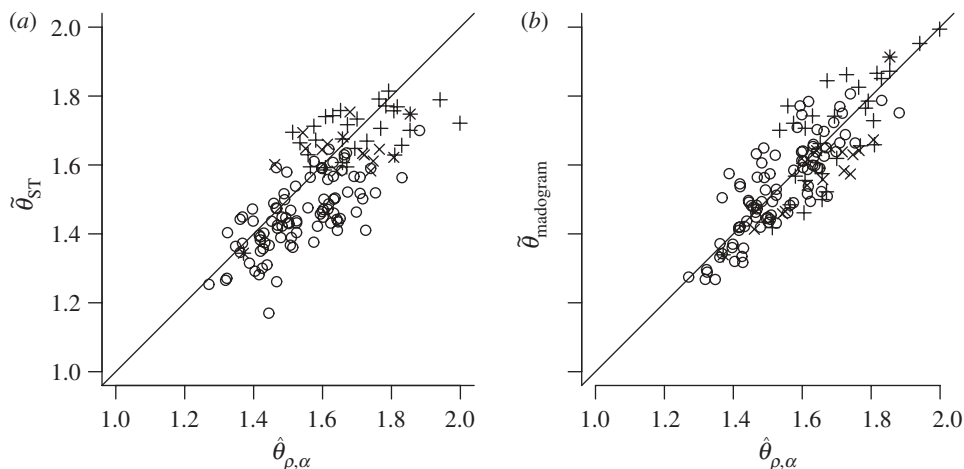
Figure 4. Non-parametric and model-based estimates of the extremal coefficients θ for all possible pairs of sites for the temperature data. The non-parametric estimators are (a) $\tilde{\theta}_{ST}$ (Schlather & Tawn 2003) and (b) $\tilde{\theta}_{madogram}$ (Poncet *et al.* 2006). The model-based estimates $\hat{\theta}_{\rho,\alpha}$ are based on the model (2.10). Pairs involving the sites Lugano and Locarno-Monti are shown as plus symbols, and pairs involving the Jungfraujoch are shown as cross symbols.

(2006). The latter two are not based on a specific parametric model. Figure 4, which shows these comparisons for the maximum temperature data, reveals nothing to cast doubt on the model (2.10).

One check on the fit of the pairwise model stems from noting from equation (2.6) that the maximum $\max(Z_1, Z_2)$ of two related variables, each with the unit Fréchet marginal distribution, has a scaled Fréchet distribution $\exp(-\theta_{1,2}/z)$, for $z > 0$, where $\theta_{1,2} = V(1,1)$. This suggests inspecting an exponential quantile plot of the sample $\hat{\theta}_{1,2}/\max(z_{1j}, z_{2j})$ $(j = 1, \ldots, n)$, where $\hat{\theta}_{1,2} = \hat{V}(1,1)$ is an estimate of $\theta_{1,2}$. We have found that the maximum-likelihood estimate based on the data pairs (Z_1, Z_2) does not work well; better results are obtained with the Schlather–Tawn or madogram-based estimators mentioned in the previous paragraph. In the present case, such plots reveal little untoward.

4. Composite marginal likelihood

(a) Introduction

Statistical inference for parametric models is ideally performed using the likelihood function, but for multi-variate extremal models, this is generally difficult for the reasons mentioned at the end of §2a. When the pairwise marginal distributions are available and identify the model parameters, it therefore seems natural to use a composite likelihood function (Lindsay 1988; Cox & Reid 2004; Varin *et al.* 2011).

Suppose that the distribution of the responses depends on a parameter vector ϑ, let $z_{\mathcal{Y}} = \{z_j : j \in \mathcal{Y}\}$ denote a subset of the overall data, and suppose that the responses may be split into mutually independent subsets $z_{\mathcal{Y}_1}, \ldots, z_{\mathcal{Y}_n}$. Also, let $\mathcal{Y}_{j,1}, \ldots, \mathcal{Y}_{j,m_j}$ denote distinct subsets of $\mathcal{Y}_j$. In the temperature data of §1b,

the $\mathcal{Y}_j$ might denote maxima for different summers $j = 1, \ldots, n$, which may be supposed independent, and the $\mathcal{Y}_{j,1}, \ldots, \mathcal{Y}_{j,m_j}$ might denote all distinct pairs of sites for which data are available in summer j. The corresponding composite marginal log likelihood is

$$\ell_{\mathcal{Y}}(\vartheta) = \sum_{j=1}^{n} \sum_{i=1}^{m_j} \log f(z_{\mathcal{Y}_{j,i}}; \vartheta) \qquad (4.1)$$

in a natural notation. Under regularity conditions like those needed for the limiting normality of the standard maximum-likelihood estimator, and if ϑ is identifiable from the marginal densities contributing to $\ell_{\mathcal{Y}}$, the maximum pairwise likelihood estimator $\tilde{\vartheta}$ has a limiting normal distribution as $n \to \infty$, with mean ϑ and covariance matrix of sandwich form estimable by $J(\tilde{\vartheta})^{-1} K(\tilde{\vartheta}) J(\tilde{\vartheta})^{-1}$, where

$$J(\vartheta) = -\sum_{j=1}^{n} \sum_{i=1}^{m_j} \frac{\partial^2 \log f(z_{\mathcal{Y}_{j,i}}; \vartheta)}{\partial \vartheta \partial \vartheta^{\mathrm{T}}}$$

and

$$K(\vartheta) = -\sum_{j=1}^{n} \sum_{i=1}^{m_j} \frac{\partial \log f(z_{\mathcal{Y}_{j,i}}; \vartheta)}{\partial \vartheta} \sum_{i=1}^{m_j} \frac{\partial \log f(z_{\mathcal{Y}_{j,i}}; \vartheta)}{\partial \vartheta^{\mathrm{T}}}$$

are the observed information and squared score statistic corresponding to $\ell_{\mathcal{Y}}$. Corresponding results for comparison of nested models fitted by maximization of equation (4.1) involve sums of independent χ^2 variables weighted by eigenvalues of matrices derived from $K(\tilde{\vartheta})$ and $J(\tilde{\vartheta})$ (Kent 1982), and so are awkward to apply in practice. In a similar situation, Chandler & Bate (2007) suggest modifications which ensure that likelihood ratio statistics retain their usual large sample distributions.

The marginal composite likelihood (4.1) is useful only for inference on model parameters that are estimable from the terms appearing therein. Thus, in the case of a pairwise likelihood, in which each of the subsets $\mathcal{Y}_{j,i}$ corresponds to pairs of observations, parameters that appear only in the joint densities for triples, quartets, etc., cannot be estimated. This is not a difficulty with the models discussed below or in other spatial contexts based on Gaussian distributions, for which the distribution is specified in terms of its first two moments, but it could be a drawback in other contexts; in that case, the use of contributions from, e.g. triplets of observations, might be indicated. There is also a potential gain of efficiency in using higher order terms (Genton *et al.* 2011).

Varin & Vidoni (2005) and Varin (2008) discuss model selection based on composite likelihoods. It is appropriate to select models that minimize the composite likelihood information criterion, CLIC $= -2[\ell_{\mathcal{Y}}(\tilde{\vartheta}) + \mathrm{tr}\{K(\tilde{\vartheta}) J(\tilde{\vartheta})^{-1}\}]$, an adaptation of the Takeuchi information criterion (Takeuchi 1976) to the present setting. One minor difficulty with this, and indeed in comparing models based on $\ell_{\mathcal{Y}}$, is that even for moderate datasets, the values of $\ell_{\mathcal{Y}}$ tend to be very large because of the large numbers of terms summed in equation (4.1). It is therefore helpful to replace $\ell_{\mathcal{Y}}$ by $\ell_{\mathcal{Y}}^* = c\ell_{\mathcal{Y}}$, where the scaling constant c is chosen to give the correct log likelihood, at least approximately, for independent observations, and to use the corresponding scaled information criterion, CLIC*.

Composite likelihood may be seen as the basis for estimating the equation $\partial\ell_Y(\vartheta)/\partial\vartheta = 0$, but has the advantage over arbitrarily defined estimating equations of stemming from an objective function. As Varin *et al.* (2011) describe, composite likelihood is increasingly being used in settings where the likelihood itself is computationally intractable. In cases where the composite maximum-likelihood estimator $\tilde{\vartheta}$ can be compared with the usual maximum-likelihood estimator, the former is often surprisingly efficient, though if n is fixed, care must be taken not to allow m_j to become too large; if so, the efficiency can drop, and in extreme cases, $\tilde{\vartheta}$ may become inconsistent (Cox & Reid 2004). In our applications, the number n of independent blocks is typically fairly large, so even if the m_j are large, inconsistency should not usually be an issue.

(b) Efficiency comparison

In the present context, it is natural to use the bivariate marginal densities corresponding to equations (2.8), (2.10), (3.2) and (3.3), thus forming the pairwise log likelihood

$$\ell_2(\vartheta) = \sum_{j=1}^{n} \sum_{c,d \in \mathcal{Y}_{j,i}} \log f(z_{j,c}, z_{j,d}; \vartheta),$$

in which the $\mathcal{Y}_{j,i}$ comprise all distinct pairs of sites. One natural concern is the efficiency of the maximum pairwise likelihood estimator $\tilde{\vartheta}$, but it is impossible to assess this directly, because the full log likelihood is unavailable for such models. Instead, we simulated data from a D-dimensional logistic model for multi-variate extremes (Tawn 1988), which has the distribution function

$$\exp\left\{ -\left(\sum_{d=1}^{D} z_d^{1/\alpha} \right)^{\alpha} \right\}, \quad z_1, \ldots, z_D > 0, \ \ 0 < \alpha \leq 1, \tag{4.2}$$

with $z_d = \{1 + \xi(y_d - \eta)/\tau\}^{-1/\xi}$; here, η, τ and ξ are location, scale and shape parameters, all taken equal to 1 for the simulation. The extremal coefficient of equation (4.2) is D^α: the extremes are independent for $\alpha = 1$, and become entirely dependent when $\alpha \to 0$. Both pairwise and full likelihood estimation can be performed for this model, with the likelihoods obtained by symbolic differentiation of equation (4.2). Table 1 shows the efficiencies for pairwise likelihood relative to full likelihood estimation for the four parameters, as the dimension D varies from 3 to 40. When there is appreciable dependence, the loss of information seems to be slight, but even for weak dependence, it is not catastrophic. We conclude that pairwise likelihood estimation is a reasonable option in the present context; in any case, there seems to be no other.

(c) Computation

Composite marginal likelihood fitting of certain of the geostatistical models for extremes described above and of the Gaussian random storm model of R. L. Smith (1990, unpublished data) may be performed using Mathieu Ribatet's R package

Table 1. Relative efficiency (%) of the pairwise likelihood estimator based on 500 datasets of size 200 simulated from a symmetric logistic distribution of dimension D with dependence parameter $\alpha = 0.1, 0.4, 0.7, 0.9$. The parameters η, τ and $\xi = 1$ are the location, scale and shape parameters of the common marginal distribution.

D	$\alpha = 0.1$				$\alpha = 0.4$			
	η	τ	ξ	α	η	τ	ξ	α
3	100	99	91	93	100	100	94	94
5	100	99	84	97	100	100	79	93
10	100	100	74	93	100	99	65	85
20	100	100	70	93	100	98	57	82
40	100	99	68	94	100	98	54	83
D	$\alpha = 0.7$				$\alpha = 0.9$			
	η	τ	ξ	α	η	τ	ξ	α
3	100	99	94	91	99	97	96	90
5	100	96	76	78	97	92	86	76
10	99	90	59	63	92	82	69	55
20	98	85	43	53	85	70	49	40
40	98	85	37	50	79	61	36	31

SPATIALEXTREMES, which allows flexible formulations of the location, scale and shape parameters of the marginal generalized extreme-value distributions, and provides appropriate plots and diagnostics.

5. Temperature data

(a) Model fitting

We now discuss the fitting of model (2.8) to the temperature data described in §1*b*. The pairwise analysis in §3*a* suggests that the sites at Locarno-Monti, Lugano and the Jungfraujoch have different meteorological behaviour than the others, and consequently we exclude them from the fitting, thus giving $D = 14$ sites (see below).

Exploratory analysis for the marginal parameters shows relations between the location parameter of the generalized extreme-value distribution (1.2) and latitude, altitude and time. We first fitted various different models with the powered exponential covariance (2.12) using the scaled pairwise likelihood

$$\ell_2^*(\vartheta) = (D-1)^{-1} \sum_{j=1}^{n} \sum_{d_1=1}^{D-1} \sum_{d_2=d_1+1}^{D} \ell\{g(y_{d_1,j}; \vartheta), g(y_{d_2,j}; \vartheta); \vartheta\}, \qquad (5.1)$$

where $y_{d,j}$ represents the maximum at site d in year j, $g(y; \vartheta) = \{1 + \xi(y - \eta)/\tau\}^{1/\xi}$ and the marginal parameters η, τ and ξ are functions of the generic parameter vector ϑ. Some of the many models that we fitted and the values of their information criteria are shown in table 2. The dependence models incorporate

Table 2. Values of information criteria CLIC* for some dependence models and Akaike information criteria for an independence model. κ is the shape parameter of the powered exponential covariance function used for the spatial fits.

model	number of marginal parameters	shape κ	information criterion
dependence			
$\eta, \tau, \xi \sim \text{time} + \text{alt}^2 + (\text{lat} + \text{lon})^2$	27	1.5	2051.0[a]
$\eta, \tau, \xi \sim \text{alt}^2 + (\text{lat} + \text{lon})^2$	24	1.5	2077.7
$\eta, \tau, \xi \sim \text{time} + (\text{lat} + \text{lon})^2$	21	1.5	3593.5
$\eta, \tau, \xi \sim \text{time} + \text{alt}^2 + \text{lat}^2$	18	1.5	2079.8
$\eta, \tau, \xi \sim \text{time} + \text{alt}^2 + \text{lon}^2$	18	1.5	2119.1
$\eta \sim \text{time} + \text{alt}^2 + (\text{lat} + \text{lon})^2; \tau, \xi \sim 1$	11	1.5	2101.4
$\eta \sim \text{alt}^2 + (\text{lat} + \text{lon})^2; \tau, \xi \sim 1$	10	1.5	2125.0
$\eta \sim \text{time} + \text{alt}^2 + \text{lat}^2; \tau, \xi \sim 1$	8	0.5	2086.1
$\eta \sim \text{time} + \text{alt}^2 + \text{lat}^2; \tau, \xi \sim 1$	8	1.0	2085.7
$\eta \sim \text{time} + \text{alt}^2 + \text{lat}^2; \tau, \xi \sim 1$	8	1.5	2073.2[b]
$\eta \sim \text{time} + \text{alt}^2 + \text{lat}^2; \tau, \xi \sim 1$	8	1.9	2085.6
independence			
$\eta \sim \text{time} + \text{alt}^2 + \text{lat}^2; \tau, \xi \sim 1$	8	—	2347.4

[a]The overall best model.
[b]Our preferred model.

the dependence among the maxima by fitting the log likelihood based on equation (3.1), whereas the independence model incorrectly treats the spatial data as independent series. There is a very large reduction in the CLIC* owing to inclusion of quadratic terms in altitude and a smaller one owing to time. The effect of latitude appears stronger than that of longitude, and this seems natural in view of the geography of the Alps. The overall best model uses quadratic terms in longitude, latitude and altitude of the sites, plus a linear term in time, to explain variation in all three marginal parameters of equation (1.2), but this seems very complex for a fit to 14 sites, and we prefer a simpler model with just eight marginal parameters. The other models vary the shape parameter κ of the exponential covariance function used for these fits. Our preferred model, denoted by superscript 'b', has constant scale and shape parameters, and

$$\eta_{d,t} = \eta_0 + \eta_1 \text{lat}(x_\text{d}) + \eta_2 \text{lat}(x_\text{d})^2 + \eta_3 \text{alt}(x_\text{d}) + \eta_4 \text{alt}(x_\text{d})^2 + \eta_5 t, \qquad (5.2)$$

where $\text{alt}(x)$ and $\text{lat}(x)$ are altitude above mean sea level (in kilometres), and latitude, measured in a coordinate system centred on the former national observatory in Bern. Time t is measured in centuries since the year 2000. The corresponding independence model has an appreciably higher information criterion value, suggesting that, as one might anticipate, the spatial dependence model better accounts for the data.

Table 3 shows the estimates and standard errors for the model with margins (5.2), when the data from the 14 sites are treated as independent, using the max-stable model with likelihood contributions (3.1) not allowing for

A. C. Davison and M. M. Gholamrezaee

Table 3. Parameter estimates (s.e.) for annual maximum temperature data from 14 sites, treated both as independent and as dependent. In the latter case, we fit a geostatistical max-stable model using the powered exponential covariance with $\kappa = 1.5$, both ignoring event times and taking them into account, using pairwise likelihood estimation based on likelihood contributions (3.1)–(3.3), respectively.

				dependence	
			independence	equation (3.1)	equations (3.2) and (3.3)
location	intercept	η_0	32.49 (0.34)	32.50 (0.41)	32.79 (0.41)
	latitude	η_1	-0.31 (0.15)	-0.20 (0.13)	-0.33 (0.13)
	latitude^2	η_2	1.21 (0.29)	1.53 (0.19)	1.46 (0.18)
	altitude	η_3	-1.42 (0.55)	-0.84 (0.38)	-1.24 (0.36)
	altitude^2	η_4	-2.04 (0.20)	-2.26 (0.13)	-2.13 (0.12)
	time	η_5	0.49 (0.52)	2.47 (1.44)	2.51 (1.38)
scale		τ	1.51 (0.05)	1.51 (0.11)	1.61 (0.11)
shape		ξ	-0.17 (0.02)	-0.15 (0.04)	-0.14 (0.04)
dependence	scale	λ	—	175.1 (67.1)	189.9 (24.9)
	nugget	$1-\nu$	—	0.37 (0.06)	0.36 (0.04)

the timing of maxima, and allowing for their timing. In the last case, we used likelihood contributions (3.2) and (3.3), with maxima considered to form part of the same episode, only if they occurred on the same day; the effect of varying this by considering that a hot episode could take place over several days was slight. Most of the estimates are similar for the independence and dependence models, with the latter giving larger standard errors for the marginal intercept parameters η_0, τ, ξ and smaller standard errors for most of the marginal regression effects. The exception is the parameter η_5 representing dependence on time, the estimate of which increases and becomes more significant when spatial dependence is taken into account, though none of the estimates differs significantly from zero. There is some reduction in uncertainty when information on the synchronicity of extreme events is included, particularly for the scale parameter λ of the correlation function.

The table provides plausible physical values for the effect of altitude, which is known to reduce temperatures by around 7°C for each kilometre of climb; the gradients implied by the quadratic curve fitted here range from -6°C to -9°C at altitudes of 1–2 km; both the linear and quadratic terms in altitude are highly significant, but the addition of a cubic term is not. The effect of climatic change is also consistent with estimates from other sources, suggesting an increase of around 2.5°C over the next century. The large uncertainty of this estimate is not surprising, as the data encompass a small geographical region and are strongly correlated. The effect of latitude is most likely a surrogate for complex effects of topography beyond that of altitude.

We also fitted equation (2.8) with correlation function (2.15) and the Whittle–Matérn correlation family, with results very similar to those described above.

Figure 5 shows the fitted extremal coefficient curves for various models; there is little to choose between them, though the model with individual
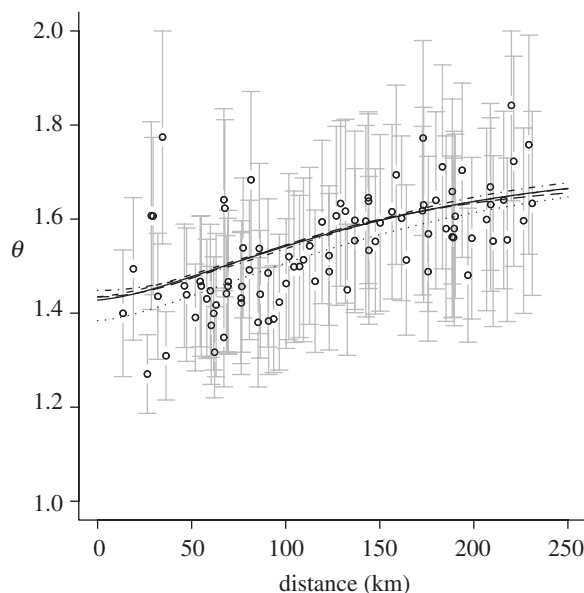
Figure 5. Fitted extremal coefficients as a function of distance for the pairs of sites, excluding Jungfraujoch, Locarno-Monti and Lugano, with Schlather–Tawn pairwise estimates and 95% CIs (shown as circles with grey vertical lines). Shown are curves for the exponential covariance function with shape $\kappa = 1.5$ (solid line) and with shape $\kappa = 1.9$ (dotted-dashed line); for the exponential covariance with individual marginal parameters for different sites (dotted line), and with 27 marginal parameters (dashed line); and for the Cauchy covariance function (long-dashed line).

marginal parameters at every site is somewhat distinct from the others. This is unsurprising, but as this model does not allow extrapolation to other sites, it is of little practical use in risk analysis.

(b) Model checking

The max-stable models are fitted using data from pairs of sites, and it is natural to assess the quality of the fit by using data from larger subsets. To do so, we perform a simulation-based test, as follows. We first note that the annual maxima $Z_{\mathcal{A}} = \max_{d \in \mathcal{A}} Z_d$ are available for any subset $\mathcal{A} \subset \mathcal{D}$ of the sites at which data are available, and that the distribution of $Z_{\mathcal{A}}$ under the max-stable model is Fréchet with parameter $\theta_{\mathcal{A}}$. The empirical values of $Z_{\mathcal{A}}$ are available for n independent years of data; call these $z_{\mathcal{A},1}, \ldots, z_{\mathcal{A},n}$. They may be compared either with the Fréchet distribution with parameter $\theta_{\mathcal{A}}$ or with maxima for datasets simulated from the fitted model. In the second case, we simulate R independent sets of data from the fitted max-stable model, and thus obtain R replicates $z_{\mathcal{A},1,r}^*, \ldots, z_{\mathcal{A},n,r}^*$ of the original data. Simple graphical tests of fit may be based on the concordance of the ordered values of $z_{\mathcal{A},1}, \ldots, z_{\mathcal{A},n}$ with the ordered values of $z_{\mathcal{A},1,r}^*, \ldots, z_{\mathcal{A},n,r}^*$.

Figure 6 shows the use of $R = 5000$ simulated max-stable processes to form simulation envelopes (Davison & Hinkley 1997, section 4.2.4) with six groups of sites, formed geographically: (a) Jungfraujoch; (b) Western Alps; (c) Ticino;
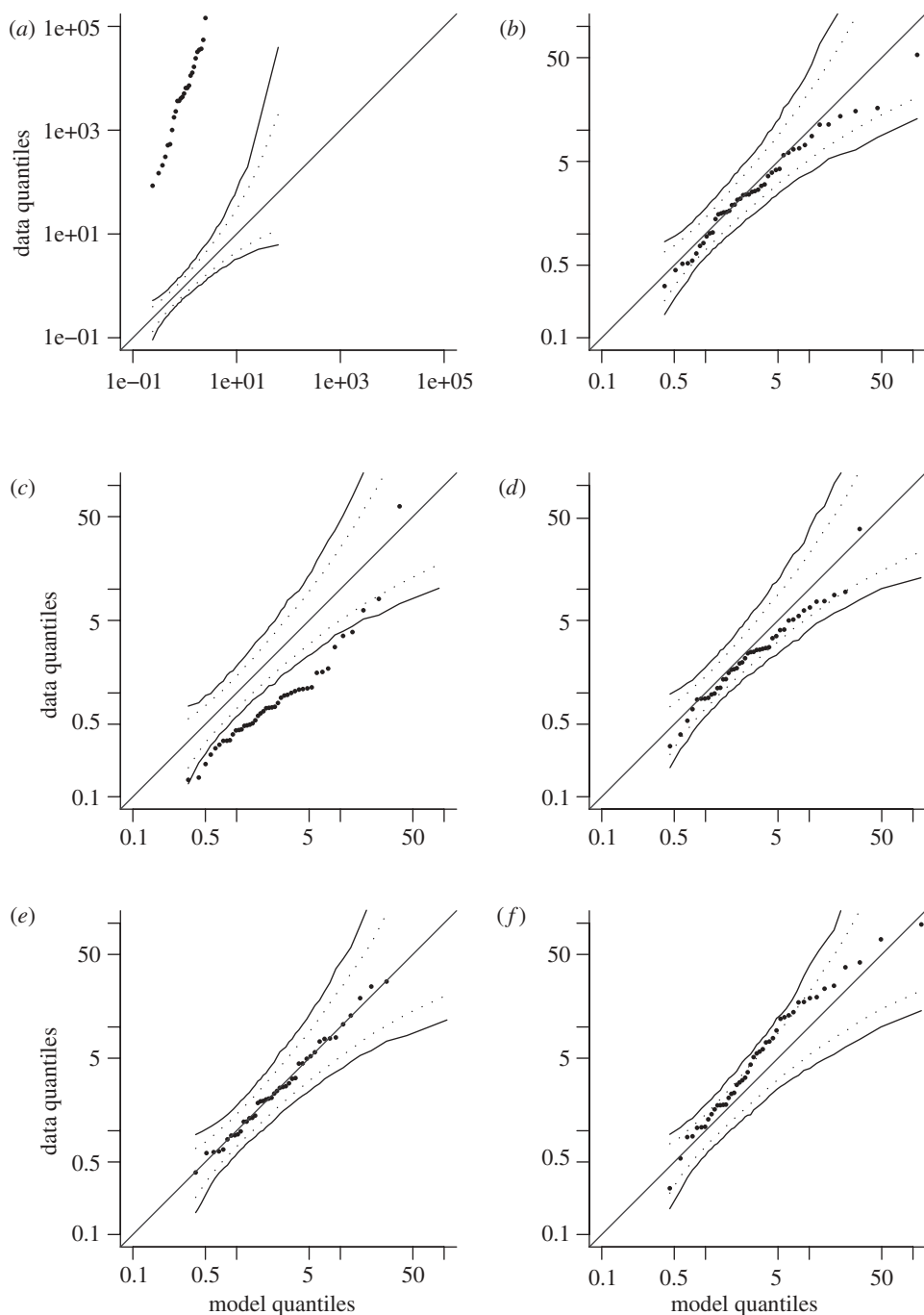
Figure 6. Comparison of groupwise annual maxima with data simulated from the fitted model. In each panel, the outer band is a 95% overall confidence band and the inner one a 95% pointwise confidence band. The groups of sites are: (*a*) Jungfraujoch; (*b*) Engelberg, Grand-St-Bernard, Montana; (*c*) Locaro-Monti, Lugano; (*d*) Bern-Liebefeld, Chateau d'Oex, Montreux-Clarens, Neuchâtel; (*e*) Basel-Binningen, Oeschberg-Koppingen, Zurich-MeteoSchweiz; and (*f*) Arosa, Bad-Ragaz, Davos-Dorf, Santis.

(d) Southern Plateau; (e) Northern Plateau; and (f) Eastern Alps. Since they were not used to fit the model, the poor fit for the Jungfraujoch and for the Ticino is not surprising. The fit is broadly satisfactory for the other groups, except perhaps group (f). Arosa, Davos-Dorf and Santis have altitudes of 1840, 1590 and 2490 m, respectively, but Bad Ragaz is much lower, at 496 m, and is unusual in this group because of its low altitude. This suggests that a more complex model accounting better for dependence on altitude would be preferable, but since the plotted data do not lie outside the overall 95% confidence band, this does not appear to be a critical issue.

Figure 7 shows simulation envelopes with five other groups of sites not in close proximity. The sites found to be badly fitting have been put together into group (a), for which the fit is very poor, but for the other groups, it is quite satisfactory. Similar computations for other groupings showed no other systematic model failures, even when Bad Ragaz was treated as an ordinary station, and we conclude that the model appears satisfactory in this respect.

(c) More complex models

We attempted to fit the random set model (2.10) by maximizing the pairwise likelihood with the best model above and $\alpha(h) = \{1 - |h|/(2r)\}_+$, corresponding to taking $\mathcal{B}$ to be a disc of constant radius r. Fitting this model is delicate, but a profile likelihood for r is maximized as $r \to \infty$. This is not surprising in view of the relatively small area that contains the sites. Quite different results would be expected with other types of data, such as annual maximum rainfall.

We also allowed for geometric anisotropy of the covariance function by fitting models in which the plane (x, y) is affinely transformed to $(\zeta^{-1}(x \cos \alpha + y \sin \alpha), \zeta(y \cos \alpha - x \sin \alpha))$, but found no evidence that $\alpha \neq 0$ or that $\zeta \neq 1$. Probably, a much larger number of sites would be needed to detect such effects.

(d) Risk analysis

A traditional approach to risk analysis is based on the pointwise estimation of high quantiles of a distribution, the so-called return levels or return values. In the univariate case and when the generalized extreme-value distribution (1.2) has been fitted to annual maxima, such quantiles are obtained as solutions to the equation $\hat{G}(y_p) = p$, with $p = 0.95$, 0.98, 0.99 and 0.999 corresponding, respectively, to 20, 50, 100 and 1000 year return levels. Similar arguments apply when exceedances over high thresholds are modelled, rather than maxima (Davison & Smith 1990; Coles 2001). When the extremal data are spatially indexed, it is now common to make the assumption of an underlying surface, and thus to obtain a smooth map of the necessary quantiles over the region of interest (see Cooley *et al.* (2007) or Sang & Gelfand (2009*a*)). A cruder and less satisfactory approach would be kriging of the relevant quantiles at different sites.

One important benefit of spatial modelling of rare events is the possibility of simulating unusual episodes, here years with high maximum temperatures, rather than merely producing pointwise maps of high quantiles, and thus allowing a more powerful risk analysis. To illustrate this, figure 8 shows four max-stable processes simulated from our fitted model, ordered by sorting 1000 simulated
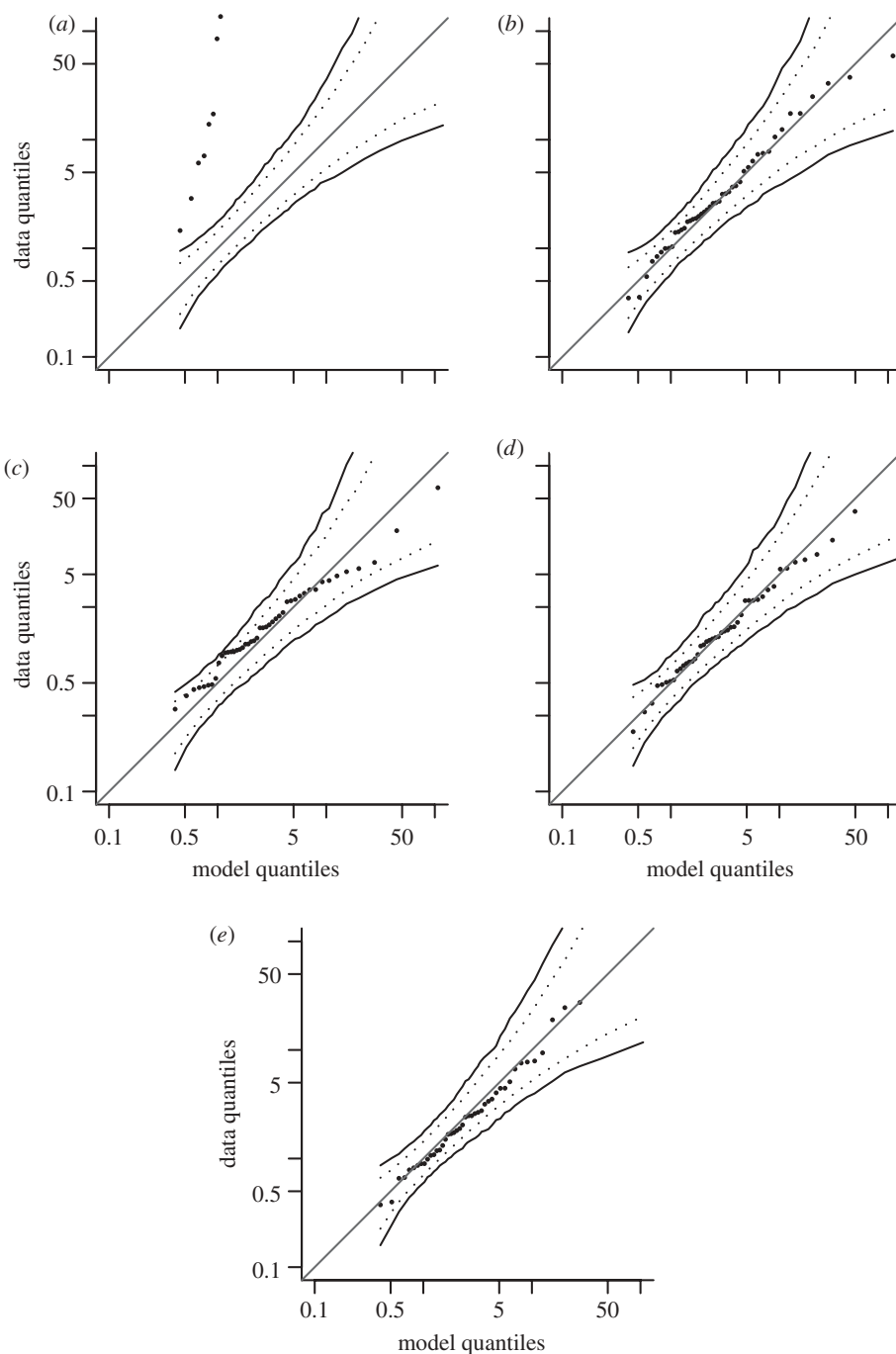
A. C. Davison and M. M. Gholamrezaee

Figure 7. Comparison of groupwise annual maxima with data simulated from the fitted model. In each panel, the outer band is a 95% overall confidence band and the inner one a 95% pointwise confidence band. The groups of sites are: (*a*) Bad Ragaz, Jungfraujoch, Locaro-Monti, Lugano; (*b*) Basel-Binningen, Oeschberg-Koppingen, Zurich-MeteoSchweiz; (*c*) Arosa, Grand-St-Bernard, Oeschberg-Koppingen; (*d*) Bern-Liebefeld, Chateau d'Oex, Engelberg, Santis; and (*e*) Davos-Dorf, Montana, Neuchâtel.
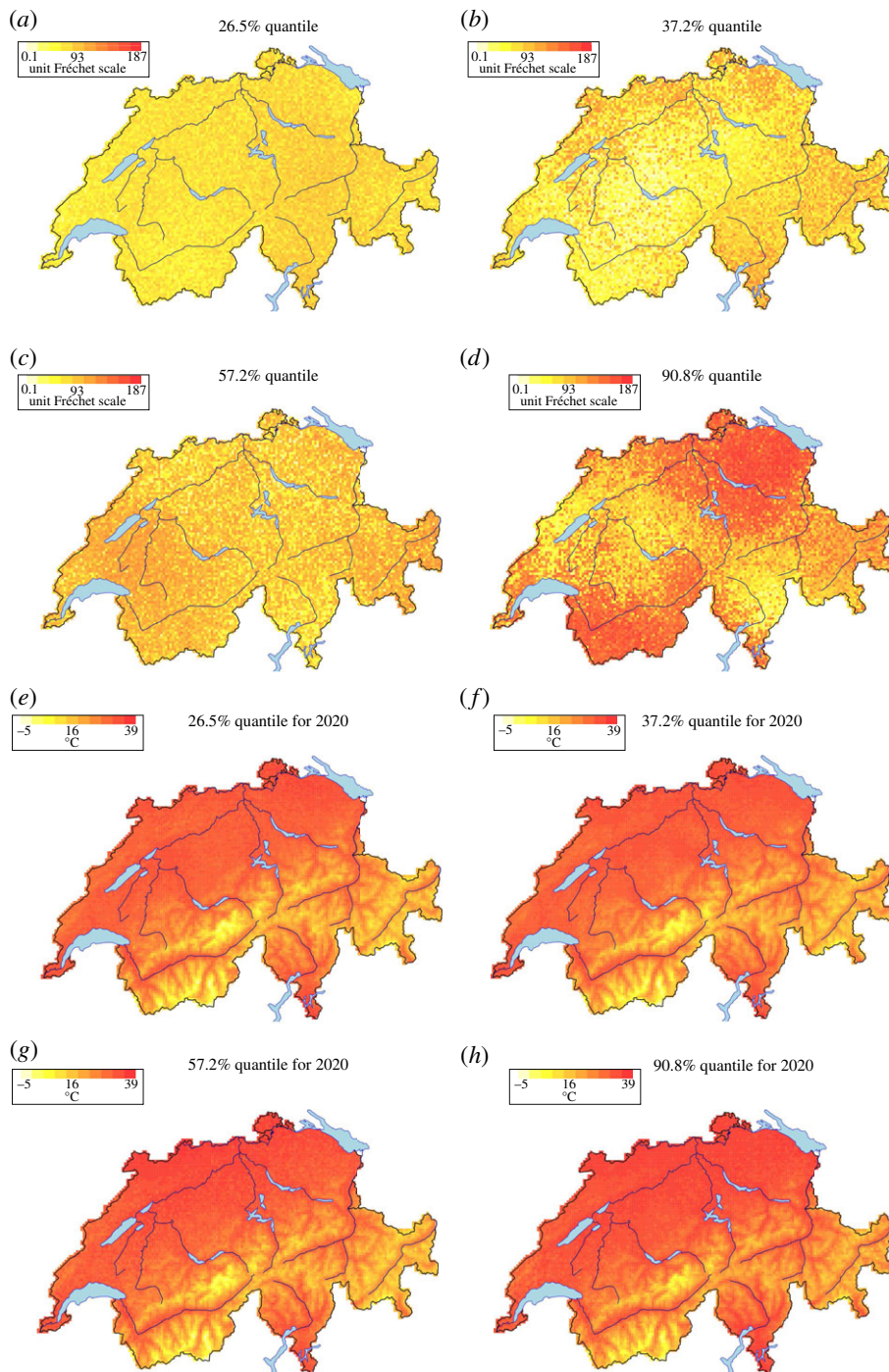
Figure 8. Four simulated max-stable processes, using the model in equation (5.2) with parameter estimates from the last column of table 3. (*a–d*) Show the simulations on the log Fréchet scale, and (*e–h*) show them on the scale of the original data, extrapolated to the year 2020.

Table 4. Simulated scenarios for years hotter than 2003. Shown are the frequencies (%) for simulated annual maxima exceeding those observed for the year 2003 at k stations, for $k = 1, \ldots, 14$, for the max-stable and independence models, for the years 2003, 2010, 2020 and 2050.

model	year	1	2	3	4	5	6	7	8	9	10	11	12	13	14
max-stable	2003	99.7	34.9	14.3	5.0	2.8	1.8	1.2	1.0	0.65	0.47	0.29	0.20	0.14	0.05
	2010	99.8	39.3	17.0	6.2	3.7	2.4	1.6	1.2	0.87	0.61	0.40	0.27	0.16	0.06
	2020	99.8	45.8	21.8	8.6	5.3	3.5	2.5	1.7	1.2	0.87	0.61	0.37	0.21	0.11
	2050	99.8	63.8	39.2	20.4	14.1	10.3	7.9	6.0	4.5	3.0	2.1	1.4	0.75	0.30
independence	2003	52.8	15.4	3.0	0.31	0.02	0								
	2010	60.3	20.7	4.7	0.52	0.06	0								
	2020	70.6	31.8	8.6	1.73	0.20	0.02	0							
	2050	96.1	80.1	53.7	27.4	10.6	3.0	0.60	0.07	0.04	0				

processes according to their overall averages on the data scale. Figure 8*a–d* shows the processes on the unit Fréchet scale, and clearly indicates the extent of the spatial dependence. Figure 8*e–h* shows the corresponding processes on the scale of the original summer maximum temperature data, for the year 2020. The effect of altitude is very strong, but close examination shows clear annual differences in the possible maximum temperatures, from which a risk analysis could be extracted using a functional such as equation (1.1). In obtaining these simulations, we made some allowance for parameter uncertainty for the parameters in table 3 by generating parameter values randomly from a multi-variate normal distribution centred on the parameter estimates and with the same covariance matrix.

Simulations can also be used to estimate the extremal coefficient (2.7) for the region $\mathcal{X}$, by averaging values of $\max_{x \in \mathcal{X} \cap \mathcal{G}} W(x)$, where $\mathcal{G}$ is a grid of points on which $W(x)$ is simulated. To estimate $\theta_{\mathcal{X}}$ for Switzerland, we used the fitted model to generate 10 000 random fields $W(x)$ on a grid of side $\delta = 0.25$ of the pixels used in figure 1 and then subsampled on grids at $0.25, 0.5, \ldots, 2.5$ of the original pixels, giving estimates of $\theta_{\mathcal{X}}$ that were extrapolated to $\delta = 0$. The resulting $\theta_{\mathcal{X}} \approx 9.5$ shows the expected high degree of dependence, corresponding to around 9.5 independent temperature maxima across the entire country. The same computations gave $\theta_{\mathcal{D}} \approx 5.5$ for the 14 sites used for the analysis.

We also used simulations to estimate the frequencies with which the temperatures observed for the exceptionally hot year 2003 at the 14 stations would be exceeded in the years 2003, 2010, 2020 and 2050, under the model of linear trend in time. Table 4 shows that under the independence model, it would effectively be impossible to observe a year like 2003 or worse at the 14 stations, whereas the probability of this event is around 0.0005 under the max-stable model: it was about a 2000 year event in 2003 and 2010, dropping to a 1000 year event in 2020, and to roughly a 333 year event in 2050. For a less extreme event in which just 7 of the 14 values were exceeded, the probability rises from around 1 per cent in 2003 to around 8 per cent in 2050. Under a model of independence between the sites, such events are much less likely; the 2003 event remains effectively impossible, even in 2050 and with the trend in temperatures.

6. Discussion

We regard the models and methods described above as a step towards realistic and flexible modelling of spatial maxima, consistent with the classical paradigm of max-stable processes and distributions. Our approach involves the following steps:

(a) construction of a marginal generalized extreme-value model for sitewise maxima that allows extrapolation to maxima across the entire spatial domain, and enables the maxima to be transformed from their original scale to a common unit Fréchet distribution;

(b) construction of a max-stable model for the transformed maxima, taking into account their spatial dependence through an appropriate covariance function and, perhaps, inclusion of a random set $\mathcal{B}$ and corresponding overlap probability $\alpha(h)$;

(c) pairwise likelihood fitting of models given by (a) and (b), with model selection using information criteria;

(d) assessment of model fit using data from pairs and from larger groups of sites; and

(e) use of the chosen model for risk analysis through simulation, entailing both extrapolation into the tails of the infinite-dimensional distribution of the random process, and perhaps extrapolation outside the period or domain for which data are available.

Below, we discuss some aspects of the approach and outline some further elements that seem to us particularly important.

Although it arises naturally from the probabilistic construction of max-stable processes, the transformation in step (a) may appear awkward because interpretation on the unit Fréchet scale is only indirectly linked to the underlying physical process. The same might be said of step (b), though the random set element may have a clear interpretation in cases such as extreme convective rainfall, where it might correspond to an exceptionally active storm cell.

Our application involved the use of a polynomial response surface for the marginal parameters in (a), but in principle, we see no difficulty in using broader classes of functions such as splines, provided the data are rich enough.

For (b), we have used a relatively simple Gaussian process model suggested by Schlather (2002), but there is wide scope for developing further classes of model for which the pairwise exponent measures can be computed and whose parameters are identifiable from the pairwise densities. In separate work, we have found that such measures for other models both stationary and non-stationary, and the work of R. L. Smith (1990, unpublished data), de Haan & Pereira (2006), Smith & Stephenson (2009) and Padoan *et al.* (2010) on the random storm formulation are also relevant. Although not an issue for our example, it is a nuisance that the pairwise extremal coefficient for the simple Gaussian process model cannot attain its maximal value of 2, and in other applications, more flexible models are needed (see Davison *et al.* in press; Wadsworth & Tawn in press).

The fitting at step (c) is sometimes delicate, and it would be useful to have fast robust algorithms. We have fitted the marginal and joint models simultaneously, but in some cases alternating algorithms seem to more be stable (Blanchet &

Davison in press). Perhaps, the fitting would be improved if higher order densities were available, though there is clearly a trade-off between computational and statistical efficiency.

Steps (d) and (e) rely on simulation from the fitted model, and therefore rule out the use of exponent measures that do not correspond to processes that can be constructed.

The use of max-stable models stems from the stability postulate underlying the derivation of the basic distributions used for univariate extremes, here generalized to equation (2.1), which enables mathematically consistent extrapolation beyond the range of the data. However, such extrapolation is fraught in several dimensions and great care should be exercised in the present infinite-dimensional situation. Even for multi-variate extremes, the classical theory has limits in modelling how dependence changes with events of increasing rarity. Important further developments to deal with these near-independence cases are owing to Ledford & Tawn (1996, 1997, 2003), who have, in an important series of papers, described diagnostics for more subtle dependencies in the joint tails for several variables than are allowed by the classical theory, and models that accommodate such dependencies. More recent work on these lines is by Heffernan & Resnick (2005, 2007). It seems important to extend our approach to cater for these more flexible and realistic forms of asymptotic behaviour (see Wadsworth & Tawn in press). Under asymptotic independence, extremely rare events would have smaller spatial extent than less rare ones; thus, our asymptotically dependent approach would tend to overstate risk if applied to data from an asymptotically independent spatial process. In order to detect such cases, it is helpful to plot measures of extremal dependence as in Blanchet & Davison (in press). It would also be helpful to find an analogue of the spatial madogram (Naveau *et al.* 2009) suitable for near-independence cases.

References

Banerjee, S., Carlin, B. P. & Gelfand, A. E. 2004 *Hierarchical modeling and analysis for spatial data.* New York, NY: Chapman & Hall/CRC.

Beirlant, J., Goegebeur, Y., Teugels, J. & Segers, J. 2004 Statistics of multivariate extremes. In *Statistics of extremes: theory and applications*, ch. 9. New York, NY: Wiley. (doi:10.1002/0470012382.ch9)

Beniston, M. *et al.* 2007 Future extreme events in European climate: an exploration of regional climate model projections. *Clim. Change* **81**, 71–95. (doi:10.1007/s10584-006-9226-z)

Blanchet, J. & Davison, A. C. In press. Spatial modelling of extreme snow depth. *Ann. Appl. Stat.*

Buishand, T. A., de Haan, L. & Zhou, C. 2008 On spatial extremes: with application to a rainfall problem. *Ann. Appl. Stat.* **2**, 624–642. (doi:10.1214/08-AOAS159)

Butler, A., Heffernan, J. E., Tawn, J. A. & Flather, R. A. 2007 Trend estimation in extremes of synthetic North Sea surges. *Appl. Stat.* **56**, 395–414. (doi:10.1111/j.1467-9876.2007.00583.x)

Cameron, P. J. 1994 *Combinatorics: topics, techniques, algorithms.* Cambridge, UK: Cambridge University Press.

Casson, E. & Coles, S. 1999 Spatial regression models for extremes. *Extremes* **1**, 449–468. (doi:10.1023/A:1009931222386)

Chandler, R. E. & Bate, S. 2007 Inference for clustered data using the independence loglikelihood. *Biometrika* **94**, 167–183. (doi:10.1093/biomet/asm015)

Chavez-Demoulin, V. & Davison, A. C. 2005 Generalized additive models for sample extremes. *Appl. Stat.* **54**, 207–222. (doi:10.1111/j.1467-9876.2005.00479.x)

Coles, S. G. 1993 Regional modelling of extreme storms via max-stable processes. *J. R. Stat. Soc. B* **55**, 797–816.

Coles, S. G. 2001 *An introduction to statistical modeling of extreme values*. New York, NY: Springer.

Coles, S. G. & Casson, E. 1998 Extreme value modelling of hurricane wind speeds. *Struct. Safety* **20**, 283–296. (doi:10.1016/S0167-4730(98)00015-0)

Coles, S. G. & Tawn, J. A. 1996 Modelling extremes of the areal rainfall process. *J. R. Stat. Soc. B* **58**, 329–347.

Coles, S. G. & Walshaw, D. 1994 Directional modelling of extreme wind speeds. *Appl. Statist.* **43**, 139–157. (doi:10.2307/2986118)

Cooley, D., Naveau, P., Jomelli, V., Rabatel, A. & Grancher, D. 2006 A Bayesian hierarchical extreme value model for lichenometry. *Environmetrics* **17**, 555–574. (doi:10.1002/env.764)

Cooley, D., Nychka, D. & Naveau, P. 2007 Bayesian spatial modeling of extreme precipitation return levels. *J. Am. Stat. Assoc.* **102**, 824–840. (doi:10.1198/016214506000000780)

Cox, D. R. & Reid, N. 2004 A note on pseudolikelihood constructed from marginal densities. *Biometrika* **91**, 211–221. (doi:10.1093/biomet/91.3.729)

Cressie, N. A. C. 1993 *Statistics for spatial data* (revised edn). New York, NY: Wiley.

Cressie, N. A. C. & Wikle, C. K. 2011 *Statistics for spatio-temporal data*. New York, NY: Wiley.

Davison, A. C. & Hinkley, D. V. 1997 *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press.

Davison, A. C. & Ramesh, N. I. 2000 Local likelihood smoothing of sample extremes. *J. R. Stat. Soc. B* **62**, 191–208. (doi:10.1111/1467-9868.00228)

Davison, A. C. & Smith, R. L. 1990 Models for exceedances over high thresholds (with discussion). *J. R. Stat. Soc. B* **52**, 393–442.

Davison, A. C., Padoan, S. & Ribatet, M. In press. Statistical modelling of spatial extremes. *Stat. Sci.*

de Haan, L. 1984 A spectral representation for max-stable processes. *Ann. Probab.* **12**, 1194–1204. (doi:10.1214/aop/1176993148)

de Haan, L. & Ferreira, A. 2006 *Extreme value theory: an introduction*. New York, NY: Springer.

de Haan, L. & Pereira, T. T. 2006 Spatial extremes: models for the stationary case. *Ann. Stat.* **34**, 146–168. (doi:10.1214/009053605000000886)

Diggle, P. J. & Ribeiro, P. J. 2007 *Model-based geostatistics*. New York, NY: Springer.

Embrechts, P., Klüppelberg, C. & Mikosch, T. 1997 *Modelling extremal events for insurance and finance*. Berlin, Germany: Springer.

Fawcett, L. & Walshaw, D. 2006 A hierarchical model for extreme wind speeds. *Appl. Stat.* **55**, 631–646. (doi:10.1111/j.1467-9876.2006.00557.x)

Genton, M. G., Ma, Y. & Sang, H. 2011 On the likelihood function of Gaussian max-stable processes. *Biometrika* **98**, 481–488. (doi:10.1093/biomet/asr020)

Hall, P. & Tajvidi, N. 2000 Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Stat. Sci.* **15**, 153–167.

Heffernan, J. E. & Resnick, S. I. 2005 Hidden regular variation and the rank transform. *Adv. Appl. Probab.* **37**, 393–412. (doi:10.1239/aap/1118858631)

Heffernan, J. E. & Resnick, S. I. 2007 Limit laws for random vectors with an extreme component. *Ann. Appl. Probab.* **17**, 537–571. (doi:10.1214/105051606000000835)

Heffernan, J. E. & Tawn, J. A. 2004 A conditional approach for multivariate extreme values (with discussion). *J. R. Stat. Soc. B* **66**, 497–546. (doi:10.1111/j.1467-9868.2004.02050.x)

Kent, J. T. 1982 Robust properties of likelihood ratio tests. *Biometrika* **69**, 19–27. (doi:10.1093/biomet/69.1.19)

Leadbetter, M. R., Lindgren, G. & Rootzén, H. 1983 *Extremes and related properties of random sequences and processes*. New York, NY: Springer.

Ledford, A. W. & Tawn, J. A. 1996 Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169–187. (doi:10.1093/biomet/83.1.169)

Ledford, A. W. & Tawn, J. A. 1997 Modelling dependence within joint tail regions. *J. R. Stat. Soc. B* **59**, 475–499. (doi:10.1111/1467-9868.00080)

Ledford, A. W. & Tawn, J. A. 2003 Diagnostics for dependence within time series extremes. *J. R. Stat. Soc. B* **65**, 521–543. (doi:10.1111/1467-9868.00400)

Lindsay, B. G. 1988 Composite likelihood methods. *Contemp. Math.* **80**, 220–241.

Naveau, P., Guillou, A., Cooley, D. & Diebolt, J. 2009 Modeling pairwise dependence of maxima in space. *Biometrika* **96**, 1–17. (doi:10.1093/biomet/asp001)

Padoan, S. A. & Wand, M. P. 2008 Mixed model-based additive models for sample extremes. *Stat. Probab. Lett.* **78**, 2850–2858. (doi:10.1016/j.spl.2008.04.009)

Padoan, S., Ribatet, M. & Sisson, S. 2010 Likelihood-based inference for max-stable processes. *J. Am. Stat. Assoc.* **105**, 263–277. (doi:10.1198/jasa.2009.tm08577)

Pauli, F. & Coles, S. G. 2001 Penalized likelihood inference in extreme value analyses. *J. Appl. Stat.* **28**, 547–560. (doi:10.1080/02664760120047889)

Poncet, P., Cooley, D. & Naveau, P. 2006 Variograms for max-stable random fields. In *Dependence in probability and statistics* (eds P. Bertail, P. Doukhan & P. Soulier). Lecture Notes in Statistics, no. 187, pp. 373–390. New York, NY: Springer.

Ramesh, N. I. & Davison, A. C. 2002 Local models for exploratory analysis of hydrological extremes. *J. Hydrol.* **256**, 106–119. (doi:10.1016/S0022-1694(01)00522-4)

Resnick, S. I. 1987 *Extreme values, regular variation and point processes*. New York, NY: Springer.

Sang, H. & Gelfand, A. E. 2009*a* Hierarchical modeling for extreme values observed over space and time. *Environ. Ecol. Stat.* **16**, 407–426. (doi:10.1007/s10651-007-0078-0)

Sang, H. & Gelfand, A. E. 2009*b* Continuous spatial process models for spatial extreme values. *J. Agric. Biol. Environ. Stat.* **15**, 49–65. (doi:10.1007/s13253-009-0010-1)

Schabenberger, O. & Gotway, C. A. 2005 *Statistical methods for spatial data analysis*. London, UK: Chapman & Hall/CRC.

Schlather, M. 2002 Models for stationary max-stable random fields. *Extremes* **5**, 33–44. (doi:10.1023/A:1020977924878)

Schlather, M. & Tawn, J. A. 2003 A dependence measure for multivariate and spatial extreme values: properties and inference. *Biometrika* **90**, 139–156. (doi:10.1093/biomet/90.1.139)

Sibuya, M. 1960 Bivariate extreme statistics. *Ann. Inst. Stat. Math.* **11**, 195–210. (doi:10.1007/BF01682329)

Smith, E. L. & Stephenson, A. G. 2009 An extended Gaussian max-stable process model for spatial extremes. *J. Stat. Plan. Inference* **139**, 1266–1275. (doi:10.1016/j.jspi.2008.08.003)

Stein, M. L. 1999 *Interpolation of spatial data: some theory for kriging*. New York, NY: Springer.

Stephenson, A. G. & Tawn, J. A. 2005 Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika* **92**, 213–227. (doi:10.1093/biomet/92.1.213)

Takeuchi, K. 1976 Distribution of informational statistics and a criterion of fitting [in Japanese]. *Suri-Kagaku* **153**, 12–18.

Tawn, J. A. 1988 Bivariate extreme value theory: models and estimation. *Biometrika* **75**, 397–415. (doi:10.1093/biomet/75.3.397)

Varin, C. 2008 On composite marginal likelihoods. *Adv. Stat. Anal.* **92**, 1–28. (doi:10.1007/s10182-008-0060-7)

Varin, C. & Vidoni, P. 2005 A note on composite likelihood inference and model selection. *Biometrika* **92**, 519–528. (doi:10.1093/biomet/92.3.519)

Varin, C., Reid, N. & Firth, D. 2011 An overview of composite marginal likelihoods. *Stat. Sinica* **21**, 5–42.

Wackernagel, H. 2003 *Multivariate geostatistics: an introduction with applications*, 3rd edn. New York, NY: Springer.

Wadsworth, J. & Tawn, J. A. In press. Dependence modelling for spatial extremes. *Biometrika*.