REVSTAT – Statistical Journal
Volume 10, Number 1, March 2012, 135–165

A SURVEY OF SPATIAL EXTREMES: MEASURING SPATIAL DEPENDENCE AND MODELING SPATIAL EFFECTS

Authors: Daniel Cooley
– Department of Statistics, Colorado State University,
Fort Collins, CO, USA (cooleyd@stat.colostate.edu)

Jessi Cisewski
– Department of Statistics and Operations Research,
University of North Carolina at Chapel Hill,
Chapel Hill, NC, USA (cisewski@email.unc.edu)

Robert J. Erhardt
– Department of Statistics and Operations Research,
University of North Carolina at Chapel Hill,
Chapel Hill, NC, USA (erhardt@email.unc.edu)

Soyoung Jeon
– Department of Statistics and Operations Research,
University of North Carolina at Chapel Hill,
Chapel Hill, NC, USA (soyoung@email.unc.edu)

Elizabeth Mannshardt
– Department of Statistics, The Ohio State University,
Columbus, OH, USA (elizabeth@stat.osu.edu)

Bernard Oguna Omolo
– Division of Mathematics and Computer Science,
University of South Carolina – Upstate,
Spartanburg, SC, USA (bomolo@uscupstate.edu)

Ying Sun
– Statistical and Applied Mathematical Sciences Institute,
Research Triangle Park, NC, USA (sunwards@samsi.info)

Abstract:

- We survey the current practice of analyzing spatial extreme data, which lies at the intersection of extreme value theory and geostatistics. Characterizations of multivariate max-stable distributions typically assume specific univariate marginal distributions, and their statistical applications generally require capturing the tail behavior of the margins and describing the tail dependence among the components. We review current methodology for spatial extremes analysis, discuss the extension of the finite-dimensional extremes framework to spatial processes, review spatial dependence metrics for extremes, survey current modeling practice for the task of modeling marginal distributions, and then examine max-stable process models and copula approaches for modeling residual spatial dependence after accounting for marginal effects.

Key-Words:

- *copula; extremal coefficient; hierarchical model; madogram; max-stable process; multivariate extreme value distribution.*

AMS Subject Classification:

- 62M30, 62H11, 62H20.

1. INTRODUCTION

Assessing the behavior of rare events such as the risk of flooding, potential crop damage from drought, or health effects of potential extreme air pollution events presents unique statistical challenges, and requires one to characterize the tail of the distribution of the quantity of interest. Since many quantities such as rainfall, temperature, or air pollution are measured at specifically-located monitors, spatial modeling is necessary. Applications such as these have motivated the development of methods and tools for analyzing and characterizing spatial extreme data. In this paper, we survey the current practice of spatial extremes. Recently, Davison *et al.* (2012) review spatial extremes methods via a case study of extreme precipitation in Switzerland. This survey can be viewed as complementary to Davison *et al.* (2012), and we aim to provide an entry point for the study of spatial extremes.

An analysis of spatial extreme data lies at the intersection of two branches of statistics: extreme value analysis and geostatistics. Below, we give brief introductions to each field. There are a number of books in each field which give comprehensive overviews. For extremes, references include de Haan & Ferreira (2006), Beirlant *et al.* (2004) and Coles (2001); geostatistics references include Schabenberger & Gotway (2005), Banerjee *et al.* (2004) and Cressie (1993). For many studies, spatial effects are often separated into large scale (i.e., regional) effects and small scale (i.e., local) effects.

In terms of a statistical model, regional spatial effects are often captured by characterizing how the marginal distribution varies over a study region. Local spatial effects are typically described by a dependence structure. Given data, the distinction between the local and regional effects is likely not obvious, and one can view the task of separating these two effects as analogous to the task of decomposing time series data into mean (trend and seasonal effects) and a stationary noise process described by a covariance structure (Brockwell & Davis, 2002, §1.3.3).

Spatial data are necessarily multivariate as they are recorded at multiple locations. Throughout, we will assume that we analyze only one quantity (e.g., rainfall) at multiple sites, although spatially analyzing multiple quantities is a possible extension of the work surveyed herein.

Spatial data are often modeled as a realization of a spatial process which is observed at a finite set of locations. We, too, will consider a spatial process, but we begin with a review of the theory and statistical practice for finite-dimensional extremes in order to introduce important concepts.

1.1. Multivariate extremes: theory and practice

We assume that the reader is familiar with univariate extreme value theory and its statistical application. For background on the univariate case see the monographs by Coles (2001), Beirlant *et al.* (2004), and de Haan & Ferreira (2006).

The notion of max-stability forms the foundation of extreme value theory. Let $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})^{\mathrm{T}}$, $i = 1, \ldots, n$, be an i.i.d. sequence of d-dimensional continuous random vectors, and let $\boldsymbol{M}_n = (M_{n,1}, \ldots, M_{n,d})^{\mathrm{T}} = \left(\bigvee_{i=1}^n X_{i,1}, \ldots, \bigvee_{i=1}^n X_{i,d}\right)^{\mathrm{T}}$, where $\bigvee$ denotes the maximum. Assume there exist normalizing sequences $\{\boldsymbol{a}_n\}$ and $\{\boldsymbol{b}_n\}$ such that

$$\Pr\left(\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{y}\right) \longrightarrow G(\boldsymbol{y}) , \qquad n \longrightarrow \infty$$

where the division is understood to be element-wise and G is non-degenerate. Then G belongs to the class of multivariate max-stable (equivalently, extreme value) distributions. We will denote by $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^{\mathrm{T}}$ a max-stable random vector; that is, $\boldsymbol{a}_n^{-1}(\boldsymbol{M}_n - \boldsymbol{b}_n) \xrightarrow{\mathrm{d}} \boldsymbol{Y}$. The i.i.d. assumption can be relaxed and the multivariate max-stable distributions continue to serve as the possible limiting distributions, if certain mixing conditions are met (Leadbetter *et al.*, 1983).

Unlike in the univariate case, no fully parametric representation exists for the multivariate max-stable distributions. The univariate marginal distributions must be univariate max-stable and therefore can be described by the generalized extreme value (GEV) distribution:

$$\Pr\big(Y_j \leq y\big) \;=\; G_j(y) \;=\; \exp\left\{-\left[\left(1 + \xi_j \, \frac{y - \mu_j}{\sigma_j}\right)_+^{-1/\xi_j}\right]\right\} , \tag{1.1}$$

for $j = 1, \ldots, d$. Here, μ_j, σ_j and ξ_j are the location, scale and shape parameters for the j^{th} component's marginal and $x_+ = \max(0, x)$. Accounting for each GEV marginal is a nuisance, and representations for multivariate max-stable distributions generally presuppose that the marginals have a common, convenient max-stable distribution. Throughout, we will assume $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^{\mathrm{T}}$ has a multivariate max-stable distribution with unit-Fréchet marginals: $\Pr(Z_j \leq z) = \exp(-z^{-1})$. Then

$$\Pr(\boldsymbol{Z} \leq \boldsymbol{z}) \;=\; G^*(\boldsymbol{z}) \;=\; \exp\big\{-V(\boldsymbol{z})\big\} , \tag{1.2}$$

$$V(\boldsymbol{z}) \;=\; d \int_{\Delta_d} \bigvee_{j=1}^{d} \frac{w_j}{z_j} \, H(\mathrm{d}\boldsymbol{w}) . \tag{1.3}$$

Here $\Delta_d = \big\{\boldsymbol{w} \in \mathbb{R}_+^d \mid w_1 + \cdots + w_d = 1\big\}$ is the $(d-1)$-dimensional simplex, and the angular (or spectral) measure H is a probability measure on Δ_d, which determines the dependence structure of the random vector. Due to the common marginals, H obeys the moment conditions $\int_{\Delta_d} w_j \, H(d\boldsymbol{w}) = 1/d$ for $j = 1, \ldots, d$.

There is no loss of generality in assuming the multivariate max-stable distribution has unit-Fréchet margins, as Resnick (1987, Prop. 5.10) states that the domain-of-attraction condition is preserved under monotone transformations of the marginal distributions. If $\boldsymbol{a}_n^{-1}(\boldsymbol{M}_n - \boldsymbol{b}_n) \xrightarrow{\mathrm{d}} \boldsymbol{Y}$ which does not have unit-Fréchet marginals, from (1.1) one can define marginal transformations

$$T_j(x) = \left(1 + \xi_j \, \frac{x - \mu_j}{\sigma_j}\right)^{-\xi_j}, \qquad j = 1, \ldots, d \; ,$$

and define

$$G^*(z_1, \ldots, z_d) \; = \; G\bigl\{T_1^{\leftarrow}(z_1), \ldots, T_d^{\leftarrow}(z_d)\bigr\} \; ,$$

where $T_j^{\leftarrow}$ is the inverse function of T_j, for $j = 1, \ldots, d$. This approach of transforming to convenient marginals is similar to copula approaches, albeit with a marginal suggested by extreme value theory rather than Uniform [0,1].

The above asymptotic theory suggests the following general statistical methodology, referred to as the block maxima approach. Choose n to be a fixed block size which is large enough such that the asymptotic theory holds approximately, and assume a sequence of i.i.d. $\boldsymbol{X}_i$, $i = 1, \ldots, nm$, are observed, where m denotes the number of blocks. Define $\boldsymbol{M}_k = \bigl(\bigvee_{i=(k-1)n+1}^{kn} X_{i,1}, ..., \bigvee_{i=(k-1)n+1}^{kn} X_{i,d}\bigr)^{\mathrm{T}}$ for $k = 1, ..., m$ (note that the dependence on n in the notation $\boldsymbol{M}_k$ has been suppressed), and fit a multivariate max-stable distribution to the $\boldsymbol{M}_k$. It is important to note that $\boldsymbol{M}_k$ will not appear in the observation record unless the occurrence times of each element's block maximum coincide.

Using representation (1.2) to fit a multivariate max-stable distribution requires that the marginals be unit Fréchet. Although transforming the marginals is a simple theoretical procedure, in practice the marginal distributions must be estimated. Subsequently, utilizing (1.2) to perform a multivariate analysis of extremes involves two tasks: (1) estimating the marginals, and (2) characterizing the dependence via a model for $V(\boldsymbol{z})$ or $H(\boldsymbol{w})$. Tasks (1) and (2) seem sequential; however, we note that inference can be performed all-at-once either in the frequentist (Padoan *et al.*, 2010) or Bayesian (Ribatet *et al.*, 2011) settings.

For spatial extremes studies, the aforementioned regional and local spatial effects each can be associated with one of the above tasks. Most study regions are large enough that the marginal distribution of the studied quantity will vary over the region. Thus, in order to transform to a common marginal, one must first account for how the distribution's tail varies by location. The local spatial effect is related to the spatial extent of individual extreme events and the resulting dependence in the data due to multiple sites being affected by the same event. In terms of (1.2), this dependence is captured by $V(\boldsymbol{z})$ or $H(\boldsymbol{w})$. We will refer to the dependence remaining after the marginal standardization as 'residual' dependence, as Sang & Gelfand (2010) termed the data after marginal transformation as 'standardized residuals'. There is a useful analogy here drawn from atmospheric science: these two types of spatial effects can be thought of as corresponding to 'climate' and 'weather' effects. Climate can be thought of as the distribution of weather (Guttorp & Xu, 2011), and climate varies with location. Performing the marginal transformation is akin to standardizing the climate across the study region. Weather events have a spatial extent which is best captured by a stochastic representation. For most applications, there is a difference in the scale of these two spatial effects. Climate varies on a larger (regional) spatial scale, and can be largely (but often not completely) characterized by covariates such as latitude and elevation. Weather spatial effects, particularly for extremes, are often more localized.

Although all data are finite-dimensional, a finite-dimensional framework can be inadequate for dealing with unobserved locations, and thus most classical spatial work assumes a stochastic process framework. Let $\mathcal{S}$ be a study region, and let $\boldsymbol{s}$ denote a location in a study region. For spatial applications, most often $\boldsymbol{s} \in \mathbb{R}^2$ and we will assume this throughout. We will assume $X_i(\boldsymbol{s})$ is a stochastic process, where it may be helpful to think of i as indexing the day of the observation. A fundamental construct for spatial extremes is the max-stable process, which is the infinite-dimensional analogue to a max-stable random vector. If for all $\boldsymbol{s} \in \mathcal{S}$ there exist normalizing sequences $a_n(\boldsymbol{s})$ and $b_n(\boldsymbol{s})$ such that

$$a_n^{-1}(\boldsymbol{s}) \left\{ \max_{i=1,\ldots,n} X_i(\boldsymbol{s}) - b_n(\boldsymbol{s}) \right\} \xrightarrow{\mathrm{d}} Y(\boldsymbol{s}) \;, \tag{1.4}$$

which has a non-degenerate distribution, then $Y(\boldsymbol{s})$ is a max-stable process. When the max-stable process has unit Fréchet margins, we will denote it by $Z(\boldsymbol{s})$. Given any finite set of locations $s_1, \ldots, s_d$, one can let $\boldsymbol{X}_i = (X_i(s_1), \ldots, X_i(s_d))^{\mathrm{T}}$ or $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_d))^{\mathrm{T}}$ and return to the finite-dimensional setting.

An alternative approach to analyzing block maxima is to instead select and analyze a subset of threshold exceedances. Although there has been work to develop methods for multivariate threshold exceedance data, most spatial extremes work to date has aimed at developing max-stable models and fitting block-maximum data. This survey will primarily focus on such work, but we briefly discuss ongoing work in developing methods for spatial threshold exceedance data in the discussion in §5.

1.2. Standard geostatistics

Let $X(\boldsymbol{s})$ be a stochastic process. The field of geostatistics provides a framework for exploring, modeling, and predicting or interpolating $X(\boldsymbol{s})$. Much of classical geostatistics tries to characterize $X(\boldsymbol{s})$ in terms of its mean and covariance function. One typically thinks of using the mean to represent large scale changes of $X(\boldsymbol{s})$ and the covariance function to capture the variability due to small- and micro-scale stochastic sources (Schabenberger & Gotway, 2005, p. 132).

A basic model can be formulated as

$$X(\boldsymbol{s}) = \alpha(\boldsymbol{s}) + e(\boldsymbol{s}) \ , \tag{1.5}$$

where $\alpha(\boldsymbol{s})$ is the (non-random) mean function and $e(\boldsymbol{s})$ is a zero-mean stochastic process. Often, a regression relation is assumed for the mean function: $\alpha(\boldsymbol{s}) = W(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\beta}$, where $W(\boldsymbol{s})$ is a vector of covariate information at location $\boldsymbol{s}$ and $\boldsymbol{\beta}$ is a vector of regression coefficients.

The process e tries to account for any behavior not captured by the mean function α. A simple geostatistical model may assume $e(\boldsymbol{s})$ is second-order stationary and isotropic. Stationarity implies that the covariance does not depend on location, i.e., $\mathrm{Cov}(e(\boldsymbol{s}), e(\boldsymbol{s}')) = \mathrm{Cov}(e(\boldsymbol{s}+\boldsymbol{h}), e(\boldsymbol{s}'+\boldsymbol{h}))$, while isotropy implies that covariance is a function of distance only, i.e., $\mathrm{Cov}(e(\boldsymbol{s}), e(\boldsymbol{s}+\boldsymbol{h})) = C(h)$, where $h = \|\boldsymbol{h}\|$. At times it is useful to further assume that $e(\boldsymbol{s})$ is a Gaussian process, which in turn implies $X(\boldsymbol{s})$ is a Gaussian process with mean $\alpha(\boldsymbol{s})$.

The second-order stationary and isotropic random field e is characterized by its covariance function $C(h)$ or equivalently its semivariogram

$$\gamma(h) \;=\; \frac{1}{2}\,\mathrm{Var}\Big[e(\boldsymbol{s}+\boldsymbol{h}) - e(\boldsymbol{s})\Big] \;=\; \frac{1}{2}\,\mathrm{E}\Big[\big\{e(\boldsymbol{s}+\boldsymbol{h}) - e(\boldsymbol{s})\big\}^2\Big] \ .$$

Since the covariance function or semivariogram must satisfy several requirements to be valid, models for $C(h)$ or $\gamma(h)$ are generally selected from parametric families (see Schabenberger & Gotway, 2005, §4.2) known to meet these requirements.

Less restrictive than second-order stationary is intrinsic stationarity, which implies that a process has stationary increments, i.e., $e(\boldsymbol{s}) - e(\boldsymbol{s}+\boldsymbol{h}) \stackrel{d}{=} e(\boldsymbol{h}) - e(\boldsymbol{0})$. An intrinsically stationary process can be viewed as akin to a random walk in time series, which is stationary after differencing. An intrinsically stationary process does not have a covariance function $C(\boldsymbol{h})$ but does have a semivariogram $\gamma(\boldsymbol{h})$.

A geostatistical analysis often begins with an exploratory phase where dependence is investigated via an empirical covariogram or semivariogram. As best as one can, one must first account for large scale effects in the mean as "much damage can be done by applying semivariogram estimators … to data from non-stationary spatial processes" (Schabenberger & Gotway, 2005, p. 135). Assuming stationarity and isotropy, the traditional sample semivariogram is

$$\hat{\gamma}(h) \;=\; \frac{1}{2\,|\mathcal{N}(h)|} \sum_{(\boldsymbol{s},\boldsymbol{s}')\in\mathcal{N}(h)} \big\{e(\boldsymbol{s}) - e(\boldsymbol{s}')\big\}^2 \ , \tag{1.6}$$

where $\mathcal{N}(h)$ denotes the number of pairs $(\boldsymbol{s}, \boldsymbol{s}')$ separated by the distance h. Applying the empirical semivariogram to observations provides insight for semivariogram model selection. Having selected a parametric family $\gamma_{\boldsymbol{\phi}}(h)$ for the semivariogram function, one often proceeds to estimate the model parameters $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$.

A primary goal of many geostatistics analyses is spatial prediction / interpolation employing an estimated semivariogram, which is known as *kriging*. The point predictor from kriging corresponds to the best linear unbiased predictor (or the conditional expectation under a Gaussian assumption) of the value of $X(\boldsymbol{s}_0)$ at unobserved location $\boldsymbol{s}_0$ given observed values $X(\boldsymbol{s}_1), \ldots, X(\boldsymbol{s}_d)$. Prediction uncertainty is typically quantified in terms of mean-square prediction error.

There is an analogy between the two tasks described in § 1.1 and the geostatistics model (1.5). If e is stationary and Gaussian, then the marginal distribution can only vary with $\alpha(\boldsymbol{s})$ which captures the regional spatial effects. After accounting for regional effects with α, the residual dependence in e is characterized by its semivariogram or covariance function.

There are important fundamental differences between geostatistics and spatial extremes. As it is based on first and second moments, geostatistics focuses on central tendencies, not on the distribution's tail. The Gaussian framework which is never far from a traditional geostatistics analysis is incorrect for data that are maxima, as the Gaussian distribution is not max-stable. Dependence in extremes is described via the exponent measure function $V(\boldsymbol{z})$ or angular measure $H(\boldsymbol{w})$ which cannot be linked to covariance. Finally, much of classical geostatistics is applied to situations where one has only one realization of the process $X(\boldsymbol{s})$, observed at multiple locations. To perform an extreme value analysis, it is necessary that multiple realizations $X_i(\boldsymbol{s})$ underlie the subset of extreme data which are eventually analyzed.

Much of the paper will study max-stable processes, and we need to extend the notion of stationarity and isotropy to these processes. Stationarity for max-stable processes is first-order, implying invariance of any finite-dimensional joint distribution to translation:

$$\Pr\Big(Y(\boldsymbol{s}_1) \le y_1, \ldots, Y(\boldsymbol{s}_d) \le y_d\Big) = \Pr\Big(Y(\boldsymbol{s}_1 + \boldsymbol{h}) \le y_1, \ldots, Y(\boldsymbol{s}_d + \boldsymbol{h}) \le y_d\Big) .$$

Isotropy will imply that all bivariate joint distributions are also invariant to rotation:

$$\Pr\Big(Y(\boldsymbol{s}) \le y_1, Y(\boldsymbol{s} + \boldsymbol{h}) \le y_2\Big) = \Pr\Big(Y(\boldsymbol{s}) \le y_1, Y(\boldsymbol{s} + \boldsymbol{h}') \le y_2\Big) ,$$

if $\|\boldsymbol{h}\| = \|\boldsymbol{h}'\|$. For simplicity, we will generally assume that a random field is stationary and isotropic.

The remainder of the paper is structured to follow the order of a possible spatial extremes analysis. We begin by reviewing tools which measure spatial dependence in §2. In §3 we survey methods for modeling the marginal tail behavior over a study region. In §4, we review two primary methods for modeling the residual dependence: max-stable process models and copula approaches. We conclude with a discussion which mentions work in development, challenges posed by applications, and open problems.

2. MEASURING SPATIAL DEPENDENCE

To completely characterize the dependence among the components of a max-stable random vector requires one to specify the angular measure $H(\boldsymbol{w})$ or the exponent measure function $V(\boldsymbol{z})$. Specification of $H(\boldsymbol{w})$ or $V(\boldsymbol{z})$ is arduous, especially as the dimension d grows, and representations are not easily compared. It is useful to have summary measures of tail dependence, and several metrics have been developed which aim to summarize the amount of tail dependence in one number.

A complication arises because tail dependence falls into two distinct categories: asymptotic dependence and asymptotic independence. Since the categories are distinct, summary measures for dependence have been developed for each category. We first focus on metrics for the asymptotic dependence case, with a particular interest in measuring dependence in terms of spatial distance. We then explain the notion of asymptotic independence (which does not imply complete independence) and briefly mention how the amount of dependence in the asymptotic independence case can be measured.

We note that the metrics all assume at least that random vectors or fields have a common marginal distribution. Like in geostatistics, one must first try to account for large-scale marginal effects before using these tools to assess (residual) dependence.

2.1. Tail dependence metrics for asymptotic dependence

There are many related metrics for quantifying tail dependence when the random vector exhibits asymptotic dependence. The list includes the metric d of Davis & Resnick (1989) and the metric χ of Coles *et al.* (1999). We focus on two metrics: the extremal coefficient (Smith, 1990; Schlather & Tawn, 2003) which is readily interpretable and the madogram (Cooley *et al.*, 2006; Naveau *et al.*, 2009) which has ties to the semivariogram.

2.1.1. Extremal coefficient

Let $\boldsymbol{Y}$ be a d-dimensional max-stable random variable with common margins. The d-dimensional extremal coefficient θ_d can be implicitly defined as

$$\Pr\Big(Y_1 \le y, \ldots, Y_d \le y\Big) = \Pr\left(\bigvee_{j=1}^{d} Y_j \le y\right) = \Pr^{\theta_d}\big(Y_1 \le y\big) ,$$

for any y in the support of Y_1. Transforming the marginals of $\boldsymbol{Y}$ to obtain $\boldsymbol{Z}$, and due to the homogeneity property of V, we have

$$\Pr\Big(Z_1 \le z, \ldots, Z_d \le z\Big) = \exp\big\{-z^{-1} V(1, \ldots, 1)\big\} \Rightarrow \theta_d = V(1, \ldots, 1) . \tag{2.1}$$

The value θ_d can be thought of as the effective number of independent random variables in the d-dimensional random vector. The coefficient takes values between 1 and d, with a value of 1 corresponding to complete dependence among the locations, and a value of d corresponding to complete independence. The extremal coefficient is studied extensively in Schlather & Tawn (2003) and relations between extremal coefficients of different orders are given in Schlather & Tawn (2002).

Given replicates of a d-dimensional random vector $\boldsymbol{Z}$, Smith (1990) and Coles & Dixon (1999) propose an estimator of the extremal coefficient θ_d. As $\boldsymbol{Z}$ has unit Fréchet margins, $1/Z_j$ is unit exponential, and $1/\bigvee_{j=1}^{d} Z_j$ is exponential with mean $1/\theta_d$. Given i.i.d. replicates $\boldsymbol{Z}_k$, $k = 1, \ldots, m$, a simple estimator is

$$\hat{\theta}_d = \frac{m}{\sum_{k=1}^{m} 1/\bigvee_{j=1}^{d}(z_{k,j})} , \tag{2.2}$$

where $z_{k,j}$ is the j^{th} component of the observation $\boldsymbol{z}_k$.

Although higher-order extremal coefficients are sometimes useful (see Erhardt & Smith, 2011), θ_2 is most widely used as it conveys the amount of dependence between a pair of components. Bivariate dependence metrics are especially useful in spatial studies as one generally wants to link the level of dependence to spatial distance. Let $Z(\boldsymbol{s})$ be a stationary and isotropic max-stable random field with unit-Fréchet margins. It is possible to extend (2.1) to be a distance-based dependence metric:

$$\theta(h) = \Pr\Big(Z(\boldsymbol{s}) \le z, Z(\boldsymbol{s} + \boldsymbol{h}) \le z\Big) .$$

One could extend (2.2) to construct a distance-based estimator for $\theta(h)$; however, to our knowledge, distance-based dependence summary measures have been primarily estimated via the madogram (for instance, see the `SpatialExtremes` package in `R`, Ribatet, 2011).

2.1.2. Madogram

The madogram, a first-order semivariogram, has its roots in geostatistics and its properties were studied by Matheron (1987). Since the madogram requires the first-moment to be finite which is not always the case in extremes studies, Cooley *et al.* (2006) proposed the F-madogram, which first transforms the random variable by applying the cdf and is finite for any distribution. If $Y(\boldsymbol{s})$ is a stationary and isotropic max-stable process with marginal distribution G, then the F-madogram is:

$$\nu(h) = \frac{1}{2}\,\mathbf{E}\Big|G\big\{Y(\boldsymbol{s})\big\} - G\big\{Y(\boldsymbol{s}+\boldsymbol{h})\big\}\Big| \tag{2.3}$$

(for consistency with above, we have used G, rather than F, to denote a max-stable marginal distribution). The F-madogram's values range from 0 to 1/6, which corresponds to complete dependence and independence respectively. The F-madogram is related to the other extremal dependence metrics, as Cooley *et al.* (2006) show that

$$\theta(h) = \frac{1+2\,\nu(h)}{1-2\,\nu(h)}\;. \tag{2.4}$$

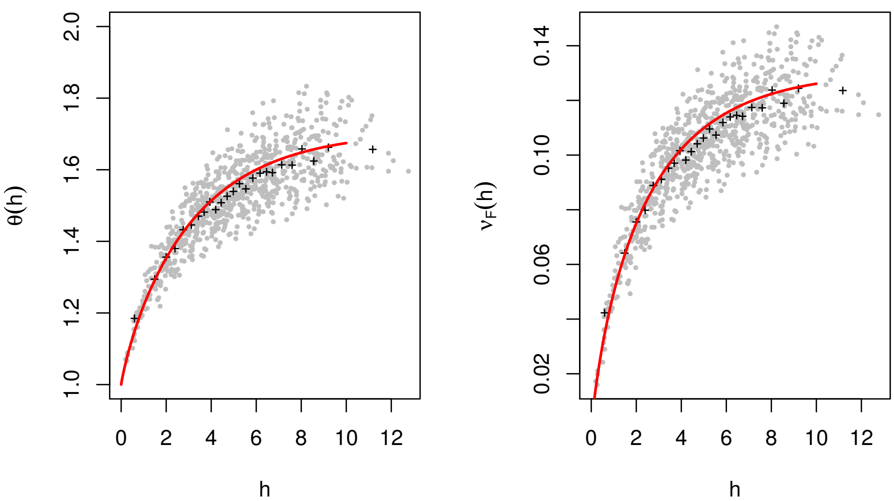

Figure 1: Extremal coefficient (left panel) and the F-Madogram (right panel) with unit Gumbel margins for the Schlather model with Whittle–Matérn correlation functions. The red lines are the theoretical extremal coefficient and F-madogram, gray points are pairwise estimates, and black crosses are binned estimates.

An advantage of the F-madogram is that its definition (2.3) suggests an estimator. Let $y_k(\boldsymbol{s})$, $k = 1, \ldots, m$, be i.i.d. replicates of a max-stable process with marginal distribution G which is observed at a finite set of locations. Then the sample F-madogram is given by

$$\hat{\nu}(h) = \frac{1}{m} \sum_{k=1}^{m} \frac{1}{2\,|\mathcal{N}(h)|} \sum_{(\boldsymbol{s},\boldsymbol{s}')\in\mathcal{N}(h)} \bigl| G\{y_k(\boldsymbol{s})\} - G\{y_k(\boldsymbol{s}')\} \bigr| , \tag{2.5}$$

analogous to the semivariogram estimator (1.6). An estimator $\hat{\theta}(h)$ can be obtained via a plug-in estimator from (2.4). Equation (2.5) assumes that the marginal distribution G is known. Naveau *et al.* (2009) discuss estimation of the madogram when the marginal distribution is not known and further define the λ-madogram, which is related to the Pickands' dependence function and which extends the notion of the madogram to completely describe the bivariate dependence structure.

2.2. Asymptotic independence

Two components X_1 and X_2 from a random vector $\boldsymbol{X}$ with common marginals are *asymptotically independent* if

$$\lim_{x\to x^*} \Pr\Bigl(X_2 > x \mid X_1 > x\Bigr) = 0 , \tag{2.6}$$

where x^* is the upper endpoint of the common marginal distribution. The components are *asymptotically dependent* if the limit in (2.6) is a non-zero constant. If asymptotically independent, then (X_1, X_2) lie in the domain of attraction of max-stable (Y_1, Y_2), whose angular measure $H(\boldsymbol{w})$ has a mass of $1/2$ at $\boldsymbol{w} = (0,1)$ and at $\boldsymbol{w} = (1,0)$, and 0 elsewhere. That is, the bivariate angular measure has mass only on the axes. The extremal coefficient of (Y_1, Y_2) is 2, correspondingly.

Asymptotic independence does not imply (complete) independence and, in fact, can mask relevant dependencies. The canonical example is that if (X_1, X_2) is standard bivariate normal with fixed correlation $\rho < 1$, then it can be shown (Sibuya, 1960) that the variables are asymptotically independent. Summarizing tail dependence with the extremal coefficient or madogram, or modeling tail dependence via $H(\boldsymbol{w})$ or $V(\boldsymbol{z})$, will ignore any dependence in an asymptotically independent couple.

Ledford and Tawn (1996) identified the problems arising with the existing modeling methodology in the case of asymptotic independence and proposed a new parameter η, the coefficient of tail dependence. If (Z_1, Z_2) has unit Fréchet marginals, Ledford and Tawn (1996) assume a joint survival function $\bar{F}$ satisfying

$$\bar{F}(z,z) \;=\; \Pr\Big(Z_1 > z, Z_2 > z\Big) \;\sim\; \mathcal{L}(z)\, z^{-1/\eta}\;, \qquad z \to \infty\;,$$

where $\mathcal{L}$ is a slowly varying function ($\mathcal{L}(tz)/\mathcal{L}(z) \to 1$ as $z \to \infty$) and $\eta \in (0,1]$.[1] The coefficient of tail dependence η is used to quantify the tail dependence in the asymptotic independent setting; $\eta = 1$ implies asymptotic dependence and $\eta < 1$ measures the degree of dependence under asymptotic independence. Ledford and Tawn (1996) spawned much subsequent work including other estimators for η (Draisma *et al.*, 2004; Peng, 1999), development of other dependence measures in the asymptotic independence setting (Coles *et al.*, 1999), and development of models in the case of asymptotic independence (Ledford & Tawn, 1997; Heffernan & Tawn, 2004; Ramos & Ledford, 2009).

Models for and measures of tail dependence for spatial extremes have thus far been limited to the asymptotically dependent case. However, an understanding of the concept of asymptotic independence is essential to fully understanding the limitations of summary dependence metrics such as $\theta(h)$ and for understanding the models for residual dependence presented in §4.

3. CAPTURING SPATIAL STRUCTURE IN MARGINAL BEHAVIOR

The representation of multivariate max-stable distributions (1.2) as well as the spatial dependence models in §§4.1, 4.2 presuppose that the univariate marginal distributions are known and common at all locations in the study region. In most applications, the study regions are large enough that the assumption of a common marginal distribution is unrealistic and the marginal distribution is not known. Therefore, it becomes essential to model the marginal distribution at all locations within the study region.

A simplistic approach would be to individually estimate the marginal distributions at each location. This approach has been used in (non-spatial) multivariate applications (e.g., Heffernan & Tawn, 2004; Cooley *et al.*, 2010). Such an approach is less-than-ideal for spatial applications for a number of reasons. First, a goal of many spatial projects is to make inference at locations where there are no data, i.e., to perform spatial interpolation. Constructing unconnected models at each location does not allow one to readily interpolate. Second, there is a desire to borrow strength across locations when estimating marginal parameters. Many spatial data have a temporal record of several decades. Such a data record

[1] We use F to denote the cdf as Ledford and Tawn (1996) do not require that (Z_1, Z_2) be max-stable.

is more-than-enough to pin down the central tendencies of the marginal distribution, but large uncertainties remain about tail quantities. It is well-documented that tail quantities, and in particular point estimates for the tail index ξ, can vary wildly over the spatial region when estimated individually (e.g., Cooley & Sain (2010)). Methods which borrow strength across locations 'trade space for time' and help to reduce uncertainties.

Below we briefly detail several methods used to capture spatial structure when describing the tail of the marginal distribution before focusing on the recent approach of hierarchical modeling.

3.1. Methods for marginal parameter estimation

As mentioned in §1.2, regression approaches are frequently used to model the mean process $\alpha(\boldsymbol{s})$ of a geostatistical model. Similarly, regression approaches have been used to model the parameters of the extreme value distributions. When modeling annual maximum precipitation in the southeast United States, Padoan *et al.* (2010) select a model in which the location parameter $\mu(\boldsymbol{s})$ and scale parameter $\sigma(\boldsymbol{s})$ are linear functions of latitude and elevation, and the shape parameter $\xi(\boldsymbol{s})$ is constant over the study region. Padoan *et al.* (2010) go on to model the residual dependence via a max-stable process (see §4.1). Mannshardt-Shamseldin *et al.* (2010) use a spatial regression approach which downscales extreme precipitation as generated by climate models to that observed at weather stations. Generally, one wishes to employ covariates which are known at all locations $\boldsymbol{s} \in \mathcal{S}$ (both observed and unobserved), and often spatial coordinates are the only such covariates. In geostatistics, regression models on spatial coordinates are known as trend surface models (Schabenberger & Gotway, 2005, §5.3.1). However simple regression models on available covariates are sometimes unable to fully capture complex spatial behavior, and Ribatet *et al.* (2011) found this to be the case in a study of extreme precipitation in Switzerland. Regional frequency analysis (RFA) (Hosking & Wallis, 1997) is a term for a statistical procedure which explicitly borrows strength across locations and which in turn characterizes the spatial nature of the tail of the marginal distribution. RFA has a long history in hydrology and traces its roots to the index flood procedure of Dalrymple (1960). RFA is a multi-step procedure which pools data over predefined subregions of $\mathcal{S}$ determined to be 'homogeneous'. Hosking & Wallis (1997) advocate an estimation method based on L-moments for RFA. A recent effort to update the precipitation-frequency atlases for the United States (Bonnin *et al.*, 2004a,b) employs an L-moment-based RFA coupled with an interpolation method based on the PRISM method (Daly *et al.*, 1994, 2002). A possible disadvantage of RFA is that it does not construct an explicit spatial model for the marginal parameters. To our knowledge, no RFA-based work has tried to account for residual dependence in the data after accounting for marginal effects.

3.2. Hierarchical modeling

Hierarchical (or multi-level) models have been extensively used in describing the relationship between observations and the complex processes that generate them. For many hierarchical models, the data collected is not well-suited for modeling within the usual Gaussian framework of geostatistics. An early example of spatial hierarchical modeling is Diggle *et al.* (1998), who model two sets of spatial data: the first data set is assumed to be Poisson distributed, the second set is assumed to be binomial. To explain the spatial variation of the data, hierarchical models typically assume that the behavior of the data over the study region is driven by an unobserved or latent process. For example, Diggle *et al.* (1998) modeled both the Poisson intensity and a risk level associated with the binomial with a Gaussian process which varied over the respective study regions. Treating the parameters of the observations' marginal distribution as spatial random effects is especially useful when the spatial behavior is too complex to capture with fixed effects. Much of the early work in spatial hierarchical modeling was done in the field of epidemiology, and the book of Banerjee *et al.* (2004) has several examples.

Hierarchical models are often devised within a Bayesian framework, and typically have three levels: (i) the data level, (ii) the process level, and (iii) the prior level. To describe the set-up of the Bayesian hierarchical model (BHM), let the vector of parameters, $\boldsymbol{\psi}$, be defined as $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)$, where $\boldsymbol{\psi}_1$ are the parameters in the data level, and $\boldsymbol{\psi}_2$ are the parameters in the process level. Then, the posterior distribution of $\boldsymbol{\psi}$ given the data $\mathbf{y}$, $\pi(\boldsymbol{\psi} \mid \mathbf{y})$, is given by

$$\pi(\boldsymbol{\psi} \mid \mathbf{y}) \;\propto\; \pi(\mathbf{y} \mid \boldsymbol{\psi}_1)\, \pi(\boldsymbol{\psi}_1 \mid \boldsymbol{\psi}_2)\, \pi(\boldsymbol{\psi}_2) \;. \tag{3.1}$$

Here $\pi(\mathbf{y} \mid \boldsymbol{\psi}_1)$ defines the likelihood function, $\pi(\boldsymbol{\psi}_1 \mid \boldsymbol{\psi}_2)$ describes the distribution of the process and $\pi(\boldsymbol{\psi}_2)$ the (hyper) priors. When applied to extremes data, the likelihood is based on an extreme-value distribution (EVD). Spatial modeling at the process level is designed to borrow strength across locations and flexibly capture spatial variation, showing how the marginal parameters of the EVD vary over the study region.

In a spatial hierarchical model, the likelihood must account for the fact that the data are observed at multiple locations. A simple approach is to assume that the data at different locations are conditionally independent given the parameters $\boldsymbol{\psi}_1$, which themselves are spatially dependent from the process level of the model. With this assumption, the likelihood becomes a product of the individual likelihoods at each location. This conditional independence assumption is widely made in hierarchical modeling, and is quite sensible in most epidemiological settings where disease counts at different locations can be assumed to be independent once the latent risk level is accounted for. However, the conditional independence assumption is questionable when modeling weather data because individual events can affect more than one location. A hierarchical model which assumes conditional independence in the likelihood ignores any residual dependence which remains after accounting for marginal effects.

Despite the aforementioned limitation of assuming conditional independence, there are several applications of BHMs in the spatial extremes literature whose primary aim is to characterize the marginal tail behavior and which make this assumption. Among the earliest work constructing BHM's for extremes is that of Casson & Coles (1999) who use a GEV-based model and study hurricane wind speed data. Fawcett & Walshaw (2006) apply a BHM to extreme wind speed data. Cooley *et al.* (2007) use a hierarchical approach to model extreme precipitation data from weather stations and then interpolate the marginal distribution over the study region to produce return level maps. Sang & Gelfand (2009) modeled annual maximum rainfall over a region of South Africa, using a spatial autoregressive model in the process level of their hierarchy, rather than a geostatistical model. Both Cooley & Sain (2010) and Schliep *et al.* (2010) use a BHM to characterize extreme precipitation as generated by climate models over spatial regions with thousands of locations. Mendes *et al.* (2010) apply BHMs to spatial extremes of large wildfire sizes.

Recent hierarchical work in extremes has sought to move beyond the conditional independence assumption and capture both regional spatial effects via a process-level spatial model on marginal parameters and local spatial effects via a dependence structure within the likelihood at the data level. Sang & Gelfand (2010) and Fuentes *et al.* (2011) respectively use a Gaussian copula model and a Dirichlet process model to try and capture residual dependence within the structure of a hierarchical model. Ribatet *et al.* (2011) use max-stable process models to formulate a likelihood designed to capture residual dependence. We discuss these approaches in depth in §4.

Inference for BHMs is done by sampling, which is often complicated by the fact that the full conditional distributions for the parameters often do not exist in closed form. Markov Chain Monte Carlo (MCMC) methods can be used to approximate the posterior distributions. One of the most popular approaches is the Gibbs sampling method (Gelfand & Smith, 1990), which is accommodated by the conditional relationships in (3.1). As conjugate priors for EVD's are not known, BHMs for extremes require a Metropolis–Hastings step to be included within each iteration of the Gibbs sampler.

Both hierarchical approaches which assume conditional independence and other methods such as RFA ignore residual dependence which arises due to the spatial extent of individual events. In instances when interest is primarily on the marginal behavior, it may be appropriate to ignore the residual dependence. There is a long history in hydrology of producing return level maps; that is, a map which depicts a high quantile (e.g., the quantile associated with a 100-year return level) at any site within a study region. The aforementioned projects by Bonnin *et al.* (2004a,b) produced such maps and NOAA's Hydrometeorological Design Studies Center maintains a website[2] which provides point-located return level estimates. Likewise, projects such as Cooley & Sain (2010), Sang & Gelfand (2009) and Cooley *et al.* (2007) aimed only to estimate pointwise return levels. It is imperative that it is recognized that such projects cannot be used to quantify the aggregated effects of a large event which occurs across multiple locations, nor can they be used to produce realistic simulated data (Davison *et al.*, 2012, Figure 4).

4. MODELS FOR SPATIAL DEPENDENCE

In this section, we consider two approaches for capturing the residual dependence assuming that the marginal effects have been accounted for. The first approach is to use a max-stable process model (§ 1.1), which will assume that the marginals have been transformed to be unit Fréchet. The second is a copula approach, a popular recent approach for modeling multivariate data which assumes the marginals are Uniform [0,1].

4.1. Max-stable processes

The definition of a max-stable process, Equation (1.4), as the infinite dimensional generalization of a multivariate max-stable distribution gives a well-defined class of models, but it does not suggest how to construct such processes. A conceptual construction of max-stable processes was first given with a spectral representation of extremal processes by de Haan (1984) and de Haan & Ferreira (2006). A general representation of max-stable processes can be described by two components, a stochastic process $\{X(\boldsymbol{s})\}$ and a Poisson process Π with intensity $\mathrm{d}\zeta/\zeta^2$ on $(0,\infty)$. Let $\{X_i(\boldsymbol{s})\}_{i\in\mathbb{N}}$ be independent realizations of a process $X(\boldsymbol{s})$ with $\boldsymbol{E}[X(\boldsymbol{s})]=1$, and let $\zeta_i \in \Pi$ be points of the Poisson process. Then

$$Z(\boldsymbol{s}) \,=\, \max_{i\geq 1} \zeta_i\, X_i(\boldsymbol{s}) \ , \qquad \boldsymbol{s}\in\mathcal{S} \ ,$$

is a max-stable process with unit-Fréchet margins and the distribution function is determined by

$$\Pr\Big(Z(\boldsymbol{s}) \leq z(\boldsymbol{s}),\, \boldsymbol{s}\in\mathcal{S}\Big) \,=\, \exp\left(-E\left[\sup_{\boldsymbol{s}\in\mathcal{S}}\left\{\frac{X(\boldsymbol{s})}{z(\boldsymbol{s})}\right\}\right]\right) ,$$

where minus the exponent is the infinite-dimensional analogue to V; see Equation (1.3). Different choices of the process $X_i(\boldsymbol{s})$ lead to different classes of max-stable processes.

[2] `http://hdsc.nws.noaa.gov/hdsc/pfds/index.htm`

Although Gaussian distributions and processes are not well-suited for modeling extremes, they can be used within the above max-stable construction to produce appropriate models. This was first proposed by Smith (1990). Let $\{(\zeta_i, \boldsymbol{s}_i);\, i \geq 1\}$ denote the points of a Poisson process on $(0,\infty) \times \mathcal{S}$ with intensity $\zeta^{-2}\, d\zeta\, d\boldsymbol{s}$. Let $\{f(x)\}$ on $\mathbb{R}^d$ denote a non-negative function such that $\int f(x)\, dx = 1$ and define

$$Z(\boldsymbol{s}) = \max_i \zeta_i\, f(\boldsymbol{s} - \boldsymbol{s}_i)\ .$$

Then $Z(\cdot)$ is a max-stable process with unit-Fréchet margins. Smith also introduced the so-called *rainfall-storms* interpretation: think of $f(\cdot)$ as the shape of a storm centered at point $\boldsymbol{s}_i$, and think of the overall magnitude of storm as ζ_i. Then the max-stable process $Z(\cdot)$ is the pointwise maximum rainfall (taken over all storms) for each location in $\mathcal{S}$. If $f(\cdot)$ is a multivariate normal density with covariance parameter Σ, then the process $Z(\cdot)$ is called a *Gaussian extreme value process* and the joint distribution at two sites is given by

$$\Pr\Big\{Z(\boldsymbol{s}_1) \leq z_1,\, Z(\boldsymbol{s}_2) \leq z_2\Big\} = \\ = \exp\left\{-\frac{1}{z_1}\,\Phi\Big(\frac{a}{2} + \frac{1}{a}\,\log\frac{z_2}{z_1}\Big) - \frac{1}{z_2}\,\Phi\Big(\frac{a}{2} + \frac{1}{a}\,\log\frac{z_1}{z_2}\Big)\right\},$$

where $a = \sqrt{(\boldsymbol{s}_1 - \boldsymbol{s}_2)^{\mathrm{T}}\, \Sigma^{-1}(\boldsymbol{s}_1 - \boldsymbol{s}_2)}$ and Φ is the standard normal cumulative distribution function. The dependence parameter a represents a transformed distance between two sites and the limits $a \to 0$ and $a \to \infty$ correspond to perfect dependence and independence, respectively. Thus the most common Smith model takes $X_i(s)$ to be a multivariate density function. Figure 2 shows a realization.

Schlather (2002) suggested a more flexible class of max-stable processes by taking $X_i(\boldsymbol{s})$ to be any stationary Gaussian random field with finite expectation. A stationary max-stable process with unit-Fréchet margins can be obtained by

$$Z(\boldsymbol{s}) = \max_i \zeta_i\ \max\big\{0, X_i(\boldsymbol{s})\big\}$$

where $\mu = E\big\{\max(0, X_i(\boldsymbol{s}))\big\} < \infty$ and $\{\zeta_i\}$ denotes the points of a Poisson process on $(0,\infty)$ with intensity measure $\mu^{-1}\zeta^{-2}\, d\zeta$. The max-stable process also allows a useful interpretation of spatial storm events. On taking a stationary random process $X_i(\boldsymbol{s})$, the spatial rainfall events have the same dependence structure but the realizations will vary stochastically, not the deterministic form $f(\cdot)$ such as Smith's model. If the random process is specified for a stationary isotropic Gaussian random field $X_i(\cdot)$ with unit variance, correlation $\rho(\cdot)$ and $\mu^{-1} = \sqrt{2\pi}$, then the process $Z(\boldsymbol{s})$ is called an *extremal Gaussian process* and the bivariate marginal distributions are given by

$$\Pr\Big\{Z(\boldsymbol{s}_1) \leq z_1,\, Z(\boldsymbol{s}_2) \leq z_2\Big\} = \\ = \exp\left\{-\frac{1}{2}\Big(\frac{1}{z_1} + \frac{1}{z_2}\Big)\left(1 + \sqrt{1 - 2\big(\rho(h) + 1\big)\,\frac{z_1 z_2}{(z_1 + z_2)^2}}\right)\right\}$$

where h is the Euclidean distance between station $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$. The correlation is chosen from one of the valid families of correlations for Gaussian processes. Figure 2 shows a realization of an extremal Gaussian process.

One drawback to the Schlather model is that it cannot attain independence of extremes, because the extremal coefficient $\theta(h) = 1 + \left[\frac{1-\rho(h)}{2}\right]^{1/2}$ takes the value in the interval $[1, 1.838]$ (and not the usual range of [1,2]) as the distance $h \to \infty$. To overcome this, the process $X_i(\boldsymbol{s})$ can be restricted to a random set $\mathcal{B}$, i.e.,

$$Z(\boldsymbol{s}) = \max_i \zeta_i \, X_i(\boldsymbol{s}) \, I_{\mathcal{B}_i}(\boldsymbol{s} - \boldsymbol{s}_i)$$

where $I_{\mathcal{B}}$ is the indicator function of a compact random set $\mathcal{B} \subset \mathcal{S}$ and $\boldsymbol{s}_i$ are the points of a Poisson process on $\mathcal{S}$. If X_i is a Gaussian process, the bivariate marginal distribution is

$$\exp\left\{-\left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left[1 - \frac{\alpha(h)}{2}\left(1 - \sqrt{1 - 2\big(\rho(h)+1\big)\,\frac{z_1 z_2}{(z_1+z_2)^2}}\right)\right]\right\}$$

where $\alpha(h) = E\big\{|\mathcal{B} \cap (h+\mathcal{B})|\big\}/E(|\mathcal{B}|) \in [0,1]$. The extremal coefficient takes on any value in the interval $[1,2]$ and thus independent extremes are available. One possible choice for the set $\mathcal{B}$ is a disc of radius r, meaning $\alpha(h) \doteq \{1 - |h|/(2r)\}_+$, which equals 0 when $|h| > 2\,r$. One could consider $\mathcal{B}$ as a random set, which means that the radius of the disk is random and all elliptical sets having the same major axis are permissible as a generalization for $\mathcal{B}$ (Davison & Gholamrezaee, 2012).

Kabluchko *et al.* (2009) proposed an alternative specification for the $X_i(\cdot)$ process, which requires weaker assumptions than second-order stationarity. Let $X_i(\boldsymbol{s}) = \exp\big\{e_i(\boldsymbol{s}) - \frac{1}{2}\,\sigma^2(\boldsymbol{s})\big\}$ where $e_i(\boldsymbol{s})$ is a Gaussian process with stationary increments and $\sigma^2(\boldsymbol{s}) = \mathrm{Var}\{e(\boldsymbol{s})\}$. Then the process, known as the *Brown–Resnick process*, can be a very general class of max-stable processes which allows the use of semivariogram from standard geostatistics. The bivariate cdf transformed to

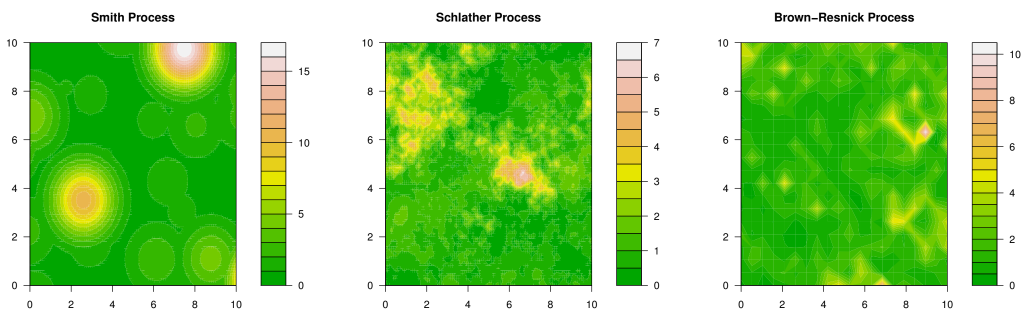

Figure 2: Realization of Gaussian extreme value process (left), extremal Gaussian process (centre), and Brown–Resnick process (right) from SpatialExtremes R package (Ribatet, 2011; R Development Core Team, 2011).

unit Fréchet margins is the same as the so-called Smith model but instead of the parameter a^2 we have $\gamma(\cdot)$, the semivariogram of $e(\cdot)$, i.e.

$$\Pr\Big\{Z(\boldsymbol{s}_1) \le z_1, Z(\boldsymbol{s}_2) \le z_2\Big\} =$$
$$= \exp\left\{-\frac{1}{z_1}\,\Phi\!\left(\frac{\sqrt{\gamma(h)}}{2} + \frac{1}{\sqrt{\gamma(h)}}\,\log\frac{z_2}{z_1}\right) - \frac{1}{z_2}\,\Phi\!\left(\frac{\sqrt{\gamma(h)}}{2} + \frac{1}{\sqrt{\gamma(h)}}\,\log\frac{z_1}{z_2}\right)\right\},$$

where Φ is the standard normal distribution function and h is the Euclidean distance between location $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$.

Realizations of the Gaussian extreme value process, the extremal Gaussian process, and the Brown–Resnick process are shown in Figure 2. The left panel shows a simulation from the Gaussian extreme value process with covariance matrix ($\sigma_{11} = \sigma_{22} = 9/8$ and $\sigma_{12} = 0$), the central panel shows shows one from the extremal Gaussian process with the Whittle–Matérn correlation function (nugget $= 0$, range $= 3$ and smooth $= 1$), and the right panel shows one from the Brown–Resnick process with parameter (range $= 3$, smooth $= 0.5$).

4.1.1. Inference for max-stable processes: composite likelihood approach

A barrier to fitting max-stable processes to data is that closed-form expressions for the joint likelihood can only be written out in low dimensional settings. The likelihood for the Smith model in $\mathbb{R}^2$ can be written out for dimension $d \le 3$ (Genton *et al.*, 2011), but the likelihood for all other max-stable processes can only be written for dimension $d \le 2$ (and we write this section for the more general case). This means that if the data are observed at $d > 2$ locations in space, the joint likelihood cannot be written in closed form. Padoan *et al.* (2010) proceeded with a likelihood-based approach to fitting max-stable processes by substituting a composite likelihood for the unavailable joint likelihood. We first introduce composite likelihoods, then show the connection to max-stable processes.

If $f(z;\psi)$ is a statistical model and we have a set of marginal or conditional events $\{\mathcal{A}_j \subseteq \mathcal{F},\ j = 1, \ldots, J\}$ where $\mathcal{F}$ is a sigma algebra on $\mathcal{Z}$ and each event A_j has associated likelihood $\mathcal{L}_j(\psi; z) \propto f(z \in \mathcal{A}_j; \psi)$, then a composite log-likelihood is a weighted sum of log-likelihoods for each event (Lindsay, 1988; Varin, 2008):

$$\ell_C(\psi; z) = \sum_j w_j \cdot \log f\big(z \in \mathcal{A}_j; \psi\big)\ ,$$

where w_j is a weight function on j^{th} event. If the weights are all equal they may be ignored, though non-equal weights may be used to improve the statistical performance in certain cases. One example of a composite log-likelihood is the pairwise log-likelihood, defined (in a spatial application) as

$$\ell_C(\psi; z) \;=\; \sum_{k=1}^{m} \sum_{j=1}^{d-1} \sum_{j'=j+1}^{d} \log f\bigl(z_k(\boldsymbol{s}_j),\, z_k(\boldsymbol{s}_{j'});\, \psi\bigr)\;,$$

where each term $f\bigl(z_k(\boldsymbol{s}_j),\, z_k(\boldsymbol{s}_{j'});\, \psi\bigr)$ is a bivariate marginal density function based on observations at locations j and j', and ψ is a spatial dependence parameter. The two inner summations sum over all unique pairs, while the outer sums over the m i.i.d. replicates. Similar to the full likelihood function, the parameter which maximizes a composite log likelihood can be found, and is termed a *maximum composite likelihood estimate*, or MCLE. When $m \to \infty$, the maximum composite likelihood estimator is consistent and asymptotically normal (Lindsay, 1988; Cox & Reid, 2004), with

$$\text{(4.1)} \qquad \hat{\psi}_{\text{MCLE}} \sim \mathcal{N}(\psi, \tilde{I}^{-1})\;, \qquad \tilde{I} = H(\psi)\, J^{-1}(\psi)\, H(\psi)\;,$$

where $H(\psi) = \mathrm{E}(-H_\psi\, \ell_C(\psi; Z))$ is the expected information matrix, $J(\psi) = V(D_\psi\, \ell_C(\psi; Z))$ is the covariance of the score, H_ψ is the Hessian matrix, D_ψ is the gradient vector, and V is the covariance matrix. When one has the full likelihood, $H(\psi) = J(\psi)$, but in the composite likelihood setting these matrices are not equal.

Padoan *et al.* (2010) used the composite likelihood to model the joint spatial dependence of extremes and accounted for regional effects with a regression model on the GEV parameters. This approach is implemented in the `R` package `SpatialExtremes` (Ribatet, 2011). The maximum composite likelihood estimator $\hat{\psi}_{\text{MCLE}}$ is found numerically. The variance is estimated using

$$\hat{H}(\hat{\psi}_{\text{MCLE}}) \;=\; -\sum_{k=1}^{m} \sum_{j=1}^{d-1} \sum_{j'=j+1}^{d} H_\psi \log f\bigl(z_k(\boldsymbol{s}_j),\, z_k(\boldsymbol{s}_{j'});\, \hat{\psi}_{\text{MCLE}}\bigr)$$

and

$$\begin{aligned}\hat{J}(\hat{\psi}_{\text{MCLE}}) \;=\; \sum_{k=1}^{m} &\left\{ \sum_{j=1}^{d-1} \sum_{j'=j+1}^{d} D_\psi \log f\bigl(z_k(\boldsymbol{s}_j),\, z_k(\boldsymbol{s}_{j'});\, \hat{\psi}_{\text{MCLE}}\bigr) \right\} \\ &\times \left\{ \sum_{j=1}^{d-1} \sum_{j'=j+1}^{d} D_\psi \log f\bigl(z_k(\boldsymbol{s}_j),\, z_k(\boldsymbol{s}_{j'});\, \hat{\psi}_{\text{MCLE}}\bigr) \right\}^{\mathrm{T}}.\end{aligned}$$

The dependence parameter ψ is generic and stands in for the matrix Σ in the Smith model, the Gaussian correlation function $\rho(h; \psi)$ in the Schlather model, and the semivariogram $\gamma(h; \psi)$ in the Brown–Resnick model. For each of these models the target parameter appears in the corresponding bivariate density functions, and thus also in the pairwise log-likelihood.

Recently, there has been work which begins to explore the use of composite likelihood methods within Bayesian inference, and much of this work has been driven by interest in spatial extremes. Both Pauli *et al.* (2011) and Ribatet *et al.* (2011) seek to employ a pairwise likelihood, rather than the unattainable true likelihood, to obtain a posterior distribution. The pairwise likelihood does not accurately represent the information in the data, as it repeatedly uses each observation when pairing it with others. Pauli *et al.* (2011) adjust the pairwise likelihood so that the first moment of the log-likelihood ratio corresponds to that of the asymptotic χ^2-distribution. Ribatet *et al.* (2011) suggest an adjustment which ensures that the curvature of the likelihood surface agrees with the asymptotic covariance matrix given in Equation (4.1). Ribatet *et al.* (2011) apply the likelihood within a spatial hierarchical model to study extreme precipitation in Switzerland.

Erhardt & Smith (2011) used approximate Bayesian computing (ABC) to obtain an approximate posterior distribution for the max-stable process dependence parameters ψ. ABC methods have been successfully implemented for problems where the joint likelihood function is intractable, but simulations are possible (Beaumont *et al.*, 2002; Sisson & Fan, 2010). Given observed data Z and prior $\pi(\psi)$, the simplest ABC algorithm is: (1) Draw $\psi' \sim \pi(\psi)$; (2) Simulate a new dataset Z' conditional on ψ'; (3) If $d(S(Z), S(Z')) \leq \epsilon$ for some summary statistic S, distance function d, and threshold ϵ, then accept ψ'; otherwise, reject. The method produces an i.i.d. sample from $\pi\big[\psi \mid d\{S(Z), S(Z')\} \leq \epsilon\big]$, which in the limit as $\epsilon \to 0$ equals $\pi(\psi \mid S(Z))$. Further, if $S(Z)$ were a sufficient statistic, this would be the exact posterior. In practice, computational costs often force concessions like in-sufficient statistics S and a non-zero threshold ϵ. Erhardt & Smith (2011) used tripletwise extremal coefficients in the construction of a summary statistic S, and then showed that the resulting ABC implementation can outperform the composite likelihood approach when estimating the spatial dependence of a max-stable process, though at an appreciably higher computational cost.

4.1.2. Spatial prediction/interpolation for max-stable processes

Kriging is a central focus of geostatistics and there has been recent work to perform spatial prediction for max-stable processes (§1.2). Wang & Stoev (2011) and Dombry *et al.* (2011) have proposed computational solutions for the prediction problem for max-stable processes.

Max-linear models are a subclass of the multivariate max-stable distributions. Let Y_i, $i = 1, \ldots, n$, be i.i.d. unit Fréchet random variables. Assume $\boldsymbol{Z} = \big(Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_d)\big)^{\mathrm{T}}$ is a max-linear combination of the Y_i's; that is,

$$Z(\boldsymbol{s}_j) = \bigvee_{i=1}^{n} c_{j,i} Y_i \ , \qquad j = 1, \ldots, d \ ,$$

where $c_{j,i}$ are non-negative constants. $\boldsymbol{Z}$ is multivariate max-stable, and further if $\sum_{i=1}^{n} c_{j,i} = 1$ for all j, then $\boldsymbol{Z}$ has unit-Fréchet marginals. Any max-linear model with finite n will have a discrete angular measure; however, max-linear models form a dense subclass of multivariate max-stable random vectors as $n \to \infty$ (Zhang & Smith, 2004). Wang & Stoev (2011) propose an algorithm for efficient and exact sampling from the conditional distributions of a spectrally discrete max-stable random field. The main idea is to first generate samples from the regular conditional probability distribution of $\boldsymbol{Y} \mid \boldsymbol{Z} = \boldsymbol{z}$, where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ and $\boldsymbol{Z}$ is the vector of values at observed locations. Then, the conditional distribution of $Z(\boldsymbol{s}_0) = \bigvee_{i=1}^{n} c_{0,i} Y_i$ can be easily obtained for any given $c_{0,i}$. The performance of the algorithm was illustrated over the discretized Smith model for spatial extremes.

Dombry *et al.* (2011) also take a computational approach to spatial prediction. Specifically working with the Brown–Resnick process, Dombry *et al.* (2011) establish a link between the conditional distribution of this process and the multivariate log-normal distribution. Like Wang & Stoev (2011), the computational method of Dombry *et al.* (2011) considers different hitting-scenarios; that is, the possible combinations of individual events which could yield the observed maxima. Dombry *et al.* (2011) illustrate their method on precipitation data from Switzerland.

4.2. Copula approaches

Copulas provide another framework for representing the dependence structure of a multivariate distribution with known marginals. Copulas are multivariate distributions with standard uniform marginal distributions, and they characterize the dependence structure of a multivariate distribution from univariate marginal distributions by defining a joining mechanism (Nelsen, 2006; Joe, 1997). Given a d-dimensional random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^{\mathrm{T}}$ with corresponding marginal cdfs F_i for $j = 1, \ldots, d$ and joint distribution function F, a copula is a function $C\colon [0,1]^d \longrightarrow [0,1]$ such that

$$F(\boldsymbol{Y}) = C\big(F_1(Y_1), \ldots, F_d(Y_d)\big) . \tag{4.2}$$

If the marginal cdfs of $\boldsymbol{Y}$ are all continuous, then the copula function C is uniquely defined by (4.2). Conversely, for a copula C and continuous margins $F_1, \ldots, F_d$, the copula C corresponds to the distribution of $F_1(Y_1), \ldots, F_d(Y_d)$, i.e.,

$$C\big(u_1, \ldots, u_d\big) = F\big(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\big) . \tag{4.3}$$

This follows from the multi-dimensional analog of Sklar's theorem (Sklar, 1959), which proves the existence of a copula in the bivariate case and uniqueness when the marginals are continuous (Nelsen, 2006).

The copula framework has appeal for modeling multivariate extremes, especially since the list of existing parametric subfamilies of the multivariate extreme value distribution (MEVDs) is limited. When working with multivariate block maximum data, extreme value theory suggests that each marginal should be approximately GEV distributed. Equation (4.3) suggests one can combine knowledge of the marginal distributions with a copula model to obtain a valid cdf. Further, (4.2) says that one can obtain a copula model from any multivariate distribution. While this approach allows one great flexibility to create multivariate distributions with GEV marginals, these distributions will not correspond to a MEVD as characterized in § 1.1 unless one uses an extremal copula model (Joe, 1997), which essentially correspond to the existing parametric MEVD subfamilies. Use of (nonextremal) copula models to describe extremes has been controversial, and Mikosch (2006) and the associated discussion details much of the argument.

For spatial data which are typically observed at many locations, one would need a copula which can handle high dimensions, and further, whose pairwise dependence can be linked to distance. Two obvious choices are to use the multivariate Gaussian or multivariate Student t distributions to generate a copula. Given a d-dimensional Gaussian random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^{\mathrm{T}}$, the Gaussian copula is defined as

$$C_\Sigma\big(u_1, \ldots, u_d\big) \;=\; \Phi_\Sigma\big(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\big) \,, \tag{4.4}$$

where $u_i \in [0,1]$ for $i = 1, \ldots, d$, Φ is the cdf for a standard normal distribution, and Φ_Σ is the joint cdf of the multivariate normal distribution with covariance matrix Σ. The multivariate Student t distribution copula is defined similarly, except the marginal and joint distributions displayed in (4.4) are replaced by marginal and joint (with scale matrix Σ) Student t distributions , i.e., $C_\Sigma(u_1, \ldots, u_d) = T_{\nu;\Sigma}\big(T_\nu^{-1}(u_1), \ldots, T_\nu^{-1}(u_d)\big)$, where $T_{\nu;\Sigma}$ and T_ν are the joint and marginal t distributions with ν degrees of freedom respectively.

The simplicity of the Gaussian and Student t copulas make them appealing, but they have some undesirable properties from the extreme value theory viewpoint. Because neither copula is extremal, the resulting cdf will not be max-stable. Additionally, the Gaussian copula is asymptotically independent, so it may underestimate joint tail probabilities if the data are actually asymptotically dependent.

Hüsler & Reiss (1989) devised an alternative approach which used the bivariate Gaussian distribution to create an extremal copula (equivalently a MEVD), and this general approach can be used with either the multivariate Gaussian distribution or multivariate t distribution as described by Davison *et al.* (2012). The approach is essentially the same as that which leads to the Smith (1990) max-stable process model, albeit in the multivariate rather than process setting. Thus, the drawback of the extremal Gauss or extremal t copulas is that the full multivariate distribution cannot be written in closed form, and composite likelihood methods must be employed. Davison *et al.* (2012) fit the Gaussian and t copulas as well as the extremal Gauss and t copulas to data which are annual maxima and conclude that the extremal copulas have improved ability to capture the dependence found in this data. Padoan (2011) discusses some further copula models potentially useful in spatial extremes.

In a desire to move away from the conditional independence assumption often made in hierarchical modeling, Sang & Gelfand (2010) use a Gaussian copula to create a likelihood function in a model structured as (3.1). Suppose that $Y(\boldsymbol{s})$ is an extremal spatial process at location $\boldsymbol{s}$, e.g., $Y(\boldsymbol{s})$ is the annual maximum of daily rainfall measurements at location $\boldsymbol{s}$. Then $Y(\boldsymbol{s}) \sim \mathrm{GEV}\big(\mu(\boldsymbol{s}), \sigma(\boldsymbol{s}), \xi(\boldsymbol{s})\big)$. Furthermore, $Y(\boldsymbol{s})$ can be represented as

$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \frac{\sigma(\boldsymbol{s})}{\xi(\boldsymbol{s})}\Big(Z(\boldsymbol{s})^{\xi(\boldsymbol{s})} - 1\Big) \tag{4.5}$$

where $Z(\boldsymbol{s})$ is unit-Fréchet. Focusing on the underlying unit-Fréchet process defined at locations $\boldsymbol{s}_1, ..., \boldsymbol{s}_d$, Sang & Gelfand (2010) induced a dependence structure on the $Z(\boldsymbol{s}_i)$ for $i = 1, ..., d$ using the Gaussian copula of (4.4). Suppose that $\mathbf{X}(\boldsymbol{s}) = \big(X(\boldsymbol{s}_1), ..., X(\boldsymbol{s}_d)\big)$ is a centered spatial Gaussian process with a correlation structure defined by the function $\rho(\boldsymbol{s}_i, \boldsymbol{s}_j; \psi)$ for $i, j = 1, ..., d$. The unit-Fréchet process $Z(\boldsymbol{s})$ is defined in terms of a transformed spatial Gaussian process $\mathbf{X}(\boldsymbol{s})$ by

$$\mathbf{Z} = \Big(G^{*-1}\Phi\big(X(\boldsymbol{s}_1)\big), \; ..., \; G^{*-1}\big(\Phi(X(\boldsymbol{s}_d))\big)\Big) \,,$$

where G^* is the unit-Fréchet distribution function. Given the spatial Gaussian process $X(\boldsymbol{s})$, the corresponding copula is defined as $C_X(u_1, ..., u_d) = F_{X,\Sigma_\rho}\big(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)\big)$ where $u_1, ..., u_d \in [0, 1]$, F_{X,Σ_ρ} is the multivariate Gaussian distribution function with a covariance matrix Σ_ρ defined by some correlation function $\rho(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\psi})$, $i, j = 1, ..., d$. Then the multivariate joint distribution of $\mathbf{Z}$ is

$$F_Z(z_1, ..., z_d) = C_X\big(G^*(z_1), ..., G^*(z_d)\big) = F_X\big(\Phi^{-1}G^*(z_1), ..., \Phi^{-1}G^*(z_d)\big) \,,$$

where $z_j \in \mathbb{R}$, $j = 1, ..., d$. Sang & Gelfand (2010) considered this the first stage of a hierarchical model, where the second stage deals with the characterizations of $\mu(\boldsymbol{s})$, $\sigma(\boldsymbol{s})$, and $\xi(\boldsymbol{s})$ found in equation (4.5). While one could be critical of the Gaussian copula that Sang & Gelfand (2010) employ, the approach does attempt to account for the residual spatial dependence which remains in the data after accounting for marginal effects, while a conditional independence assumption ignores this dependence entirely.

An alternative copula approach for spatial extremes data is proposed by Fuentes *et al.* (2011), who use a Dirichlet process (DP) copula model which is a Bayesian, nonparametric generalization of the spatial Gaussian copula model.

The DP copula model is flexible and capable of dealing with high dimensional data, but when combined with GEV marginal distributions does not yield a MEVD. The DP prior is a special case of the stick-breaking prior (Sethuraman, 1994). If a distribution function is such that

$$F \overset{d}{=} \sum_{i=1}^{m} p_i \, \delta_{\psi_i} \ , \tag{4.6}$$

where δ_x denotes a Dirac measure at x, $p_1 = V_1$, $p_i = V_i \prod_{j<i}(1 - V_j)$ with $V_i \overset{\text{iid}}{\sim} \text{Beta}(1, \nu)$, $\nu > 0$ and $\psi_1, \ldots, \psi_m \overset{\text{iid}}{\sim} H_0$, which is a centering distribution. In the spatial setting, the spatial DP copula introduces a latent process $\boldsymbol{X}$ such that the joint density of $\boldsymbol{X} = \big(X(\boldsymbol{s}_1), \ldots, X(\boldsymbol{s}_d)\big)$ is almost surely of a countable mixture of normals,

$$f\big(\boldsymbol{X} \,|\, H^d, \tau^2\big) \;=\; \sum_{i=1}^{\infty} p_i \, \phi_d\big(\boldsymbol{X} \,|\, \boldsymbol{\psi}_i, \tau^2 \boldsymbol{I}_d\big) \ , \tag{4.7}$$

where $\phi_d(\,\cdot \,|\, \boldsymbol{\mu}, \Sigma)$ denotes the d-dimensional multivariate normal density with mean $\boldsymbol{\mu}$ and covariance matrix Σ, p_i has the same distribution as in (4.6), and $\boldsymbol{\psi}_i = \big(\psi_i(\boldsymbol{s}_1), \ldots, \psi_i(\boldsymbol{s}_d)\big)$ with $\boldsymbol{\psi}_i \,|\, H^d \overset{\text{iid}}{\sim} H^d$ and $H^d \overset{d}{=} DP(\nu H_0^d)$, where $H_0^d \overset{d}{=} \phi_d(\cdot\,|\mathbf{0}, \Sigma)$. Let $F_{\boldsymbol{X}}$ be the cdf associated with the density in (4.7). Then the multivariate joint distribution of $\boldsymbol{Y}$ given by the copula

$$F(\boldsymbol{Y}) \;=\; C_{\boldsymbol{X}}\big(G^*(y_1), \ldots, G^*(y_d)\big) \;=\; F_{\boldsymbol{X}}\big(H_{\boldsymbol{s}_1}^{-1} G^*(y_1), \ldots, H_{\boldsymbol{s}_d}^{-1} G^*(y_d)\big) \ ,$$

where $H_{\boldsymbol{s}}$ is the cdf corresponding to the density in Equation (4.7) for $Z(\boldsymbol{s})$. As with the spatial Gaussian copula in Sang & Gelfand (2010), the unit-Fréchet spatial process is the transformed process defined as $Y(\boldsymbol{s}) = G^{*-1} H_{\boldsymbol{s}}(Z(\boldsymbol{s}))$ but with space-dependent parameters. The joint distribution function given by a spatial Gaussian copula can be expressed explicitly for any given set of locations, whereas the DP copula represents the joint distribution implicitly in a Bayesian framework.

5. DISCUSSION

The study of spatial extremes is young, as evidenced by the fact that the majority of the work referenced in this survey article has been done in the last 10 years, and nearly all of it in the last 20 years. The study has a well-developed foundation in probability theory, and the field has made great strides in developing realistic models based on this theory.

We anticipate further development of methods and models for spatial extremes, and we imagine this work will follow similar themes to recent work in geostatistics. It remains a challenge to fit max-stable process models to data recorded at many locations, for example climate model output which has thousands of locations. Interest in large data sets has spawned geostatistics work in Gaussian Markov random fields (Rue & Held, 2005; Lindgren *et al.*, 2011), fixed-rank kriging (Cressie & Johannesson, 2008), and predictive processes (Banerjee *et al.*, 2008). Another interest in geostatistics has been space-time modeling (Cressie & Wikle, 2011), an area which spatial extremes modeling is only beginning to address (Huser & Davison, 2012). Likewise, models with nonstationary dependence structure (e.g., Sampson & Guttorp, 1992), models which can handle spatial misalignment (e.g., Berrocal *et al.*, 2010), and models for multivariate spatial data (e.g., Wackernagel, 2003) have been of interest in spatial statistics for some time, but these topics are thus far unexplored for spatial extremes.

This article has largely focused on modeling data which are block maxima, as the spatial extremes literature to date has concentrated on such data. There is work in progress (Jeon, 2012) to extend max-stable process models to appropriately model threshold exceedance data. Spatial threshold exceedance modeling would address the first issue that when restricting one's attention to block maximum data, one is likely to discard other extreme data which could be useful in describing the spatial extent of extreme events. Although from the classical extreme value theory point-of-view it is natural to consider vectors of componentwise maxima (see §1.1), practitioners can view these vectors as artificial since they likely do not appear in the data record.

Finally, we would be remiss in this survey if we did not mention the available software for analyzing spatial extremes data. The `R` package `SpatialExtremes` (Ribatet, 2011) can be used both to estimate spatial dependence and to fit max-stable process models. The `RandomFields` package (Schlather, 2011) is useful for simulating max-stable processes. The conditional simulation method of Wang & Stoev (2011) can be found in the package `maxLinear` (Wang, 2010).

ACKNOWLEDGMENTS

This work was supported by the 2011–12 Program on Uncertainty Quantification at the Statistical and Applied Mathematical Sciences Institute. Additionally, D. Cooley was partially supported by National Science Foundation grant DMS-0905315.

REFERENCES

Banerjee, S.; Carlin, B. & Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Chapman, Boca Raton.

Banerjee, S.; Gelfand, A.; Finley, A. & Sang, H. (2008). Gaussian predictive process models for large spatial data sets, *J. R. Statist. Soc. B*, **70**, 825–848.

Beaumont, M.; Zhang, W. & Balding, D. (2002). Approximate Bayesian computing in population genetics, *Genetics*, **162**, 2025–2035.

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

Berrocal, V.; Gelfand, A. & Holland, D. (2010). A bivariate space-time downscaler under space and time misalignment, *Ann. Appl. Statist.*, **4**, 1942–1975.

Bonnin, G.M.; Todd, D.; Lin, B.; Parzybok, T.; Yekta, M. & Riley, D. (2004a). *NOAA Atlas 14, Precip. Freq. Atlas United States*, Vol. 1, U.S. Dep. Comm., Nat. Oceanic Atmos. Administ, Nat. Weather Serv., Silver Spring, Maryland.

Bonnin, G.M.; Todd, D.; Lin, B.; Parzybok, T.; Yekta, M. & Riley, D. (2004b). *NOAA Atlas 14, Precip. Freq. Atlas United States*, Vol. 2, U.S. Dep. Comm., Nat. Oceanic Atmos. Administ, Nat. Weather Serv., Silver Spring, Maryland.

Brockwell, P. & Davis, R.A. (2002). *Introduction to Time Series and Forecasting* (2nd ed.), Springer, New York.

Casson, E. & Coles, S.G. (1999). Spatial regression models for extremes, *Extremes*, **1**, 449–468.

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

Coles, S.G. & Dixon, M.J. (1999). Likelihood-based inference for extreme value models, *Extremes*, **2**, 5–23.

Coles, S.G.; Heffernan, J. & Tawn, J.A. (1999). Dependence measures for extreme value analysis, *Extremes*, **2**, 339–365.

Cooley, D.; Davis, R.A. & Naveau, P. (2010). The pairwise beta: A flexible parametric multivariate model for extremes, *J. Mult. Anal.*, **101**, 2103–2117.

Cooley, D.; Naveau, P. & Poncet, P. (2006). Variograms for spatial max-stable random fields, *Lect. Notes Statist.*, **187**, 373–390, Springer, New York.

Cooley, D.; Nychka, D. & Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels, *J. Am. Statist. Assoc.*, **102**, 824–840.

Cooley, D. & Sain, S.R. (2010). Spatial hierarchical modeling of precipitation extremes from a regional climate model, *J. Agric. Bio. Environ. Statist.*, **15**, 381–402.

Cox, D.R. & Reid, N. (2004). A note on pseudo-likelihood constructed from marginal densities, *Biometrika*, **91**, 729–737.

Cressie, N.A.C. & Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets, *J. R. Statist. Soc. B*, **70**, 209–226.

Cressie, N.A.C. & Wikle, C. (2011). *Statistics for Spatio-Temporal Data*, Wiley, New York.

Cressie, N.A.C. (1993). *Statistics for Spatial Data*, Wiley, New York.

Dalrymple, T. (1960). *Flood frequency analyses*, Water supply paper 1543-a, U.S. Geological Survey, Reston, VA.

Daly, C.; Gibson, W.P.; Taylor, G.H.; Johnson, G.L. & Pasteris, P. (2002). A knowledge-based approach to the statistical mapping of climate, *Clim. Res.*, **23**, 99–113.

Daly, C.; Neilson, R.P. & Phillips, D.L. (1994). A statistical topographic model for mapping climatological precipitation in mountainous terrain, *J. Appl. Meteo.*, **33**, 140–158.

Davis, R.A. & Resnick, S.I. (1989). Basic properties and prediction of max-ARMA processes, *Adv. Appl. Prob.*, **21**, 781–803.

Davison, A.C. & Gholamrezaee, M.M. (2012). Geostatistics of extremes, *Proc. R. Soc. Lond. A*, **468**, 581–608.

Davison, A.C.; Padoan, S. & Ribatet, M. (2012). Statistical modeling of spatial extremes (with Discussion), *Statist. Sci.*, to appear.

de Haan, L. (1984). A spectral representation for max-stable processes, *Ann. Prob.*, **12**, 1194–2004.

de Haan, L. & Ferreira, A. (2006). *Extreme Value Theory*, Springer, New York.

Diggle, P.J.; Tawn, J.A. & Moyeed, R.A. (1998). Model-based geostatistics (with Discussion), *Appl. Statist.*, **47**, 299–350.

Dombry, C.; Éyi-Minko, F. & Ribatet, M. (2011). *Conditional simulations of Brown–Resnick processes*, submitted.

Draisma, G.; Drees, H.; Ferreira, A. & de Haan, L. (2004). Bivariate tail estimation: dependence in asymptotic independence, *Bernoulli*, **10**, 251–280.

Erhardt, R.J. & Smith, R.L. (2011). Approximate Bayesian computing for spatial extremes, *Comp. Statist. Data Anal.*, to appear.

Fawcett, L. & Walshaw, D. (2006). A hierarchical model for extreme wind speeds, *J. R. Statist. Soc. C*, **15**, 631–646.

Fuentes, M.; Henry, J. & Reich, B. (2011). Nonparametric spatial models for extremes: Application to extreme temperature data, *Extremes*, to appear.

Gelfand, A. & Smith, F. (1990). Sampling-based approaches to calculating marginal densities, *J. Am. Statist. Assoc.*, **85**, 398–409.

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

Guttorp, P. & Xu, J. (2011). Climate change, trends in extremes, and model assessment for a long temperature time series from Sweden, *Environmetrics*, **22**, 456–463.

Heffernan, J. & Tawn, J.A. (2004). A conditional approach for multivariate extreme values, *J. R. Statist. Soc. B*, **66**, 497–546.

Hosking, J.R.M. & Wallis, J.R. (1997). *Regional Frequency Analysis: An approach based on L-Moments*, Cambridge University Press, Cambridge.

Huser, R. & Davison, A.C. (2012). *Space-time modeling of extreme events*, submitted.

Hüsler, J. & Reiss, R. (1989). Maxima of normal random vectors: between independence and complete dependence, *Statist. Prob. Lett.*, **7**, 283–286.

Jeon, S. (2012). *Max-stable Processes for Threshold Exceedances in Spatial Extremes*, Ph. D. thesis, University of North Carolina at Chapel Hill.

Joe, H. (1997). *Multivariate Models and Dependence Concepts* (1st ed.), Chapman, New York.

Kabluchko, Z.; Schlather, M. & de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions, *Ann. Prob.*, **37**, 2042–2065.

Leadbetter, M.R.; Lindgren, G. & Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.

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

Ledford, A.W. & Tawn, J.A. (1997). Modelling dependence within joint tail regions, *J. R. Statist. Soc. B*, **59**, 475–499.

Lindgren, F.; Rue, H. & Lindstrom, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, *J. R. Statist. Soc. B*, **73**, 423–498.

Lindsay, B. (1988). Composite likelihood methods, *Contemp. Math.*, **80**, 221–239.

Mannshardt-Shamseldin, E.C.; Smith, R.L.; Sain, S.R.; Mearns, L.O. & Cooley, D. (2010). Downscaling extremes: A comparison of extreme value distributions in point-source and gridded precipitation data, *Ann. Appl. Statist.*, **4**, 484–502.

Matheron, G. (1987). Suffit-il, pour une covariance, d'être de type positif? *Sciences Terre, Série Inform. Géo.*, **26**, 51–66.

Mendes, J.M.; de Zea Bermudez, P.C. & Pereira, J. (2010). Spatial extremes of wildfire sizes: Bayesian hierarchical models for extremes, *Environ. Ecol. Statist.*, **17**, 1–28.

Mikosch, T. (2006). Copulas: Tales and facts, *Extremes*, **9**, 3–20.

Naveau, P.; Guillou, A.; Cooley, D. & Diebolt, J. (2009). Modeling pairwise dependence of maxima in space, *Biometrika*, **96**, 1–17.

Nelsen, R.B. (2006). *An Introduction to Copulas* (2nd ed.), Springer, New York.

Padoan, S.A. (2011). Multivariate extreme models based on underlying skew-t and skew-normal distributions, *J. Mult. Anal.*, **102**, 977–991.

Padoan, S.A.; Ribatet, M. & Sisson, S.A. (2010). Likelihood-based inference for max-stable processes, *J. Am. Statist. Assoc.*, **105**, 263–277.

Pauli, F.; Racugno, W. & Ventura, L. (2011). Bayesian composite marginal likelihoods, *Statist. Sinica*, **21**, 149–164.

Peng, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes, *Statist. Prob. Lett.*, **43**, 399–409.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Ramos, A. & Ledford, A.W. (2009). A new class of models for bivariate joint tails, *J. R. Statist. Soc. B*, **71**, 219–241.

Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*, Springer, New York.

Ribatet, M. (2011). *SpatialExtremes: Modelling Spatial Extremes*, R package version 1.8-1.

Ribatet, M.; Cooley, D. & Davison, A.C. (2011). Bayesian inference for composite likelihood models and an application to spatial extremes, *Statist. Sinica* (doi:10.5705/ss.2009.248).

Rue, H. & Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, Chapman, Boca Raton.

Sampson, P. & Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure, *J. Am. Statist. Assoc.*, **47**, 108–119.

Sang, H. & Gelfand, A.E. (2009). Hierarchical modeling for extreme values observed over space and time, *Environ. Ecol. Statist.*, **16**, 407–426.

Sang, H. & Gelfand, A.E. (2010). Continuous spatial process models for spatial extreme values, *J. Agric. Bio. Environ. Statist.*, **15**, 49–65.

Schabenberger, O. & Gotway, C.A. (2005). *Statistical Methods for Spatial Data Analysis*, Chapman, Boca Raton.

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

Schlather, M. (2011). *RandomFields: Simulation and Analysis of Random Fields*, R package version 2.0.53.

Schlather, M. & Tawn, J.A. (2002). Inequalities for the extremal coefficients of multivariate extreme value distributions, *Extremes*, **5**, 87–102.

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

Schliep, E.; Cooley, D.; Sain, S. & Hoeting, J. (2010). A comparison study of extreme precipitation from six regional climate models via spatial hierarchical modeling, *Extremes*, **13**, 219–239.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors, *Statist. Sinica*, **4**, 639–650.

Sibuya, M. (1960). Bivariate extremal distribution, *Ann. Inst. Statist. Math.*, **11**, 195–210.

Sisson, S. & Fan, Y. (2010). *Likelihood-free Markov chain Monte Carlo.* In "Handbook of Markov Chain Monte Carlo" (S. Brooks, A. Gelman, G. Jones and X. Meng, Eds.), pp. 319–341, Chapman, Boca Raton.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges, *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229–231.

Smith, R.L. (1990). *Max-stable processes and spatial extremes*, unpublished manuscript.

Varin, C. (2008). On composite marginal likelihoods, *Adv. Statist. Anal.*, **92**, 1–28.

Wackernagel, H. (2003). *Multivariate Geostatistics, An Introduction with Applications* (3rd ed.), Springer, Berlin.

Wang, Y. (2010). *maxLinear: Conditional Sampling for Max-Linear Models*, R package version 1.0.

Wang, Y. & Stoev, S.A. (2011). Conditional sampling for spectrally discrete max-stable random fields, *Adv. Appl. Prob.*, **43**, 461–483.

Zhang, Z. & Smith, R.L. (2004). The behavior of multivariate maxima of moving maxima processes, *J. Appl. Prob.*, **41**, 1113–1123.