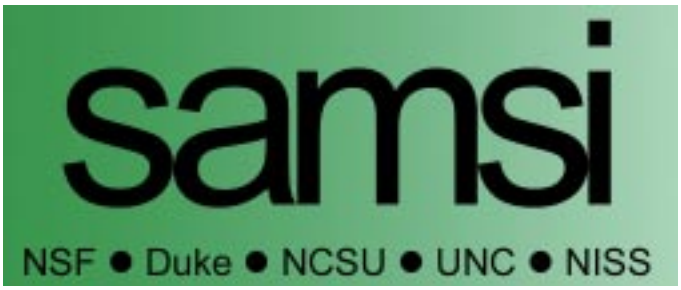

A Statistical Characterization of a Simulated Canadian Annual Maximum Rainfall Field.

Audrey Fu, Nhu D. Le and James V. Zidek

Technical Report #2003-17
June 17, 2003

This material was based upon work supported by the National Science Foundation under Agreement No. DMS-0112069. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Statistical and Applied Mathematical Sciences Institute
PO Box 14006
Research Triangle Park, NC 27709-4006
www.samsi.info

A Statistical Characterization of a Simulated Canadian Annual Maximum Rainfall Field.

Audrey Fu[1], Nhu D Le[1,2], James V Zidek[1]

[1] Department of Statistics, University of British Columbia
[2] Cancer Control Research, BC Cancer Agency

June 17, 2003

Abstract

The paper explores the use of a joint log Gaussian measurement distribution to represent a field of extreme values. A hierarchical Bayes prior distribution with estimated hyperparameters models uncertainty in that distribution's unknown parameters. The resulting multivariate - t distribution is fitted to the random maximum annual rainfall field simulated from a Canadian Climate model with 319 grid cells. To assess its performance in that application, we find its predictive credibility ellipsoids and confirm that their coverage fractions are close to their nominal credibility levels, using cross validation. Return values are estimated for cell marginal distributions and their overall accuracy is assessed by an index of discrepancy.

1 Introduction

This paper presents a joint distribution for random extreme values over a discrete geographical field. We apply the method to simulated annual maximum precipitation generated by the first generation Canadian Global Coupled Model (CGCM1) (Kharin and Zwiers, 2000) for the 319 grid cells into which it partitions Canada.

That application stems from the pervasive interest in extreme weather and its associated risks. For designers, those risks can translate into return values for the annual maximum precipitation. More precisely, if X denotes the annual maximum precipitation in a specific geographical region (or cell in the simulation), X's T year return value, X_T, would be the value X exceeds

once every T years on average. That is,

$$P(X > X_T) \;=\; \frac{1}{T}. \tag{1}$$

In this paper, "extreme value" or "extreme" for short, refers to the maximum (or minimum) of a possibly autocorrelated series of random responses (although we recognize that terminology sometimes refers to values in the tails of a distribution above some threshold). It addresses questions about the return values of those extremes, here the annual maximum hourly precipitation at each of a number of locations (cells). Because of spatial correlation in this context, we cannot calculate the probability of simultaneously exceeding such marginal return values as the product of their marginal exceedance probabilities, thereby limiting their usefulness. That observation, leads to the study of extremes within the domain of spatial statistics and of the joint distribution of a field of random responses, X, representing geographical subdomains (grid cells). In general, we need the joint extreme value distribution for a variety of purposes, for example, to characterize uncertainty about aggregate events such as the number of locations at which exceedances occur on any given future year.

In Section 2, we describe current approaches to modelling the joint distribution of extremes. Section 3 presents our new theory based on the multivariate Gaussian distribution. Section 4 applies that theory and, in particular, assesses its multivariate as distinct from marginal validity. Finally, Section 5 presents our conclusions.

2 Survey of Current Approaches.

Multivariate extreme value distributions, the subject of much current research, are described in this section. Of specific concern are environmental time series such as autocorrelated hourly precipitation levels at given locations that possess temporal patterns such as seasonality. Generally, the autocorrelations are not important in modelling the distribution of extreme values (Coles and Tawn 1996). Moreover, trends can be handled by removing them from the original record, calculating such things as the return values for the joint residual distribution, and then reinstalling that trend. Sun et al (2000), for example, use that idea although in a slightly different context.

Single-Site Models. The distribution of the annual maximal precipitation at a single site can be modelled using amongst other things, the generalized extreme value (GEV) distribution (Kharin and Zwiers, 2000),

the generalized Pareto distribution (GPD; Smith, 2001), and the Poisson process model for the precipitation (Coles and Tawn, 1996, Ledford and Tawn, 1997). Smith (2001) provides a good description of these distribution models.

If $X_1, X_2, \cdots, X_n$ are iid random variables (for example, hourly precipitation levels) with distribution F, the distribution of the maximum $M_n = max\{X_1, X_2, \cdots, X_n\}$ can be found explicitly in terms of n-th powers of F from which the limit distribution can be found. More precisely,

$$P(\frac{M_n - b_n}{a_n} \leq x) \quad \rightarrow \quad H(x), \quad \text{as } n \rightarrow \infty$$

where a_n and b_n are normalizing constants that keep H(x) from being degenerate. A celebrated result of extreme value theory tells us that H must be one of three types:

1. Gumbel type:

$$H(x) = \exp\{-\exp(-x)\}, \quad -\infty < x < \infty,$$

2. Fréchet type:

$$H(x) = \begin{cases} 0 & \text{if } x < 0 \\ \exp(-x^{-\alpha}) & \text{if } 0 < x < \infty, \end{cases}$$

3. Weibull type:

$$H(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & \text{if } -\infty < x < 0, \\ 1 & \text{if } x > 0. \end{cases}$$

In the above, $\alpha > 0$ is a constant.

The GPD models exceedances over some threshold, treating the maxima or minima as a special case. Point process theory combines the GEV and the GPD distributions, taking time into account. That theory takes the exceedances or annual maxima to be a heterogeneous stochastic process. It models the number of exceedances or values of annual maxima during each time interval. Assuming the number of exceedances in each time interval are independently and identically distributed (*i.i.d.*), it can be shown that it is a GPD, the annual maxima, a GEV distribution.

Multi-Site Models. As noted above, many models for multivariate extreme value distributions obtain from their univariate counterparts.

In their initial analysis of precipitation data, Kharin and Zwiers (2000) use the marginal GEV distribution and incorporate spatial correlation in parameter estimation. First, they estimate the parameters of the marginal GEV distributions separately. Regarding these parameter estimates as indexed by their associated spatial site coordinates makes them a spatial field in their own right. The authors then predict that field at a particular site, by averaging all the parameter estimates in a neighborhood of it. [In another manuscript currently in preparation, the first author and her co - investigators, extend the method of Kharin and Zwiers (2000) using a variety of smoothing methods including the weighted MLE method and the thin-plate splines method.]

Reiss and Thomas (1997) present a number of other univariate to multivariate approaches. Like univariate extreme value distributions, multivariate extreme value distributions are introduced as limiting distributions of componentwise extreme variables. Three models will be summarized here: the Marshall-Olkin model, the Gumbel-McFadden Model and the Hüsler-Reiss model. These models share a notable feature, a parameter, λ, that indexes dependence. Although that parameter has different ranges for the different models, dependence increases with λ. (Bivariate versions of the three models are given with higher dimensional extensions indicated.)

1. The Marshall-Olkin model: The standard version, M_λ, is a bivariate extreme value distribution with Weibull marginals. (See the previous section for a definition of the Weibull with parameter, $\alpha = 1$.) The dependence parameter, λ, ranges from 0 to 1. It can be expressed as follows:

   - For the case of total dependence ($\lambda = 1$),

     $$M_\lambda(x, y) = \exp\Big((1-\lambda)(x+y) + \lambda \min\{x, y\}\Big), \quad x, y < 0.$$

   - For $0 < \lambda < 1$,

     $$M_\lambda(x, y) = P\left\{ \max\left\{\frac{Z_1}{1-\lambda}, \frac{Z_0}{\lambda}\right\} \le x, \max\left\{\frac{Z_2}{1-\lambda}, \frac{Z_0}{\lambda}\right\} \le y \right\},$$

     where X and Y are extreme variables, while Z_0, Z_1 and Z_2 are iid random variables with common Weibull distribution having $\alpha = 1$.

   Falk et al (1994) extend the second case to p dimensions.

2. The Gumbel-McFadden model: This extreme value distribution has Weibull marginals with $\alpha = 1$. However, the dependence parameter λ ranges from 1 (independence) to ∞ (total dependence). The distribution has the form

$$L_\lambda(x,y) = \exp\bigg(-((-x)^\lambda + (-y)^\lambda)^{1/\lambda}\bigg), \quad x, y < 0.$$

The p-dimensional extension can be found in McFadden (1978).

3. The Hüsler-Reiss model: This model represents the limiting distribution of the maxima of standard normal random vectors, λ ranging between 0 (independence) and infinity (total dependence). The bivariate form is

$$H_\lambda(x,y) = \exp\bigg(-\Phi\bigg(\frac{1}{\lambda} + \frac{\lambda(x-y)}{2}\bigg)e^{-y} - \Phi\bigg(\frac{1}{\lambda} + \frac{\lambda(y-x)}{2}\bigg)e^{-x}\bigg),$$

where Φ is the univariate standard normal distribution. The p - dimensional form is given by Joe (1994):

$$H_\Lambda(\mathbf{x}) = \exp\bigg(-\sum_{k\le p}\int_{x_k}^{\infty} \Phi_{\sum(k)}\bigg((\lambda_{i,k}^{-1} + \lambda_{i,k}(x_i - z)/2)_{i=1}^{k-1}\bigg)e^{-z}dz\bigg),$$

where $\Lambda = (\lambda_{i,j})$ is a symmetric $d \times d$ matrix with $\lambda_{i,j} > 0$ if $i \neq j$ and $\lambda_{i,i} = 0$, while $\Phi_{\sum(k)}$ is a $(k-1)$-dimensional normal distribution. The mean of $\Phi_{\sum(k)}$ is zero and $\sum(k) = 3D\bigg(\sigma_{i,j}(k)\bigg)$ is the correlation matrix given by

$$\sigma_{i,j}(k) = \begin{cases} \lambda_{i,k}\lambda_{j,k}(\lambda_{i,k}^{-2} + \lambda_{j,k}^{-2} - \lambda_{i,j}^{-2})/2, & \text{if } 1 \le i < j \le k-1; \\ 1, & \text{if } i \neq j. \end{cases}$$

This section has reviewed what could be considered "distribution modelling" rather than "process modelling" approaches. Such approaches have the advantage of being well - founded on underlying assumptions that embrace the special nature of extremes and thereby provide some assurance that they will have good properties particularly when dealing with extremes so far out that little or no data will be available to validate them. Such assurance would seem vital when dealing for example with return periods of 1000 years or more.

However, these models for multivariate extremes also have shortcomings. For one thing, they tend to be mathematically intractable making explicit

calculation difficult. In particular, developing inferential methods can be challenging. That makes an approach like that of Kharin and Zwiers (2000) appealing. However, approaches like the latter can also be criticized for the ad hoc way in which spatial dependence is superimposed through the smoothing of the parameter estimates from univariate site - marginals. For example, we see no compelling basis for selecting one smoothing method over another. Moreover, the resulting joint distributions need not yield tractable expressions for the conditional and marginal distributions needed for simulating the distributions of complex metrics calculated from these extremes. That leads us to the approach in the next section.

3 A Process Modelling Approach.

This section proposes a hierarchical Bayesian approach based on a conditional multivariate log Gaussian response distribution. A richness of existing theory makes inference for our model relatively easy. Moreover, we can readily estimate return values.

However, our method does depend on successfully transforming the responses to make their assumed joint distribution valid, something that may not be feasible for extremes of high order. In any case, assessing the validity of that approximation is necessary and we develop a operational method for making that assessment in this section. Our theory is applied in Section 4 and shown to work reasonably well.

3.1 Bayesian Framework

We follow Le and Zidek (1992) in our construction of a hierarchical Bayesian model. Let $\mathbf{Y}_j : p \times 1$ denote the vector of annual precipitation maxima over p cells in year $j = 1, \ldots, n$. Suppose that conditional on $\boldsymbol{\mu}$ and Σ, the spatial mean vector and covariance matrix, respectively, the log-transformed maxima $\mathbf{X}_j = \log \mathbf{Y}_j$ are independent and identically distributed with a multinormal distribution

$$\mathbf{X}_j \sim MVN_p(\boldsymbol{\mu}, \Sigma).$$

Note that the discussion in Section 2 justifies our ignoring temporal patterns and correlation here.

Since all precipitation maxima are positive, they can be log-transformed to make their joint distribution more nearly multinormal. Assessing the quality of that approximation is another matter and we develop below a

procedure based on calibrating the credibility ellipsoids generated by that joint distribution.

To construct our hierarchical distribution, we adopt the conjugate distribution that has been shown to work well in modelling environmental processes (see, for example, Sun et al 2000). More precisely,

$$\boldsymbol{\mu}|\Sigma \sim MVN\Big(\boldsymbol{\nu}, F^{-1}\Sigma\Big);$$

$$\Sigma \sim W^{-1}\Big(\Sigma|\Psi, m\Big).$$

where $\boldsymbol{\nu}$ is the hypermean vector while Ψ and m are respectively, the scale matrix and degrees of freedom in the inverted Wishart distribution. We use F^{-1}, to rescale Σ from the level of response uncertainty to that of the mean while retaining the simplicity of conjugacy in the prior model.

Le and Zidek (1992) show that the resulting posterior joint distribution (here of the extremes) to be a multivariate t distribution:

$$\mathbf{X}_{p\times 1}|D \quad \sim \quad t\Big(\tilde{\mathbf{x}}, \tilde{\Sigma}, l\Big), \tag{2}$$

where

$$\begin{aligned}
\tilde{\mathbf{x}} &= \boldsymbol{\nu} + (\bar{\mathbf{x}} - \boldsymbol{\nu})\hat{E} && (3)\\
\tilde{\Sigma} &= \frac{1 + nF^{-1} - n\hat{E}F^{-1}}{l}\hat{\Psi} && (4)\\
\hat{E} &= F^{-1}(n^{-1} + F^{-1})^{-1}\\
l &= m + n - p + 1\\
\hat{\Psi} &= \Psi + (n-1)S + (\bar{\mathbf{x}} - \boldsymbol{\nu})(n^{-1} + F^{-1})^{-1}(\bar{\mathbf{x}} - \boldsymbol{\nu})'.
\end{aligned}$$

In the equations following (2), $\tilde{\mathbf{x}}$ is the "posterior" mean, a linear combination of the sample averages and the hypermean. At the same time, $\tilde{\Sigma}$ is the "posterior" covariance matrix, $\hat{\Psi}$, a linear combination of the sample and hypercovariance matrix. Finally, l represents the degrees of freedom in the t distribution, F^{-1}, the variability between samples.

3.2 Hyperparameter Specification

We use an empirical Bayes method and avoid the need to fully specify the hierarchical prior distribution. Nevertheless, the sample information will

update the prior estimates, even though it does not fully reflect their uncertainty.

Estimating $\boldsymbol{\nu}$. Recall, that over a one dimensional domain, the spline, f, obtains from minimizing the objective function,

$$f \quad = \quad \arg\min_g \frac{1}{n}\sum_{i=1}^{n}(y_i - g(t_i))^2 + \lambda \int_0^1 (g^{(m)}(u))^2 du, \tag{5}$$

where the $\{y_i\}$'s are the measured responses (data), $g^{(m)}$ is the m-th derivative of g, and λ is the smoothing parameter. Over the two dimensional domains addressed in this paper, the objective function becomes

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - g(x_1(i), x_2(i)))^2 + \lambda \sum_{v=0}^{m} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\begin{array}{c} m \\ v \end{array} \right) \left(\frac{\partial^m f}{\partial x_1^v \partial x_2^{m-v}} \right)^2 dx_1 dx_2,$$

where $x_1(i)$ and $x_2(i)$ are the coordinate-pair that identify the site where y_i was measured, for all i.

We estimate the smoothness parameter, λ, by "generalized cross validation", that simplicity suggests we explain for one dimension rather than two, the latter being just a formalistic extension. Ordinary cross validation leaves out one data point in each successive iteration. With the k-th point removed, define $f_\lambda^{|k|}$ to be the minimizer of the objective function in Equation (5). Whereas the ordinary cross validation estimator of λ minimizes

$$V_0(\lambda) = \frac{1}{n}\sum_{k=1}^{n}\left(y_k - f_\lambda^{|k|}(x_k)\right)^2,$$

its generalized counterpart minimizes a weighted version of $V_0(\lambda)$:

$$V(\lambda) = \frac{1}{n}\sum_{k=1}^{n}\left(y_k - f_\lambda^{|k|}(x_k)\right)^2 w_{kk}(\lambda),$$

where

$$w_{kk}(\lambda) = \frac{(1 - a_{kk}(\lambda))^2}{1 - \bar{a}(\lambda))^2}$$

and

$$\bar{a}(\lambda) = \frac{1}{n}\sum_{i=1}^{n} a_{ii}(\lambda).$$

The $\{a_{ii}\}$ are the diagonal terms of the influence matrix $\mathbf{A}(\lambda)$, that satisfy

$$\begin{pmatrix} f_\lambda(x_1) \\ \vdots \\ f_\lambda(x_n) \end{pmatrix} = \mathbf{A}(\lambda)\mathbf{y}.$$

Here, $f_\lambda(x_k)$ is a linear combination of the components of $\mathbf{y} = (y_1, \cdots, y_n)'$ for fixed λ and each x_k. More details can be found in Wahba (2000).

Estimating Ψ and m. We estimate the inverted Wishart's scalar matrix, Ψ, by exploiting the spatial dependence in the data. The fact that Ψ's estimate plays the role of the "sum-of-squares" matrix in Σ's estimate [see Equation (4)], suggests the representation

$$\Psi = c \times \Phi,$$

Φ being a covariance matrix that is estimated using a semivariogram technique from geostatistics. More precisely, we represent elements of Φ as

$$Cov(X_i, X_j) \quad = \quad \sigma^2 - \gamma(h_{ij}), \tag{6}$$

where σ^2 is a common sample variance, h_{ij} is the Euclidean distance between the two sites X_i and X_j, and $\gamma(h)$ is an isotropic semivariogram model fitted from the data. Our analysis is not unduly sensitive to this assumption, since the Bayes estimate of the covariance matrix, a linear combination of prior and sample estimates, will account for anisotropy in the data.

We modified the EM algorithm developed by Sun (1994) to estimate the coefficient c and the degrees of freedom m. When F^{-1} is assumed known, the likelihood function of the extremes is proportional to

$$f(\mathbf{X}, \boldsymbol{\mu}, \Sigma|\Psi, m) \propto c_0^{-1}|\Psi|^{\frac{m}{2}}|\Sigma|^{\frac{m+p+1}{2}} \exp\left[-\frac{1}{2}tr(\Psi\Sigma^{-1})\right],$$

where

$$c_0 = 2^{pm/2}\pi^{(p-1)p/4}\prod_{i=1}^{p}\Gamma\left(\frac{m-i+1}{2}\right).$$

Since $(\Sigma, \log|\Sigma|)$ is a sufficient statistic for the family indexed by (Ψ, m), with $\Psi = c \times \Phi$, the algorithm will increase the likelihood at each iteration, that algorithm being:

E-step Given the current value of m and c,

$$E(\Sigma^{-1}|D) = (m+n-1)\hat{\Psi}^{-1},$$

where

$$\hat{\Psi} = c\Phi + (n-1)S + (\bar{\mathbf{X}} - \nu)'(n^{-1} + F^{-1})^{-1}(\bar{\mathbf{X}} - \nu)$$

and

$$E(\log|\Sigma||D, m, c) = -p\log 2 - \sum_{i=1}^{p} \Psi(\frac{m+n-i}{2}) + \log|\hat{\Psi}|.$$

Here Φ is the covariance matrix obtained from the semivariogram model, and S is the sample covariance matrix. So $\hat{\Psi}$ is the Bayes estimator of the scalar matrix Ψ at each step. $\Psi(\cdot)$ is the digamma function.
M-step Given the current values of Σ^{-1} and $\log|\Sigma|$, the log-likelihood function is proportional to

$$m\log|c\Psi| - tr(c\Psi\Sigma^{-1}).$$

When m is fixed, c can be found by maximizing the first derivative of the above equation,

$$c = \frac{mp}{tr(\Psi\Sigma^{-1})}.$$

When c is fixed, we may take the first derivative with respect to m so that m is the solution to the following equation

$$\sum_{i=1}^{p}\left[\Psi(\frac{m+n-i}{2}) - \Psi(\frac{m-i+1}{2})\right] = \log|\hat{\Psi}| - \log|c\Phi|.$$

Estimation of F^{-1}. Here we propose an estimator like that of Le et al (1997). Recall that conditional on μ and Σ, $X_j : p\times 1 \sim MVN_p(\mu, \Sigma)$ for all $j = 1, \ldots, n$. The independence of these responses implies $\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i\cdot})^2 \sim \Sigma_{ii}\chi_{n-1}$. Thus, conditional on Σ, $E[\hat{\Sigma}_{ii}^{-1}] = \Sigma_{ii}^{-1}$ where $\hat{\Sigma}_{ii} = \sum_{j=1}^{n}(X_{ij} - \bar{X}_{i\cdot})^2(n-3)^{-1}$.

At the same time, conditional on υ and Σ, $\bar{X} : p\times 1 \sim MVN_p(\upsilon, (F^{-1} + n^{-1})\Sigma)$. Hence, $E[(\bar{X}_{i\cdot} - \upsilon_i)^2]\Sigma_{ii}^{-1} = (F^{-1} + n^{-1})$ for all $i = 1, \ldots, p$. However, by standard Gaussian linear model theory, $\hat{\Sigma}_{ii}$ and $(\bar{X}_{i\cdot} - \upsilon_i)^2$ are independent. It therefore follows that $E[(\bar{X}_{i\cdot} - \upsilon_i)^2\hat{\Sigma}_{ii}^{-1}] = (F^{-1} + n^{-1})$ for all i conditional on Σ, and the result does not depend on Σ, conditional only the hyperparameters including υ. This result suggests an unbiased estimator for F^{-1}:

$$\hat{F}^{-1} = p^{-1}\sum_{i=1}^{p}(\bar{X}_{i\cdot} - \upsilon_i)^2\hat{\Sigma}_{ii}^{-1} - n^{-1}. \tag{7}$$

In practice, υ will also need to be specified. As one possibility, we might suppose $\upsilon_i \equiv \upsilon_o$ for all i and then plug-in an estimator for υ_o, say

$$\hat{\upsilon}_o = \frac{\sum_{i=1}^{p} \bar{X}_{i.} \hat{\Sigma}_{ii}^{-1}}{\sum_{i=1}^{p} \hat{\Sigma}_{ii}^{-1}},$$

an estimate of the BLUE estimator. The estimator obtained from Equation (7) might nevertheless be approximately unbiased if n were reasonably large. Finally, Fu (2002) shows that if $\Sigma_{ii} \equiv \Sigma_o\}$ for all, taking $\hat{\upsilon}_o$ to be the grand sample average yields an unbiased estimator.

3.3 Semivariogram

We used SAS's PROC VARIOGRAM to calculate the "sample semivariogram". To be precise, we label site i as P_i and assume the quantity being measured has a value there of V_i. SAS's method partitions the sites, $\{P_i\}$, into "angle/distance classes" $\{N(\theta_k, L)\}$, according to their geographical distribution. For every pair, (P_i, P_j), within each such (so-called lag) class, it then calculates $(V_i - V_j)^2$, $|P_i P_j|^2$ (their squared Euclidean distance), and the total number of such distinct pairs in that class, say $P(\theta_k, L)$. Finally, it computes:

$$\gamma(h_k) = \frac{1}{|P(\theta_k, L)|} \sum_{P_i \neq P_j \in N(\theta_k, L)} (V_i - V_j)^2,$$

where h_k is calculated by:

$$h_k = \frac{1}{|N(\theta_k, L)|} \sum_{P_i, P_j \in N(\theta_k, L)} |P_i P_j|.$$

The result is SAS's sample semivariogram to which we can fit a parametric semivariogram model from any one of a number of classes such as the exponential, Gaussian, wave, spherical, and linear (see Cressie, 1993). An understanding of the physical processes involved can help select that class. However, the estimated covariance matrix $\mathbf{\Psi}$ must be positive definite matrix

We take the semivariogram to be 0 when the distance between any two sites is 0, that is we assume no "nugget" effect, in agreement with several semivariogram models proposed by Cressie (1993). One is the exponential semivariogram model defined as

$$\gamma(h) = \begin{cases} c_0 + c_W \left(1 - \exp\left(\frac{-h^2}{a_W^2}\right)\right), & \text{if } h > 0 \\ 0, & \text{if } h = 0 \end{cases}$$

where c_0, c_W and a_W are parameters of nugget, sill, and shape, respectively. The nugget effect, c_0, (or discontinuity in the semivariogram at $h = 0$) derives from a combination of microscale- and measurement-error (see Cressie (1993):

$$c_0 = c_{MS} + c_{ME}.$$

When using Equation (6) in conjunction with

$$Cov(X_i, X_j) = \sigma^2 - \gamma(h_{ij})$$

to calculate the variance at each site, σ^2 accounts only for the measurement error. Using c_0 as the value at $h = 0$ may result in a negative variance and thus a non-positive definite Ψ.

Note that we obtain the parametric semivariogram model from the sample semivariogram that is estimated by the method of moments. We can also use that method to estimate the diagonal terms in Φ, and thus base our estimate of that common variance on all samples and sites under our assumption of a common variance.

3.4 Model Assessment.

Before adopting multivariate t obtained in Subsection 3.1 as the joint posterior extreme response distribution , we need to see if agrees with the data in any given application. This subsection offers a variant of cross-validation for making that assessment. For any given year, the method removes data from selected sites and uses that from the remaining sites, in conjunction with the predictive t-distribution model, to "predict" the missing values using

$$\hat{\mathbf{x}}_\mathbf{u} = \boldsymbol{\nu}_u + (\hat{\mathbf{x}}_g - \boldsymbol{\nu}_g)' \Psi_{gg}^{-1} \Psi_{gu},$$

u denoting "ungauged" and g, "gauged". Here, $\boldsymbol{\nu}_u$ and $\boldsymbol{\nu}_g$ partition the "prior mean" according to whether the site is ungauged or gauged respectively, and Ψ_{gg}^{-1} and Ψ_{gu} come from conformably partitioning Ψ.

Given the data at the remaining sites, a $(1-\alpha)$-level credibility ellipsoid can readily be found and it is provided in the next theorem. Appendix A offers a proof.

Theorem 3.1 *The $(1-\alpha)$th credibility ellipsoid is*

$$\{\mathbf{X_u} : (\mathbf{X_u} - \hat{\mathbf{x}}_\mathbf{u})' \Psi_{u|g}^{-1} (\mathbf{X_u} - \hat{\mathbf{x}}_\mathbf{u}) < b\}$$

where

$$
\begin{aligned}
b &= (u \times P_{u|g} \times F_{1-\alpha,u,m-u+1}) \times (m-u+1)^{-1}, \\
\Psi_{u|g} &= \Psi_{uu} - \Psi_{ug}\Psi_{gg}^{-1}\Psi_{gu}, \text{ and} \\
P_{u|g} &= 1 + F^{-1} + (\mathbf{x_g} - \boldsymbol{\nu}_g)'\Psi_{gg}^{-1}(\mathbf{x_g} - \boldsymbol{\nu}_g).
\end{aligned}
$$

Note that in Theorem 3.1, $P_{u|g}$ denotes the posterior mean of the conditional multivariate-t posterior distribution.

Repeatedly removing at random, site subsets of fixed size as above, and determining whether or not the missing data vector lies in the ellipsoid each time, enables us to estimation their coverage probabilities. An estimate of about $(1-\alpha) * 100$ would assure, in operational form, the quality of the multivariate - t approximation. The analysis, also enables "outlier" detection, that is, the identification of sites with undue influence on coverage probability. We can even explore the 'coverage probability' - 'number of sites removed' relationship.

3.5 Return Values

As noted in Section 1, the return value X_T is defined for a marginal extreme response distribution by

$$P(X > X_T) = \frac{1}{T}. \tag{8}$$

Here, T is a return period, for example, 10, 50, 100, or even 1000 years, depending on the context. If the logarithmic response transformation is used, the return values transform in the obvious way. In any case, the probabilities in Equation (8) would be computed *a posteriori*, that is conditionally on available data.

Return values at each site can readily be estimated from the posterior multivariate t distribution by repeatedly sampling from it. The (1-1/T)-th percentile of the simulated data for the given marginal response, estimates that value. However, we can also simplify this task by approximating the log multivariate t with the multinormal distribution giving a T-year return value for site or region i, of

$$x_{1-T,j} = \tilde{x}_i + \Phi(1-T) \times \tilde{\sigma}_i.$$

If a log transform has been used, this percentile needs to be transformed back to the original scale to get the return value estimate of

$$y_{1-T,i} = e^{x_{1-T,i}}.$$

3.6 Assessment of Return Value Estimates

When the return values are defined marginally, we use the following Measure-of-Discrepancy (MOD) to index the proximity of the estimated and true return values:

$$\text{Measure of Discrepancy} \quad = \quad \frac{1}{p}\sum_{i=1}^{p}\left|\frac{1}{T} - \frac{\sum_{j=1}^{n} I(X_{ij} > X_{i,T})}{n}\right|. \qquad (9)$$

Here p is the number of sites (cells in the application addressed in this paper); X_{ij} is the annual maximum precipitation for year j and region i; 1/T is the true exceedance probability; the second (empirical) fraction estimates fraction of exceedances of estimated return values, $\{X_{i,T}\}$.

While the MOD has intuitive appeal, other discrepancy summaries may be more appropriate. In any case, it measures performance only crudely and finer analyses would generally be needed. For example, the distribution of discrepancies can point to locations that deviate markedly from the rest.

4 Application.

We now apply the theory developed above to model Canada's daily precipitation data as described in Kharin and Zwiers (2000).

4.1 Daily Precipitation Simulated by CGCM1

The data derive from the first generation Canadian Global Coupled Model (CGCM1) (Kharin and Zwiers, 2000). Three independent simulations were run. These simulated hourly precipitation (mm/day) in three 21-year windows (1975-1995, 2040-2060, 2080-2100) for all cells in a grid that geographically partitions the whole of Canada. That 26×12 grid is uniform along lines of longitude, with a cell size of 3.75°, and nearly uniform for latitude (approx. 3.75°). The 26 lines of longitude are ordered eastwards from 146.25°W to 52.50°W while the 12 lines of latitude are ordered from south (42.68°N) to north (83.48°N). We computed annual precipitation maxima to get $21 \times 3 = 63$ values in each grid cell in each time window.

4.2 Preliminary Data Analysis

To apply the theory developed above, we need to assess the applicability of the assumed underlying distributions.

Preliminary data analysis began with examination of the empirical marginal distributions. Normal Q-Q plots of annual maximum precipitation levels for many selected regions among the 319 were drawn, two examples being shown in Figure 1. The top left plot shows a long right tail, that is quite successfully removed by log-transformation. In contrast, log transforming the data displayed in the top right plot overcorrects, leading to the expectation that return values would be underestimated. (Note that for the data analysis, we simplified the geographical coordinates by replacing the exact longitude and longitude with sequences of consecutive integers starting from 1 and going up to 26 and 12, respectively.)

Autocorrelations plots (not shown here) for the simulated 1975-1995 data period suggest we can treat the series as auto-uncorrelated.

We did similar preliminary marginal analysis for the periods, 2040-2060 and 2080-2100. Although extreme levels generally shift upwards over time, the shapes of the Q-Q and autocorrelation plots are generally similar.

To make computation feasible, our analyses are restricted to data from 1975-1995 since that will meet our objective of illustrating our approach. In the sequel, "data", we mean log-transformed data, unless stated otherwise. Also note that, since three independent simulations have been run for each time period, there are $21 \times 3 = 63$ samples in each time window.

4.3 Estimating Hyperparameters

To estimate the hyperparameter $\boldsymbol{\nu}$, the "prior mean vector" of 312 dimensions, we smooth the sample mean vector $\bar{\mathbf{X}}$ using thin-plate splines as described in the previous section. The sample mean vector is calculated by taking the average of the 63 samples. The smoothing parameter in that method is given by generalized cross-validation (GCV) that makes use of the information provided by data and therefore takes the spatial correlation into account. The smoothing parameter chosen this way proves to be rather small. So the surface of the averages after smoothing remains fairly rough.

To estimate the hyperparameter Ψ, the 312×312 "prior sum of squares matrix", we follow the procedure in Section 2.2.2 to estimate the covariance Φ and the coefficient c separately.

To estimate Φ, we start with Equation 6:

$$Cov(X_i, X_j) = \sigma^2 - \gamma(h_{ij}),$$

$\gamma(h)$ being fitted to the sample semivariogram. The initial estimates of the three parameters, nugget, sill, and shape, in the semivariogram function are modified by the Gauss-Newton algorithm, to minimize the residual sum of

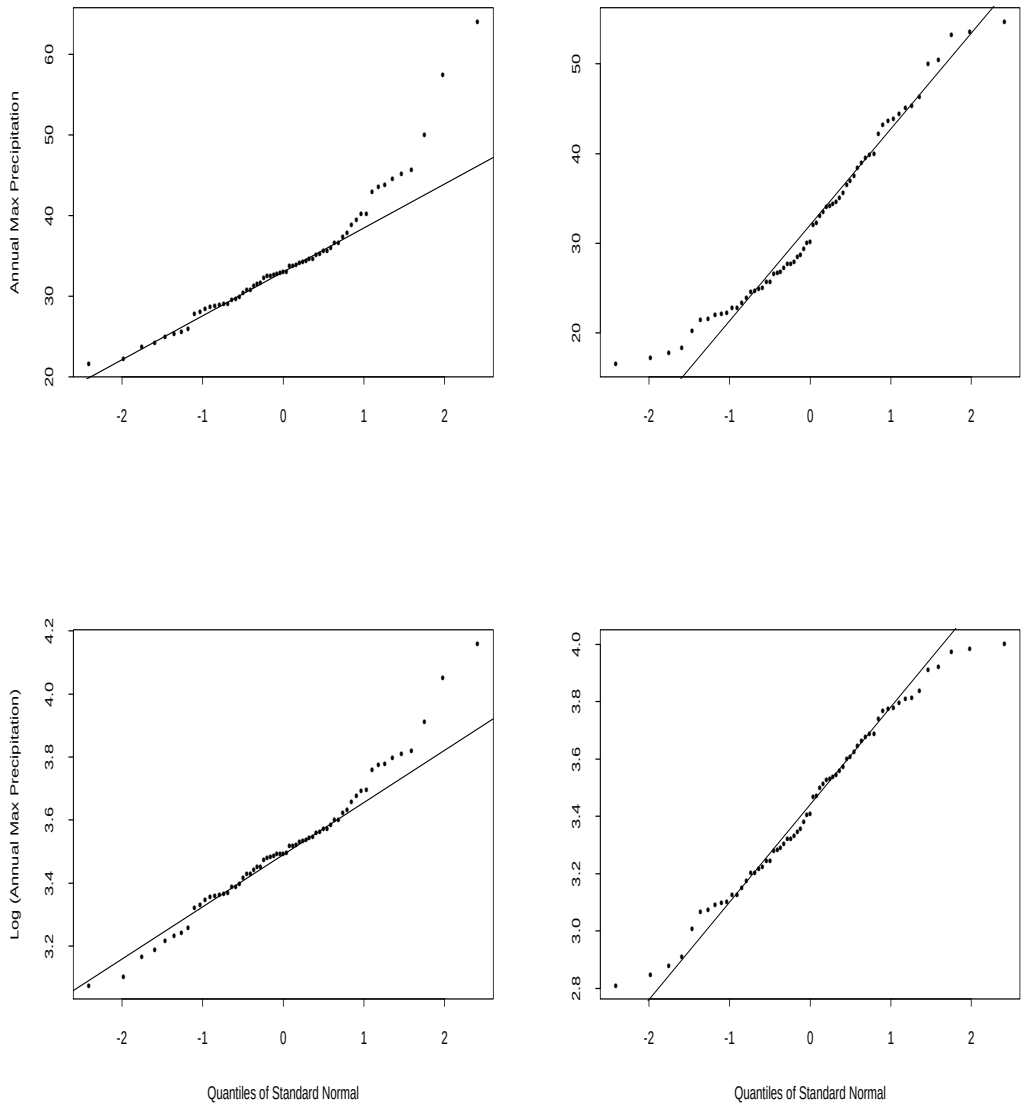

Figure 1: Q-Q plots of annual maximum precipitation data at selected sites, (i=5, j=1) on the left corresponding to(131.25°W, 42.68°N) and (i=18, j=8) on the right corresponding to (82.50°W, 68.64°N). The upper two plots are for raw data while the bottom two, for log-transformed data. Letters i and j code longitude and latitude, respectively.

squares and achieve positive definiteness of Ψ. (See Section 2.3 for further details.)

The SAS procedure described in Section 3.3 estimated the sample variogram for 1975-1995 . It indicated a significant nugget effect, had a wavy shape and therefore pointed to the wave model that we fitted. To ensure the estimated Ψ's positive definiteness, we successively tried various values of the three parameters based on the initial estimates until success was realized. The best model obtained by this procedure is

$$\gamma(h) = \begin{cases} 0.09 + 0.15\left(1 - 3\left(\frac{\sin(h/3)}{h}\right)\right), & \text{if } h > 0 \\ 0, \quad \text{if } h = 0. \end{cases} \tag{10}$$

In Formula (6), σ^2 is estimated by calculating the overall variance across all 312 grid points and 63 samples. The result, 0.23, is significantly larger than any covariance of any two grid points, that may result in a larger estimated covariance matrix. Consequently, the constructed credibility region may be inflated along with the resulting estimated coverage probability.

We used the estimate of the hyperparameter F^{-1} proposed in Subsection 3.2. In our hierarchical Bayes model, the "true mean" is the unknown hypermean and we estimated it with that obtained by smoothing the sample mean using a thin-plate spline, to get an approximately unbiased estimator of F^{-1}. The is -0.016, that is essentially zero. Incidentally, we are not justified in using the estimator of F^{-1} suggested in that subsection for the case of a constant mean. That assumption is not tenable in this application.

Using the EM algorithm, we estimated the degrees of freedom in the inverted Wishart distribution m to be 355 and the coefficient c in the Ψ matrix to be 49. The degrees of freedom m, being large, yields avoids tails of moderate weight in the estimated distribution, as demonstrated by using an equation from Le et al (1997)

$$var(\mathbf{X}^u|\mathbf{X}^g = \mathbf{x}^g) = (m - u - 3)P_{u|g}\Psi_{u|g},$$

where

$$\begin{aligned} P_{u|g} &= 1 + F^{-1} + (\mathbf{x}_g - \boldsymbol{\nu}_g)'\Psi_{gg}^{-1}(\mathbf{x}_g - \boldsymbol{\nu}_g) \\ \Psi_{u|g} &= \Psi_{gg} - \Psi_{ug}\Psi_{gg}^{-1}\Psi_{gu} \\ \Psi &= \begin{pmatrix} \Psi_{uu} & \Psi_{ug} \\ \Psi_{gu} & \Psi_{gg} \end{pmatrix} \end{aligned}$$

When all the sites are gauged, u in the above equations becomes 0. The variance relies heavily on m and increases as m increases.

Meanwhile, c=49 makes the estimated matrix $\hat{\Psi}$ comparable to the sample sum-of-squares matrix which is the product of 62 and the sample covariance matrix. Therefore, the approach does not put too much weight on either the prior knowledge expressed through the inverted Wishart distribution or the sample covariance matrix.

4.4 Assessing the Multivariate t Distribution

Following the approach described in Section 3.4, we assess the suitability of the joint multivariate t posterior derived above using a variant of cross - validation:

1. Select a credibility level, 30%).
2. Randomly remove 30 sites (without replacement).
3. Use Sample 1 for the remaining 282 (=312-30) sites to obtain the credibility region for those 30 using Theorem 3.1.
4. See if that region includes the 30 dimensional vector of removed site values.
5. Repeat Steps 2 and 3 above for each of the remaining 62 sample records and find the fraction of times those vectors are included. This completes one run.
6. Replace the 30 sites and repeat Steps 2 - 4.
7. Repeat Steps 2 - 5 100 times.
8. Repeat Step 1 - 5 for levels 95% and 99.9%.

We can now compare the nominal and empirical credibility levels obtained above. (The intended purpose of our predictive distribution suggests putting more emphasis on the largest 2 of the 3 levels used above.) Figure 2 and Table 1 display the results and show reasonable agreement.

As an ideal, coverage probabilities should remain close to their nominal levels as the number of removed sites increases. This ideal leads us to repeat the cross - validation study above with 30 replaced systematically by $1, 2, \ldots$ and a level of 95% while fixing the degrees of freedom in our model, m, at 355 and the parameter, c, at 49. The empirical coverage fractions

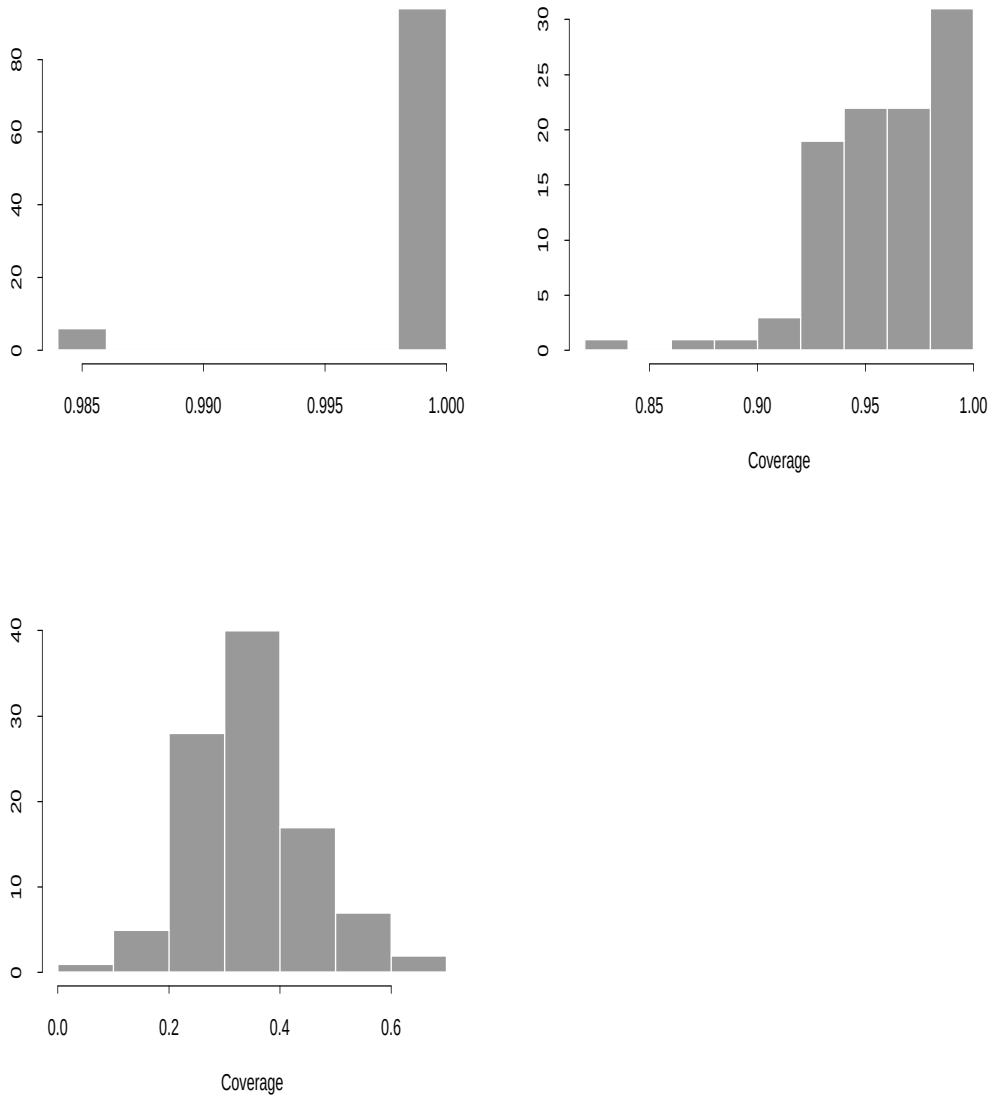

Figure 2: Histograms of credibility ellipsoid coverage probabilities for annual maximum precipitation levels, 1975 - 1995. Clockwise from the upper left, they correspond to credibility levels 99.9%, 95%, and 30%.

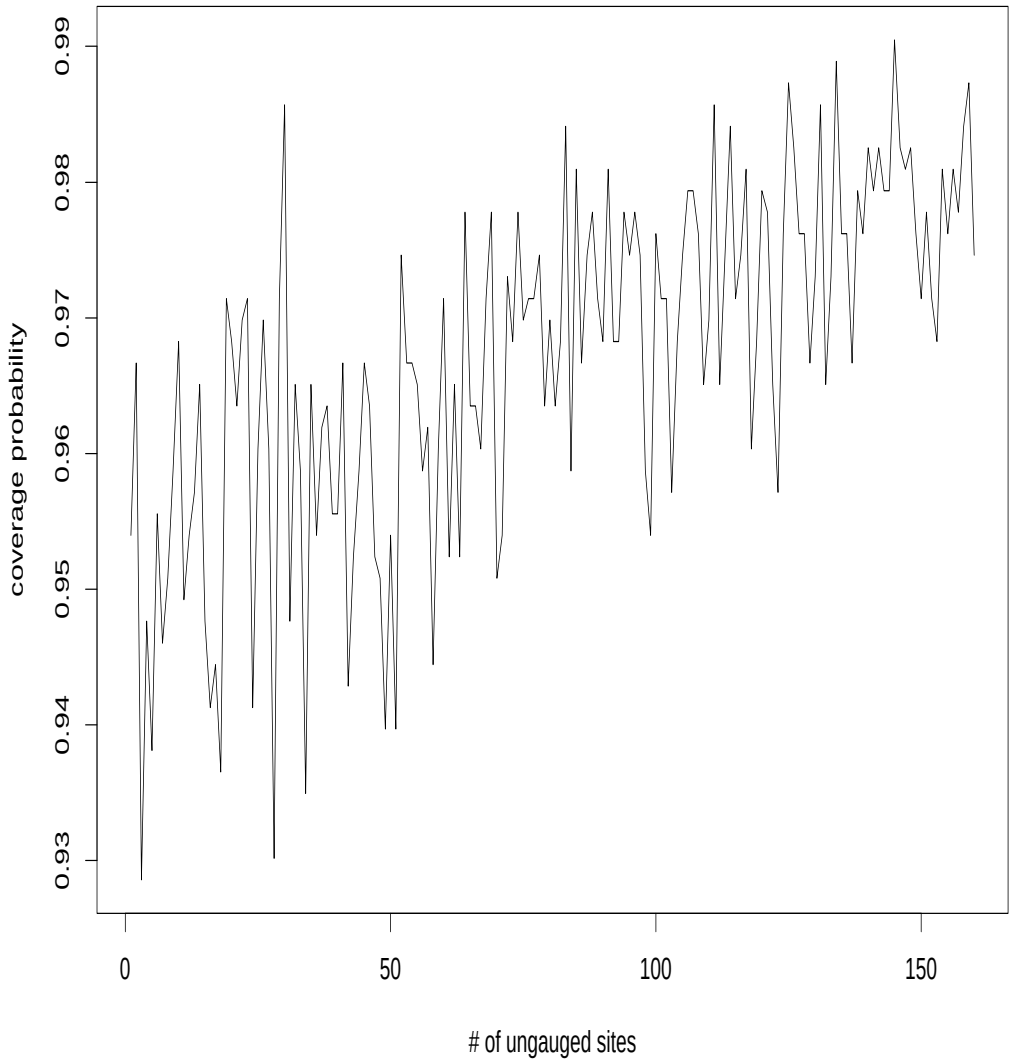

Figure 3: Coverage probabilities versus numbers of removed sites for the annual maximum precipitation data with degrees of freedom fixed at m = 355 and parameter c = 49.

Credibility Level	Mean	Median
30%	35	35
95%	96	97
99.9%	99.9	1

Table 1: Summary of credibility ellipsoid coverage probabilities for different levels of annual maximum precipitation data based on 100 runs.

are plotted in Figure 3 against the number removed. Below about 60, the coverage probability remains stable at about 95%. However above that, the coverage probability trends upwards to about 98%. This finding indicates a conservative inflation of the actual credibility level when the number of simultaneously predicted site responses grows beyond about 20% of the total. That inflation may derive from the generally small inter-site spatial correlations in this application. Removing a large number of sites results in a paucity of information from the combination of the few remaining and hence the inflation noted above.

Overall, our assessment suggests the multivariate -t models the annual maximum precipitation data fields reasonably well.

4.5 Assessment of Return Value Estimates

We followed the procedures described in Section 2.5 to calculate approximate return values. Figure 4 depicts the estimated 10-year return values, i.e. the 90% quantiles. That surface is shifted above that for averages as expected. Moreover, the return value surface preserves trends in that for averages.

Formula 9 yields the Measure of Discrepancy (MOD) score for the estimated 10-, 20-, 50- and 100-year return values in time window 1975-1995. In Table 2, we see that the MOD drops dramatically as the return period increases. (The smallest feasible level included there derives from the granularity of the relative proportions, making 0 unattainable.) However, none of the scores based on the estimated return values (the middle column) are close to 0 even though we see above that the multivariate -t models the response distribution well. A possible explanation lies in our use of multivariate normal approximation of the t and subsequent use of just the covariance's diagonal terms. Thus, in effect we failed to borrow strength through spatial dependence. The first method in Section 2.5 that uses the t directly, may give a better estimate. That will be the subject of a future investigation.

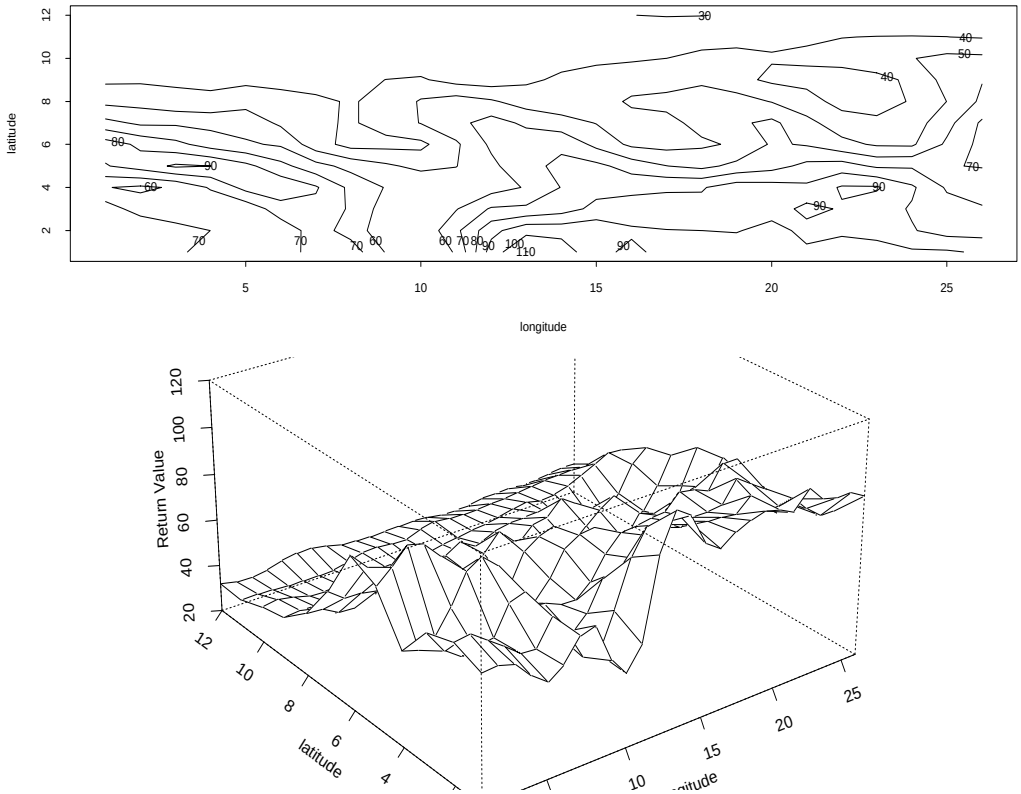

Figure 4: Contour and perspective plots of estimated 10-year return values for the precipitation data. Units are in mm/day.

Return Period	100*MOD (from data)	100*Smallest MOD
10	8.35	0.48
20	4.62	0.24
50	1.96	0.41
100	1.00	0.59

Table 2: Comparison of the return value MOD for the annual maximum precipitation data along with the smallest feasible MOD.

5 Discussion

This paper points to a promising direction for modelling the joint response distribution for extreme values of not unduly high order. That distribution builds on a log multivariate normal (multi - normal). Its richness of theory makes it an ideal foundation for hierarchical modelling. In conjunction with the Inverted Wishart, a conjugate distribution for the multi - normal, it yields a log multivariate - t posterior. The latter's tractability enables, among other things, the construction of credibility regions for the prediction of unmeasured responses.

The log multivariate - t approximation does not depend explicitly on the distribution of their componentwise counterparts (hourly precipitation in our example), unlike the GEV, GPD and the point process theories that are limiting distributions of those components. In the case of the GEV, for example, the componentwise distributions need to meet specific criteria to be eligible for a role in the limit even though at most one of three limiting types obtain. Verifying those criteria poses a problem (Smith, 2001). For example, an F whose tail has the form

$$1 - F(x) \sim cx^{-\alpha}, \quad x \to \infty$$

for any constant $c > 0$ and $\alpha > 0$, has a Fréchet type of limiting distribution. Thus, this "domain of attraction" requirement may confine the generality of the GEV. Similar problem exists for the GPD and point process theories. In contrast, no such limitations apply to the multivariate -t.

Furthermore, the multivariate - t approximation of Le and Zidek (1992) has been substantially extended to include covariates (*e.g.* location or time) as well as multiple responses (*e.g.* chemical species; Sun, 1994, Le et al, 1997). Thus, for example, by including a time covariate, we could embrace the three sampling time windows in one model rather than relying on three. More generally, the extended theory offers a good deal of flexibility for modelling extreme responses.

Finally, our multivariate distribution approximation enables the calculation of joint exceedance probabilities for several regions simultaneously, a calculation not be possible with marginals alone. In fact, this suggests extending the return value in Equation 1 as follows:

$$P(X_1 > x_R, X_2 > x_R, ..., X_p > x_R) = \frac{1}{T}.$$

That extension may be useful in making policy and more generally, environmental risk analysis although it seems unduly stringent. Such a simultaneous

exceedance would seem an unreasonable criterion. A more plausible choice would be:

$$P(\text{the exceedance occurs at least 2 or 3 sites}) = \frac{1}{T},$$

However, with the advantage of a tractable joint distribution, many other definitions are possible and the ultimate choice could be made context dependent.

As the aggregation level increases, say from weekly maximum to yearly maximum for example, the inter-site correlation commonly decreases. That phenomenon is seen in Chang et al (2003) where correlations among sites at different aggregation levels are calculated for de-trended hourly PM_{10} data from nine stations in the Greater Vancouver Regional District during 1997 to 2001. A simulation study of Chang et al (2003) confirms that finding.

On the other hand, Chang et al (2003) find a few site pairs with correlations that persistent under increasing aggregation levels and consider the implications of their findings for network design as well as the setting of urban air quality criteria.

Acknowledgements. This material was based upon work supported partly by the Natural Science and Engineering Research Council of Canada and partly by the National Science Foundation under Agreement No. DMS-0112069. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] Anderson, T. W. (1984) *An Introduction to Multivariate Statistical Analysis.* New York: Wiley.

[2] Chang, H, Le, ND, Zidek, JV and Fu, A. (2002). Perspectives on Designing Environmental Monitoring Networks for Measuring Extremes. In preparation.

[3] Coles, S. G. and Tawn, J. A. (1996) Modelling Extremes of the Areal Rainfall Process. *J.R. Statis. Soc.* **B**, **58**, No. 2, 329-347.

[4] Cressie, N. (1993) *Statistics for Spatial Data*. New York: Wiley.

[5] Dempster, A.P. (1969). *Elements of Continuous Multivariate Analysis.* Addison-Wesley Publishing Company. 307.

[6] Falk, M., Hüsler, J. and Reiss, R.-D. (1994). *Laws of Small Numbers: Extremes and Rare Events.* DMV-Seminar Bd 23. Birkhäuser, Basel.

[7] Fu, A (2002). Case Study: Inference for Extreme Spatial Random Rainfall Fields. M.Sc Thesis. Department of Statistics, University of British Columbia.

[8] Joe, H. (1994) Multivariate Extreme-Value Distributions with Applications to Environmental Data. *Can. J. Statis.*, **22**, No. 1, 47-64.

[9] Kharin, V. V., and Zwiers, F. W. (2000). Changes in the Extremes in an Ensemble of Transient Climate Simulations with a Coupled Atmosphere-Ocean GCM. *Journal of Climate*. **13** 3760-3788.

[10] Le, N. D. and Zidek, J. V. (1992) Interpolation with uncertain spatial covariances: a bayesian alternative to kriging. *Journal of Multivariate Analysis*, **43**, No. 2, 351-374.

[11] Le, N. D., Sun, W. and Zidek, J. V. (1997) Bayesian multivariate spatial interpolation with data missing by design. *J. R. Statis. Soc.* **B**, **59**, No.2, 501-510.

[12] Ledford, A. W. and Tawn, J. A. (1997) Modelling Dependence with Joint Tail Regions. *J. R. Statis. Soc.* **B**, **59**, No. 2, 475-499.

[13] McFadden, D. (1978) Modelling the choice of residential location. *Spatial Interaction Theory and Planning Models.* Karlquist, A. et al (eds.), North Holland, Amsterdam, 75-96.

[14] Reiss, R.-D. and Thomas, M. (1997) *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields.* Birkhäuser. 167-173.

[15] Smith, R. L. (2001) *Environmental Statistics.* Notes at the Conference Board of the Mathematical Sciences (CBMS) course at the University of Washington, June 25-29, 2001. http://www.stat.unc.edu/postscript/rs/envnotes.ps

[16] Sun, L., Zidek, J.V., Le, N.D. and Özkaynak, H. (2000) Interpolating Vancouver's daily ambient PM_{10} field. *Environmetrics*, **11**, 651-663.

[17] Sun, W. (1994) Bayesian multivariate interpolation with missing data and its applications. Ph.D Dissertation. Dept. of Statistics. Univ. of British Columbia. G

[18] Wahba, G. (2000) Splines in nonparametric regression. Technical Report 1024, Department of Statistics, University of Wisconsin, Madison WI.

A The Credibility Ellipsoid.

Theorem A.1 *Given data from the gauged sites, a $(1-\alpha)$-level credibility ellipsoid is as follows:*

$$\{\mathbf{X}_u : (\mathbf{X}_u - \hat{\mathbf{x}}_u)' \Psi_{u|g}^{-1} (\mathbf{X}_u - \hat{\mathbf{x}}_u) < b\}$$

where,

$$b = (u \times P_{u|g} \times F_{1-\alpha,u,m-u+1}) \times (m-u+1)^{-1},$$

and

$$\Psi_{u|g} = \Psi_{uu} - \Psi_{ug}\Psi_{gg}^{-1}\Psi_{gu}.$$

u and g denote the numbers of ungauged sites and gauged sites, respectively. $P_{u|g}$ is the mean in the a posteriori multivariate t distribution.

$$P_{u|g} = 1 + F^{-1} + (\mathbf{x}_g - \boldsymbol{\nu}_g)' \Psi_{gg}^{-1} (\mathbf{x}_g - \boldsymbol{\nu}_g).$$

Proof. Le et al (1997) show that

$$\mathbf{X}_u \sim \boldsymbol{\nu}_u + (\mathbf{x}_g - \boldsymbol{\nu}_u)' \Psi_{gg}^{-1} \Psi_{gu} + \boldsymbol{\tau}(\delta + g, P_{u|g}, \Psi_{u|g}),$$

where

$$P_{u|g} = 1 + F^{-1} + (\boldsymbol{\nu}_g - \boldsymbol{\nu}_g)' \Psi_{gg}^{-1} (\mathbf{x}_g - \boldsymbol{\nu}_g).$$

Here, δ is the degrees of freedom as defined in Dawid (1981) rather than Anderson (1984). The following equation relates the two notations and enables us to switch to m:

$$\delta = m - p - 1 = m - g - u - 1.$$

Here, p is the dimension of the vector. In the example of precipitation data, it is the sum of the numbers of gauged and ungauged sites.

The above multivariate t distribution can be written as:

$$\mathbf{Y} = \frac{(\mathbf{X}_u - \hat{\mathbf{x}}_u)}{\sqrt{P_{u|g}}} \sim \boldsymbol{\tau}(\delta + g, 1, \Psi),$$

which means

$$\mathbf{Y}|\mathbf{V} \sim MVN(\mathbf{0}, \mathbf{V}),$$

$$\mathbf{V} \sim IW(\delta + g, \Psi),$$

$$\mathbf{V^{-1}} \sim W(\delta + g + u - 1, \Psi^{-1}),$$

where

$$\delta + g + u - 1 = m.$$

We seek the distribution of $\mathbf{Y}'\Psi^{-1}\mathbf{Y}$. Since Ψ is known, we need to introduce the random variable $\mathbf{A}$ to use the definition of the F distribution. Suppose $\mathbf{Z}_i,\ i = 1, \ldots, u$ are independent vectors, each with a multivariate normal distribution, $MVN_m(\mathbf{0}, \mathbf{I})$. Let $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z_u})$. Then

$$\mathbf{Z}'\mathbf{Z} \sim W_u(m, \mathbf{I}).$$

We may transform $\mathbf{Z}$ to $\mathbf{T_{u\times u}}$ so that

$$\mathbf{Z}'\mathbf{Z} = \mathbf{T}'\mathbf{T} \sim W_u(m, \mathbf{I}).$$

To see this, observe that for any orthogonal matrix $\mathbf{Q}_{m\times m}$,

$$\mathbf{Z}'_{p\times m}\mathbf{Z}_{m\times p} = \mathbf{Z}'\mathbf{Q}'\mathbf{Q}\mathbf{Z} = (\mathbf{Q}\mathbf{Z})'\mathbf{Q}\mathbf{Z}.$$

Select $\mathbf{Q} = \mathbf{Q}_1$ so that

$$\mathbf{Q}_1\mathbf{Z}^1 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ ||Z^1|| \end{pmatrix},$$

where

$$||Z^1||^2 \equiv t_{m,1} = (Z_1^1)^2 + (Z_2^1)^2 + \cdots + (Z_m^1)^2 \sim \chi^2_m.$$

Also,

$$\mathbf{Q}_1\mathbf{X}^i \sim N_m(\mathbf{0}, \mathbf{Q}\mathbf{I}_m\mathbf{Q}') = 3DN_m(\mathbf{0}, \mathbf{I}_m).$$

So,

$$\mathbf{Q}_1\mathbf{Z} = \begin{pmatrix} 0 & & & & \\ \vdots & \mathbf{X}^2 & \cdots & \mathbf{X}^p \\ t_{m,1} & & & \end{pmatrix}.$$

Now select $\mathbf{Q}_2$ to transform the above matrix into

$$\begin{pmatrix} 0 & & & & \\ \vdots & 0 & \mathbf{X}^3 & \cdots & \mathbf{X}^p \\ 0 & t_{m-1,2} & & \cdots & \\ t_{m,1} & X_m^2 & & \cdots & \end{pmatrix}$$

and

$$t^2_{m-1,2} = (X_1^2)^2 + (X_2^2)^2 + \cdots + (X^2_{m-1})^2 = \chi^2_{m-1}.$$

Repeating these steps, we transform $\mathbf{Z}$ into

$$\begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & t_{m-p+1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & t_{m-1,2} & \cdots & X^p_{m-p+2} \\ t_{m,1} & X^2_m & \cdots & \cdots \end{pmatrix} \equiv \begin{pmatrix} 0 \\ \mathbf{T}_{p\times p} \end{pmatrix}.$$

Let $\Psi^{-1} = \boldsymbol{\gamma}'\boldsymbol{\gamma}$, where $\boldsymbol{\gamma}_{u\times u}$ is the decomposition of Ψ^{-1}. Then

$$\begin{aligned} \mathbf{V}^{-1} &\equiv \boldsymbol{\gamma}'\mathbf{T}'\mathbf{T}\boldsymbol{\gamma} \\ &= \boldsymbol{\gamma}' \sum_{i=1}^{m} \mathbf{Z}_i\mathbf{Z}_i'\boldsymbol{\gamma} \\ &= \sum_{i=1}^{m} (\boldsymbol{\gamma}'\mathbf{Z}_i)(\mathbf{Z}_i'\boldsymbol{\gamma}) \end{aligned}$$

where

$$\mathbf{Z}_i \sim N(\mathbf{0}, \mathbf{I}),$$

and

$$\boldsymbol{\gamma}'\mathbf{Z}_i \sim N(\mathbf{0}, \boldsymbol{\gamma}'\mathbf{I}\boldsymbol{\gamma}) = N(\mathbf{0}, \Psi^{-1}).$$

Therefore,

$$\mathbf{V}^{-1} \equiv \boldsymbol{\gamma}'\mathbf{T}'\mathbf{T}\boldsymbol{\gamma} \sim W_u(m, \Psi^{-1}).$$

Now we have

$$\begin{aligned} \mathbf{Y}'\Psi^{-1}\mathbf{Y} &= \mathbf{Y}'\boldsymbol{\gamma}'\boldsymbol{\gamma}\mathbf{Y} \\ &= (\mathbf{Y}'\boldsymbol{\gamma}'\mathbf{T}')\mathbf{T}'^{-1}\mathbf{T}^{-1}(\mathbf{T}\boldsymbol{\gamma}\mathbf{Y}) \\ &= \mathbf{U}'\mathbf{T}'^{-1}\mathbf{T}^{-1}\mathbf{U} \\ &= (\mathbf{U}\mathbf{T}^{-1})'(\mathbf{U}\mathbf{T}^{-1}) \\ &= ||\mathbf{U}\mathbf{T}^{-1}||^2 \end{aligned}$$

Since,

$$\begin{aligned} \mathbf{U} &= \mathbf{T}\boldsymbol{\gamma}\mathbf{Y} \\ &\sim N_u(\mathbf{0}, \boldsymbol{\gamma}'\mathbf{T}'\Psi\mathbf{T}\boldsymbol{\gamma} \\ &\sim N_u(\mathbf{0}, \mathbf{I}_u), \end{aligned}$$

the distribution of $\mathbf{U}$ does not depend on $\mathbf{T}$. On the other hand, we have shown that $\mathbf{T}'\mathbf{T} \sim W_u(m, I)$. So from Theorem 13.6.2 in Dempster (1969), we conclude that $[(m-u+1)/u]\mathbf{Y}'\Psi^{-1}\mathbf{Y} = 3D\mathbf{U}'(\mathbf{T}'\mathbf{T})^{-1}\mathbf{U}$ has the $F(u, m-u+1)$ distribution. We state that theorem here for completeness.

Theorem A.2 *Suppose that* $\mathbf{X}$ *and* $\mathbf{Q}$ *are independent with* $N(\boldsymbol{\nu}, \Sigma)$ *and* $W(n, \Sigma)$ *distributions, respectively, where* Σ *has full rank* p *and* $n \geq p$. *Then* $\mathbf{X}\mathbf{Q}^{-1}\mathbf{X}'$ *has the* $G(p, n-p+1, 1, \boldsymbol{\nu}\Sigma^{-1}\boldsymbol{\nu}')$ *distribution.*

In the above theorem, when $\boldsymbol{\nu} = \mathbf{0}$, the G distribution centers at 0. Furthermore, according to the definition of G distribution given in Dempster (1969, p. 281), $G(r, s, \theta) \equiv \theta R/S$, where R and S follow χ^2_r and χ^2_s distributions respectively. Thus, $G(r, s, s/r)$ is equivalent to the $F(r, s)$ distribution.