Prediction for Max-Stable Processes via an Approximated Conditional Density

Daniel Cooley[1]
Richard A. Davis[2]
Philippe Naveau[3]

[1]Department of Statistics
Colorado State University, Fort Collins, CO USA

[2]Department of Statistics
Columbia University, New York, NY USA

[3]Laboratiore des Sciences du Climat et de l'Environnement
IPSL-CNRS, Gif-sur-Yvette, France

cooleyd@stat.colostate.edu
phone: (970) 491-5269
fax: (970) 491-7895

September 18, 2007

Abstract

The dependence structure of a max-stable random vector is characterized by its spectral measure. Using only the spectral measure, we present a method for approximating the conditional density of an unobserved component of a max-stable random vector given the other components of the vector. The approximated conditional density can be used for prediction. Additionally, we present a new parametric model for the spectral measure of a multivariate max-stable distribution. This model is used to perform prediction for time series and interpolation for spatial applications.

Keywords: spectral measure, spatial interpolation, time series prediction, pairwise beta model

1 Introduction

Often, a primary goal of a time series analysis or a spatial analysis is that of prediction (or interpolation in the spatial context). The data in Figure 1 are the maximum flow rate (cubic feet per second) for each week taken from a time series of flow rates recorded at 15 minute intervals for a river in western Colorado. Figure 2 displays the maximum ozone readings at 29 sites in a region of North Carolina for the year 1999. These examples differ from typical time series and geostatistical data in that they each represent a maximum over a block of time (respectively one week and one year). We would like to perform prediction or interpolation for data such as these.

Maximum data such as that illustrated in Figures 1 and 2 are often analyzed to assess risk; that is, to answer some question associated with the distribution's upper tail. Given the river data, one might wish to approximate an upper quantile of the next week's maximum flow. Given the ozone pollution data, one might wish to assess the probability that the annual maximum ozone level at an unmonitored location exceeds some level of interest set by government or health officials. Prediction in both the time series and spatial context is often done by producing a point estimate, but a point estimate alone would be inadequate to answer questions such as the ones above. Instead of producing just a point estimate, our goal is to approximate the entire conditional distribution of the unobserved value given the observed data, and we focus on how well the upper tail is approximated.

As will be explained in Section 2, both the river flow data and the ozone data could be approximated by a max-stable model. The dependence structure of a max-stable random vector is characterized by its spectral measure. We propose a method to use the spectral measure to approximate the conditional density of an unobserved component given the other components of the random vector. Additionally, we introduce a new parametric model for a spectral measure which we use to approximate unknown spectral measures in both time series and spatial applications.

The paper is organized as follows. In Section 2 we provide an overview of multivariate max-stable distributions and max-stable processes and the role played by the spectral measure. In Section 3 we review the parametric max stable models which have been proposed and introduce the pairwise beta model. In Section 4 we present our method for approximating the conditional density. We follow with Sections 5 and 6 where we analyze time series and spatial data respectively. We conclude with a discussion section.

2 Multivariate Max–stability and Max-stable Processes

Our goal is to approximate the conditional density of an unobserved value in a process which is approximately max-stable. We presume that the process is observed at a finite number of locations $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ (either in time or space) and therefore model a max-stable random vector. More specifically, let $\boldsymbol{Y}_m$, $m = 1, 2, \ldots$ be iid copies of a random vector $\boldsymbol{Y} = (Y(\boldsymbol{x}_1), \ldots, Y(\boldsymbol{x}_p))^T$ and let the component maxima be denonted by $\boldsymbol{M}_n = \left(\bigvee_{m=1,\ldots,n} Y_m(\boldsymbol{x}_1), \ldots, \bigvee_{m=1,\ldots,n} Y_m(\boldsymbol{x}_p)\right)^T$, where $\bigvee$ denotes max. We assume $\boldsymbol{Y}$ is in the domain of attraction of a max-stable distribution G, that

is there exist sequences $\boldsymbol{a}_n$ and $\boldsymbol{b}_n$ such that

$$\mathbb{P}\left(\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{z}\right) = \mathbb{P}^n\left(\frac{\boldsymbol{Y} - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{z}\right) \xrightarrow{d} G(\boldsymbol{z}). \tag{1}$$

We observe the first $p-1$ components of $\boldsymbol{M}_n$ and wish to predict $M_n(x_p)$.

Characterizing the multivariate max-stable distributions is difficult, but the task is more manageable if the random vector has convenient common marginal distributions. We will henceforth assume that G has unit Fréchet marginals; that is $G_i(z_i) = \exp(-z_i^{-1})$ for $i = 1, \ldots, p$. There is no loss of generality in this assumption, as the marginals of G must be univariate max-stable and Resnick (1987) shows that the max-stable property and the domain of attraction are preserved under a monotone transformation between max-stable marginals.

Assuming $\boldsymbol{Y}$ is in the domain of attraction of a multivariate max-stable distribution with unit Fréchet marginals, then $\boldsymbol{b}_n = \boldsymbol{0}$. Taking logarithms and applying Taylor series approximations to (1), we get

$$n \log\left[1 - \mathbb{P}\left(\frac{\boldsymbol{Y}}{\boldsymbol{a}_n} > \boldsymbol{z}\right)\right] \approx -n\left[\mathbb{P}\left(\frac{\boldsymbol{Y}}{\boldsymbol{a}_n} \in [\boldsymbol{0}, \boldsymbol{z}]^c\right)\right] \approx \log G(\boldsymbol{z}). \tag{2}$$

The random vector $\boldsymbol{Y}$ is known to be regularly varying with index 1. Multivariate regular variation can be characterized by the relation

$$n\mathbb{P}\left(\frac{\boldsymbol{Y}}{\boldsymbol{a}_n} \in \cdot\right) \xrightarrow{v} \mu(\cdot), \tag{3}$$

where v denotes vague convergence, and μ is a positive Radon measure on $[0, \infty]^p \setminus \{\boldsymbol{0}\}$ (Resnick, 1987). Furthermore it is known that μ is a product measure in terms of "radial" and "angular" components. Let $\mathcal{B}$ be the unit ball, i.e. $\{\boldsymbol{z} \in \mathbb{R}_+^p : \|\boldsymbol{z}\| = 1\}$. If $A = \{\boldsymbol{z} \in \mathbb{R}_+^p : \|\boldsymbol{z}\| > r; \boldsymbol{z}/\|\boldsymbol{z}\| \in B\}$, where B is a set on the unit ball $\mathcal{B}$, then $\mu(A) = r^{-1} \times H(B)$ where H is a positive measure on $\mathcal{B}$ and is known as the spectral measure. Many parametric models for max-stable subfamilies use the norm $\|\boldsymbol{z}\| = |z_1| + \ldots + |z_p|$, and we henceforth follow this convention. Hence $\mathcal{B}$ becomes the $(p-1)$-dimensional unit simplex $S_p = \{\boldsymbol{w} \in [0,1]^p : w_1 + \ldots + w_p = 1\}$.

Putting together equations (2) and (3), the limiting distribution for component-wise maxima can be written in terms of μ:

$$G(\boldsymbol{z}) = \exp(-\mu[\boldsymbol{0}, \boldsymbol{z}]^c). \tag{4}$$

To convert the radial and angular components into cartesian coordinates, the exponent measure function

$$V(\boldsymbol{z}) = \mu\{[\boldsymbol{0}, \boldsymbol{z}]^c\} = \int_{S_p} \bigvee_{i=1,\ldots,p} \frac{w_i}{z_i} dH(\boldsymbol{w}) \tag{5}$$

is defined. The only conditions on $H(\boldsymbol{w})$ are that

$$\int_{S_p} w_i dH(\boldsymbol{w}) = 1 \tag{6}$$

for all $i = 1, \ldots, p$, which enforces the marginals to be unit Fréchet. Condition (6) has a nice geometric interpretation. Notice that from (6) we obtain $\int_{S_p} dH(\boldsymbol{w}) = p$ and therefore the center of mass ($\int_{S_p} w_i dH(w) / \int_{S_p} dH(w)$ for $i = 1, \ldots, p$) of H must be at the point $\boldsymbol{w} = (1/\mathrm{p}, \ldots 1/\mathrm{p})$.

If V is differentiable, then H can be decomposed into densities which exist on the interior and on each of the lower-dimensional subspaces of S_p. Following Coles and Tawn (1991), let $S_{l,c}$ denote the lower-dimensional subspaces of S_p where $l = 1, \ldots, p$ is the dimension of the subspace and $c = \{v_1, \ldots, v_l\}$ is an index variable of the subspaces of size l. The subspaces $S_{l,c} = \{\boldsymbol{w} \in S_p : w_i = 0, i \notin c, w_j > 0, j \in c\}$ produce a disjoint partition of S_p. $S_{l,c}$ is isomorphic to the interior of the $(l-1)$-dimensional unit simplex, and $S_p = \bigcup_{l,c} S_{l,c}$. Let $h_{l,c}$ be the density on $S_{l,c}$. Given a spectral measure $H(\boldsymbol{w})$, one must evaluate the integral in (5) to find $V(\boldsymbol{z})$, which is often impossible analytically. However, it is relatively easy to obtain $H(\boldsymbol{w})$ given $V(\boldsymbol{z})$. Coles and Tawn (1991) showed that the density $h_{l,c}$ can be obtained via the relation

$$\frac{\partial V}{\partial z_{v_1}, \ldots, \partial z_{v_l}} = -\frac{1}{\|z\|^{p+1}} h_{l,c}(\boldsymbol{z}/\|\boldsymbol{z}\|). \tag{7}$$

We refer to the $h_{l,c}$ as the spectral densities, and if all the mass of H is on the interior of S_p, then $h_{l,c} = 0$ for $l < p$ and we denote $h_{p,c}$ as h_p.

A max-stable process is simply the infinite-dimensional analogue to a max-stable random vector. As in the finite-dimensional case, max-stable processes are the limit process of maxima of iid random processes, and we say that $Y(\boldsymbol{z})$ is in the domain of attraction of $Z(\boldsymbol{z})$ (see de Haan (1984)) if

$$\frac{M_n(\boldsymbol{x}) - b_n(\boldsymbol{x})}{a_n(\boldsymbol{x})} \xrightarrow{d} Z(\boldsymbol{x}), \tag{8}$$

where $M_n(\boldsymbol{x}) = \bigvee_{m=1,\ldots,n} Y_m(\boldsymbol{x})$, $Y_m(\boldsymbol{x}), m = 1, 2, \ldots$ are independent copies of $Y(\boldsymbol{x})$, and $\xrightarrow{d}$ denotes convergence of all finite-dimensional distributions. As the data in Figures 1 and 2 represent (respectively weekly and annual) maxima, we assume they can be well-approximated by a max-stable process.

Several related dependence metrics have been suggested to summarize the amount of dependence in a multivariate max-stable distribution (Coles et al., 1999; Joe, 1993; Cooley et al., 2006); we use the extremal coefficient (Smith, 1990; Schlather and Tawn, 2003). The pairwise extremal coefficient is given by $\phi_{i,j} = V(\boldsymbol{c}_{i,j})$, where $\boldsymbol{c}_{i,j}$ is the p-dimensional vector with components $c_i = 1, c_j = 1$ and $c_k = \infty$ for $k \neq i$ or j. The dependence metric can be understood via the relation

$$\mathbb{P}(\max(Z_i, Z_j) \leq z) = \exp[-V(z\boldsymbol{c}_{i,j})] = \exp[-z^{-1}V(\boldsymbol{c}_{i,j})] = (\exp[-z^{-1}])^{\phi_{i,j}} = (\mathbb{P}(Z_1 \leq z))^{\phi_{i,j}};$$

that is, the pairwise extremal coefficient is the effective number of independent random variables in the bivariate couple. Hence $\phi_{i,j} \in [1, 2]$ and as $\phi_{i,j}$ increases, the amount of dependence decreases. Schlather and Tawn (2003; 2002) describe higher-order extremal coefficients, but we focus only on pairwise dependence.

3 Parametric Models

Since any spectral measure H which satisfies equation (6) will generate a valid multivariate max-stable distribution, it is impossible to characterize the entire family of distributions with a parametric model. Nevertheless, there have been some parametric subfamilies suggested which can capture important behavior and which have been used to successfully model many applications.

While several models have been suggested for the bivariate case such as the bilogistic (Joe et al., 1992) and polynomial (Nadarajah, 1999), we concentrate on models which can (theoretically) handle any finite dimension. The book by Kotz and Nadarajah (2000) gives a collection of multivariate extreme value distributions. The parametric models can be divided into two classes. The first class gives a parametric model for the exponent measure function, $V(\boldsymbol{z};\boldsymbol{\theta})$, from which one can obtain the spectral measure via (7). The second class models the spectral density $h_p(\boldsymbol{w};\boldsymbol{\theta})$ directly.

The most widely known multivariate extreme value distribution is the logistic (Gumbel, 1960) which has exponent measure function $V(\boldsymbol{z};\alpha) = (\sum_{i=1}^{p} z_i^{-1/\alpha})^{\alpha}$ for $0 < \alpha \leq 1$. Dependence between the components increases as α decreases. However, because it is characterized by a single parameter α, it is inadequate to model situations where dependence between components differs, such as most any spatial application or time-series application where more than one lag is modeled.

The asymmetric logistic (Tawn, 1988) and the negative logistic (Joe, 1990) are two parametric multivariate max-stable models which are defined by an exponent measure function and which can handle different levels of dependence between the components. One difficulty of these models is that they have a large number of parameters; in the three dimensional case the asymmetric logistic has sixteen parameters, twelve of which can be freely chosen. Another possible disadvantage of these models is that both achieve the center-of-mass condition by putting mass on the edges and vertices of the simplex S_p, resulting in a discontinuous spectral measure.

Alternatively to defining a parametric model for $V(\boldsymbol{z};\boldsymbol{\theta})$, parametric multivariate max-stable distributions can be defined in terms of the spectral density $h_p(\boldsymbol{w};\boldsymbol{\theta})$. Coles and Tawn (1991) showed that if h^* is a positive function on S_p with finite first moments $m_i = \int_{S_p} w_i h^*(\boldsymbol{w};\boldsymbol{\theta}) d(\boldsymbol{w})$, then

$$h_p(\boldsymbol{w};\boldsymbol{\theta}) = (\boldsymbol{m}\cdot\boldsymbol{w})^{-(p+1)} \prod_{j=1}^{p} m_j h^* \left(\frac{m_1 w_1}{\boldsymbol{m}\cdot\boldsymbol{w}}, \ldots, \frac{m_p w_p}{\boldsymbol{m}\cdot\boldsymbol{w}}; \boldsymbol{\theta} \right) \tag{9}$$

is a valid spectral density which has all its mass on the interior of S_p. In effect, if one thinks of h^* as a (perhaps unnormalized) density on S_p, then (9) alters the density so that it has center of mass at $(1/p, \ldots, 1/p)$ and total mass of p. Coles and Tawn (1991) used their technique to create a max-stable model from the Dirichlet density. The p-dimensional Dirichlet spectral density model has p parameters, $\alpha_i, i = 1, \ldots, p$ and can be asymmetric. Coles (1993) found the Dirichlet model preferable to the logistic and negative logistic models when fitting simulated spatial rainfall extremes. A disadvantage of this model is that because one must apply (9), the spectral density's parameters are largely uninterpretable.

The time-series and spatial applications we have in mind led us to search for a new spectral density parametrization. In both time-series and geostatistics where Gaussian processes play a dominant role, there is a tradition of modeling pairwise dependences. In time series, pairwise dependence is modeled as a function of the lag between the pair of interest. In geostatistics, pairwise dependence is modeled as a function of the distance between the two locations. Although a max-stable distribution cannot be uniquely specified by its pairwise dependences, we aim to construct a model which sensibly models the pairwise dependences and which is useful in a variety of applications.

In contrast to the approach of Coles and Tawn (1991), we directly construct a spectral density

$h_p(\boldsymbol{w})$. We call the model given by

$$\begin{aligned} h_p(\boldsymbol{w};\alpha,\boldsymbol{\beta}) &= K_p(\alpha)\sum_{1\le i<j\le p} h_{i,j}(w_i,w_j;\alpha,\beta_{i,j}), \qquad (10)\\ \text{where } h_{i,j}(w_i,w_j;\alpha,\beta_{i,j}) &= (w_i+w_j)^{(p-1)(\alpha-1)}(1-(w_i+w_j))^{\alpha-1}\times \\ &\quad \frac{\Gamma(2\beta_{i,j})}{(\Gamma(\beta_{i,j}))^2}\left(\frac{w_i}{w_i+w_j}\right)^{\beta_{i,j}-1}\left(\frac{w_j}{w_i+w_j}\right)^{\beta_{i,j}-1} \end{aligned}$$

the pairwise beta model. The function $h_{i,j}$ in (10) consists of two parts. The first part, determined by the parameter $\alpha > 0$, simply draws the mass of the density toward the center of the simplex, i.e., the point $(1/p,\ldots,1/p)$. So α has some control over the total amount of dependence in the model. The second part is basically a symmetric bivariate beta distribution determined by the parameter $\beta_{i,j} > 0$. The $\binom{p}{2}$ parameters, $\beta_{i,j}$, can be used to increase or decrease the pairwise dependence between the ith and jth components. The factor $K_p(\alpha) = \frac{2(p-3)!}{(p-1)\sqrt{p}}\frac{\Gamma(\alpha p)}{\Gamma(\alpha p-\alpha-p+3)\Gamma(\alpha+p-3)}$ is a normalizing constant. It is shown in Appendix A.1 that

$$\int_{S_p} w_i h_p(\boldsymbol{w};\alpha,\boldsymbol{\beta})d\boldsymbol{w} = 1 \qquad (11)$$

and therefore h_p is a valid spectral density. On a side note, $p^{-1}h_p(\boldsymbol{w})$ provides a new model for a density on the unit simplex.

The pairwise beta spectral measure model has several nice properties. The spectral measure allows for asymmetry, has all its mass on the interior of S_p, and is continuous. Because (9) is not used, its parameters are interpretable as described in the previous paragraph. Additionally, the pairwise beta model has $\binom{p}{2}$ β-parameters which relate to the amount of pairwise dependence, which is often of primary interest in spatial and time series applications.

4 Prediction

There has been little attention in the literature paid to the prediction problem for extremes. In a time series setting, Davis and Resnick (1989, 1993) define a distance d between the components of a bivariate max-stable random variable, and suggest a method of prediction which minimizes the distance between the observed component and the predictor. In fact, their distance metric can be related to the extremal coefficient by $d = 2(\phi - 1)$. Rather than relying on extreme value theory, Craigmile et al. (2006) comes at the problem of spatial interpolation for high values from the geostatistical point-of-view and adjusts the loss function of the kriging predictor. de Haan and Pereira (2006) state that the theory they develop for spatial max-stable processes can be used to solve problems in spatial prediction, but the authors do not tackle the problem directly within the article. It is important to note that the problem studied here differs from the one studied by Heffernan and Tawn (2004), among others. Whereas their goal was to assess the probability of an unlikely multivariate event (given no conditional information), our the goal is to perform prediction for an unobserved location given other nearby observations. Because a point estimate may be insufficient to answer questions associated with the upper tail, we propose estimating the conditional distribution of the unobserved component.

If $V(\boldsymbol{z})$ is known and is differentiable, then the conditional density can be calculated exactly. As seen in Section 3, the models where $V(\boldsymbol{z})$ is defined explicitly are limited and may not handle asymmetry well. Given models for which only $h_p(\boldsymbol{w})$ is explicitly defined, one should still be able to approximate the conditional density as all the dependence information is contained in the spectral measure. Provided that $\|\boldsymbol{z}\|$ is large, then $V(\boldsymbol{z}) = \mu\{(0, \boldsymbol{z}]^c\}$ is small and

$$G(\boldsymbol{z}) = \exp(-V(\boldsymbol{z})) \approx 1 - V(\boldsymbol{z}). \tag{12}$$

Then, for $\|\boldsymbol{z}\|$ large, (7) relates the joint density directly to the spectral density:

$$g(\boldsymbol{z}) \approx \frac{\partial^p}{\partial z_1, \ldots, \partial z_p}[1 - V(\boldsymbol{z})] = \frac{1}{||\boldsymbol{z}||^{(p+1)}} h\left(\frac{\boldsymbol{z}}{||\boldsymbol{z}||}\right), \tag{13}$$

and the conditional density can be approximated by

$$g_{Z_p|Z_1,\ldots,Z_{p-1}}(z_p|z_1, \ldots, z_{p-1}) \approx \frac{\frac{1}{||\boldsymbol{z}||^{(p+1)}} h\left(\frac{\boldsymbol{z}}{||\boldsymbol{z}||}\right)}{\int_0^\infty \frac{1}{||\boldsymbol{z}^*||^{(p+1)}} h\left(\frac{\boldsymbol{z}^*}{||\boldsymbol{z}^*||}\right) d\zeta}, \tag{14}$$

where $\boldsymbol{z}^* = (z_1, \ldots, z_{p-1}, \zeta)$. The denominator in (14) must be evaluated numerically, but it is easily done as it is a one-dimensional integral. When assessing risk, our interest lies in cases when $\|\boldsymbol{z}\|$ is large, and thus we proceed using the approximation in (14).

We test the approximation in (14) by simulating trivariate max-stable random variables with a symmetric logistic dependence structure. Because both $V(\boldsymbol{z})$ and $h(\boldsymbol{w})$ are known for this model, we can compare the actual conditional density of the third component given the first two to the approximated conditional density. Figure 3 illustrates how the approximation in (14) improves with $\|\boldsymbol{z}\|$. Figure 4 shows results for repeated simulations: 10000 trivariate symmetric logistic random vectors were simulated and the conditional density was approximated for the largest 3500 triples (as determined by the sum of the first two components). The rank histogram in Figure 4 records which decile of the approximated conditional density the third component fell. Additionally a qq plot is constructed; since each conditional density is unique, we calculate the quantile of the hidden component as approximated by our conditional density and create the plot from these values. As the histogram is nearly flat and the qq plot closely tracks the 45-degree line, it shows that when the observed values are large and when the spectral measure is known, the conditional density is well approximated with the method described in this section.

5 Time Series Application

We analyze river flow rate data for the Crystal River, located in the mountains of western Colorado. The US Geological Survey (USGS) records the data from a nationwide network of stations which monitor surface water levels and additionally provides a web interface to the national water information system (NWIS). The Crystal is a relatively unaltered river and the remarks associated with this particular monitoring station (#09081600) state that there are only a "few small diversions for irrigation upstream from station". Consequently, there is little if any protection from potential flood damage.

The river flow time series (Figure 1) is obviously influenced by seasonal effects, most notably, snowmelt in the early summer. In Appendix A.2 we outline the method we use to create the deseasonalized time series shown in Figure 5.

Several max-stable time series models have been proposed. One can construct a max-stable process by using moving maxima. Davis and Resnick (1989; 1993) investigated properties of max-ARMA time series models, which are infinite moving maxima processes and max-stable analogues to linear time series models. Deheuvels (1983) and Zhang and Smith (2004) proposed max-stable processes based on multiple moving maxima sequences. A perceived shortcoming of all these moving maxima processes is that the spectral measure has mass only at discrete locations. Alternatively, one can easily construct a first-order Markov process that is max-stable using any model for the exponent measure function or spectral measure. Coles and Tawn (1991) use the first-order Markov approach based on the bivariate logistic model and Smith et al. (1997) use the approach to model exceedances over a threshold. Extending the Markov approach beyond first order is not simple as additional constraints must be met to generate a stationary sequence.

We transform the marginal of the deseasonalized time series to unit Fréchet by fitting a generalized extreme value (GEV) distribution. The GEV is the parametric form for the family of univariate extreme value distributions and has distribution function $G(z) = \exp\left[-\left(1+\xi\frac{z-\mu}{\sigma}\right)^{-1/\xi}\right]$. Ignoring dependence, we use all the weekly data to estimate the GEV parameters, but fitting every fourth observation yields similar parameter estimates. We obtain parameter estimates $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (0.21, 0.91, 0.072)$; the small positive ξ value indicates a slightly heavy tail.

Based on acf and pacf plots of the deseasonalized Crystal River data, it appears that prediction should be based on the previous two observations. This rules out the first-order Markov approach, and as our goal is to model the conditional density, we choose not to use a moving maximum model because of its spectral measure properties. Rather than fitting a stationary model to the entire series, we simply model any given triple of the series with a trivariate pairwise beta model.

The three dimensional pairwise beta has parameters ($\alpha, \beta_{1,2}, \beta_{1,3}$, and $\beta_{2,3}$), and it can be shown that if $\beta_{1,2} = \beta_{2,3}$, then $\phi_{1,2} = \phi_{2,3}$ which we assume. We denote the lag 1 parameters ($\beta_{1,2}$ and $\beta_{2,3}$) by β_{-1} and similarly denote the lag 2 parameter $\beta_{1,3}$ by β_{-2}. To estimate the parameters, we use a method-of-moments type approach. We estimate the lag 1 and lag 2 extremal coefficients, ϕ_{-1} and ϕ_{-2}, using the madogram (Cooley et al., 2006; Naveau et al., 2006). We then use a plug-in method to find parameters $[\alpha, \beta_{-1}, \beta_{-2}]$ which yield a trivariate pairwise beta model that has these extremal coefficients. Each pairwise extremal coefficient is related to the parameters via the integral

$$\phi_{i,j} = \int_{S_p} \max(w_i, w_j) h_p(\boldsymbol{w}; \alpha, \boldsymbol{\beta}) d\boldsymbol{w}. \tag{15}$$

This integral cannot be evaluated analytically; however, it can be approximated via Monte Carlo (MC) methods. For the trivariate pairwise beta, we have three such integrals. Parameters $(\alpha, \boldsymbol{\beta})$ must be found which agree with the set of extremal coefficients; this optimization problem requires the set of integrals corresponding to the set of extremal coefficients to be evaluated repeatedly via MC. We have thus far limited our investigations to three dimensional cases (2 observed values + 1 value to predict) because of the computational challenges associated with fitting a spectral density model, but there is nothing in the prediction method that is limited by the dimension of the problem.

The estimated extremal coefficients for the unit-Fréchet transformed time series were $\phi_{-1} = 1.36$ and $\phi_{-2} = 1.49$. A pairwise beta model with $[\alpha, \beta_{-1}, \beta_{-2}] = [1, 16, 0.7]$ yielded the extremal coefficient estimates. The spectral density of this model is shown in the left panel of Figure 6, and the strong dependence between the two pairs of lag-1 components is clearly illustrated.

Turning our attention now to prediction, we break the time series up into non-overlapping triples, and use the first two observations to predict the third which we hide from view. We use the method described in Section 4 to approximate the conditional density of the third component. Because we are interested in large events, we apply the method to the 75 largest (as determined by the sum of the first two components) triples. We test the fit of our approximated conditional densities by again constructing a rank histogram and qq plot (Figure 7). As expected, there is some variation in the histogram and some deviation from the 45-degree line in the qq plot as the pairwise beta spectral density fit to the time series triples is not the true spectral density. Nevertheless, as the rank histogram is neither dramatically u-shaped nor spiked and the qq plot seems to track the 45-degree line well, Figure 7 shows that the conditional densities are reasonably approximated.

We compare our results to those obtained from a traditional time series model. We fit an AR(2) model using Yule-Walker estimation to the deseasonalized weekly maximum data. This model also uses only the previous 2 observations to predict the third. If we compare the point estimate predictor of the AR(2) model to a point estimate predictor (median) obtained from our approximated conditional distribution, the AR(2) point estimate slightly outperforms the one obtained from our model. However, our primary interest is in comparing the upper tail estimates of the two models. Using our method, we are easily able to estimate the 95% quantile of the approximate conditional density on the unit Fréchet scale which is then transformed back to the original marginal distribution. For the AR(2) model, we make three different assumptions for the noise: (1) that it is Gaussian distributed with the estimated variance of .88, (2) that it is t-distributed with 14 degrees of freedom to mimic the tail weight (also with variance 0.88), and (3) that it has the empirical distribution given by the residuals. For each of the three AR(2) models, we also find the 95% quantile for the predicted value. We found these values for the 75 largest triples used to make Figure 7, and compared them to the actual observations. Of the 75 observations, 6 (8%) exceeded the estimated 95% quantile of our model, whereas 9 (12%) exceeded the 95% quantile of each of the three AR(2) models. This seems to indicate that our model better predicts the probability of an occurrence in the upper tail of the river flow time series than does the traditional time series model. Additionally, we note that this data is only slightly heavy-tailed ($\hat{\xi} = 0.072$), and we conjecture that if the data had a heavier tail, the improvement may be even more significant.

We also compare our results to an alternative max-stable time series model. Davis and Resnick (1993; 1989) proposed max-ARMA time series models which are given by the recursion

$$X_t = \varphi_1 X_{t-1} \vee \ldots \vee \varphi_p X_{t-p} \vee Z_t \vee \vartheta_{t-1} Z_{t-1} \vee \ldots \vee \vartheta_q Z_{t-q},$$

where $\{Z_t\}$ is iid with common distribution function $\mathbb{P}(Z_t \leq z) = \exp(-\alpha x^{-1})$. We fit a max-AR(2) model to the unit-Fréchet transformed data. If the Crystal River data were in fact generated by a max-AR(2) process, the model parameters could be estimated exactly. Since the data were not generated in this manner, fitting a max-AR (or more generally, any moving maxima) model is not straightforward. We again use a method-of-moments type approach based on our earlier estimates for ϕ_{-1} and ϕ_{-2} to obtain $\hat{\varphi}_1 = 0.36$ and $\hat{\varphi}_2 = 0.51$. Using results from Davis and Resnick (1989) and a plug-in approach, our predictor is $\hat{X}_t = \hat{\varphi}_1 X_{t-1} \vee \hat{\varphi}_2 X_{t-2}$.

	c = 1.5	c = 2	c = 3
max-AR(2)	28 (37%)	10 (13%)	3 (4%)
condt'l median	14 (19%)	7 (9%)	2 (3%)

Table 1: The values for $1_{\{X_t/\hat{X}_t>c\}}$ for both the max-AR(2) predictor and the predictor obtained by taking the median of the approximated conditional density. It appears that the point predictor obtained from the approximated conditional density outperforms the max-AR predictor by this loss function.

The max-AR approach yields only a point predictor, which we compare to the median of the approximated conditional density. We compare the predictors' skill by examining the prediction errors (absolute and squared) as well as the indicator function $1_{\{X_t/\hat{X}_t>c\}}$ for $c = 1.5, 2, 3$ which we use as a measure of the predictors' ability to foresee large occurrences. We again look at the same "largest" 75 triples in the time series as before. The mean absolute error of the max-AR predictor and for the conditional density median are 1.11 and 1.03 respectively, while the mean squared error for the two models are 2.22 and 2.29 respectively, indicating that the two predictors are very comparable by these measures. However, Table 5 shows the risk scores for the two predictors. It appears that the point predictor obtained from the approximated conditional density outperforms the max-AR predictor by this loss function. Perhaps most importantly, our method yields a continuous approximated conditional density whereas a conditional density obtained from a max-AR model would be discontinuous because of the nature of its spectral density.

6 Spatial Applications

6.1 Simulated Max-stable Random Fields

We now turn our attention to the spatial prediction problem. We first perform prediction on simulated fields and use a procedure by Schlather (2002) to generate fields with unit Fréchet marginals on a 40×40 grid. The fields have a known bivariate dependence structure where the exponent measure function for $(Z(\boldsymbol{x}_1), Z(\boldsymbol{x}_2))$ is given by

$$V(z_1, z_2) = 1/2(z_1^{-1} + z_2^{-1})(1 + \sqrt{1 - 2(\exp(-h/\rho) + 1)z_1 z_2 (z_1 + z_2)^{-2}},$$

where h is the Euclidean distance between x_1 and x_2. The exponent measure function is not known for higher dimensions.

We apply our method to a simple spatial prediction problem. We assume that we observe the values at two locations $Z(x_1)$ and $Z(x_2)$ and try to predict the value at a third location $Z(x_3)$. In our examples $\boldsymbol{x}_1 = (20, 27)$, $\boldsymbol{x}_2 = (17, 23)$, and $\boldsymbol{x}_3 = (20, 20)$. Because of the known bivariate dependence, the pairwise extremal coefficients are $\phi_{1,2} = 1.34$, $\phi_{1,3} = 1.28$, and $\phi_{2,3} = 1.22$. We approximate the three-dimensional spectral measure given by the three locations in the field using the pairwise beta model and use the same method-of-moments approach as in the time series example to obtain $(\alpha; \beta_{1,2}, \beta_{1,3}, \beta_{2,3}) = (4.3; 0.87, 4.4, 74)$. The corresponding spectral measure in the right panel of Figure 6 shows the strongest dependence between components 2 and 3 which are the closest of the pairs of points.

Figure 8 shows the approximated conditional densities for the third component given the first two components for three simulated fields. Figure 9 is a histogram and qq plot for the largest 300 of 1000 simulated fields (largest once again determined by the sum of the observed components). The rank histogram shows that the the model captures most of the behavior of the conditional density as it is neither dramatically spiked nor u-shaped which indicates a poor representation of the conditional distribution. The height of the leftmost bar of the histogram indicates that the conditional density is poorly approximated for low values. This model underrepresents the probability of a very low observation, as the histogram shows that more of the points than would be expected occurred in the first decile of the conditional distribution. Consequently, the qq-plot is also shifted. Despite the fact that the pairwise beta model under-represents the probability of a small value, the upper tail of the conditional distribution appears to be well modeled.

As mentioned in the introduction, one of the advantages to approximating the entire conditional distribution is that it allows one to assess the probability of exceeding some specified value at the unmonitored location given the values of the other two. From each approximate conditional density, we estimate the probabilities that the hidden value is greater than 20 and 50 which roughly correspond to the 0.95 and 0.98 quantiles of the marginal distribution. This was done for the largest 300 of the 1000 simulated fields (as measured by the sum of the observed locations $Z(\boldsymbol{x}_1)$ and $Z(\boldsymbol{x}_2)$). We use an ROC (receiver operating characteristic) plot (Figure 10) to show the skill of the conditional density at predicting exceedances. ROC measures the ability of the forecast to discriminate between two alternative outcomes, in this case an exceedance or non-exceedance. The ROC plot is constructed by first ordering the simulations by their estimated probability of exceedance. A threshold is then moved through the ordered list and at each level the proportion of exceedances above the threshold is plotted verses the proportion of false positives above the threshold. The ROC plot has endpoints (0,0) and (1,1) and predictive ability increases as area under the curve increases. Figure 10 shows that the conditional density does a good job of predicting exceedances of both levels as the ROC curves for both have sharp corners.

6.2 Ozone Data

The last data set we examine is a ground-level ozone data set. The data consists of ozone measurements for the summer months of the years 1995-1999, and we focus on 29 stations found in North Carolina, many of which have incomplete data. This is a subset of data studied by Gilleland et al. (2006) who modeled how the parameters of an extreme value distribution varied over an area of the eastern United States which encompasses North Carolina. Gilleland et al. estimated a region-wide ξ parameter and used thin-plate splines to model how the other parameters of the Generalized Pareto distribution (GPD) varied. Since we are interested in annual maxima, we use Gilleland et al.'s GPD parameter estimates to obtain GEV parameter estimates (μ_i, σ_i) for each of the monitoring stations. Combining these with the regional estimate $\hat{\xi} = -0.31$ (bounded tail), we then transform the annual maximum measurements at each station to have a common unit Fréchet distribution.

We assume that the transformed measurements come from a field which is stationary, has unit Fréchet marginals, and is isotropic. We use the madogram (Cooley et al., 2006; Naveau et al., 2006) to estimate dependence between the annual maxima as a function of distance. An exponential model was fit to the madogram estimates by least-squares, and then this model was transformed

to yield an estimate for the extremal coefficient as a function of distance.

We apply our prediction method to three stations located near the city of Charlotte which are marked with an 'X' in Figure 2. These stations were chosen as they were the three closest stations which had sufficient records for all five years of the study period. Even between these three closely located stations, there is weak dependence in the annual maxima ($\hat{\boldsymbol{\phi}} = (1.95, 1.84, 1.67)$).

Since we have only 5 sets of annual maxima, we are not able to restrict our attention to the largest observations as we did with the earlier time series and spatial examples. Because of this, an slight adaptation must be made to our method to obtain a reasonable approximation for the conditional density. To obtain a density, we must take partial derivatives of our joint distribution $G(\boldsymbol{z}) = \exp(-V(\boldsymbol{z}))$. In contrast to (12), rather than approximating before we take the partials, we proceed by differentiating to obtain (in our three-dimensional case):

$$\begin{aligned} g(\boldsymbol{z}) &= \exp(-V(\boldsymbol{z}))\left[\frac{\partial V(\boldsymbol{z})}{\partial z_1}\frac{\partial V(\boldsymbol{z})}{\partial z_2}\frac{\partial V(\boldsymbol{z})}{\partial z_3} + \frac{\partial^2 V(\boldsymbol{z})}{\partial z_1 \partial z_2}\frac{\partial V(\boldsymbol{z})}{\partial z_3} + \right. \\ &\quad \left. \frac{\partial^2 V(\boldsymbol{z})}{\partial z_1 \partial z_3}\frac{\partial V(\boldsymbol{z})}{\partial z_2} + \frac{\partial^2 V(\boldsymbol{z})}{\partial z_2 \partial z_3}\frac{\partial V(\boldsymbol{z})}{\partial z_1} + \frac{\partial^3 V(\boldsymbol{z})}{\partial z_1 \partial z_2 \partial z_3}\right]. \end{aligned} \tag{16}$$

When $\boldsymbol{z}$ is large enough in all components, $\exp(-V(\boldsymbol{z})) \approx 1$ and the partials other than $\frac{\partial^3 V(\boldsymbol{z})}{\partial z_1 \partial z_2 \partial z_3}$ are near zero, resulting in the earlier approximations. However, when $\boldsymbol{z}$ is small, $\exp(-V(\boldsymbol{z}))$ is near zero.

For the ozone application, we need to bound the lower tail of the approximated conditional density to keep it from overwhelming the normalizing integral. In this three dimensional application

$$V(\boldsymbol{z}) = \int_{S_p} \max\left(\frac{w_1}{z_1}, \frac{w_2}{z_2}, \frac{w_3}{z_3}\right) dH(\boldsymbol{w}) d\boldsymbol{w}.$$

As $z_3 \to 0$, $z_3 \max(w_i/z_i)$ increases to w_3, so that by an application of the dominated convergence theorem, $z_3 V(\boldsymbol{z}) \to \int_{S_p} w_3 dH(w) = 1$, thus $V(\boldsymbol{z}) \sim z_3^{-1}$. In the ozone application we approximate the joint density with

$$g(\boldsymbol{z}) \approx \exp(-1/z_3)\frac{1}{||\boldsymbol{z}||^{(p+1)}} h\left(\frac{\boldsymbol{z}}{||\boldsymbol{z}||}\right)$$

and approximate the conditional density by normalizing as before.

We again use the pairwise beta model, with $\alpha = .15$ and $\beta = (0.01, 0.2, 0.8)$ which approximates the estimated extremal coefficient values. The conditional density (on the unit Fréchet scale) of the annual maximum measurement at the central station given the measurements at the northern and southern stations (see Figure 2) is approximated and then is transformed back to the original scale of the observations. Figure 11 shows the approximated conditional densities for the annual maxima for each of the five years. As would be expected with weak dependence, the conditional densities do not differ dramatically from the marginal density. However we see that the conditional density shifts upward for years where the observed stations' measurements are relatively high (e.g., 1996 and 1998) and shifts downward when the observed measurements are relatively low (e.g., 1995).

7 Discussion and Conclusions

We have proposed a method for approximating the conditional density of an unobserved component of a max-stable vector given that the other components are observed. The approximation method relies on a model for the spectral measure of the max-stable distribution rather than a model for the distribution itself. As few max-stable distributions are known explicitly and these are rather restrictive, we feel that specifying only a model on the spectral measure is a strength of the method. In both time series and spatial contexts, we have shown that a conditional density based on an approximated spectral measure largely captures the behavior of the unobserved component. By approximating an entire conditional density, we obtain more information than a simple point estimate, especially since we are most interested in quantities in the upper tail. We have shown that the method estimates well a high quantile of the conditional distribution and has good skill at predicting whether the unobserved component exceeds some level of interest.

A current challenge of modeling multivariate extremes is finding an adequate model for either the exponent measure function, or alternatively, the spectral measure. In this work, we have introduced a new model for a spectral measure, the pairwise beta. The pairwise beta has all its mass on the interior of the spectral measure, has parameters that have some intuitive understanding, and allows some control over the amount of dependence between a pair of components. In extremes, it is relatively well understood how to measure pairwise dependences, but it is not as well understood how to measure higher-order dependences, so having a model largely based on pairwise dependencies has advantages. In both the time series and spatial examples, the pairwise beta model proved useful and we were able to use pairwise dependences to obtain a reasonable approximation to the spectral measure. It is important to note however that the proposed prediction method is in no way reliant on the pairwise beta model and could be used with any spectral measure model.

Finally, one might argue that we are attempting to solve an artificial problem. It is true that both the data in Figures 1 and 2 are "artificial" in that we have more than just the weekly and annual maximum readings. There are examples where only the maximum data are recorded, but as our ability to store data increases exponentially, these instances are becoming more rare. To this potential criticism, we argue that our hope is that our block-maximum methodology can be extended to an exceedance-over-threshold methodology, as has been the case for extremes historically. Recall from Section 2 that the spectral measure relates to the intensity measure of the limiting point process. As we are modeling the spectral measure and using it to obtain our conditional density, it seems likely that this methodology can be extended.

Acknowledgements: This work has been supported in part by the PRIMES program at Colorado State University (NSF-IGERT grant DGE-0221595), by the Geophysical Statistics Project at the National Center for Atmospheric Research (NSF DMS-03055474) and by the National Science Foundation grants DMS-0308109 and DMS-0743459.

A Appendix

A.1 Center of mass condition for the pairwise beta model

Theorem 1 *The pairwise beta model, $h_p(\boldsymbol{w};\alpha,\boldsymbol{\beta})$ as defined (10), is a valid spectral measure.*

It suffices to show that $\int_{S_p} w_k h_p(\boldsymbol{w};\alpha,\boldsymbol{\beta})d\boldsymbol{w} = 1$ for all $k = 1,\ldots,p$. Without loss of generality, we only consider the case $k = 1$. From (10) we have

$$\begin{aligned}
&\int_{S_p} w_1 h_p(\boldsymbol{w};\alpha,\boldsymbol{\beta})d\boldsymbol{w} \qquad (17)\\
=\ &\int_{S_p} w_1\ K_p(\alpha) \sum_{1\le i<j\le p} h_{i,j}(w_i,w_j;\alpha,\beta_{i,j})d\boldsymbol{w}\\
=\ &K_p(\alpha) \sum_{1\le i<j\le p} \int_{S_p} w_1 h_{i,j}(w_i,w_j;\alpha,\beta_{i,j})d\boldsymbol{w}\\
=\ &K_p(\alpha)\left[\sum_{j=2}^{p}\int_{S_p} w_1 h_{1,j}(w_1,w_j;\alpha,\beta_{1,j})d\boldsymbol{w} + \sum_{2\le i<j\le p}\int_{S_p} w_1 h_{i,j}(w_i,w_j;\alpha,\beta_{i,j})d\boldsymbol{w}\right]\\
=\ &K_p(\alpha)\left[\sum_{j=2}^{p} I_{1,1,j} + \sum_{2\le i<j\le p} I_{1,i,j}\right], \qquad (18)
\end{aligned}$$

where $I_{1,1,j} = \int_{S_p} w_1 h_{1,j}(w_1,w_j;\alpha,\beta_{1,j})d\boldsymbol{w}$ and $I_{1,i,j} = \int_{S_p} w_1 h_{i,j}(w_i,w_j;\alpha,\beta_{i,j})d\boldsymbol{w}$. We first examine $I_{1,1,j}$ and consider the case $j = 2$. Suppressing the dependence of $h_{1,2}$ on α and $\beta_{1,2}$, we have

$$\begin{aligned}
I_{1,1,2} &= \sqrt{p}\int_{w_1=0}^{1}\int_{w_2=0}^{1-w_1}\int_{w_3=0}^{1-(w_1+w_2)}\cdots\int_{w_{p-1}=0}^{1-(w_1+w_2+\ldots w_{p-2})} w_1 h_{1,2}(w_1,w_2)dw_{p-1}\ldots dw_2 dw_1\\
&= \sqrt{p}\int_{w_1=0}^{1}\int_{w_2=0}^{1-w_1} w_1 h_{1,2}(w_1,w_2)\int_{w_3=0}^{1-(w_1+w_2)}\cdots\int_{w_{p-1}=0}^{1-(w_1+w_2+\ldots w_{p-2})} dw_{p-1}\ldots dw_2 dw_1\\
&= \sqrt{p}\int_{w_1=0}^{1}\int_{w_2=0}^{1-w_1} w_1(w_1+w_2)^{(p-1)(\alpha-1)}(1-(w_1+w_2))^{\alpha-1}\\
&\quad\times\frac{\Gamma(2\beta_{1,2})}{(\Gamma(\beta_{1,2}))^2}\left(\frac{w_1}{w_1+w_2}\right)^{\beta_{1,2}-1}\left(\frac{w_2}{w_1+w_2}\right)^{\beta_{1,2}-1}\\
&\quad\times\frac{(1-(w_1+w_2))^{p-3}}{(p-3)!}dw_2 dw_1.
\end{aligned}$$

After the change of variables $\theta = w_1/(w_1+w_2)$ and $r = w_1+w_2$, which has Jacobian $|J| = \left|\begin{smallmatrix}\partial w_1/\partial\theta & \partial w_1/\partial r\\ \partial w_2/\partial\theta & \partial w_2/\partial r\end{smallmatrix}\right| = r$, we obtain

$$\begin{aligned}
I_{1,1,2} &= \sqrt{p}\int_{\theta=0}^{1}\int_{r=0}^{1} r\theta r^{(p-1)(\alpha-1)}(1-r)^{\alpha-1}\frac{\Gamma(2\beta_{1,2})}{(\Gamma(\beta_{1,2}))^2}\theta^{\beta_{i,j}-1}(1-\theta)^{\beta_{i,j}-1}\frac{(1-r)^{p-3}}{(p-3)!}rdrd\theta\\
&= \frac{\sqrt{p}}{(p-3)!}\int_{\theta=0}^{1}\frac{\Gamma(2\beta_{1,2})}{(\Gamma(\beta_{1,2}))^2}\theta^{\beta_{i,j}}(1-\theta)^{\beta_{i,j}-1}d\theta\int_{r=0}^{1} r^{(p-1)(\alpha-1)+2}(1-r)^{\alpha+p-4}dr\\
&= \frac{\sqrt{p}}{(p-3)!}\frac{\Gamma(2\beta_{1,2})}{(\Gamma(\beta_{1,2}))^2}\frac{\Gamma(\beta_{1,2})\Gamma(\beta_{1,2}+1)}{(\Gamma(2\beta_{1,2}+1))^2}\frac{\Gamma(\alpha p-\alpha-p+4)\Gamma(\alpha+p-3)}{\Gamma(\alpha p+1)}\\
&= \frac{\sqrt{p}}{2(p-3)!}\frac{\Gamma(\alpha p-\alpha-p+4)\Gamma(\alpha+p-3)}{\Gamma(\alpha p+1)}. \qquad (19)
\end{aligned}$$

Similarly, for the case where $i = 2$ and $j = 3$,

$$\begin{aligned}
I_{1,2,3} &= \sqrt{p}\int_{w_2=0}^{1}\int_{w_3=0}^{1-w_2}\int_{w_4=0}^{1-(w_2+w_3)}\cdots \\
&\quad \int_{w_{p-1}=0}^{1-(w_2+w_3+\ldots w_{p-2})}\int_{w_1=0}^{1-(w_2+w_3+\ldots w_{p-1})} w_1 h_{2,3}(w_2,w_3)dw_1 dw_{p-1}\ldots dw_3 dw_2 \\
&= \sqrt{p}\int_{w_2=0}^{1}\int_{w_3=0}^{1-w_2} h_{2,3}(w_2,w_3)\int_{w_4=0}^{1-(w_2+w_3)}\cdots \\
&\quad \int_{w_{p-1}=0}^{1-(w_2+w_3+\ldots w_{p-2})}\int_{w_1=0}^{1-(w_2+w_3+\ldots w_{p-1})} w_1 dw_1 dw_{p-1}\ldots dw_3 dw_2 \\
&= \sqrt{p}\int_{w_2=0}^{1}\int_{w_3=0}^{1-w_2}(w_2+w_3)^{(p-1)(\alpha-1)}(1-(w_2+w_3))^{\alpha-1}\frac{\Gamma(2\beta_{2,3})}{(\Gamma(\beta_{2,3}))^2}\left(\frac{w_2}{w_2+w_3}\right)^{\beta_{2,3}-1}\left(\frac{w_3}{w_2+w_3}\right)^{\beta_{2,3}-1} \\
&\quad \times\frac{(1-(w_2+w_3))^{p-2}}{(p-2)!}dw_3 dw_2 \\
&= \frac{\sqrt{p}}{(p-2)!}\int_{\theta=0}^{1}\frac{\Gamma(2\beta_{2,3})}{(\Gamma(\beta_{2,3}))^2}\theta^{\beta_{2,3}-1}(1-\theta)^{\beta_{2,3}-1}d\theta\int_{r=0}^{1} r^{(p-1)(\alpha-1)+1}(1-r)^{\alpha+p-3}dr \\
&= \frac{\sqrt{p}}{(p-2)!}\frac{\Gamma(\alpha p-\alpha-p+3)\Gamma(\alpha+p-2)}{\Gamma(\alpha p+1)}. \qquad (20)
\end{aligned}$$

Putting (19) and (20) into (17) and substituting for $K_p(\alpha)$, we obtain

$$\begin{aligned}
&= \frac{2(p-3)!}{(p-1)\sqrt{p}}\frac{\Gamma(\alpha p)}{\Gamma(\alpha p-\alpha-p+3)\Gamma(\alpha+p-3)}\left[(p-1)\frac{\sqrt{p}}{2(p-3)!}\frac{\Gamma(\alpha p-\alpha-p+4)\Gamma(\alpha+p-3)}{\Gamma(\alpha p+1}\right. \\
&\quad \left.+\left(\binom{p}{2}-(p-1)\right)\frac{\sqrt{p}}{(p-2)!}\frac{\Gamma(\alpha p-\alpha-p+4)\Gamma(\alpha+p-3)}{\Gamma(\alpha p+1)}\right] \\
&= \frac{\alpha p-\alpha-p+3}{\alpha p}+\frac{\alpha+p-3}{\alpha p} \\
&= 1.
\end{aligned}$$

A.2 Deseasonalizing the river flow time series

We have two time series for this monitoring site. The first is a relatively short time series of instantaneous flows and the second is a much longer time series which records only daily mean flows. The instantaneous flow time series (USGS, 2007a) records the flow rate at 15 minute intervals. It covers the time period from Oct 1, 1990 to Oct 1, 2005, and 82 days (1.5%) do not have data recorded. The second data set only records the mean flow rate for each day (USGS, 2007b) but covers a much longer time interval (1955-2007). This is a complete time series with no missing data, although a few (less than 0.5%) of the readings are marked as being estimated. Both time series exhibit the same seasonal behavior. As we are interested in maximum flow, we primarily focus on the time series of instantaneous flows, but use summary information from the second data set to deseasonalize the measurements.

Let $z_{t,(mean)}$ be the time series of observed daily mean flows, and for the time being, let the index $t = (y, d)$ where $y \in \{1955, \ldots, 2007\}$ and $d \in \{1, \ldots, 365\}$. We characterize the seasonal component of a particular day d by taking the data from all years within the window $[d-6, d+6]$ to calculate a sample mean m_d and variance v_d. Since the time of the peak runoff can vary from year to year, using the data from all years in the 13-day window creates smoother mean and variance functions. We then use these daily mean and variance values to create an approximately stationary time series of weekly maxima. The daily maxima of the instantaneous flow rate series $z_{t,(inst)}, t = (y, d)$ is centered and scaled: $(z_{y,d,(inst)} - m_d)/\sqrt{v_d}$. Missing

values are filled in by regressing on the daily mean values ($r^2 = .97$). We then take the weekly maximum of these values to create the deseasonalized time series of weekly maxima shown in Figure 5.

References

Coles, S. (1993). Regional modelling of extreme storms via max-stable processes. *Journal of the Royal Statistical Society, Series B*, 1993:797–816.

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analysis. *Extremes*, 2:339–365.

Coles, S. and Tawn, J. (1991). Modeling multivariate extreme events. *Journal of the Royal Statistical Society, Series B*, 53:377–92.

Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. In Bertail, P., Doukhan, P., and Soulier, P., editors, *Dependence in Probability and Statistics*, Springer Lecture Notes in Statistics. Springer, New York.

Craigmile, P., Cressie, N., Santner, T., and Rao, Y. (2006). A loss function approach to identifying environmental exceedances. *Extremes*, 8:143–159.

Davis, R. and Resnick, S. (1989). Basic properties and prediction of max-ARMA processes. *Advances in Applied Probability*, 21:781–803.

Davis, R. and Resnick, S. (1993). Prediction of stationary max-stable processes. *Ann. of Applied Prob*, 3:497–525.

de Haan, L. (1984). A spectral representation for max-stable processes. *Annals of Probability*, 12:1194–2004.

de Haan, L. and Pereira, T. (2006). Spatial extremes: Models for the stationary case. *The Annals of Statistics*, 34:146–168.

Deheuvels, P. (1983). Point processes and multivariate extreme values. *Journal of Multivariate Analysis*, 13:257–272.

Gilleland, E., Nychka, D., and Schneider, U. (2006). Spatial models for the distribution of extremes. In Clark, J. and Gelfand, A., editors, *Hierarchical Modelling for the Environmental Sciences: Statistical Methods and Applications*, pages 170–183. Oxford University Press, New York.

Gumbel, E. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statis. Univ. Paris*, 9:171–173.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society, Series B*, 66:497–546.

Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics and Probability Letters*, 9:75–81.

Joe, H. (1993). Parametric family of multivariate distributions with given margins. *Journal of Multivariate Analysis*, 35:262–282.

Joe, H., Smith, R., and Weissmann, I. (1992). Bivariate threshold methods for extremes. *Journal of the Royal Statisitcal Society, Series B*, 54:171–183.

Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions*. Imperial College Press, London.

Nadarajah, S. (1999). A polynomial model for bivariate extreme value distributions. *Probability letters*, 42:15–25.

Naveau, P., Guillou, A., Cooley, D., and Diebolt, J. (2006). Modeling pairwise dependence of maxima in space. *Submitted*.

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

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44.

Schlather, M. and Tawn, J. (2002). Inequalities for the extremal coefficients of multivariate extreme value distributions. *Extremes*, 5(1):87–102.

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

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

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

Tawn, J. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75:397–415.

USGS (2007a). Instantaneous data archive. http://ida.water.usgs.gov/ida/.

USGS (2007b). National water information system. http://waterdata.usgs.gov/nwis/uv.

Zhang, Z. and Smith, R. (2004). The behavior of multivariate maxima of moving maxima processes. *Journal of Applied Probability*, 41:1113–1123.

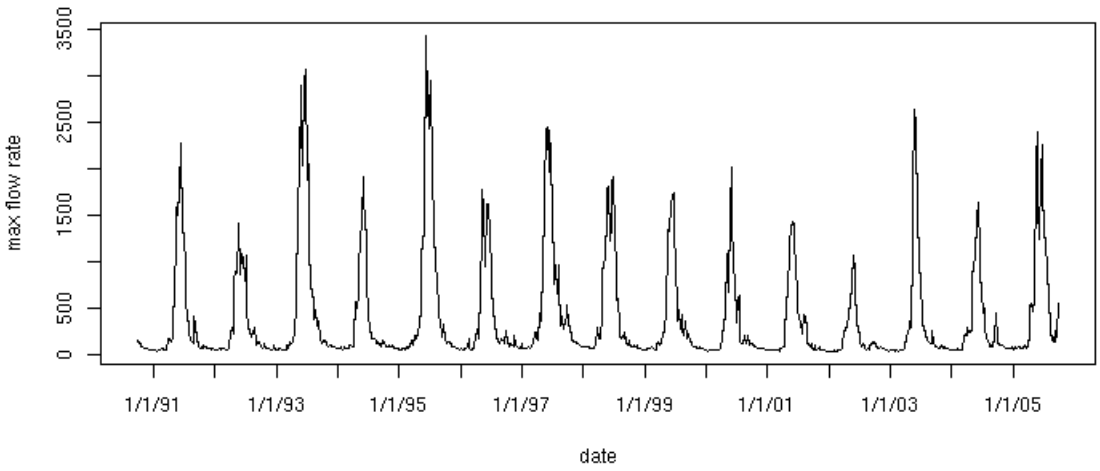

Figure 1: Weekly maximum flow readings (ft^3/sec) for the Crystal River from 10/1/1990 to 10/1/2005.

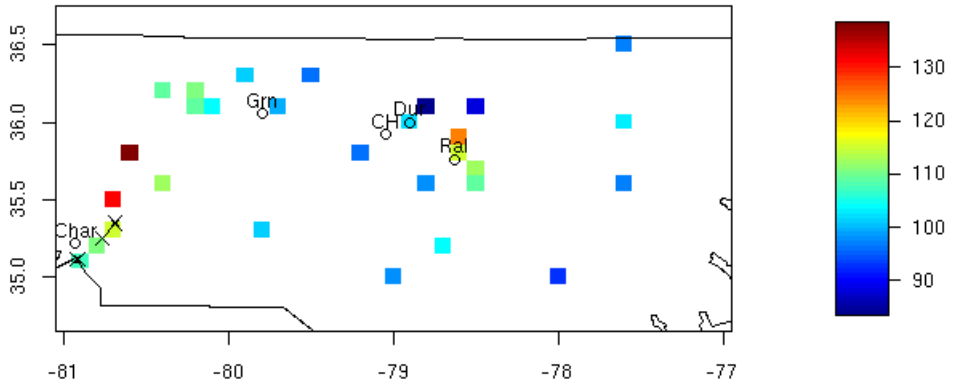

Figure 2: Maximum eight-hour ozone readings (ppb) for a region of North Carolina for the year 1999. X's indicate the three stations used for the prediction study in Section 6.

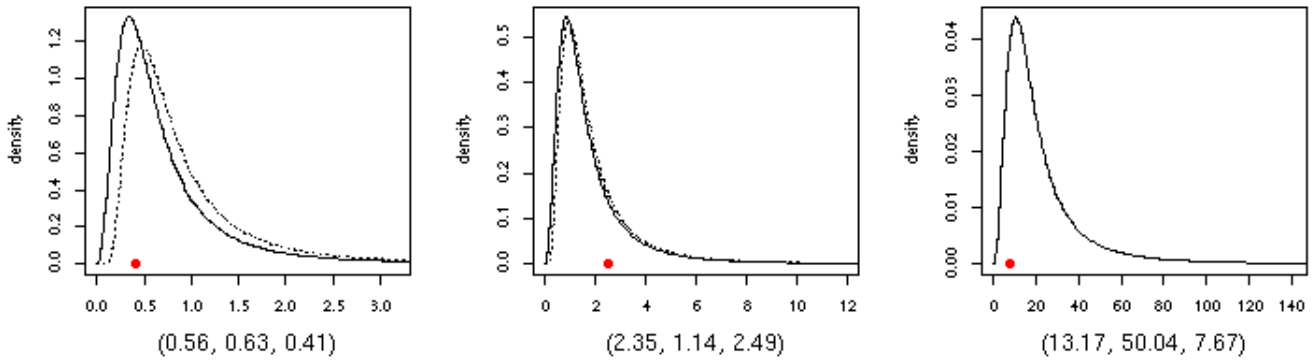

Figure 3: Three realizations from a trivariate symmetric logistic distribution. The solid curve is the actual conditional density of the third component given the first two, and the dotted curve is the approximated conditional density from (14). The left figure shows the realization $\boldsymbol{z} = [0.56, 0.63, 0.41]; (\|\boldsymbol{z}\| = 1.60)$, the center shows $\boldsymbol{z} = [2.35, 1.14, 2.49]; (\|\boldsymbol{z}\| = 5.98)$, and the right shows $\boldsymbol{z} = [13.17, 50.04, 7.67]; (\|\boldsymbol{z}\| = 70.88)$. One can clearly see that the approximation improves with $\|\boldsymbol{z}\|$.

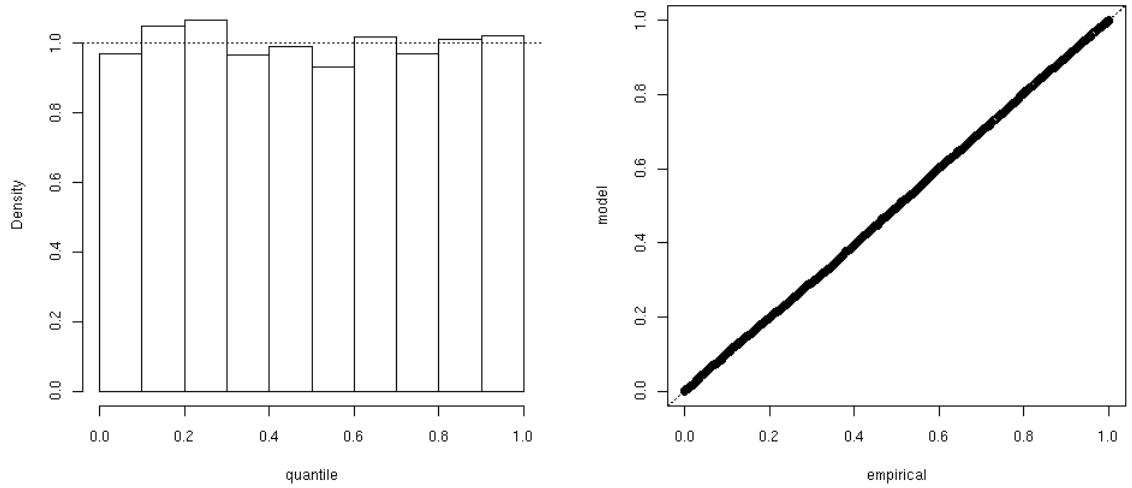

Figure 4: Results of the largest (as measured by the sum of the observed components $z_1 + z_2$) 3500 of 10000 simulated trivariate logistic random variables ($\alpha = .3$). Histogram shows which decile of the approximated conditional density the hidden observation occurred. QQ plot is for the uniform-transformed conditional density values. As the histogram is nearly flat and the qq plot follows the 45-degree line, it shows the approximated condtional density is very well represented.

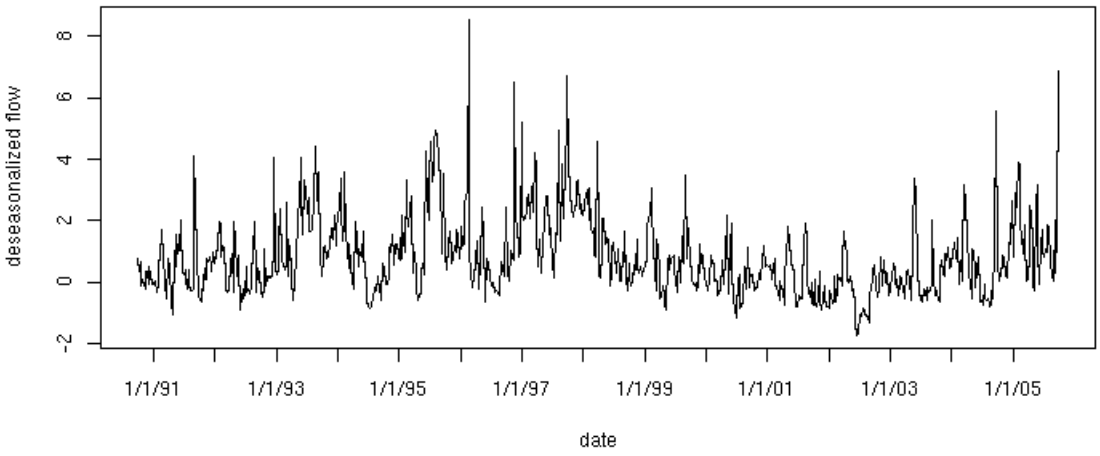

Figure 5: Deseasonalized weekly maximum flows for the Crystal River.

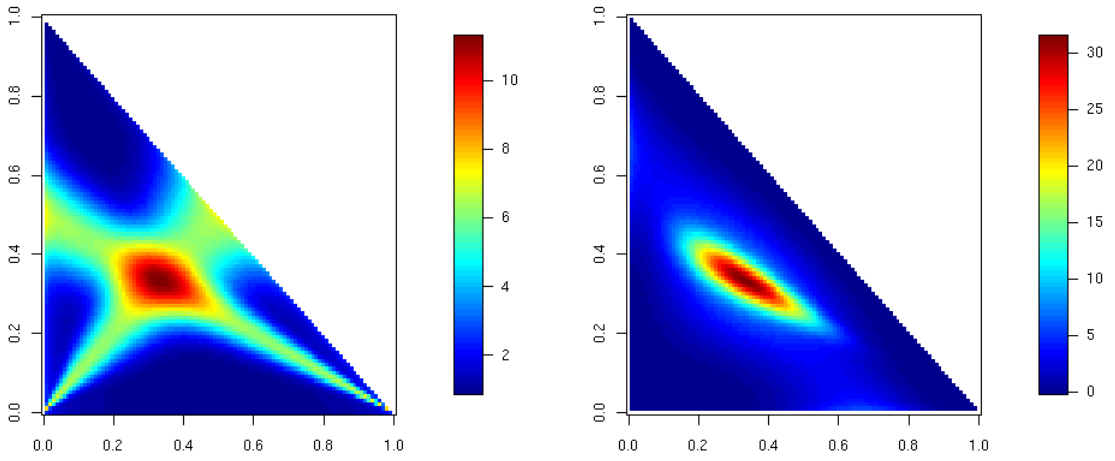

Figure 6: Pairwise beta spectral measure fit to the transformed river flow time series (left) and to the three-point spatial application (right). In both, the lower right, upper, and lower left corners correspond to the first, second, and third components of the random vector. The left panel shows the strong dependence between the lag-1 components, while the right shows the strongest dependence between the 2nd and 3rd components which are the closest pair of points.

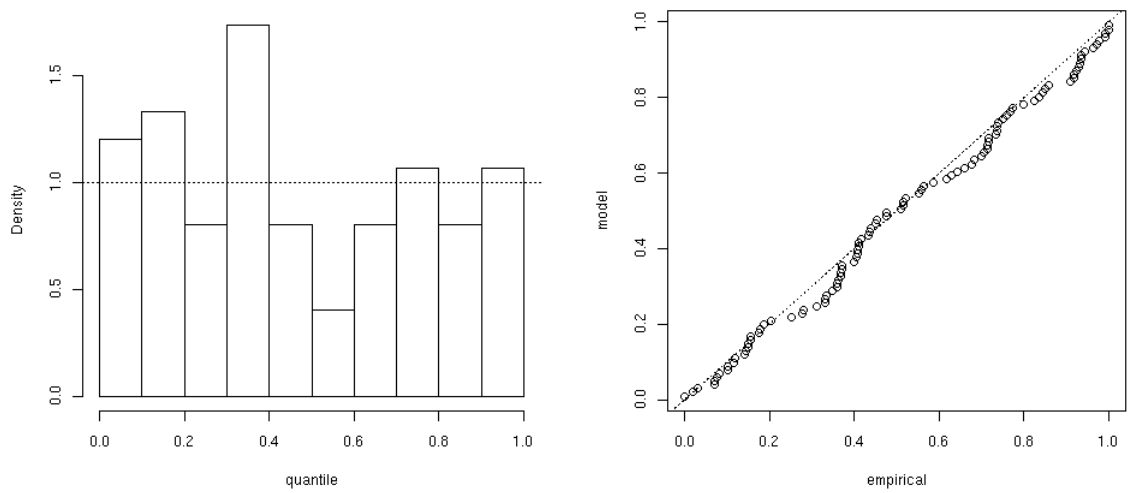

Figure 7: Histogram and qq plot for the river flow data. Shows the results of the largest (as measured by the sum of the observed components $Z(x_i) + Z(x_{i+1})$) 75 triples in the time series. Since there are only 75 observations, there is some variation in the histogram, but the QQ plot shows that the conditional density of the third component given the first two is being reasonably approximated.

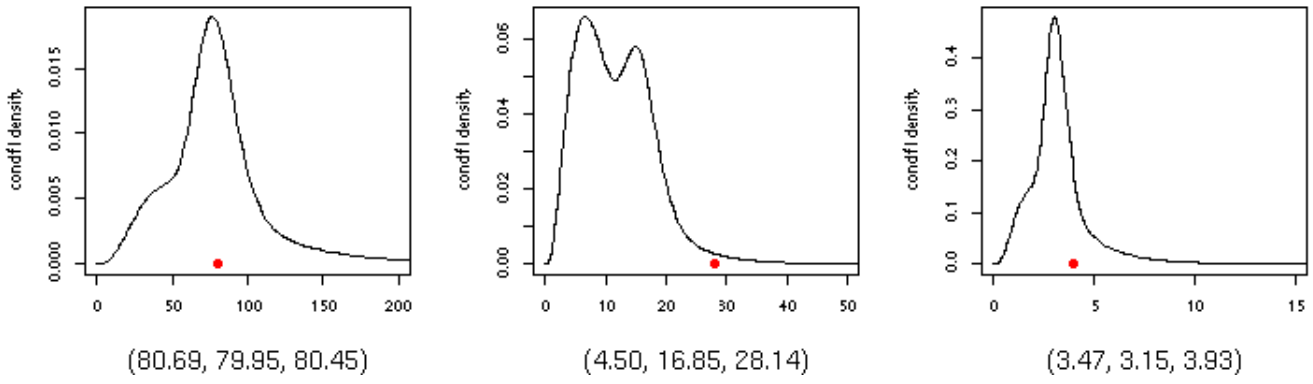

Figure 8: The black curve shows the approximated conditional density of the value at the third location given the values at the other two locations of three simulated fields. The dot is the actual value at the third location. The vector below each graph gives respectively the values at the two observed locations and the hidden value at the third location.

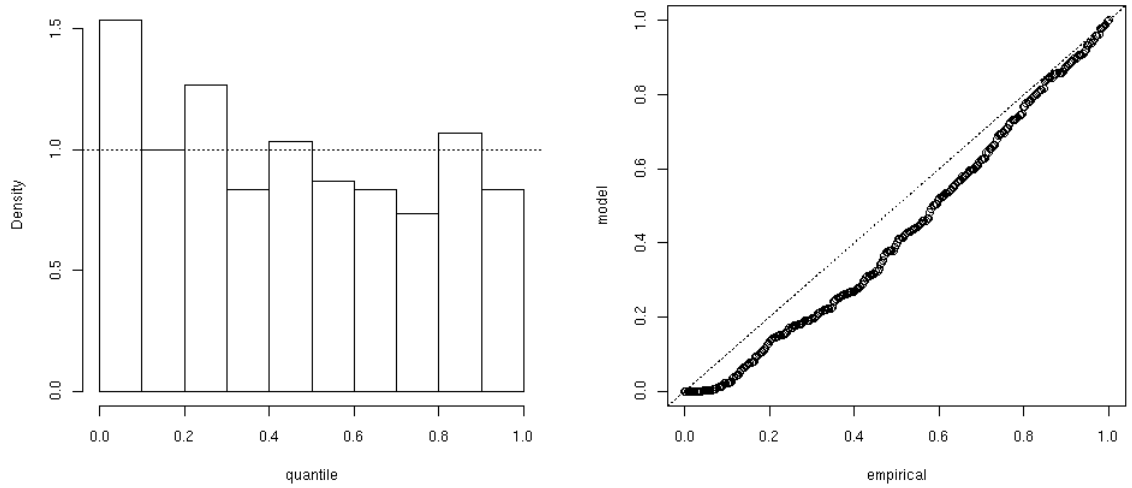

Figure 9: Histogram and qq plot for the largest 300 of 1000 fields. Histogram shows that the lowest decile is being under-approximated, and consequently the qq plot is shifted by these excess low values. Aside from the lowest observations, the conditional density is well-approximated.

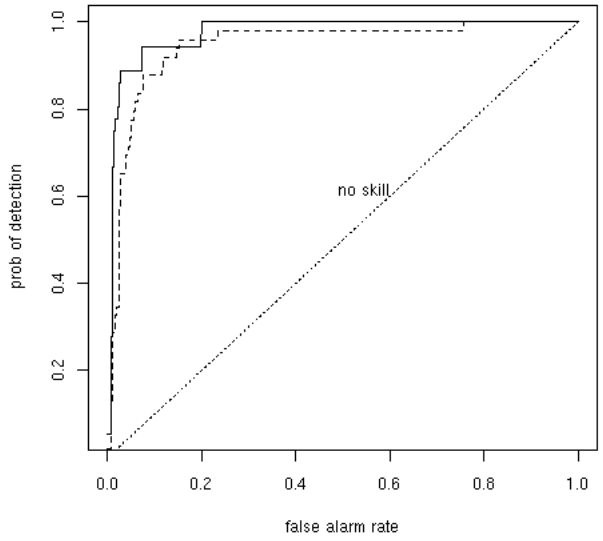

Figure 10: ROC plot of an exceedance over a level of 50 (solid) and 20 (dashed) for the simulated spatial data.

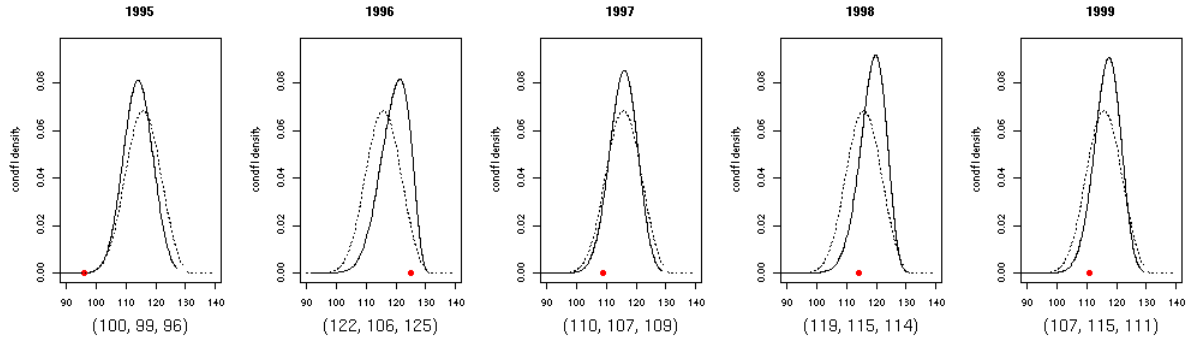

Figure 11: Shows the estimated conditional density (solid line) along with the marginal density (dotted line) for the annual maximum value given the annual maximum value at two nearby locations. Below each plot is a vector in which the first two components are the observed measurements used to approximate the conditional density and the third is the hidden component.