Spatial methods for extreme value analysis

Samuel A. Morris

17 Mar 2016

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

Defining extremes

- 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

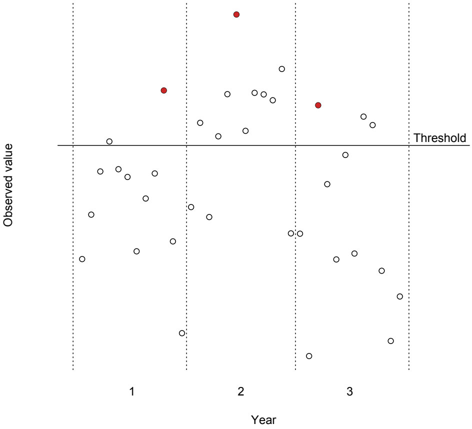

Figure: Hypothetical monthly data

Non-spatial analysis: Block maxima

Fisher-Tippett-Gnedenko theorem

- 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 $T \to \infty$

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

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 $\xi < 0$
- 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 representation for the multivariate 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

Max-stable processes: A hierarchical representation (Reich & Shaby, 2012)

- Let $\widetilde{\mathbf{Y}} \sim \mathrm{GEV}_n[\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})]$ be a realization from multivariate generalized extreme value distribution

- Consider a set of L knots, $\mathbf{v}_1, \ldots, \mathbf{v}_L$

- Model the spatial dependence using

$$\theta(\mathbf{s}) = \left[\sum_{l=1}^{L} A_l w_l(\mathbf{s})^{1/\alpha} \right]^{\alpha}$$

where

  - A_l are i.i.d. positive stable random effects
  - $w_l(\mathbf{s})$ are a set of non-negative scaled kernel basis functions, scaled so that $\sum_{l=1}^{L} w_l(\mathbf{s}) = 1$
  - $\alpha \in (0,1)$ is a parameter controlling strength of spatial dependence (0: high, 1: independent)

Max-stable processes: A hierarchical representation (Reich & Shaby, 2012)

- When conditioning on θ

$$\widetilde{Y}(\mathbf{s}_i) \mid A_l \overset{ind}{\sim} \text{GEV}[\mu^*(\mathbf{s}_i), \sigma^*(\mathbf{s}_i), \xi^*(\mathbf{s}_i)]$$
$$A_l \overset{iid}{\sim} \text{PS}(\alpha)$$

where

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

Dimension reduction for spatial extremes

- Reich and Shaby (2012) can be computationally challenging
- Computing time is driven by the positive stable random effects
- By default, knots may be placed at spatial locations
- One possibility is to use fewer knots
  - Need to decide how many knots to use
  - Need to decide where to place them
- Another possibility is a new basis representation

Dimension reduction for spatial extremes

- Another measure of spatial dependence is the pairwise extremal coefficient: ϑ_{ij}.

$$P(Z_i < c, Z_j < c) = P(Z_i < c)^{\vartheta_{ij}} \in (1, 2).$$

- In the positive stable random effects model, the extremal coefficient has the form

$$\vartheta_{ij} = \sum_{l=1}^{L} \left(w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha} \right)^{\alpha}.$$

- What if we could use a low-dimensional representation for the w_l terms?

Empirical basis functions

- Generate L basis functions, $B_1(\mathbf{s}_i), \ldots, B_L(\mathbf{s}_i)$, and use these as $w_l(\mathbf{s}_i)$
- Three steps:
  1. Obtain an initial estimate of the extremal coefficient for each pair of locations, $\hat{\vartheta}_{ij}$
  2. Spatially smooth these initial estimates $\hat{\vartheta}_{ij}$ using kernel smoothing to obtain $\tilde{\vartheta}_{ij}$
  3. Estimate the spatial dependence parameters α and $B_1, \ldots, B_L$ by minimizing the difference between model-based coefficients, ϑ_{ij}, and smoothed coefficients, $\tilde{\vartheta}_{ij}$

Empirical basis functions

- We can describe the contribution of the lth basis function to the pairwise extremal coefficients as

$$v_l = \sum_{i=1}^{n} B_{il}/n$$

  where n is the number of sites

- This approach speeds up computation in two ways.
  1. Reduction in number of parameters being fit by MCMC
  2. Typically $L << n$
     - e.g. Wildfire, can perform close to full model with $L = 15$ knots for $n = 159$ counties

Data application

- Wildfire acreage burned in GA, 1965 – 2014

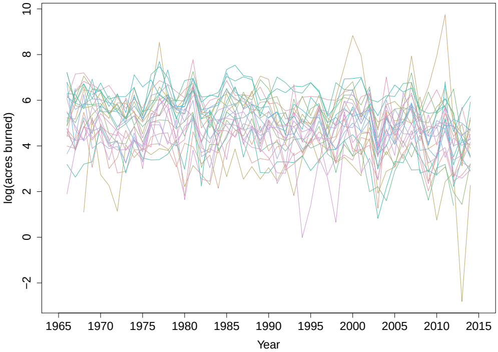

Figure: Time series of log(acres burned) for 25 randomly selected counties.

Data application

- Data are not max-stable, so we use a site-specific threshold
- Threshold originally selected using a spatially smoothed $\hat{q}(0.95)$

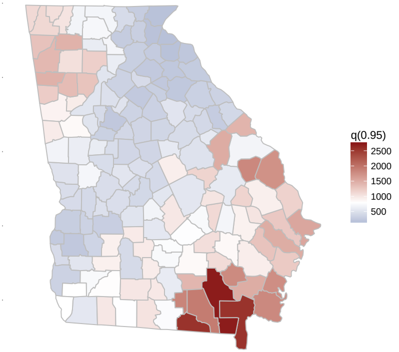

Figure: Spatially smoothed $\hat{q}(0.95)$

Analysis

- Spatio-temporal model for GEV parameters including a linear time trend and interaction with basis functions
  - $\mu(\mathbf{s}, t) = \beta_{0,\mu} + \beta_{1,\mu} t + \gamma_{1,\mu} B_1 + \cdots \gamma_{L,\mu} B_L + \delta_{1,\mu} B_1 t + \cdots \delta_{L,\mu} B_L t$
  - $\log(\sigma)(\mathbf{s}, t) = \beta_{0,\sigma} + \beta_{1,\sigma} t + \gamma_{1,\sigma} B_1 + \cdots \gamma_{L,\sigma} B_L + \delta_{1,\sigma} B_1 t + \cdots \delta_{L,\sigma} B_L t$
  - ξ is constant across space
- Prior distributions:
  - $\mu(\mathbf{s}, t)$: coefficients $\stackrel{iid}{\sim} N(0, \sigma_\mu^2)$
  - $\log(\sigma)(\mathbf{s}, t)$: coefficients $\stackrel{iid}{\sim} N(0, \sigma_\sigma^2)$
  - $\xi \sim N(0, 0.25)$
- Independent IG(0.1, 0.1) priors on σ_μ^2 and σ_σ^2

Model comparisons

- Comparing two different spatial process constructions
  - Extremal coefficient basis functions (ECB)
  - Gaussian kernel basis functions (GKB)
- Comparing two basis function structures for marginal distributions
  - Extremal coefficient basis functions (ECB)
  - 2-dimensional B splines (2BS) (in progress)
- Considering $L = 4, 9, 16, 25$ knots
  - When $L = 4$ for 2BS, we use a 2nd order spatial model for $\mu(\mathbf{s}, t)$ and $\log(\sigma)(\mathbf{s}, t)$

2d B splines

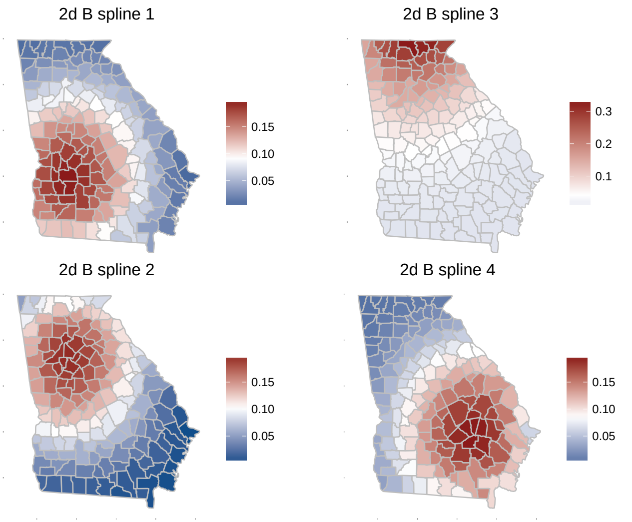

Figure: Four 2-dimensional B splines with $L = 9$

Results

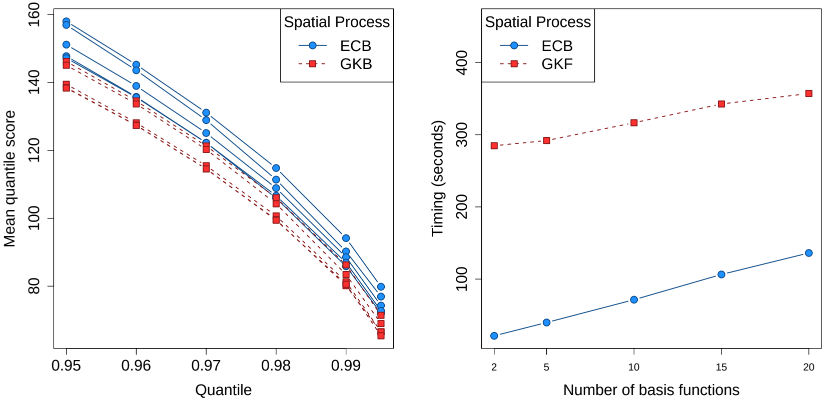

Figure: Average quantile score for selected quantiles and timing comparisons

Remaining questions

- What's the best way to select the threshold for this application?
  - Mean residual plots are helpful, but can be challenging since not all sites have same marginal parameters
  - Cross-validation is time intensive
- Are there better options for the spatial aspect of the marginal parameters?
- Are there better ways to pick the number of knots?
  - Potentially add knots until the smallest v_l is less than some threshold