Spatial methods for extreme value analysis

Samuel A. Morris

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 exceeding a certain threshold level
- Spatial methods borrow information across space to estimate spatial correlation and make predictions by Kriging at unknown locations
- Want to explore similar methods for extremes

Motivation

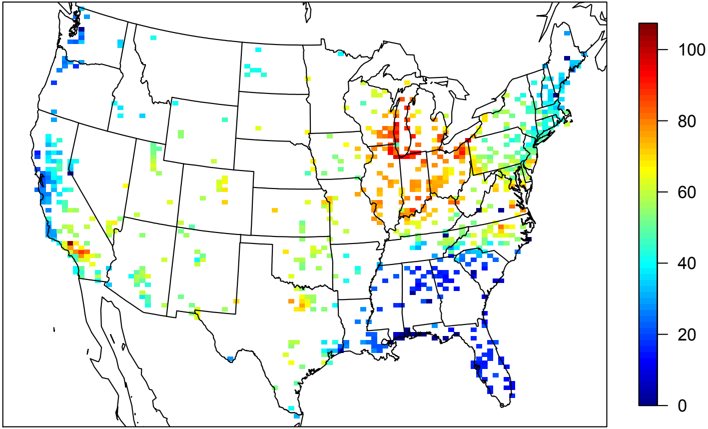

Figure: Max 8-hour ozone measurements on July 10, 2005

Defining extremes

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

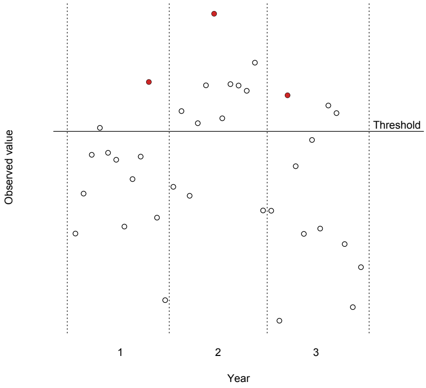

Figure: Monthly maximums recorded over a three years.

Standard analysis - Block maxima

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

$$a_n^{-1}(M_n - b_n) \xrightarrow{d} G(z)$$

then $G(z)$ follows a generalized extreme value distribution

Standard analysis - Block maxima

- Generalized extreme value distribution has three parameters:
  - $\mu \in \mathcal{R}$ is a location parameter
  - $\sigma > 0$ is a scale parameter
  - $\xi \in \mathcal{R}$ is a shape parameter
    - Unbounded above if $\xi \geq 0$
    - Bouded above by $(\mu - \sigma)/\xi$ when $\xi < 0$
- Challenge:
  - Lose information by only considering maximum in a block

Standard analysis - Block maxima

- Generalized extreme value distribution

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

- Standardized distribution is unit Fréchet or GEV(1, 1, 1)

$$\Pr(Z < z) = exp(-z^{-1})$$

Standard analysis - Peaks over threshold

- Generalized Pareto distribution has two parameters:
  - $\sigma > 0$ is a scale parameter
  - $\xi \in \mathcal{R}$ is a shape parameter
    - Unbounded above if $\xi \geq 0$
    - Bouded above by $(\mu - \sigma)/\xi$ when $\xi < 0$
- Challenges:
  - Sensitive to threshold selection
  - Temporal dependence between observations (e.g. flood levels don't dissipate overnight)

Standard analysis - Peaks over threshold

- Select a threshold, T, and use the Generalized Pareto distribution to model the exceedances

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

- Related to GEV distribution through

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

Max-stable processes

- For a spatial analysis, max-stable processes give an appropriate limiting distribution (Cooley et al., 2012):
  - Consider a spatial process $x_t(\mathbf{s})$, $t = 1, \ldots, T$.
  - Let $M_T(\mathbf{s}) = \left\{ \bigvee_{t=1}^{T} x_t(\mathbf{s}_1), \ldots, \bigvee_{t=1}^{T} x_t(\mathbf{s}_n) \right\}$
  - If there exists normalizing sequences $a_T(\mathbf{s})$ and $b_T(\mathbf{s})$ such that for all sites, $\mathbf{s}_i, i = 1, \ldots, d$,

$$a_T^{-1}(\mathbf{s}) \left\{ M_T(\mathbf{s}) - b_T(\mathbf{s}) \right\} \xrightarrow{d} Y(\mathbf{s})$$

    which has a non-degenerate distribution, then $Y(\mathbf{s})$ is a max-stable process.

Multivariate representations

- Marginally at each site, observations follow a generalized extreme value distribution
- Finite collection of sites:
  - The multivariate representation for the GEV is

$$\Pr(\mathbf{Z} \leq \mathbf{z}) = G^*(\mathbf{z}) = \exp(-V(\mathbf{z}))$$
$$V(\mathbf{s}) = d \int_{\Delta_d} \bigvee_{i=1}^{d} \frac{w_i}{z_i} H(\mathrm{d}w)$$

where

- $\Delta_d = \{\mathbf{w} \in \mathcal{R}_+^d \mid w_1 + \cdots + w_d = 1\}$
- 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 (Stephenson, 2003)
- Two common forms for $V(\mathbf{z})$:
  - Symmetric logistic

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

  - Asymmetric logistic

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

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

Multivariate peaks over threshold

- Not a lot of 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

- The χ statistic is a measure of extremal dependence in the tails
- Specifically, we focus on $\chi(h)$ for the upper tail 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 (Falk et al., 2011)
    - Asymmetric logistic has $2^{n-1}(n+2)-(2n+1)$ free parameters
  - joint distribution only available in low dimension
- 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 exceedances using a spatially dependent generalized extreme-value link function
3. A Bayesian hierarchical model to allow for non-stationary covariance in extreme value models.

Spatiotemporal modeling for extreme values

- Model to analysis spatiotemporal extreme values
- Model objectives:
  - Has marginal distribution with a flexible tail
    - allow for asymmetric distributions
    - allow for heavy tails
  - Has asymptotic spatial dependence
  - Has 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 $\lim_{c \to \infty} \chi(h) = 0$ regardless of the strength of the correlation in the bulk of the distribution
  - Tail is not flexible:
    - light tailed
    - symmetric

Spatial skew-t distribution

- Assume observed data $Y(\mathbf{s})$ come from a skew-t (Zhang and El-Shaarawi, 2012)

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

  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 and 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
- Can use standard geostatistical methods to fit this model
- Predictions can be made through Kriging

Spatial skew-t distribution

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

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

where

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

$\chi(h)$ plot

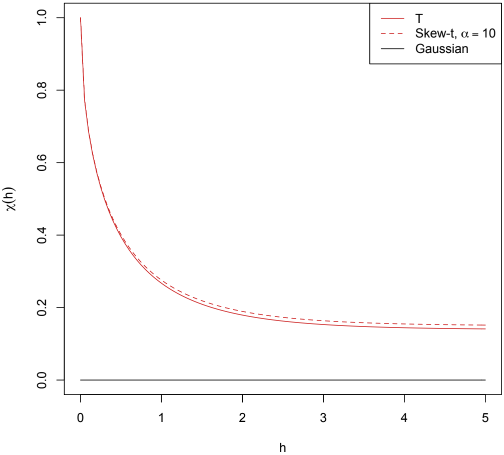

Figure: χ plot for symmetric t, skew-t, and Gaussian

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.

Spatial skew-t distribution properties

- Model properties
  - Has flexible tail:
    - Skewness controlled by λ
    - Weight of tails controlled by degrees of freedom $2a$
  - For a skew-t distribution $\lim_{c \to \infty} \chi(h) > 0$ (Padoan, 2011)
  - Computation that is on the order of Gaussian computation
- Challenge: $\chi(h) > 0$ as $h \to \infty$ (Padoan, 2011)
  - This occurs because all observations (near and far) share the same z and σ^2
  - We deal with this through a daily random partition (similar to Kim et al., 2005)

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 $\mathbf{w} = (w_1, w_2)$

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

$$\begin{aligned} z(\mathbf{s}) &= z_k \\ \sigma^2(\mathbf{s}) &= \sigma^2_k \end{aligned}$$

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 $v(\mathbf{s})$

Example partition

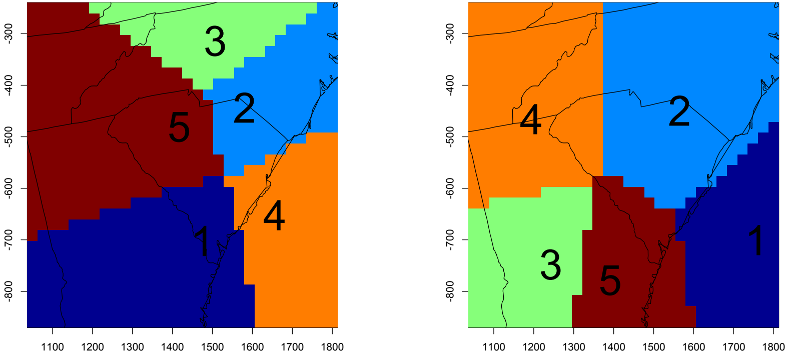

Figure: Two sample partitions (number is at partition center)

$\chi(h)$ plot

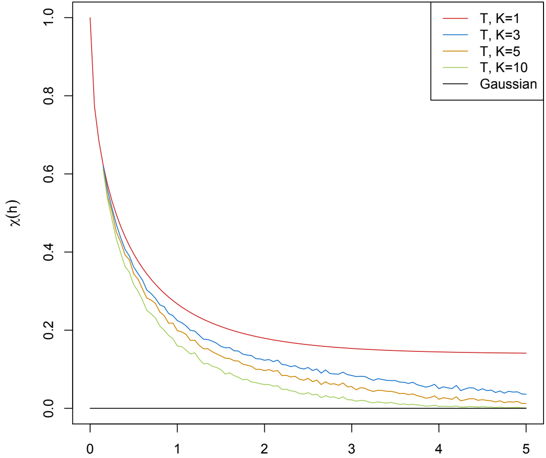

Figure: χ plot for different data settings

Random partition skew-t model

- This new model is called a random partition skew-t model, and it has the properties we desire
  - Marginal distribution with flexible tails
    - λ term allows for asymmetry
    - d.f. controls heavy vs light tails
  - Asymptotic spatial dependence for that decays with distance between sites via partition
  - Computation is on the order of Gaussian models for large space-time datasets

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 $t = 1, \ldots, T$ 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)
- Our model incorporates the time series because the $z_t(\mathbf{s})$ and $\sigma_t(\mathbf{s})$ terms dictate the tail behavior.

MCMC details

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

Simulation study

- 6 different data settings:
  - Gaussian, $K = 1$ partition
  - T, $K = 1$ partition
  - T, $K = 5$ partitions
  - Skew-t, $K = 1$ partition
  - Skew-t, $K = 5$ partitions
  - Max-stable
    - Marginally: GEV($\mu = 1, \sigma = 1, \xi = 0.2$)
    - Dependence function: asymmetric logistic with $\alpha = 0.5$

Simulation study

- 5 different models:
  - Gaussian
  - Skew-t with $K = 1$ partition, no thresholding
  - Skew-t with $K = 1$ partition, thresholding at $q(0.80)$
  - Skew-t with $K = 5$ partitions, no thresholding
  - Skew-t with $K = 5$ partitions, thresholding at $q(0.80)$

Brier score

- Brier score used to determine model that gives best fit (Gneiting and Raftery, 2007)
- 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

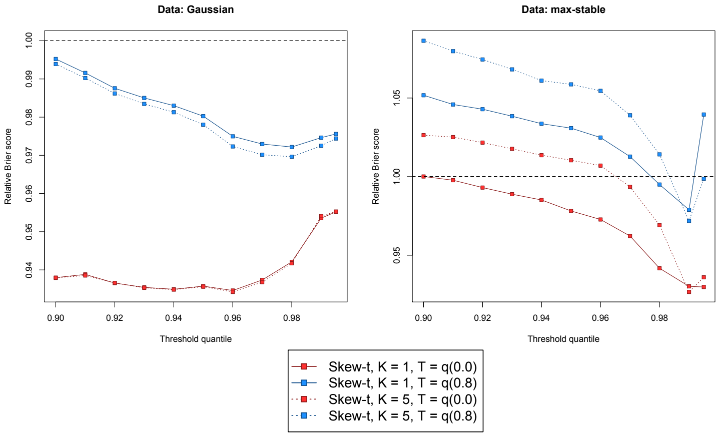

Figure: Relative Brier score results

Simulation study results

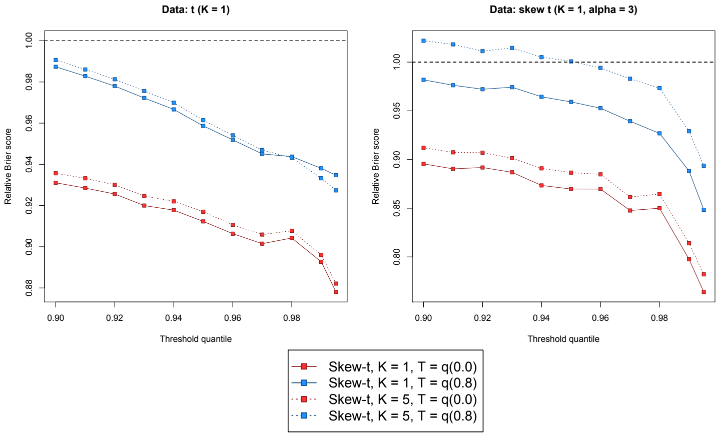

Figure: Relative Brier score results

Simulation study results

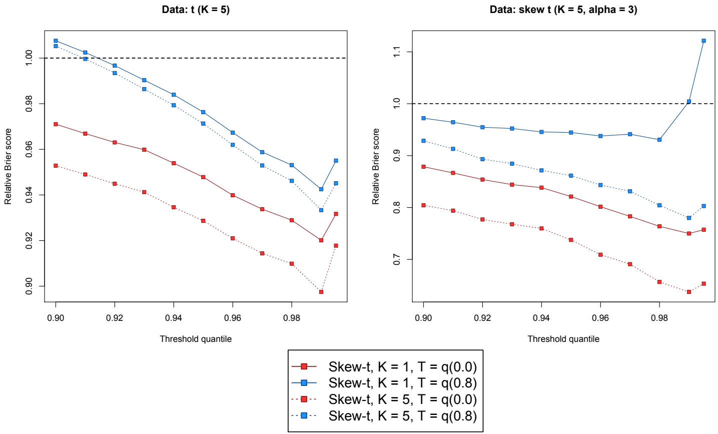

Figure: Relative Brier score results

Simulation study results

- Key findings:
  - Improvement over Gaussian methods when partitioning
  - Underestimating the number of 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
  - 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

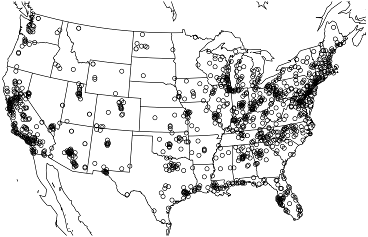

Figure: Ozone monitoring station locations

Data analysis

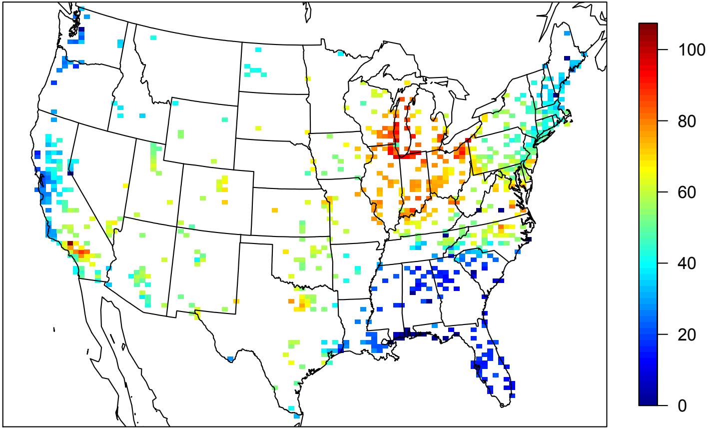

Figure: Max 8-hour ozone measurements on July 10, 2005

Exploratory data analysis

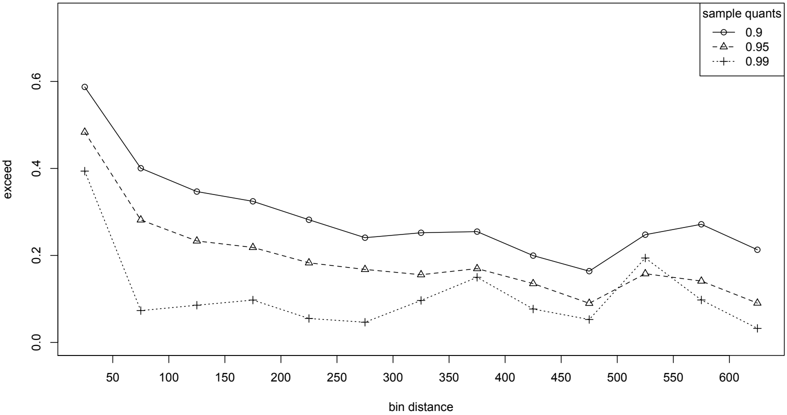

Figure: $\widehat{\chi}$-plot for sample quantiles of ozone observations

Model comparisons

- 9 different analysis methods incorporating
  - Gaussian vs t vs skew-t marginal distribution
  - $K = 1, 5, 6, 7, 8, 9, 10, 15$ partitions
  - Thresholding at
    - $T = 0$
    - $T = 50$ppb, $q(0.48)$
    - $T = 75$ppb, $q(0.92)$
    - $T = 85$ppb, $q(0.97)$
- Compare Brier scores using two-fold cross validation

Model comparisons

- All methods use a Matérn or exponential covariance ($\nu = 0.5$)
- Covariate data from the Environmental Protection Agency's Community Multiscale Air Quality (CMAQ) system.
- Mean function modeled as

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

Two-fold cross-validation results

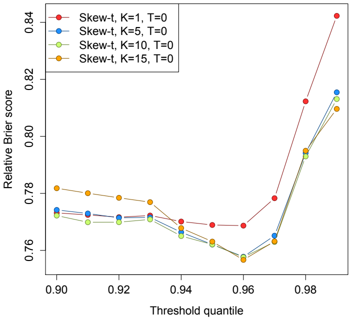

Figure: Relative Brier score results

Two-fold 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 and future work

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

- We extend our previous model to be

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

where $t = 1, \dots, T$ denotes the day of each observation.

- We incorporate an AR(1) time series on $\mathbf{w}^*_{tk} = (w^*_{tk1}, w^*_{tk2})$, z_{tk}, and σ^*_{tk} where

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

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

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

Rare spatial binary regression

- Movivation:
  - Want to incorporate spatial dependence when modeling rare events.
  - Examples:
    - Diseased trees
    - Disease outbreak
- We observe

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

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

Rare spatial binary regression

- Common examples with non-spatial analysis
  - Logistic regression

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

  - Probit regression

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

    where Φ is the standard normal distribution function
  - Wang and Dey (2010): Generalized extreme value link function

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

Rare spatial binary regression

- 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 fit parameters ξ and β in order to transform the data to GEV(1, 1, 1) marginal distributions.
- Using the link function

$$p_i = 1 - \exp\left[-(1 + \xi \mathbf{X}(\mathbf{s}_i)\boldsymbol{\beta})^{-1/\xi}\right]$$

- Latent variable $z_i = -\frac{1}{\log(1-p_i)}$ used to evaluate the likelihood.

Likelihood function

- We use a multivariate generalized extreme value distribution with 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

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
  - $Y_1, \dots, Y_K$ are the observations where $Y(\mathbf{s}_i) = 1$
  - $Y_{K+1}, \dots, Y_n$ are the observations where $Y(\mathbf{s}_i) = 0$
- Then for $K = 0, 1, 2$

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

  where $G(\mathbf{z}_{(1)} = \Pr(Z_2 < z_2, \dots, Z_n < z_n)$.
- $K > 2$ can be derived similarly

Joint likelihood

- For small K, we can evaluate the likelihood directly.
- For large K, we use the hierarchical model of Reich and Shaby (2012).

Non-stationary covariance for extreme values

- Knot-specific bandwidth

Thesis outline

- Chapter 1: Extreme value theory August 2015
- Chapter 2: Spatiotemporal model for extreme value analysis based on the skew-t distribution February 2015
- Chapter 3: Spatial binary regression May / June 2015
- Chapter 4: Non-stationary covariance through knot-specific bandwidth August 2015

Questions

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

References

- Demarta, S. and McNeil, A. J. (2007) The *t* copula and related copulas. *International Statistical Review*, **73**, 111–129.
- 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.
- Zhang, H. and El-Shaarawi, A. (2010) On spatial skew-Gaussian processes and applications. *Environmetrics*, **21**, 33–47.