Spatial methods for extreme value analysis

Samuel A. Morris

North Carolina State University

January 9, 2015

Motivation

- Average behavior is important to understand, but it does not paint the whole picture
  - e.g. When constructing river levees, engineers need to be able to estimate a 100-year or 1000-year flood levels
  - e.g. Probability of ambient air pollution exceeding a certain threshold level
- Estimating the probability of rare events is challenging because these events are, by definition, rare
- Spatial extremes is promising because it borrows information across space
- Spatial extremes is also useful for estimating probability of extremes at sites without data

Motivation

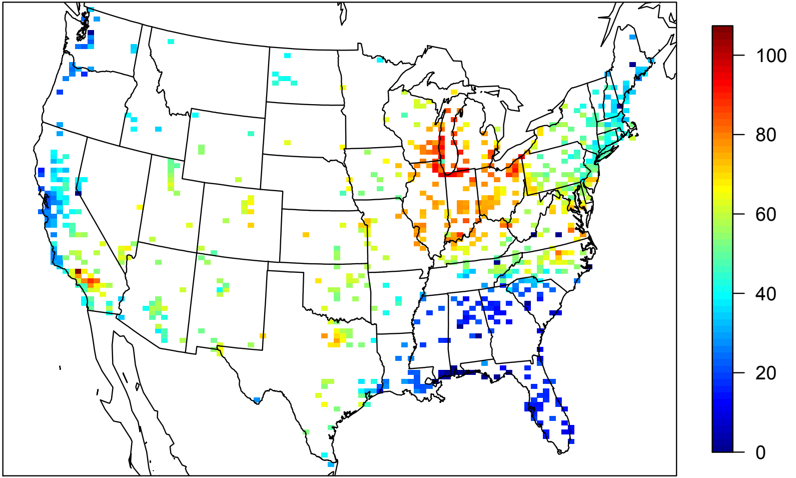

Figure: Max 8-hour ozone measurements on July 10, 2005

Motivation

Ozone compliance for Clean Air Act (EPA)

- Annual fourth-highest daily maximum 8-hour concentration, averaged over 3 years, not to exceed 75 ppb
- Annual fourth-highest is the 99th percentile for the year
- Common objectives are
  - To interpolate to unmonitored sites
  - Detect changes in extremes over time
  - Study meteorological conditions that lead to extreme events

- Key in extreme value analysis is to define extremes
- Typically done in one of two ways
    - Block maxima (red dots)
    - Values over threshold considered extreme

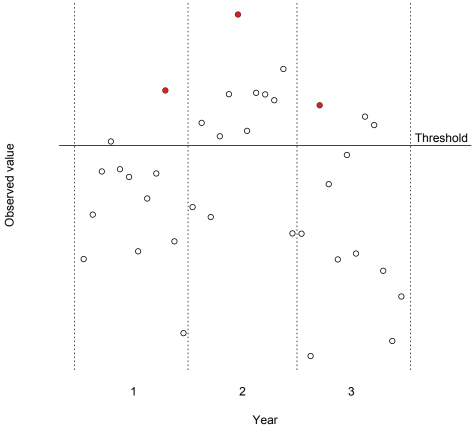

Figure: Hypothetical monthly data

Non-spatial analysis: Block maxima

Fisher-Tippett-Gnedenko

- Let $X_1, \ldots, X_n$ be i.i.d.
- Consider the block maximum $M_n = \max(X_1, \ldots, X_n)$
- If there exist normalizing sequences $a_n > 0$ and $b_n \in \mathcal{R}$ such that

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G(z)$$

  then $G(z)$ follows a generalized extreme value distribution (GEV) (Gnedenko, 1943)

- This motivates the use of the GEV for block maximum data

Non-spatial analysis: Block maxima

- GEV distribution

$$G(y) = \Pr(Y < y) = \begin{cases} \exp\left\{-\left[1+\xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} & \xi \neq 0 \\ \exp\left\{-\exp\left(-\frac{y-\mu}{\sigma}\right)\right\} & \xi = 0 \end{cases}$$

  where
  - $\mu \in \mathcal{R}$ is a location parameter
  - $\sigma > 0$ is a scale parameter
  - $\xi \in \mathcal{R}$ is a shape parameter
    - Unbounded above if $\xi \geq 0$
    - Bounded above by $(\mu - \sigma)/\xi$ when $\xi < 0$

- Challenges:
  - Lose information by only considering maximum in a block
  - Underlying data may not be i.i.d.

Non-spatial analysis: Peaks over threshold

Pickands-Balkema-de Haan theorem

- Let $X_1, \ldots, X_n \overset{iid}{\sim} F$
- If there exist normalizing sequences $a_T > 0$ and $b_T \in \mathcal{R}$ such that for any $x \geq 0$, as $T \to \infty$

$$\Pr\left(\frac{X - b_T}{a_T} > x \mid X > T\right) \xrightarrow{d} H(x),$$

where T is a thresholding value, then $H(x)$ follows a generalized Pareto distribution (GPD) (Balkema and de Haan, 1974)

Non-spatial analysis: Peaks over threshold

Select a threshold, T, and use the GPD to model the exceedances

$$H(y) = P(Y < y) = \begin{cases} 1 - \left[1 - \xi\left(\frac{y-T}{\sigma}\right)\right]^{-1/\xi} & \xi \neq 0 \\ 1 - \exp\left\{\frac{y-T}{\sigma}\right\} & \xi = 0 \end{cases}$$

where

- $\sigma > 0$ is a scale parameter
- $\xi \in \mathcal{R}$ is a shape parameter
  - Unbounded above if $\xi \geq 0$
  - Bounded above by $(T - \sigma)/\xi$ when $\xi < 0$

Non-spatial analysis: Peaks over threshold

- The GPD is related to GEV distribution through

$$H(y) = 1 + \log[G(y)], \quad y \geq T$$

- Challenges:
    - Sensitive to threshold selection
    - Temporal dependence between observations (e.g. flood levels don't dissipate overnight)

Max-stable processes for spatial data

- Consider i.i.d. spatial processes $x_j(\mathbf{s})$, $j = 1, \ldots, J$
- Let $M_J(\mathbf{s}) = \bigvee_{j=1}^{J} x_j(\mathbf{s}_i)$ be the block maximum at site $\mathbf{s}$
- If there exists normalizing sequences $a_J(\mathbf{s})$ and $b_J(\mathbf{s})$ such that for all sites, $\mathbf{s}_i$, $i = 1, \ldots, d$,

$$\frac{M_J(\mathbf{s}) - b_J(\mathbf{s})}{a_J(\mathbf{s})} \xrightarrow{d} G(\mathbf{s})$$

then $G(\mathbf{s})$ is a max-stable process (Smith, 1990)

- Therefore, max-stable processes are the standard model for block maxima

Multivariate representations

- Marginally at each site, observations follow a GEV distribution
- For a finite collection of sites the multivariate representation for the GEV (mGEV) is

$$\Pr(\mathbf{Z} \leq \mathbf{z}) = G^*(\mathbf{z}) = \exp[-V(\mathbf{z})]$$

$$V(\mathbf{z}) = d \int_{\Delta_d} \bigvee_{i=1}^{d} \frac{w_i}{z_i} H(\mathrm{d}w)$$

where

- $V(\mathbf{z})$ is called the exponent measure
- $\Delta_d = \{\mathbf{w} \in \mathcal{R}_+^d \mid w_1 + \cdots + w_d = 1\}$
- H is a probability measure on Δ_d
- $\int_{\Delta_d} w_i H(\mathrm{d}w) = 1/d$ for $i = 1, \ldots, d$

Multivariate GEV challenges

- Only a few closed-form expressions for $V(\mathbf{z})$ exist
- Two common forms for $V(\mathbf{z})$
  - Symmetric logistic (Gumbel, 1960)

$$V(\mathbf{z}) = \left[\sum_{i=1}^{n} \left(\frac{1}{z_i}\right)^{1/\alpha}\right]^{\alpha}$$

  - Asymmetric logistic (Coles and Tawn, 1991)

$$V(\mathbf{z}) = \sum_{l=1}^{L} \left[\sum_{i=1}^{n} \left(\frac{w_{il}}{z_i}\right)^{1/\alpha_l}\right]^{\alpha_l}$$

where $w_{il} \in [0,1]$ and $\sum_{l=1}^{L} w_{il} = 1$

Multivariate peaks over threshold

- Few existing methods
- Often use max-stable methods due to the relationship between GEV and GPD
- Joint distribution function given by Falk et al. (2011)

$$H(\mathbf{z}) = 1 - V(\mathbf{z})$$

where $V(\mathbf{z})$ is defined as in the GEV

Extremal dependence: χ statistic

- Correlation is the most common measure of dependence
  - Focuses on the center and not tails
  - This makes it irrelevant for extreme value analysis
- Extreme value analysis focuses on the χ statistic (Coles et al., 1999), a measure of extremal dependence given by

$$\chi(h) = \lim_{c \to \infty} \Pr[Y(\mathbf{s}) > c \mid Y(\mathbf{t}) > c]$$

  where $h = ||\mathbf{s} - \mathbf{t}||$
- If $\chi(h) = 0$, then observations are asymptotically independent at distance h

Existing challenges

- Multivariate max-stable and GPD models have nice features, but they are
    - Computationally challenging (e.g, the asymmetric logistic has $2^{n-1}(n+2) - (2n+1)$ free parameters)
    - Joint density only available in low dimensions
- Some recent approaches
    - Bayesian hierarchical model (Reich and Shaby, 2012)
    - Pairwise likelihood approach (Huser and Davison, 2014)
- Many opportunities to explore new methods

Three principal contributions

1. A spatio-temporal model with flexible tails, asymptotic spatial dependence, and computation on the order of Gaussian models for large space-time datasets

2. Predicting rare binary events with a spatially dependent generalized extreme value link function

3. A Bayesian hierarchical model to allow for non-stationary dependence in extreme value models

Spatiotemporal modeling for extreme values

Model objectives:

- Marginal distribution at each site with a flexible tail
  - Allow for asymmetric distributions
  - Allow for heavy tails
- Asymptotic spatial dependence
- Computation on the order of Gaussian models for large space-time datasets

Gaussian spatial model

- In geostatistics, $Y(\mathbf{s})$ are often modeled using a Gaussian process with mean function $\mu(\mathbf{s})$ and covariance function $\rho(h)$
- Model properties
  - Nice computing properties (closed-form likelihood)
  - For a Gaussian spatial model $\chi(h) = 0$ regardless of the strength of the correlation in the bulk of the distribution
  - Tail is not flexible
    - Light-tailed
    - Symmetric

Spatial skew-t distribution

A more flexible alternative is the spatial skew-t process (Zhang and El-Shaarawi, 2012)

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + \lambda|z| + v(\mathbf{s})$$

where

- $\lambda \in \mathcal{R}$ controls the skewness
- $z \sim N(0, \sigma^2)$ is a random effect
- $v(\mathbf{s})$ is a Gaussian process with variance σ^2, Matérn correlation, and γ is the proportion of spatial variation in the Matérn correlation
- $\sigma^2 \sim \text{IG}(a, b)$

Spatial skew-t distribution

- Conditioned on z and σ^2, $Y(\mathbf{s})$ is a Gaussian spatial model
- Standard geostatistical methods apply
- Predictions can be made through Kriging

Spatial skew-t distribution

Marginalizing over z and σ^2 (via MCMC),

$$Y(\mathbf{s}) \sim \text{skew-t}(\mathbf{X}(\mathbf{s})\boldsymbol{\beta}, \boldsymbol{\Omega}, \alpha, \text{df} = 2a)$$

where

- $\mathbf{X}(\mathbf{s})\boldsymbol{\beta}$ is the location
- $\boldsymbol{\Omega} = \frac{1}{ab}\bar{\boldsymbol{\Omega}}$ is a correlation matrix
- $\bar{\boldsymbol{\Omega}} = (\boldsymbol{\Sigma} + \lambda^2 \mathbf{1}\mathbf{1}^T)$
- $\boldsymbol{\Sigma}$ is a positive definite correlation matrix
- $\alpha = \lambda(1 + \lambda^2 \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})^{-1/2} \mathbf{1}^T \boldsymbol{\Sigma}^{-1}$ controls the skewness

$\chi(h)$ plot

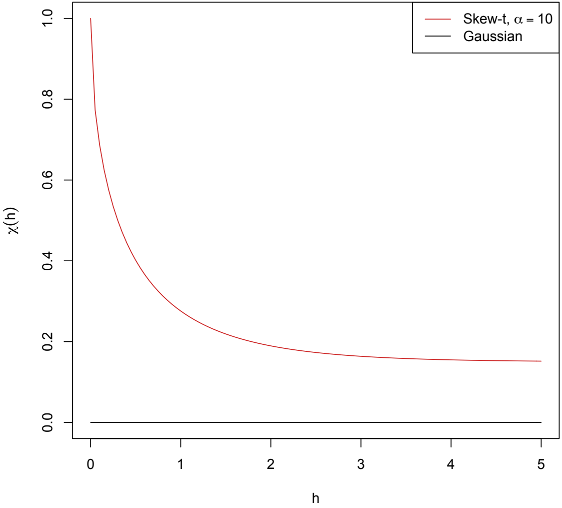

Figure: χ plot for skew-t, and Gaussian

Spatial skew-t distribution properties

- Model properties:
  - Flexible tail
    - Skewness controlled by λ
    - Weight of tails controlled by degrees of freedom $2a$
  - Computation that is on the order of Gaussian computation
- Challenge: For a skew-t distribution $\lim_{h \to \infty} \chi(h) > 0$ (Padoan, 2011)
  - Long-range dependence occurs because all observations (near and far) share the same z and σ^2

Extension of the skew-t distribution

- Skew-t distribution addresses two modeling concerns
  - Extremal dependence
  - Reasonable computing
- Our contribution is to extend the skew-t
  - Censoring to focus on extreme observations
  - Partitioning to address long-range dependence

Censoring data to focus on tail behavior

- We censor the observed data at a high threshold T
- Censored data

$$\tilde{Y}_t(\mathbf{s}) = \begin{cases} Y_t(\mathbf{s}) & \delta(\mathbf{s}) = 1 \\ T & \delta(\mathbf{s}) = 0 \end{cases}$$

where $\delta(\mathbf{s}) = I[Y(\mathbf{s}) > T]$

- Allows tails of the distribution to speak for themselves

Random partition

- Daily random partition allows z and σ^2 to vary by site

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + \lambda z(\mathbf{s}) + \sigma(\mathbf{s})v(\mathbf{s})$$

- Consider a set of knots $\mathbf{w}_k \sim$ Uniform that define a random partition $P_1, \ldots, P_K$ such that

$$P_k = \{s : k = \arg\min_\ell ||\mathbf{s} - \mathbf{w}_\ell||\}$$

where $\mathbf{w} = (w_1, w_2)$ (similar to Kim et al., 2005 for non-extreme modeling)

- For $\mathbf{s} \in P_k$

$$z(\mathbf{s}) = z_k$$
$$\sigma^2(\mathbf{s}) = \sigma^2_k$$

Example partition

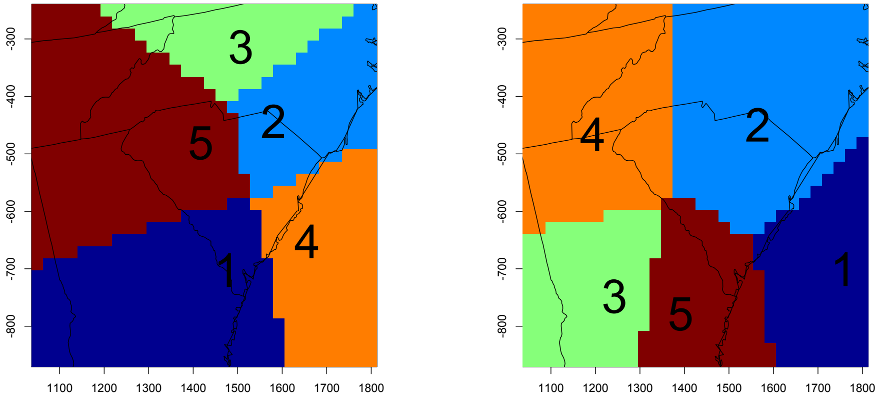

Figure: Two sample partitions (number is at partition center)

Random partition

- Within each partition, $Y(\mathbf{s})$ has the same MV skew-t distribution as before
- Across partitions $Y(\mathbf{s})$ are asymptotically independent, but still correlated through $v(\mathbf{s})$
- New expression for $\chi(h)$

$$\chi(h) = \pi(h)\chi_{\text{skew-}t}(h)$$

where $\pi(h)$ is the probability two sites are in the same partition

Proof that $\lim_{h \to \infty} \chi(h) = 0$

- Let A be the area between two sites $\mathbf{s}$ and $\mathbf{t}$
- Let $N(A)$ be the number of knots in A
- Assume that $N(A) \sim \text{HPP}[\mu(A)]$, where
  - $\text{HPP}[\mu(A)]$ is a homogeneous Poisson process with intensity measure $\mu(A)$ defined on A
  - $\lim_{h \to \infty} \mu(A) = \infty$

$$\Pr[N(A) = k] = \frac{\mu(A)^k \exp\{-\mu(A)\}}{k!}$$

Proof that $\lim_{h\to\infty} \chi(h) = 0$

- For finite k, $\lim_{h\to\infty} P[N(A) = k] = 0$
- As $N(A)$ increases, $\pi(h)$ decreases because partition is defined by closest knots, so

$$\lim_{h\to\infty} \chi(h) = \lim_{h\to\infty} \pi(h)\chi_{\text{skew-}t}(h) = 0$$

$\chi(h)$ plot

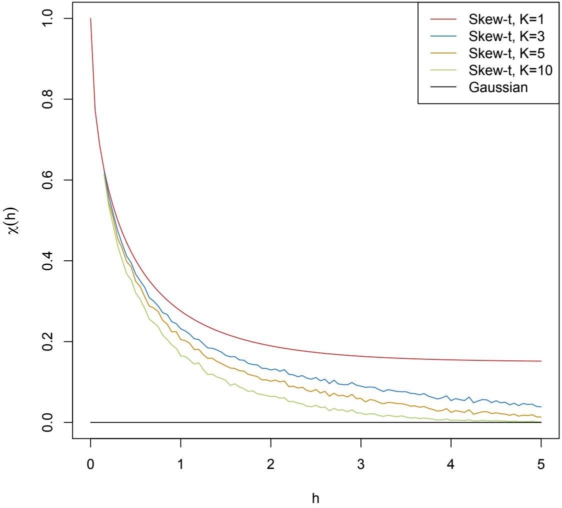

Figure: χ plot for different data settings

Random partition skew-t model

This new model is called a random partition skew-t model, and it has all the properties we desire

- Marginal distribution with flexible tails
  - λ term allows for asymmetry
  - Degrees of freedom control heavy vs light tails
- Asymptotic spatial dependence for that decays with distance between sites through partitioning
- Computation is on the order of Gaussian models for large space-time datasets

MCMC details

- Three main steps
  1. Impute censored data below T
  2. Update parameters with Metropolis Hastings or Gibbs sampling
  3. Make spatial predictions
- Priors are selected to be conjugate when possible

Simulation study

6 different data settings

1. Gaussian, $K = 1$ partition
2. Symmetric-t, $K = 1$ partition
3. Symmetric-t, $K = 5$ partitions
4. Skew-t, $K = 1$ partition
5. Skew-t, $K = 5$ partitions
6. Max-stable
    - Marginally: GEV($\mu = 1, \sigma = 1, \xi = 0.2$)
    - Dependence function: asymmetric logistic with $\alpha = 0.5$

Simulation study

- 50 datasets for each setting
  - 144 sites in $[0, 10] \times [0, 10]$
    - 100 training
    - 44 testing
- Model parameters
  - Spatial range: $\rho = 1$
  - Skew parameter: $\lambda = 3$
  - Degrees of freedom: 6 for t distributions
  - Proportion of spatial variation: $\gamma = 0.9$

Simulation study

5 different models fit to each data set

1. Gaussian
2. Skew-t with $K = 1$ partition, no thresholding
3. Skew-t with $K = 1$ partition, thresholding at $q(0.80)$
4. Skew-t with $K = 5$ partitions, no thresholding
5. Skew-t with $K = 5$ partitions, thresholding at $q(0.80)$

Brier score

- Brier score (Gneiting and Raftery, 2007) used to compare fits
  - Lower scores indicate better fits
- The Brier score for predicting exceedance of threshold c is

$$[e(c) - P(c)]^2$$

  where

  - y is a test set value
  - $e(c) = I[y > c]$
  - $P(c)$ is the predicted probability of exceeding c
- Relative Brier scores:

$$\mathrm{BS}_{\mathrm{rel}} = \frac{\mathrm{BS}_{\mathrm{method}}}{\mathrm{BS}_{\mathrm{Gaussian}}}$$

Simulation study results

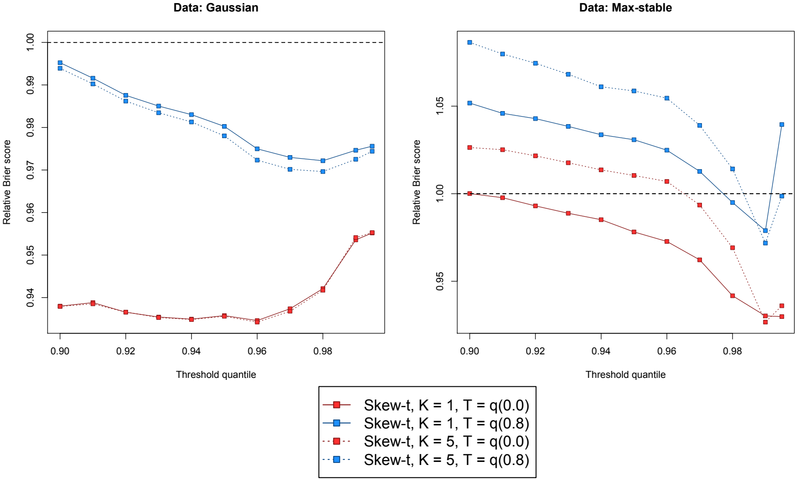

Figure: Relative Brier score results

Simulation study results

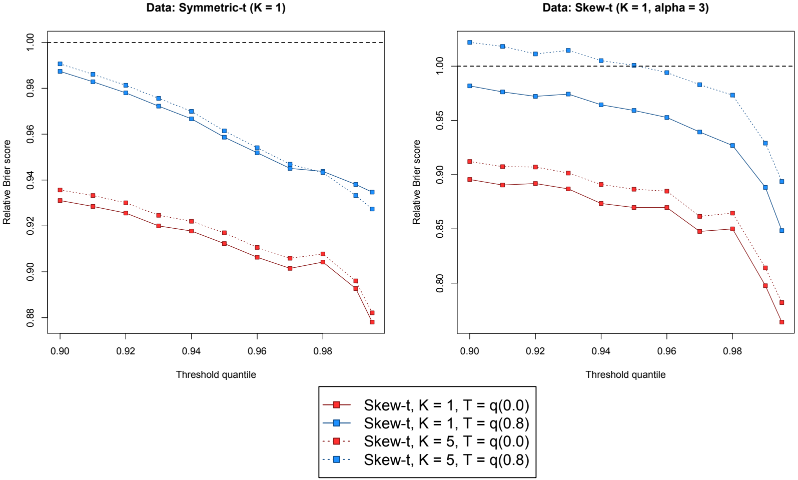

Figure: Relative Brier score results

Simulation study results

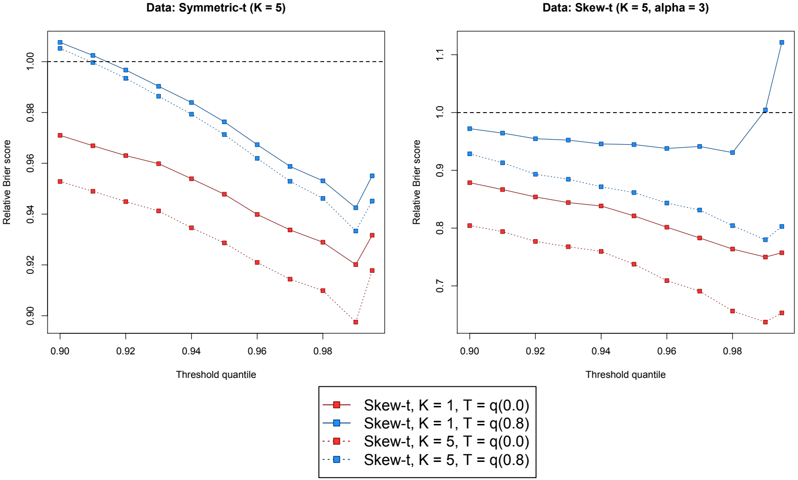

Figure: Relative Brier score results

Simulation study results

Key findings

- Improvement over Gaussian methods when partitioning
- Specifying too few knots has a detrimental impact
- In all cases, non-thresholded models perform better than thresholded models

Data analysis

- Ozone measurements
  - max 8-hour ozone measurements
  - daily data from 1089 sites
  - July 2005
- We take a stratified sample of $n = 800$ sites
  - 271 from northeast
  - 96 from northwest
  - 269 from southeast
  - 164 from southwest
- Conduct two-fold cross-validation on 800 sites

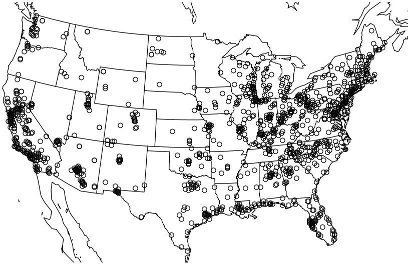

Figure: Ozone monitoring station locations

Data analysis

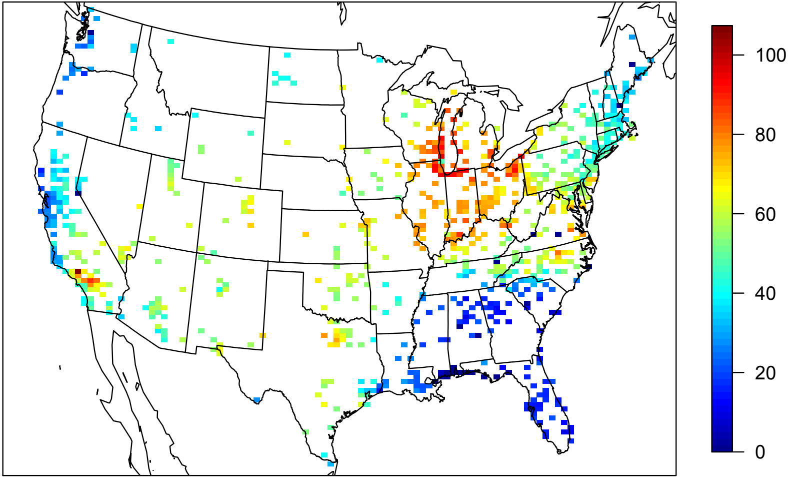

Figure: Max 8-hour ozone measurements on July 10, 2005

Model comparisons

- Many different analysis methods incorporating
  - Gaussian, symmetric-t, skew-t, and max-stable marginal distributions
  - $K = 1, 5, 6, 7, 8, 9, 10, 15$ partitions
  - 4 threshold levels for t marginals
    - $T = 0$
    - $T = 50\text{ppb}$, $q(0.48)$
    - $T = 75\text{ppb}$, $q(0.92)$
    - $T = 85\text{ppb}$, $q(0.97)$
  - Thresholded at $T = 75$ for max-stable
- Compare Brier scores from two-fold cross validation

Model comparisons

- The Community Multiscale Air Quality (CMAQ) system provides fine-resolution simulated values for multiple air pollutants
- We use the tropospheric ozone output from the corresponding days in the CMAQ model as a covariate
- Mean function modeled as

$$\mathbf{X}_t(\mathbf{s})\boldsymbol{\beta} = \beta_0 + \beta_1 \cdot \mathrm{CMAQ}_t(\mathbf{s})$$

- All methods use a Matérn covariance

Cross-validation results

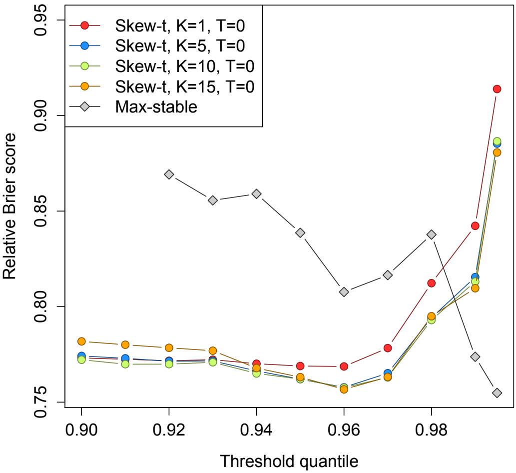

Figure: Relative Brier score results ($K = 6, \dots, 9$ are similar to $K = 5, 10$)

Probability of exceedance

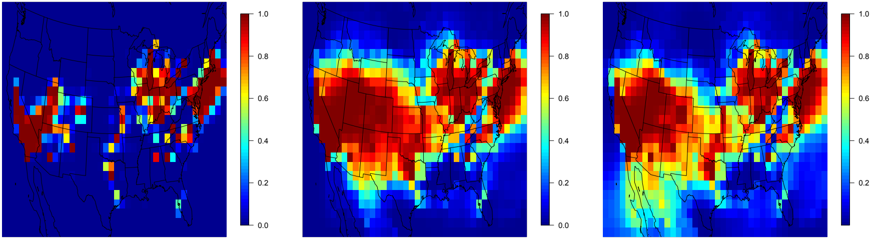

Figure: Probability of exceeding 75ppb at least two days for Gaussian (left), skew-t with 1 partition (center), skew-t with 6 partitions (right)

Quantile plots

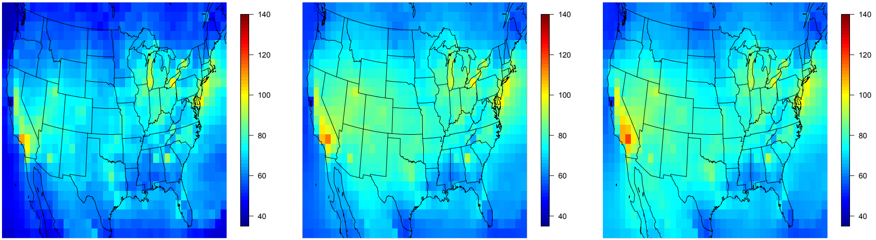

Figure: $\widehat{q}(0.95)$ for Gaussian (left), skew-t with 1 partition (center), skew-t with 6 partitions (right)

Quantile plots

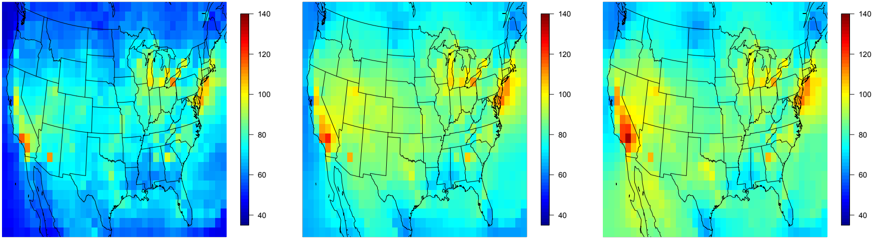

Figure: $\widehat{q}(0.99)$ for Gaussian (left), skew-t with 1 partition (center), skew-t with 6 partitions (right)

Cross-validation results

Key findings

- Partitioning improves performance across all high thresholds
- Models with anywhere from $K = 5$ to $K = 10$ partitions perform similarly
- In all cases, non-thresholded models perform better than thresholded models

Discussion

- Improvement of model performance when using partitioned models
- Thresholding makes results worse
    - Possible numerical instability due to truncated normal distribution

Future work: Knots and their impact

- Different partition structure
  - Distance weighting for each knot vs indicator functions
- Knot selection
  - Possible prior on the probability a knot is in the spatial domain

Future work: Temporal dependence

- Temporal dependence should be accounted for when using daily data
- For multiple days of observations the model becomes

$$Y_t(\mathbf{s}) = \mathbf{X}_t(\mathbf{s})^T\beta + \lambda\sigma_t(\mathbf{s})|z_t(\mathbf{s})| + \sigma_t(\mathbf{s})v_t(\mathbf{s})$$

  where t denotes the day of each observation

- Different ways to incorporate the temporal dependence
  - Time series on $\mathbf{w}_t$, $z_t(\mathbf{s})$, and $\sigma_t(\mathbf{s})$
  - Three dimensional covariance model for $v_t(\mathbf{s})$ (e.g. Huser and Davison, 2014)

Future work: Temporal dependence

- We choose the time series approach because the $z_t(\mathbf{s})$ and $\sigma_t(\mathbf{s})$ terms dictate the tail behavior
- We incorporate an AR(1) time series on $\mathbf{w}^*_{tk} = (w^*_{tk1}, w^*_{tk2})$, z_{tk}, and σ^*_{tk} where

$$w^*_{tki} = \Phi^{-1}\left[\frac{w_{tki} - \min(\mathbf{s}_i)}{\text{range}(\mathbf{s}_i)}\right] \qquad i = 1, 2$$

$$\sigma^{2*}_t(\mathbf{s}) = \Phi^{-1}\{\text{IG}[\sigma^2_t(\mathbf{s})]\}$$

are transformations to $\mathcal{R}^2$

Rare binary regression

- Motivation
  - Want to incorporate spatial dependence when modeling rare events (e.g. Diseased trees, Disease outbreak, Crimes)
- We observe

$$Y_i = \begin{cases} 1 & \text{event occurred} \\ 0 & \text{no event occurred} \end{cases}$$

- We model $\Pr[Y_i = 1]$

Rare binary regression

Common examples with non-spatial analysis

- Logistic regression

$$\Pr[Y_i = 1] = \frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta})}$$

- Probit regression

$$\Pr[Y_i = 1] = \Phi(\mathbf{X}_i\boldsymbol{\beta})$$

where Φ is the standard normal distribution function

Rare binary regression

- Generalized extreme value link function (Wang and Dey, 2010)

$$\Pr[Y_i = 1] = 1 - \exp\left[-(1 - \xi \mathbf{X}_i \beta)^{-1/\xi}\right]$$

- Link function allows for greater positive skew than existing methods
  - When $\xi = 0$, the link is the Cloglog link
  - When $\xi > 0$, the link allows for greater positive skew than Cloglog link

Rare spatial binary regression

- We propose to develop a spatial model
- Objectives are spatial prediction and to borrow strength across sites to estimate covariate effects
- Proposed method will
  - Use the GEV link function
  - Use the hierarchical method for spatially dependent extremes from Reich and Shaby (2012)
- Model parameters fit using MCMC

Rare spatial binary regression

- We model $Y_i = I(Z_i > 0)$ where $Z_i \sim$ multivariate GEV (mGEV) is a latent variable
- Hierarchical model for mGEV (Reich and Shaby, 2012)

$$Z(\mathbf{s}) = U(\mathbf{s})\theta(\mathbf{s})$$

where

- $U(\mathbf{s}) \overset{iid}{\sim} \text{GEV}(1,\, \alpha,\, \alpha)$ is a nugget effect
- $\theta(\mathbf{s}) = [\sum_{l=1}^{L} A_l w_l(\mathbf{s})^{1/\alpha}]^\alpha$ is the spatial process
- $A_l \overset{iid}{\sim}$ Positive Stable(α) is a random effect representing the intensity
- $w_l(\mathbf{s})$ gives the weight of the intensity of the lth random effect on site $\mathbf{s}$
- $\alpha \in (0, 1)$ controls strength of nugget relative to spatial dependence

Likelihood function

- After marginalizing out the A_l terms, we have the asymmetric logistic dependence function (Reich and Shaby, 2012)

$$G(\mathbf{z}) = \Pr[Z_1 < z_z, \ldots, Z_n < z_n] = \exp\left\{ -\sum_{l=1}^{L} \left[\sum_{i=1}^{n} \left(\frac{w_l(\mathbf{s}_i)}{z_i} \right)^{1/\alpha} \right]^{\alpha} \right\}$$

where

  - w_l is a weighting function subject to the constraint that $\sum_{l=1}^{L} w_l = 1$
  - α controls spatial dependence
    - $\alpha = 0$ is strong dependence
    - $\alpha = 1$ is joint independence

Weighting function

We use the Gaussian weights proposed by Reich and Shaby (2012) given by

$$w_l(\mathbf{s}_i) = \frac{\exp\left[-0.5\left(\frac{||\mathbf{s}_i - \mathbf{v}_l||}{\rho}\right)^2\right]}{\sum_{l=1}^{L}\exp\left[-0.5\left(\frac{||\mathbf{s}_i - \mathbf{v}_l||}{\rho}\right)^2\right]}$$

where

- $\mathbf{v}_l$ are spatial knots
- ρ is a bandwidth term for the kernel function

Illustrating asymmetric logistic dependence

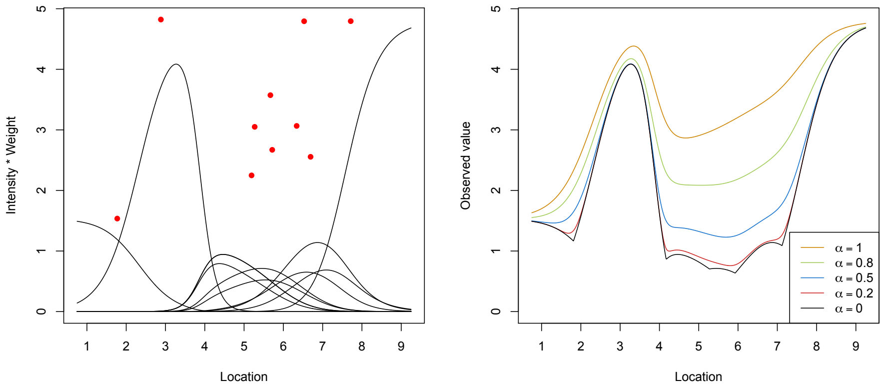

Figure: Knot intensity $\times$ weight ($\rho = 0.5$), red dots give intensity of random effects (left) Impact of α (right).

Joint likelihood

- Let $K_t = \sum_{i=1}^{n} Y_{it}$ be the number of exceedances that occur on day t.
- Rearrange the sites so
  - $Y_1, \ldots, Y_K$ are the observations where $Y(\mathbf{s}_i) = 1$
  - $Y_{K+1}, \ldots, Y_n$ are the observations where $Y(\mathbf{s}_i) = 0$
- For small K, we can evaluate the likelihood directly
- For large K, we use the hierarchical model of Reich and Shaby (2012)

Joint likelihood: K small

- For $K = 0, 1, 2$

$$\Pr(Y_1 = y_1, \ldots, Y_n = y_n) = \begin{cases} G(\mathbf{z}) & K = 0 \\ G(\mathbf{z}_{(1)}) - G(\mathbf{z}) & K = 1 \\ G(\mathbf{z}_{(12)}) - G(\mathbf{z}_{(1)}) - G(\mathbf{z}_{(2)}) + G(\mathbf{z}) & K = 2 \end{cases}$$

where $G\left(\mathbf{z}_{(1)}\right) = \Pr(Z_2 < z_2, \ldots, Z_n < z_n)$

- $K > 2$ can be derived similarly

Joint likelihood: K large

Hierarchical model: If $Z(\mathbf{s}) \sim$ mGEV with marginal distribution GEV(μ, σ, ξ), then

$$\begin{aligned} Z(\mathbf{s}) \mid A_1, \ldots, A_L &\overset{ind}{\sim} \text{GEV}[\mu^*, \sigma^*, \xi^*] \\ A_l &\overset{iid}{\sim} \text{Positive Stable}(\alpha) \end{aligned}$$

where

- $\mu^* = \mu + \frac{\sigma}{\xi}[\theta(\mathbf{s})^\xi - 1]$
- $\sigma^* = \alpha\sigma\theta(\mathbf{s})^\xi$
- $\xi^* = \alpha\xi$

Future simulation study and data application

- Simulation study
  - Data generated using logistic, Cloglog, and GEV links
    - Exploring how rarity of event impacts prediction
  - Models fit using
    - mGEV
    - Random effects Gaussian distribution
- Data application: Modeling crime data
  - Homicides, car theft, vandalism

Non-stationary dependence for extreme values

- Stationary covariance functions are a function of distance between two sites.

$$\rho[Y(\mathbf{s}), Y(\mathbf{t})] = \rho(h)$$

  where $h = ||\mathbf{s} - \mathbf{t}||$

- This assumes the covariance is the same everywhere, e.g. east vs west, mountains vs desert

- Misspecifying the covariance can impair spatial prediction and statistical inference

Non-stationary dependence for extreme values

- In extremes, stationary extremal dependence means

$$\chi(h) = \Pr[Y(\mathbf{s}) > c | Y(\mathbf{t}) > c]$$

- Currently, there are no methods to model non-stationarity in spatial extremes
- Semiparametric approach using spectral density ratios (de Carvalho and Davison, 2014)
  - Vector of observations can be transformed to pseudo-polar coordinates
  - Pairwise analysis
- New approach extending Reich and Shaby (2012)
  - Current model uses a single bandwidth term ρ for all knots
  - Proposed idea is to implement a knot-specific ρ to induce non-stationarity

Thesis outline

- Chapter 1: Review of extreme value theory August 2015
- Chapter 2: Spatiotemporal model for extreme value analysis based on the skew-t distribution February 2015
- Chapter 3: Rare spatial binary regression May / June 2015
- Chapter 4: Non-stationary extremal dependence through knot-specific bandwidth August 2015

Questions

- Questions?
- Thank you for your attention.
- Acknowledgment: This work was funded by EPA STAR award R835228

References I

- Balkema, A. A. and de Haan, L. (1974). Residual life time at great age, *Annals of Probability*, **2**, 792–804.
- de Carvalho, M. and Davison, A. C. (2014). Spectral density ratio models for multivariate extremes, *Journal of the American Statistical Association*, **109**, 764–776.
- Coles, S. G., Heffernan, J., and Tawn, J. A. (1999). Dependence measures for extreme value analysis, *Extremes*, **2**, 339–365.
- Coles S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Methodological)*, **53**, 337–392.
- Demarta, S. and McNeil, A. J. (2007). The t copula and related copulas. *International Statistical Review*, **73**, 111–129.

References II

- Falk. M., Hüsler, J., and Reiss, R. D. (2010). *Laws of Small Numbers: Extremes and Rare Events*. Springer Basel.
- Kim, H.-M., Mallik, B. K., and Holmes, C. C. (2005). Analyzing nonstationary spatial data using piecewise Gaussian processes. *Journal of the American Statistical Association*, **100**, 653–668.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, **44**, 423–453.
- Gneiting, T. and Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359–378.
- Gumbel, E. J. (1960). Multivariate extremal distributions. *Bulletin de l'Institut International de Statistique*, **37**, 471–475

References III

- Huser, R. and Davison, A. C. (2014). Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**, 439–461.
- Padoan, S. A. (2011). Multivariate extreme models based on underlying skew-t and skew-normal distributions. *Journal of Multivariate Analysis*, **102**, 977–991.
- Reich, B. J. and Shaby, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. *Annals of Applied Statistics*, **6**, 1430–1451.
- Smith R. L. (1990). Max-stable processes and spatial extremes, unpublished manuscript.
- Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika*, **77**, 245–253.

References IV

- Zhang, H. and El-Shaarawi, A. (2010). On spatial skew-Gaussian processes and applications. *Environmetrics*, **21**, 33–47.