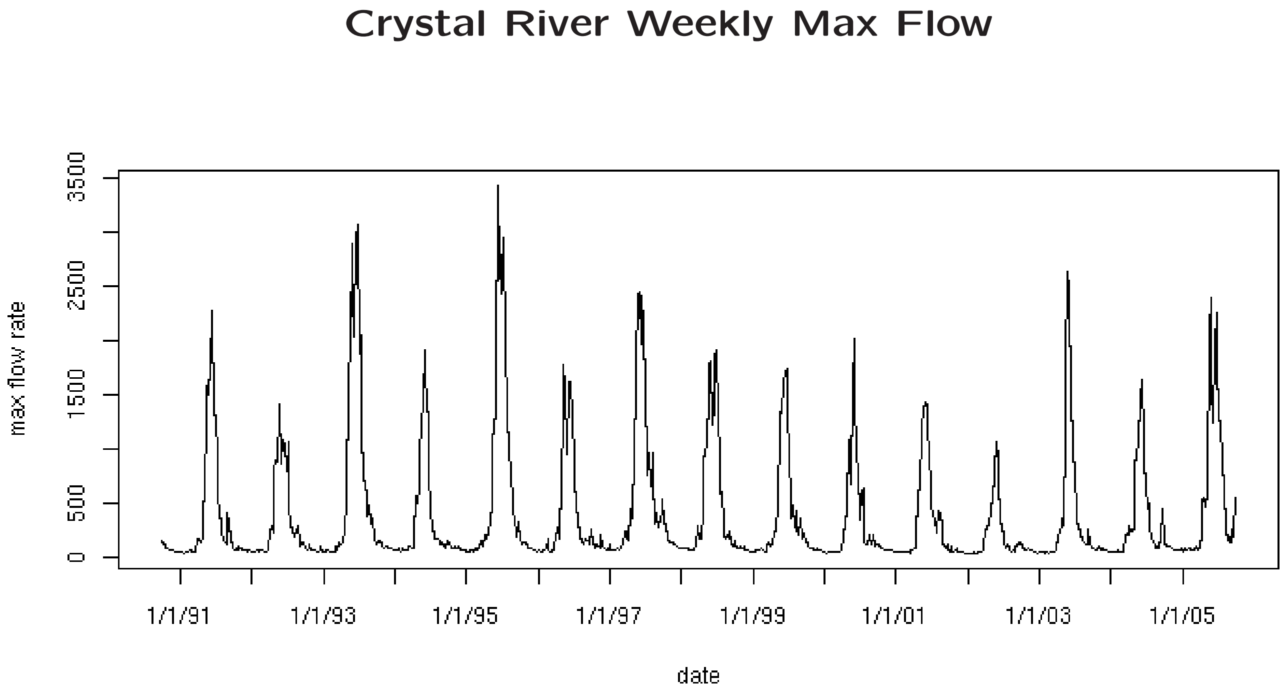
Estimating Dependence and Performing Prediction for Max-Stable Processes

Daniel Cooley
Colorado State University

Richard A. Davis
Columbia University

Philippe Naveau
LSCE, IPSL-CNRS, France

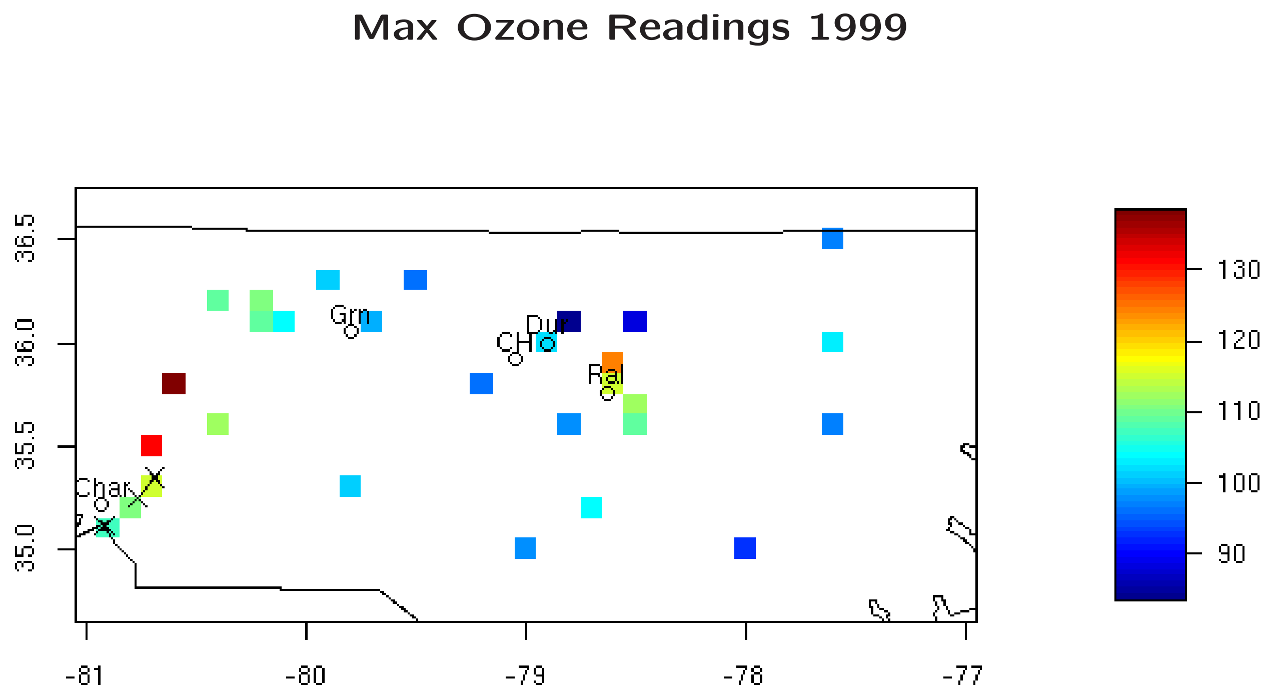
Motivating Example 1: Time Series Data



Q: How can estimate the temporal dependence between observations that are weekly maxima?

Q: Can we use the previous weekly maximum flows to predict the next weekly maximum flow?

Motivating Example 2: Spatial Data



Q: How can we estimate the spatial dependence in the data

Q: Can we use the observed annual maxima to predict (interpolate) the annual max at an unobserved location?

Outline

- Multivariate extremes background
- Measuring dependence
  - Madogram
- Performing Prediction
  - Models for MMSD's: Pairwise Beta
  - Approximating the conditional density

Multivariate Max-stablity

The foundation for extreme value theory is max-stability.

Let $\boldsymbol{Z}_k = [Z_{k,1}, \ldots, Z_{k,p}]$ be iid and define $\boldsymbol{M}_n = [\max_{k=1,\ldots,n} Z_{k,1}, \ldots, \max_{k=1,\ldots,n} Z_{k,p}]$.

- *Max stable*: $P\left(\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{z}\right) = P(\boldsymbol{Z} \leq \boldsymbol{z})$
- If $P\left(\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{z}\right) \rightarrow G(\boldsymbol{z})$, then G is max-stable

Characterizing the multivariate max-stable distributions is difficult. The task is made easier if a common marginal distribution is assumed. We assume that $\boldsymbol{Z}_k$ has unit Fréchet marginals:

$$P(Z_{k,i} \leq z) = \exp(-z^{-1}).$$

Point process representation

Consider the point process $N_n = \{\boldsymbol{Z}_k/n, k = 1, \dots n\}$.

Let $E = [0, \infty]^p \setminus \{\boldsymbol{0}\}$ and let $\mathcal{B} = \{\boldsymbol{w} \in E : \|\boldsymbol{w}\| = 1\}$. N_n converges to a Poisson process N with intensity measure

$$\Lambda\{\boldsymbol{z} : \|\boldsymbol{z}\| > r, \boldsymbol{z}/\|\boldsymbol{z}\| \in A\} = r^{-1} H(A)$$

where H is a finite measure on $\mathcal{B}$, $A \subset \mathcal{B}$ and Λ is a finite measure on E. H is termed the "angular" measure.

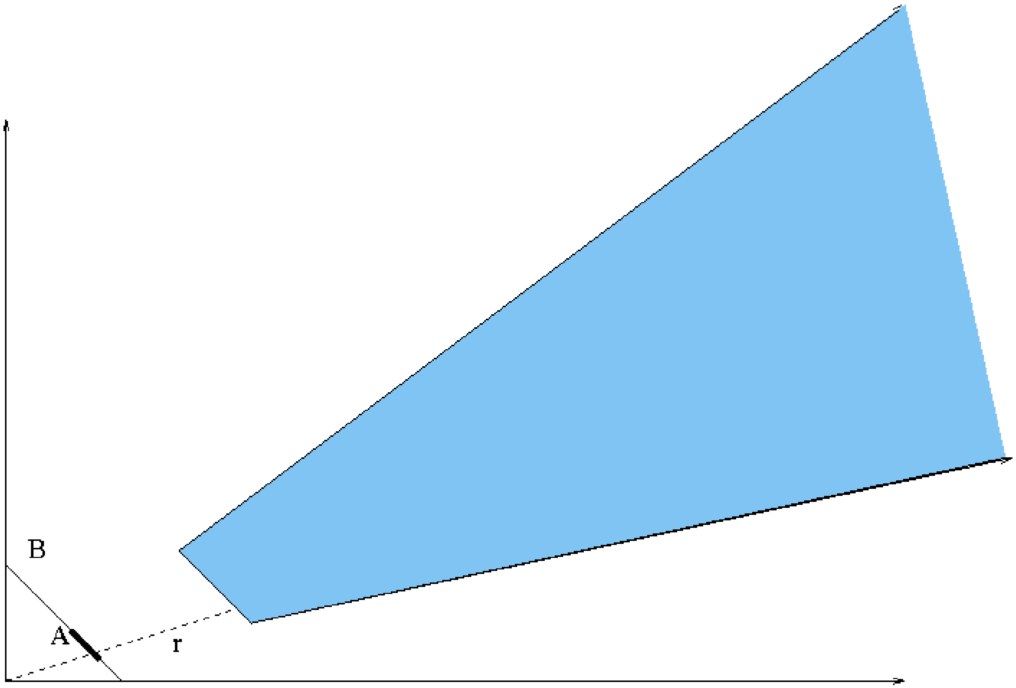
I will use the L_1 norm: $\|\boldsymbol{z}\| = |z_1| + \dots + |z_p|$.

Hence the "unit ball" $\mathcal{B}$ is unit simplex:

$$S_p = \{\boldsymbol{w} \in E : w_1 + \dots + w_p = 1\}.$$

Intensity Measure

$$\Lambda\{z : \|z\| > r, z/\|z\| \in A\} = r^{-1}H(A)$$

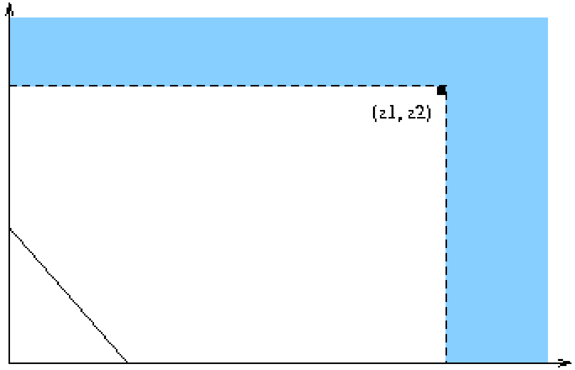


"Nice" sets aren't easily described by Cartesian coordinates which makes writing down the distribution function more difficult.

Multivariate Extreme Value Distributions

Exponent measure function relates intensity measure to Cartesian coordinates:

$$V(\boldsymbol{z}) = \Lambda\{(\boldsymbol{0}, \boldsymbol{z}]^c\}.$$



From this, multivariate max-stable distribution is defined:

$$G(\boldsymbol{z}) = \exp\left(-V(\boldsymbol{z})\right).$$

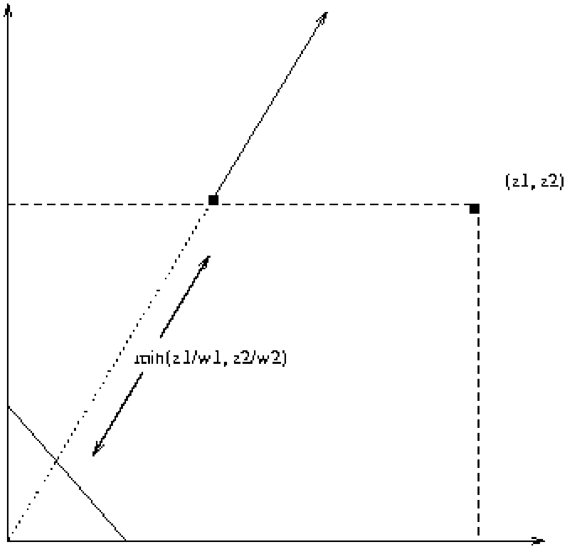
Dependence structure is completely determined by $V(\boldsymbol{z})$ which in turn is specified by $H(\boldsymbol{w})$.

$V(\boldsymbol{z})$ and $H(\boldsymbol{w})$

$$G(\boldsymbol{z}) = \exp(-V(\boldsymbol{z}))$$

$$V(\boldsymbol{z}) = \Lambda\{(0, \boldsymbol{z}]^c\}$$

$$\Lambda(dr \times dw) = r^{-2} dr dH(\boldsymbol{w})$$



$$V(\boldsymbol{z}) = \int_{S_p} \max_i \left(\frac{w_i}{z_i} \right) dH(\boldsymbol{w})$$

What do we know about H?

To have unit Fréchet margins, H must be such that

$$\int_{S_p} w_j dH(\boldsymbol{w}) = 1.$$

Consequently,

$$\int_{S_p} dH(\boldsymbol{w}) = p$$

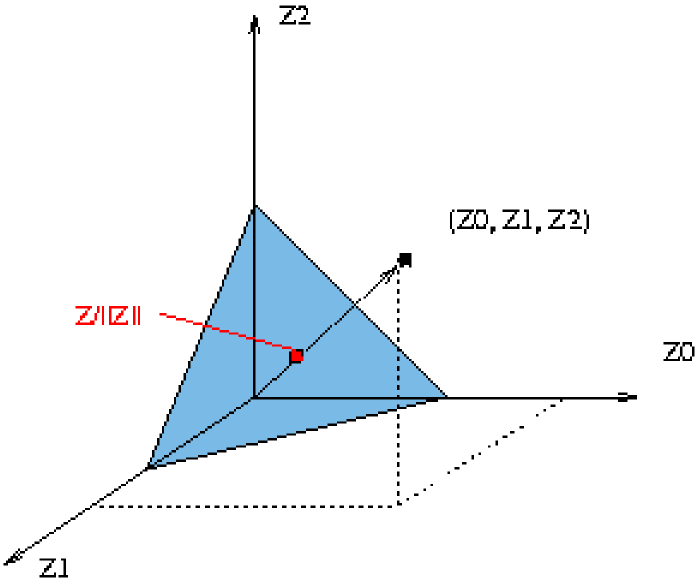
If H is differentiable, let $dH(\boldsymbol{w})$ be $h(\boldsymbol{w})$ – a angular “density”.

From the above conditions the center of mass of the density must be

$$(1/p, 1/p, \ldots, 1/p)$$

The more independent, the more mass is at the corners of the angular density.

So what should you know at this point?

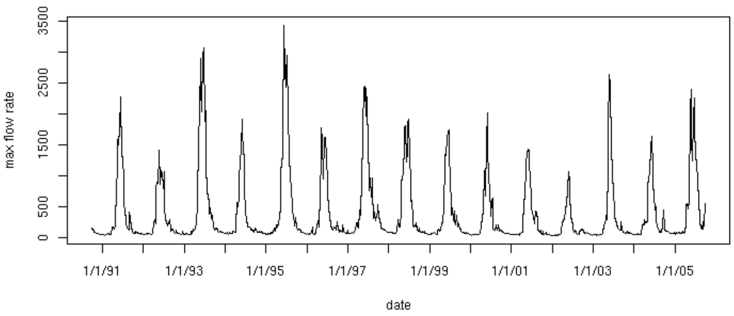


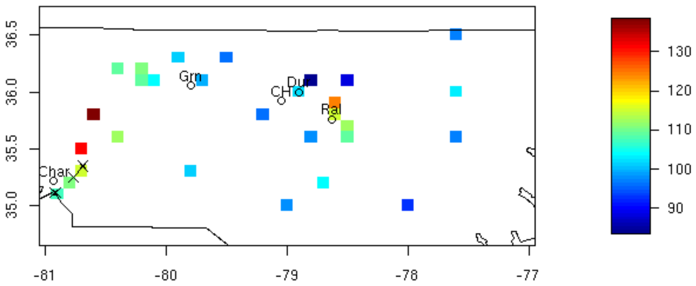
- $G(\boldsymbol{z}) = \exp[-V(\boldsymbol{z})] = \exp[-\Lambda\{[0,\boldsymbol{z}]^c\}]$
- $\Lambda(B)$ can be described in terms of a radius and a angular measure H which lives on the unit simplex
- $H(\boldsymbol{w})$ determines dependence; meets COM condition
- observed block (e.g., annual) maxima can be approximated by a multivariate max-stable distribution

Max Stable Processes

Max-stable processes are the infinite-dimensional analogue and take on values $Z(\boldsymbol{x})$, where $\boldsymbol{x}$ is some index.

Max-stable processes are the limiting distribution of the pointwise maxima of independent copies of a process.





We presume that we observe a finite number of locations of a process which is approximately max-stable.

Outline

- Multivariate extremes background
- Measuring dependence
  - Madogram
- Performing Prediction
  - Models for MMSD’s: Pairwise Beta
  - Approximating the conditional density

How is dependence measured for extremes?

To completely describe dependence one would have to describe the angular measure. But some metrics have been suggested.

Extremal Coefficient

$$\phi_{1,2} = V(1, 1, \infty, \ldots, \infty)$$

- $\phi_{i,j}$ defined analogously for $i, j \in 1, \ldots, p$
- $\mathbb{P}\left(\max(Z_i, Z_j) < z\right) = [\mathbb{P}(Z_i \leq z)]^{\phi_{i,j}}$
- $\phi_{i,j} \in [1, 2]$; $\phi_{i,j} = 2 \Leftrightarrow$ indep.; $\phi_{i,j} = 1 \Leftrightarrow$ complete dep.
- measures strength of dependence (like correlation), but does not completely characterize bivariate dependence
- has lots of 'cousins'
- estimators are complicated

Madogram

$$\nu(h) = \frac{1}{2}\mathbb{E}\left|F(Z(\boldsymbol{x}+h)) - F(Z(\boldsymbol{x}))\right|$$

- A new dependence measure for extremes.
- Applying F standardizes and allows calculation.
- A first-order variogram.
- Has a natural spatial interpretation.
- $\nu \in [0, 1/6]$
- Has a straightforward, natural estimator:

$$\widehat{\nu}(h) = \frac{1}{|\mathcal{N}_h|} \sum_{(x_i, x_j) \in \mathcal{N}_h} |F_m(Z(x_i)) - F_m(Z(x_j))|$$

where F_m is the empirical distribution function.

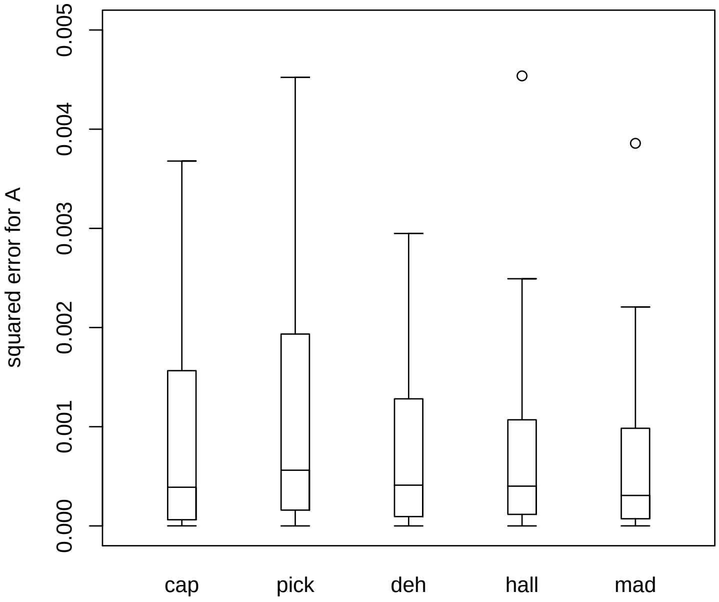
Why is the madogram well-suited for extremes?

$$\phi(h) = \tfrac{1+2\nu(h)}{1-2\nu(h)}$$

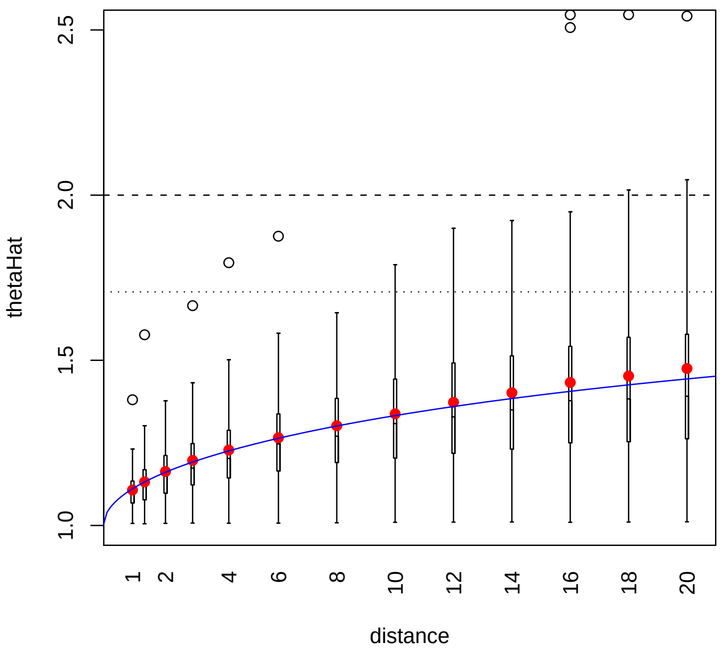
$$\begin{aligned}
\nu(h) &= \frac{1}{2} E\left|F(Z(x)) - F(Z(x+h))\right| \\
\nu(h) &= E\left[\max\left(F(Z(x), F(Z(x+h))\right)\right] - \frac{1}{2} E\left[F(Z(x)) + F(Z(x+h))\right] \\
&= \frac{\phi(h)}{1+\phi(h)} - \frac{1}{2} E\left[F(Z(x)) + F(Z(x+h))\right] \\
&= \frac{\phi(h)}{1+\phi(h)} - \frac{1}{2}
\end{aligned}$$

Madogram results

Bivariate Simulations



Spatial Simulations



Extension: λ-madogram

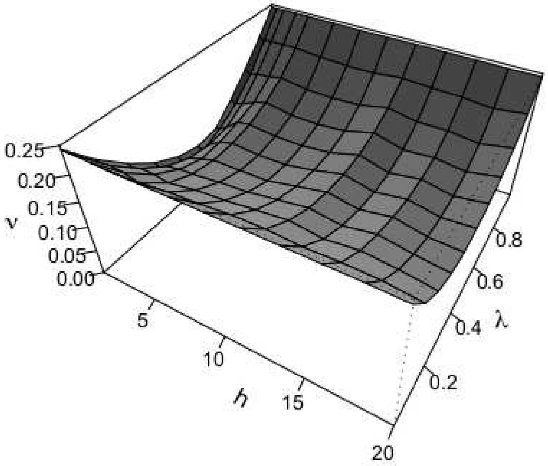
- $\nu(\lambda, h) = \frac{1}{2} E \left| F^{\lambda}(Z(x)) - F^{1-\lambda}(Z(x+h)) \right|; \quad \lambda \in (0,1)$
- extends idea of madogram to describe complete *pairwise* dependence function
- see: Naveau, Guillou, Cooley, Diebolt (2008)

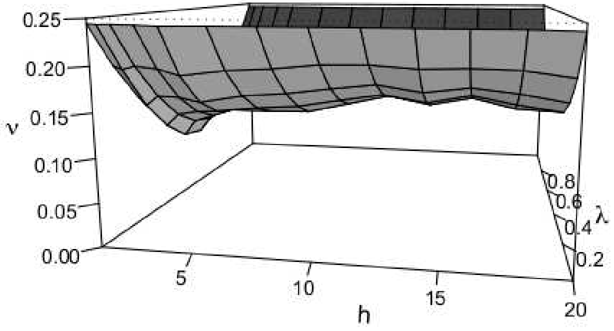
Estimation:

$$\begin{aligned}\tilde{\nu}(h, \lambda) = & \frac{1}{|\mathcal{N}_h|} \sum_{(x_i, x_j) \in \mathcal{N}_h} \left| F_m^{\lambda}(Z(x_i)) - F_m^{1-\lambda}(Z(x_j)) \right| \\ & - \frac{\lambda}{|\mathcal{N}_h|} \sum_{(x_i, x_j) \in \mathcal{N}_h} \left(1 - F_m^{\lambda}(Z(x_i))\right) \\ & - \frac{\lambda}{|\mathcal{N}_h|} \sum_{(x_i, x_j) \in \mathcal{N}_h} \left(1 - F_m^{1-\lambda}(Z(x_j))\right) \\ & + \frac{1 - \lambda + \lambda^2}{2(2-\lambda)(1+\lambda)}\end{aligned}$$

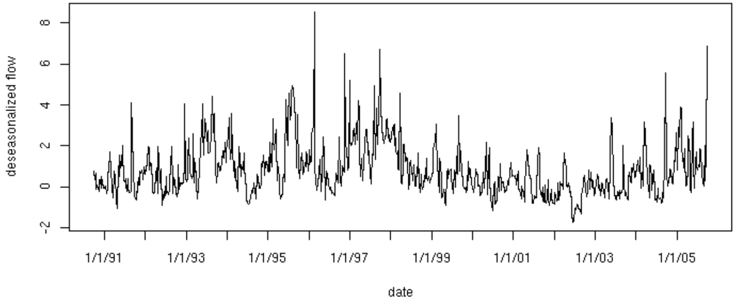
Extension: λ-madogram

- $\nu(\lambda, h) = \frac{1}{2} E \left| F^{\lambda}(Z(x)) - F^{1-\lambda}(Z(x+h)) \right|$
- extends idea of madogram to describe complete *pairwise* dependence function
- see: Naveau, Guillou, Cooley, Diebolt (2008)

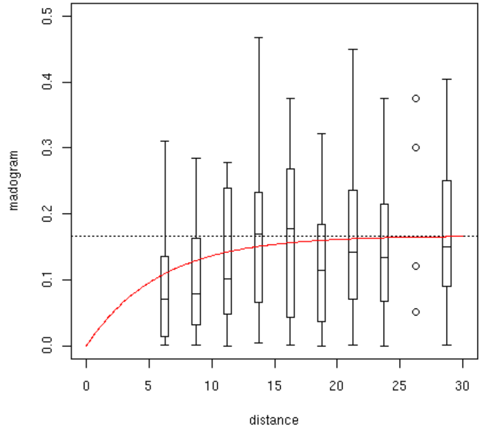




Estimating Dependence in Our Examples



- data deseasonalized
- madogram applied to lag 1 and lag 2 data
- $\phi_{\text{lag 1}} = 1.36$
  $\phi_{\text{lag 2}} = 1.49$



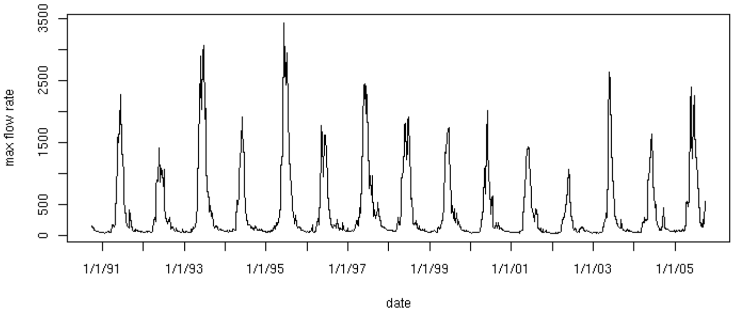
- marginals standardized
- madogram applied to annual maxima
- model fit by least squares
- $\nu(h) = \frac{1}{6}(1 - e^{-h/5.858})$

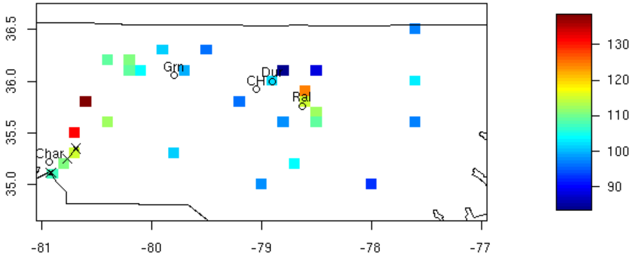
Outline

- Multivariate extremes background
- Measuring dependence
  - Madogram
- Performing Prediction *via a Conditional Density*
  - Models for MMSD's: Pairwise Beta
  - Approximating the conditional density

Conditional density for prediction?

Why not get a point estimate?





Q: What is a measure of risk (e.g., an estimated conditional 0.95 quantile) given that observed values are high?

Q: What is probability that an unmonitored location exceeds some standard?

Models for Multivariate MSD's

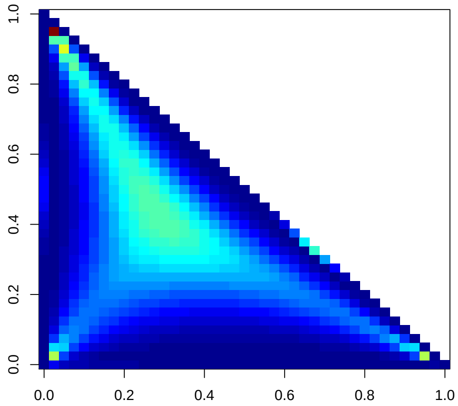
Cannot characterize the entire familiy of MMSD's parametrically, but a *few* parametric subfamilies have been suggested.

Exponent measure function $V(\boldsymbol{z})$	Angular density $h(\boldsymbol{w})$
• Logistic	• Dirichlet (Coles & Tawn, 91)
• Asymmetric Logistic (Tawn, 88)	
• Negative Logistic (Joe, 90)	

Parametric models for $V(\boldsymbol{z})$

Asymmetric Logistic:

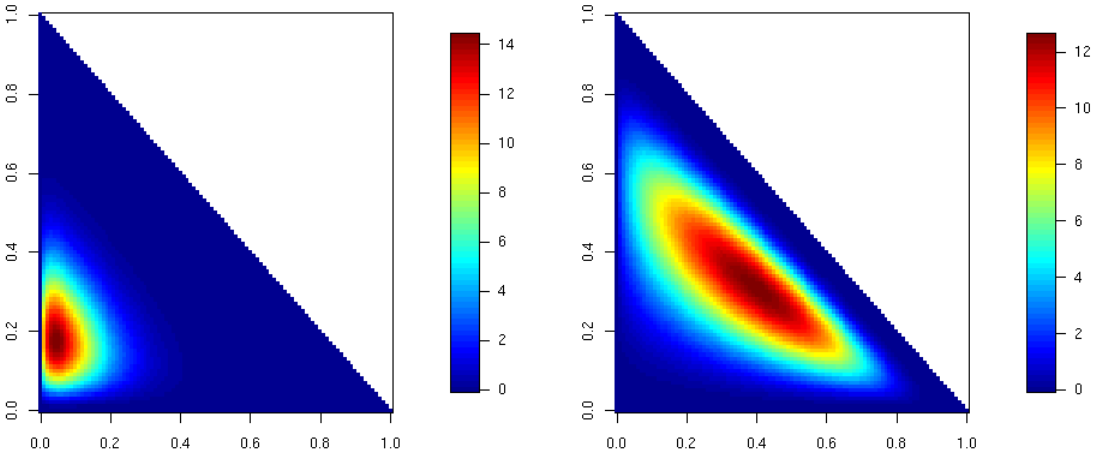
$$\begin{aligned} V_{\psi,\alpha}(z_1, z_2, z_3) \;=\; & \frac{\psi_{1,1}}{z_1} + \frac{\psi_{2,2}}{z_2} + \frac{\psi_{3,3}}{z_3} + \left[\left(\frac{\psi_{1,(1,2)}}{z_1}\right)^{1/\alpha_{1,2}} + \left(\frac{\psi_{2,(1,2)}}{z_2}\right)^{1/\alpha_{1,2}}\right]^{\alpha_{1,2}} + \\ & \left[\left(\frac{\psi_{1,(1,3)}}{z_1}\right)^{1/\alpha_{1,3}} + \left(\frac{\psi_{3,(1,3)}}{z_3}\right)^{1/\alpha_{1,3}}\right]^{\alpha_{1,3}} + \\ & \left[\left(\frac{\psi_{2,(2,3)}}{z_2}\right)^{1/\alpha_{2,3}} + \left(\frac{\psi_{3,(2,3)}}{z_3}\right)^{1/\alpha_{2,3}}\right]^{\alpha_{2,3}} + \\ & \left[\left(\frac{\psi_{1,(1,2,3)}}{z_1}\right)^{1/\alpha_{1,2,3}} + \left(\frac{\psi_{2,(1,2,3)}}{z_2}\right)^{1/\alpha_{1,2,3}} + \left(\frac{\psi_{3,(1,2,3)}}{z_3}\right)^{1/\alpha_{1,2,3}}\right]^{\alpha_{1,2,3}} \end{aligned}$$



Parametric models for $h(\boldsymbol{w})$

Only previous model is the *Dirichlet* (Coles and Tawn 1991).

$$
\begin{aligned}
h^*_\alpha(\boldsymbol{w}) &= \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} w_1^{\alpha_1-1} w_2^{\alpha_2-1} w_3^{\alpha_3-1} \\
h_\alpha(\boldsymbol{w}) &= \frac{\alpha_1\alpha_2\alpha_3\Gamma(\alpha_1+\alpha_2+\alpha_3+1)}{(\alpha_1 w_1+\alpha_2 w_2+\alpha_3 w_3)^4\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} \times \\
&\quad \left(\frac{\alpha_1 w_1}{\boldsymbol{\alpha}\cdot\boldsymbol{w}}\right)^{\alpha_1-1} \left(\frac{\alpha_2 w_2}{\boldsymbol{\alpha}\cdot\boldsymbol{w}}\right)^{\alpha_2-1} \left(\frac{\alpha_3 w_3}{\boldsymbol{\alpha}\cdot\boldsymbol{w}}\right)^{\alpha_3-1}
\end{aligned}
$$



$\boldsymbol{\alpha} = (1.5, 3, 10)$

Parametric models for $h(\boldsymbol{w})$

Only previous model is the *Dirichlet* (Coles and Tawn 1991).

$$
\begin{aligned}
h^*_\alpha(\boldsymbol{w}) &= \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} w_1^{\alpha_1-1} w_2^{\alpha_2-1} w_3^{\alpha_3-1} \\
h_\alpha(\boldsymbol{w}) &= \frac{\alpha_1\alpha_2\alpha_3\Gamma(\alpha_1+\alpha_2+\alpha_3+1)}{(\alpha_1 w_1+\alpha_2 w_2+\alpha_3 w_3)^4\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} \times \\
&\quad \left(\frac{\alpha_1 w_1}{\boldsymbol{\alpha}\cdot\boldsymbol{w}}\right)^{\alpha_1-1} \left(\frac{\alpha_2 w_2}{\boldsymbol{\alpha}\cdot\boldsymbol{w}}\right)^{\alpha_2-1} \left(\frac{\alpha_3 w_3}{\boldsymbol{\alpha}\cdot\boldsymbol{w}}\right)^{\alpha_3-1}
\end{aligned}
$$

Drawbacks:

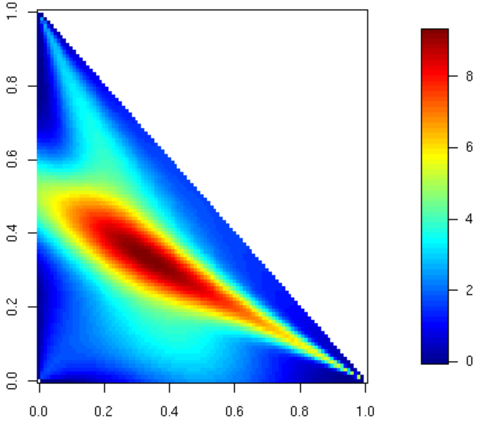
- Parameters lose intuitive meaning with transformation.
- Too few parameters to model pairwise dependencies.

Pairwise Beta Model

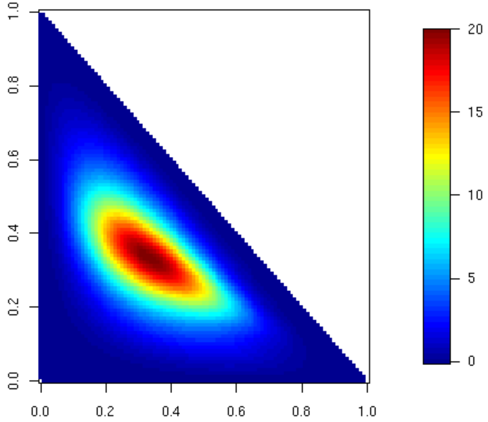
$$\begin{aligned}
h_p(\boldsymbol{w};\alpha,\boldsymbol{\beta}) &= K_p(\alpha)\sum_{i\neq j} h_{i,j}(w_i,w_j;\alpha,\beta_{i,j}), \text{ where} \\
h_{i,j}(w_i,w_j;\alpha,\beta_{i,j}) &= (w_i+w_j)^{(p-1)(\alpha-1)}(1-(w_i+w_j))^{\alpha-1}\times \\
&\quad \frac{\Gamma(2\beta_{i,j})}{(\Gamma(\beta_{i,j}))^2}\left(\frac{w_i}{w_i+w_j}\right)^{\beta_{i,j}-1}\left(\frac{w_j}{w_i+w_j}\right)^{\beta_{i,j}-1} \\
K_p(\alpha) &= \frac{2(p-3)!}{(p-1)\sqrt{p}}\frac{\Gamma(\alpha p)}{\Gamma(\alpha p-\alpha-p+3)\Gamma(\alpha+p-3)}
\end{aligned}$$

Advantages:

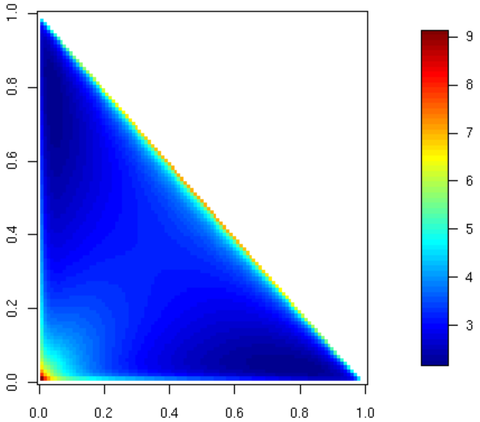
- no adjustment necessary to get center of mass condition
- parameters have some interpretation: α controls overall dependence, $\beta_{i,j}$'s control pairwise dependence
- largely specified by pairwise parameters



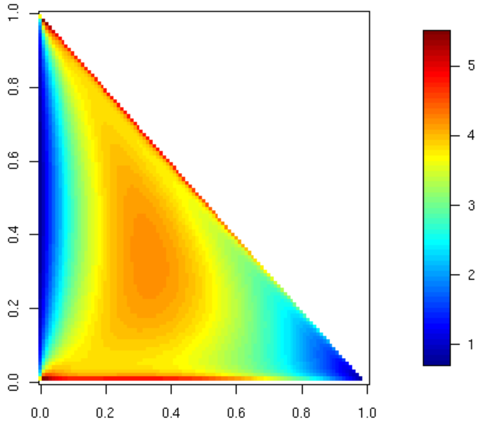
$\alpha = 1, \beta = (2, 4, 15)$



$\alpha = 4, \beta = (2, 4, 15)$



$\alpha = 1, \beta = (2, .5, .5)$

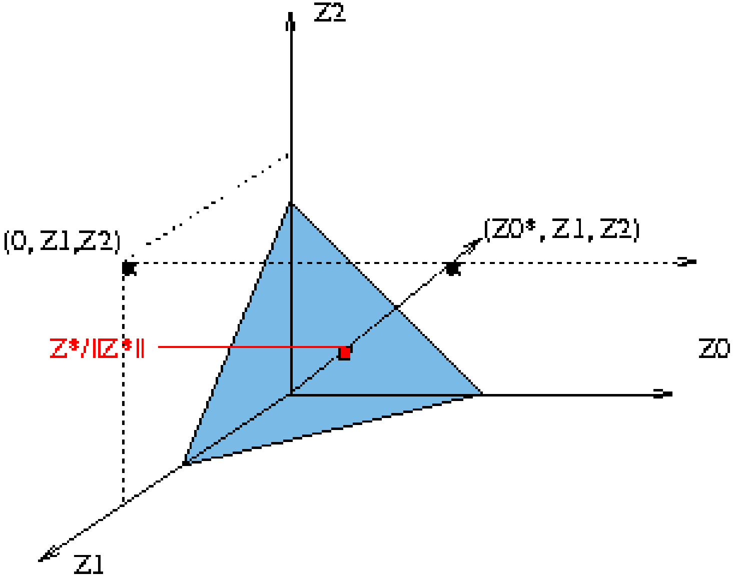


$\alpha = 1, \beta = (2, 2, .5)$

Models for Multivariate MSD's

Exponent measure function $V(\boldsymbol{z})$	angular density $h(\boldsymbol{w})$
• Logistic • Asymmetric Logistic (Tawn, 88) • Negative Logistic (Joe, 90)	• Dirichlet (Coles & Tawn, 91) • *Pairwise Beta* ∗ meets COM condition ∗ parameters interpretable ∗ pairwise specification
+ Can obtain $G(\boldsymbol{z})$ – Overparametrized? – Less flexible?	+ More flexibility? – Cannot directly get $G(\boldsymbol{z})$

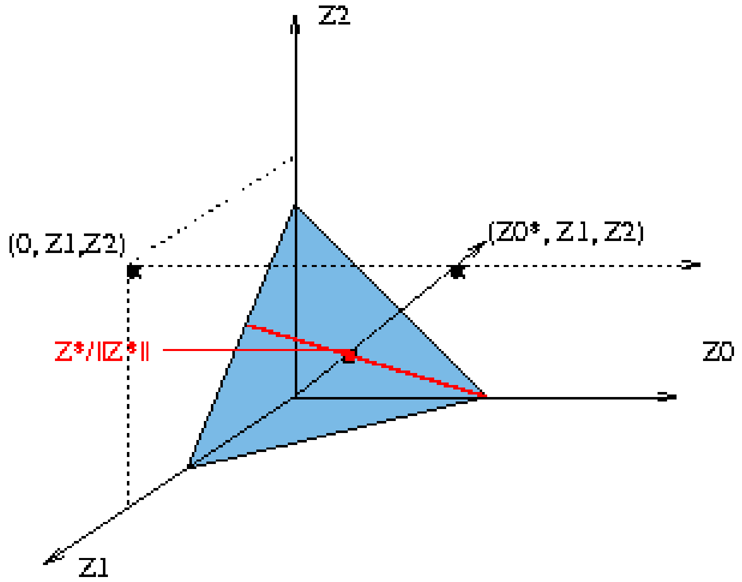
Approximating the conditional density?



If $V(\boldsymbol{z})$ is known and differentiable, then joint density can be obtained exactly. However, we are modeling $h(\boldsymbol{w})$.

Assume Z_1, Z_2 are observed and Z_0 is unobserved. Any predictor Z_0^* will yield a point $\boldsymbol{Z}^* = (Z_0^*, Z_1, Z_2)$ which can be mapped back to S_p as $\frac{\boldsymbol{Z}^*}{\|\boldsymbol{Z}^*\|_1}$.

Approximating the conditional density?



Given the radius is large, by knowing the values of the angular density at $\frac{Z^*}{\|Z^*\|_1}$ and the value of the “radius” $\|Z^*\|_1$, we can approximate the values of the joint density and in turn the *conditional density*.

Approximating the conditional density?

If $V(\boldsymbol{z}) = \mu\{(0, z]^c\}$ is small (i.e. the radius is large), then

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

Using Coles and Tawn (91) result to estimate the density at $\boldsymbol{z}$:

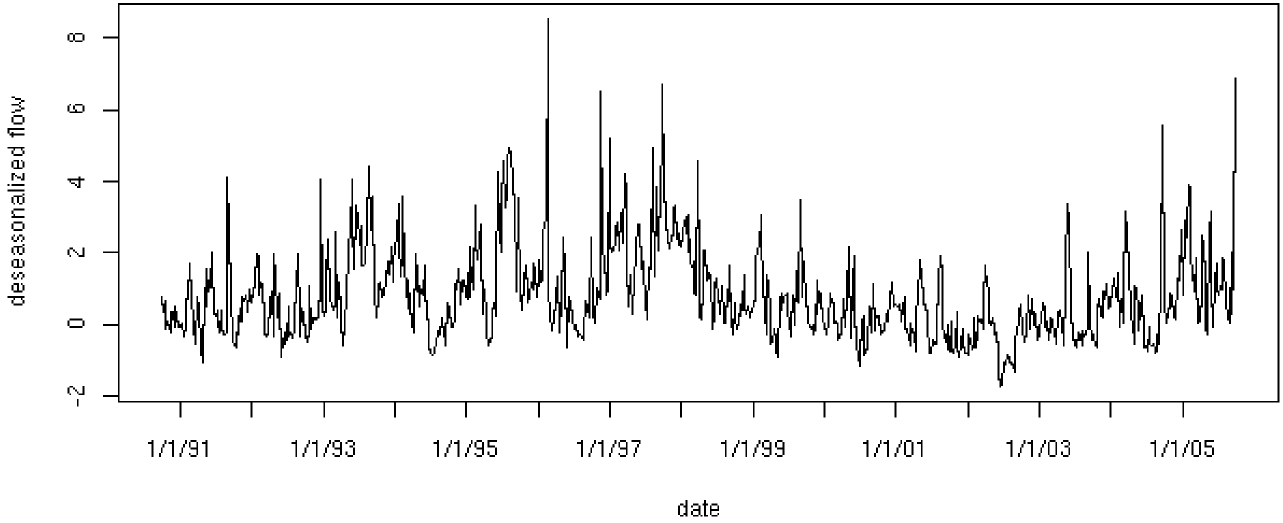
$$g(\boldsymbol{z}) \approx \frac{\partial}{\partial z_1, \ldots, \partial z_p}[1 - V(\boldsymbol{z})] = \frac{1}{||\boldsymbol{z}||^{-(p+1)}} h\left(\frac{\boldsymbol{z}}{||\boldsymbol{z}||}\right)$$

So conditional density can be approximated by

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

where $\boldsymbol{z}^* = (z_1, \ldots, z_{p-1}, \zeta)$.

Time Series Example



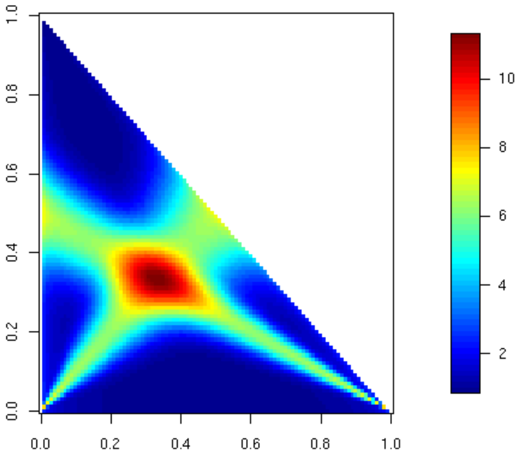
1. Data are deseasonalized.
2. ACF/PACF plots → base prediction on previous two observations.
3. GEV is fit to marginal distribution, then transformed.

Fitting the angular density model

Time series broken into non-overlapping triples. Dependence measured by the extremal coefficient (Schlather & Tawn 03).

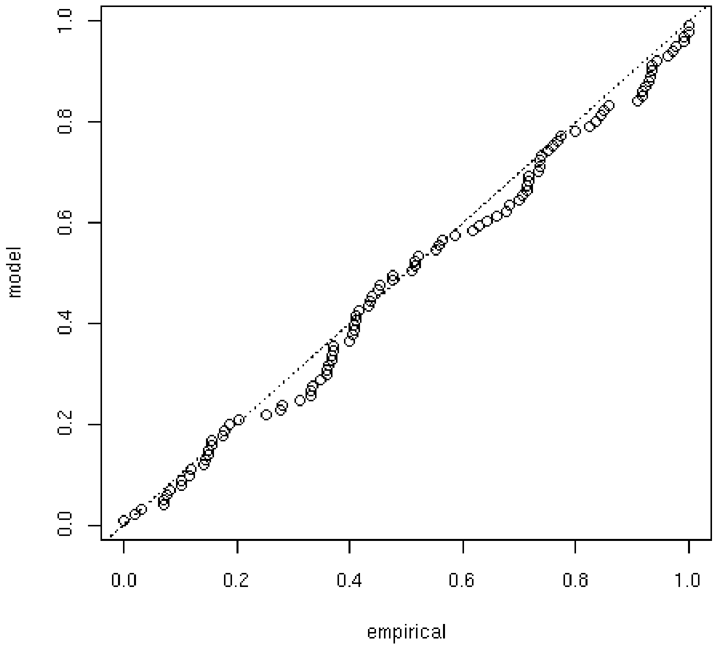
$$\phi_{1,2} = V(1, 1, \infty, \ldots, \infty); \ \phi_{i,j} \in [1, 2]$$

Ext. coefficients estimated at lags 1 and 2, $(\widehat{\phi}_{-1}, \widehat{\phi}_{-2}) = (1.36, 1.49)$, and pairwise beta model parameters found to match the extremal coefficient estimates. $(\widehat{\alpha}; \hat{\beta}) = (1, 16, 0.7, 16)$.



Time series prediction

75 largest triples selected for prediction. Conditional density of 3rd component given 1st and 2nd components is approximated.



Time series prediction

75 largest triples selected for prediction. Conditional density of 3rd component given 1st and 2nd components is approximated.

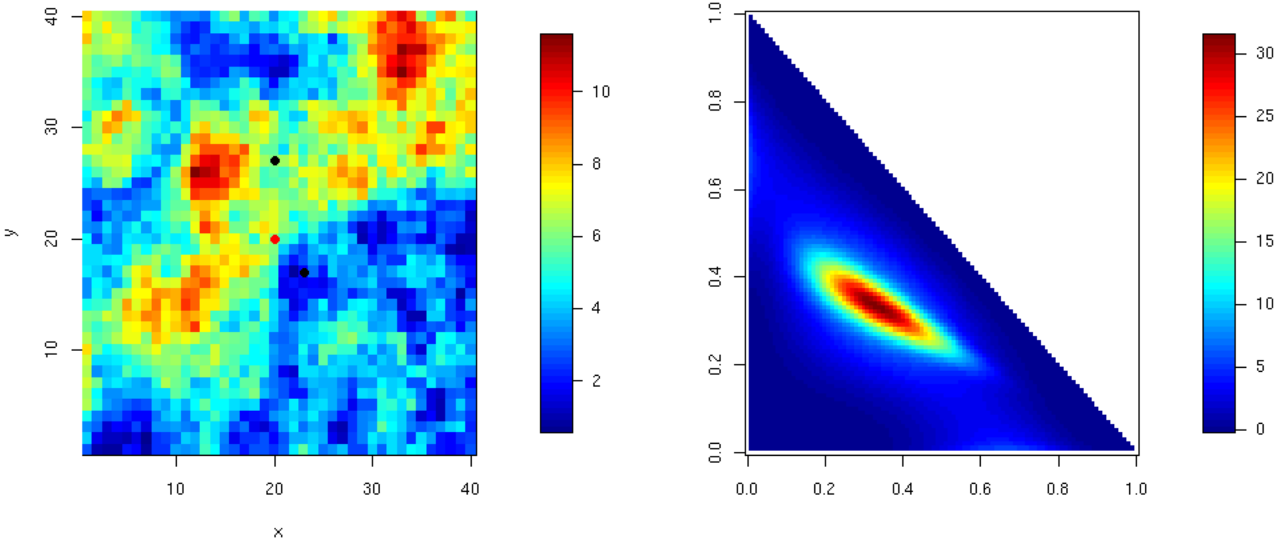
Assessing Risk

How well is the 95% quantile predicted?

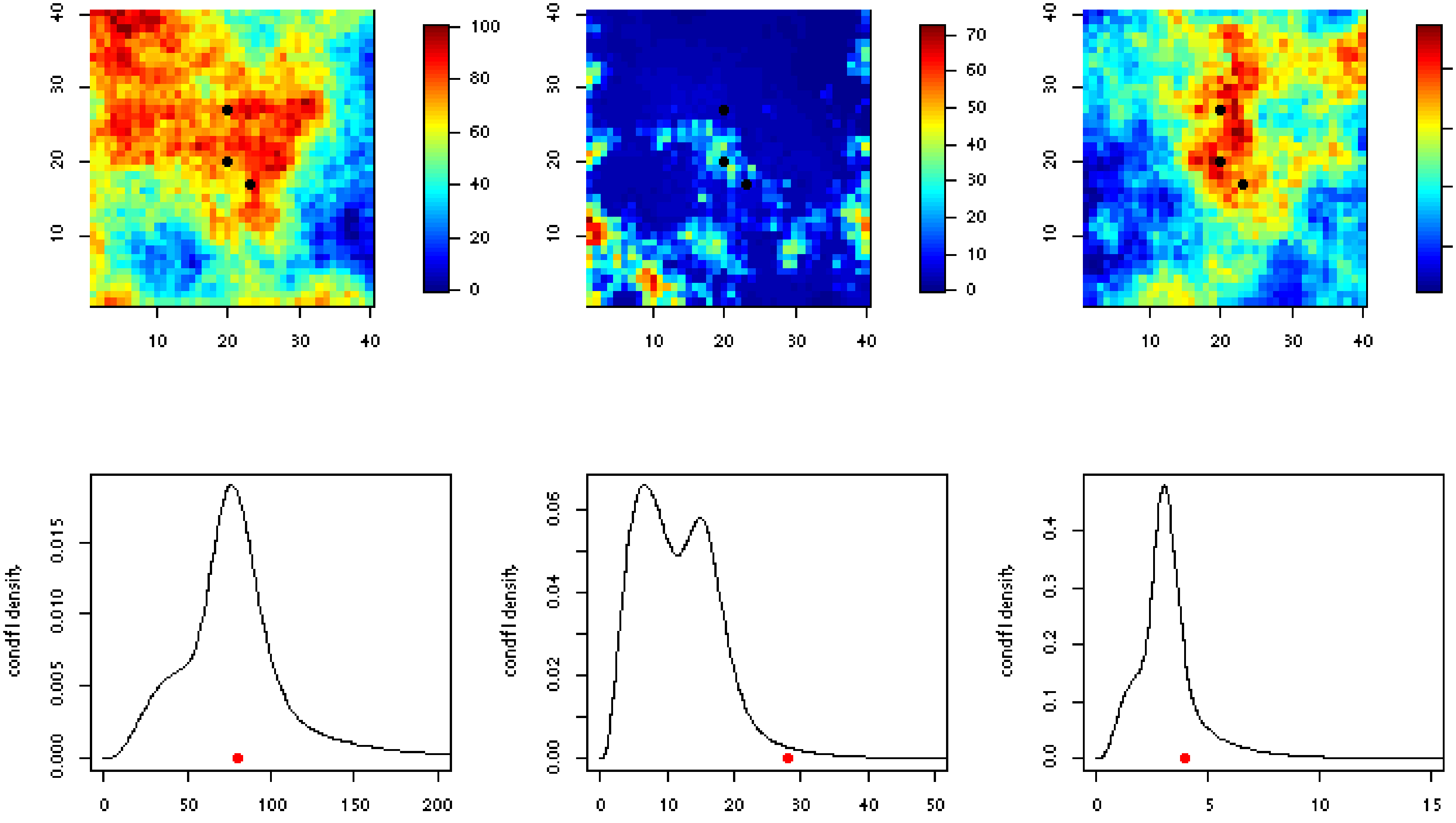
- Our method: 6 observations (8%) exceed the estimated quantile
- AR(2): 9 observations (12%) exceed the estimated quantile

Spatial interpolation

- MS fields w/ Fréchet marginals simulated (Schlather 02).
- Known bivariate dependence structure: $\phi = (1.34, 1.28, 1.22)$.
- Pairwise beta model fit as before: $(\alpha; \beta) = (4.3; 0.87, 4.4, 74)$.
- Conditional density approximated for largest 300 of 1000 simulated fields.



Spatial interpolation examples

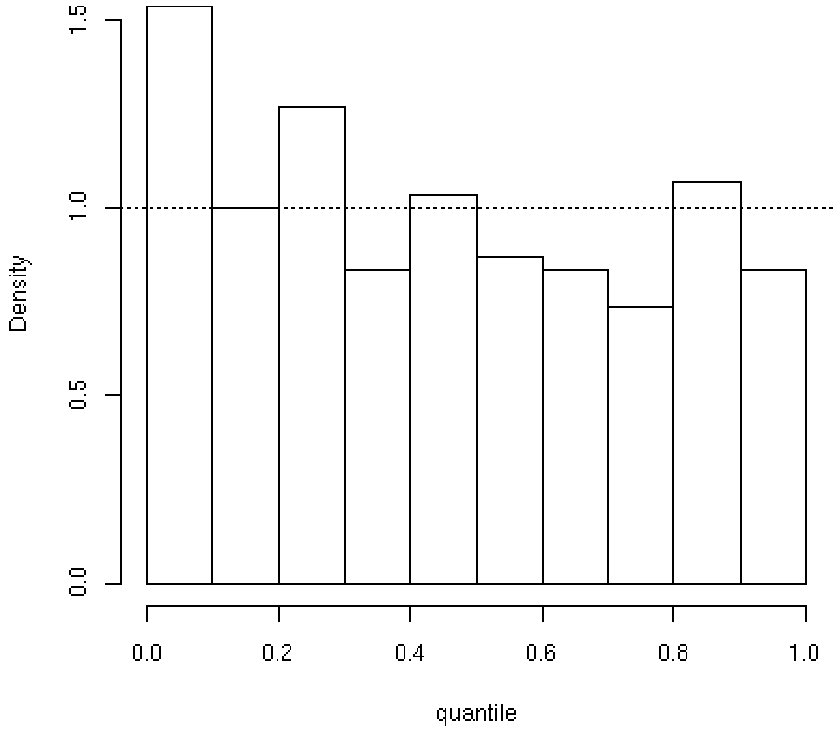


(80.69, 79.95, 80.45)

(4.50, 16.85, 28.14)

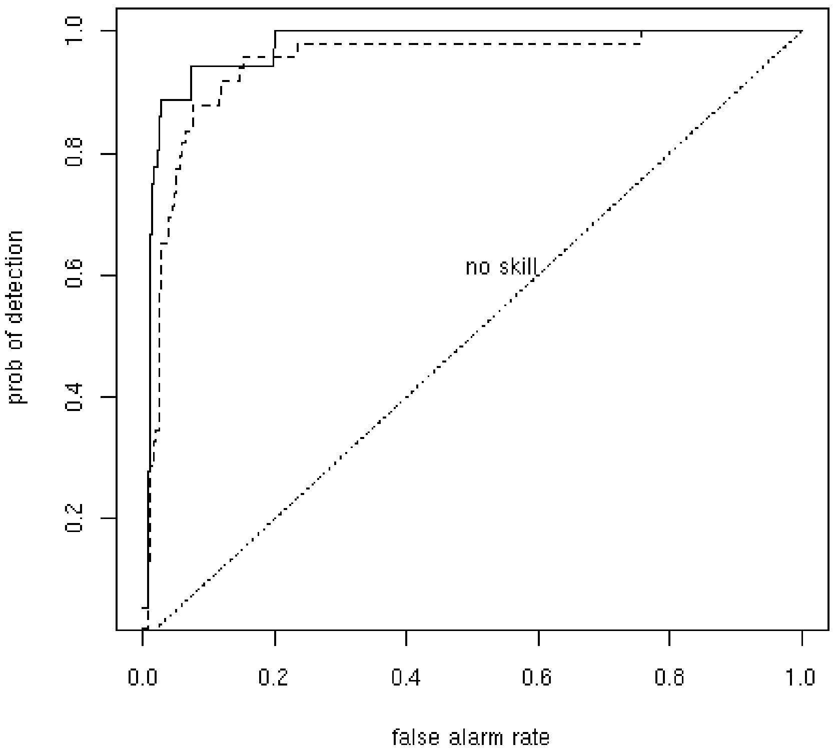
(3.47, 3.15, 3.93)

Repeated Simulation Results

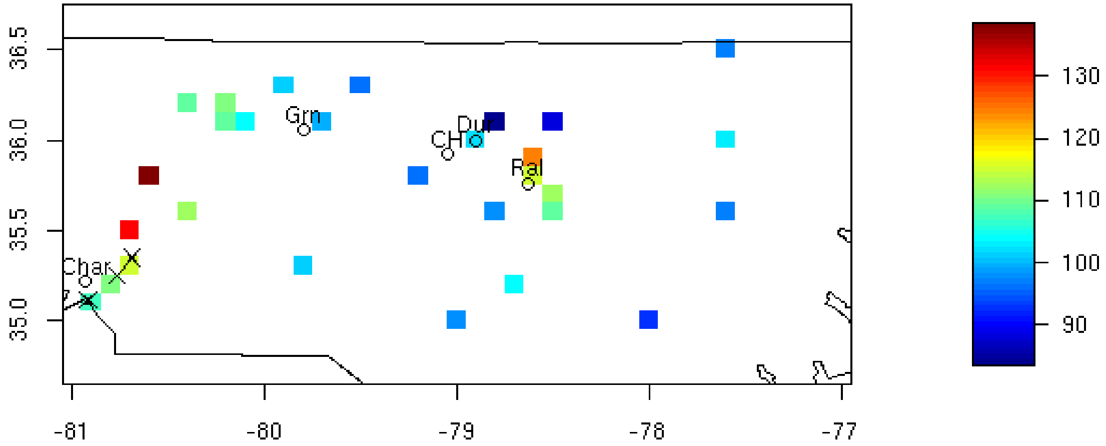


Spatial interpolation results

How well does the method assess exceeding some standard?



Ground level ozone

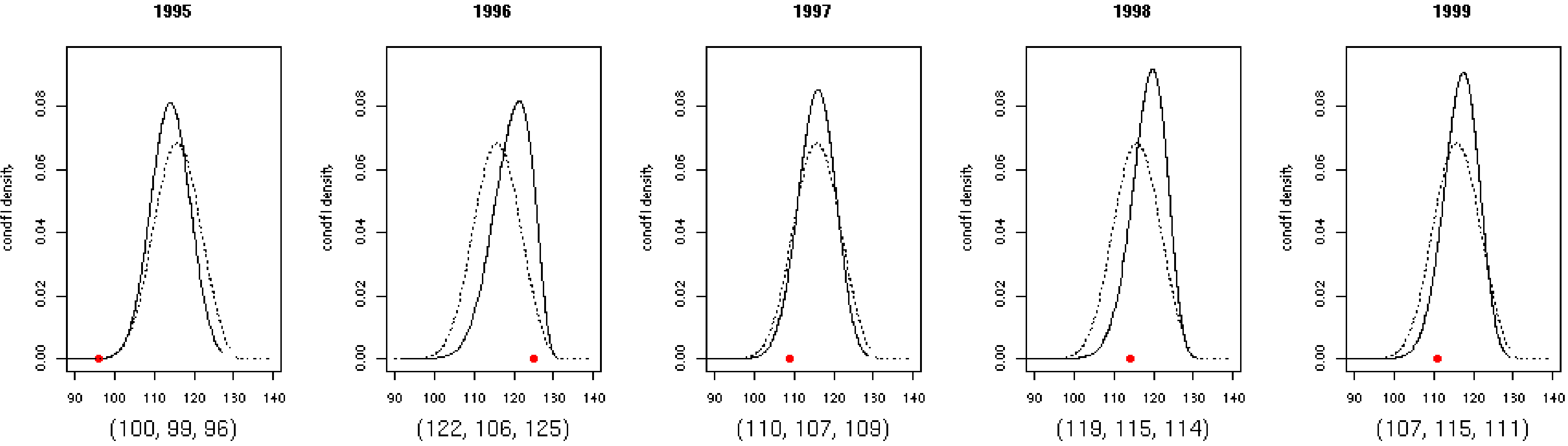


- Only 5 years of data.
- Marginal distributions from (Gilleland, et al 06).
- Dependence estimated as function of distance using madogram.
- Weak dependence estimated $\phi = (1.95, 1.84, 1.67)$.

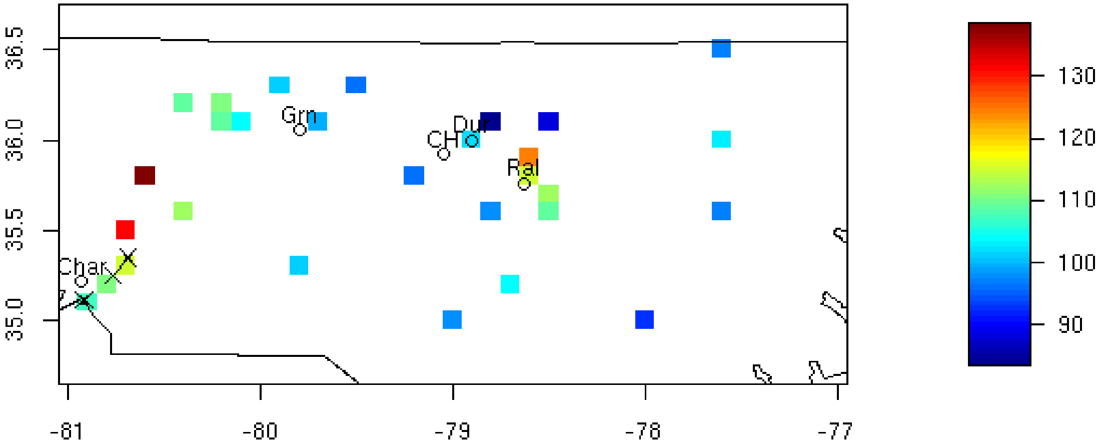
Ground level ozone

Unable to restrict attention to the "large" years. An adjustment is made to earlier approximations.

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



So what?



Is this an artificial problem?

Our hope is to be able to extend this work to the more practical problem of prediction for exceedances or high observations.

References

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.

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.

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

Naveau, P., Guillou, A., Cooley, D., and Diebolt, J. (2008). Modeling pairwise dependence of maxima in space. *Submitted*.

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

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

Why only three dimensions?

In extremes literature, parametric models are generally fit using maximum likelihood, but this technique cannot be used for the spatial prediction problem.

Instead, we have used a method-of-moments type approach, where parameters are found which match the estimated pairwise extremal coefficients.

If $V_\theta(z)$ known then find θ s.t.

$$\phi_{1,2} = V_\theta(1, 1, \infty, \ldots, \infty);$$

if only $h_\theta(\boldsymbol{w})$ is known then find θ s.t.

$$\phi_{i,j} = \int_{S_p} \max\left(w_i, w_j\right) h_\theta(\boldsymbol{w})$$