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

Introduction to extremes

- Max-stable processes (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.

Standard analysis - Block maxima

- Uses yearly maxima
- Discards many observations
- Models are fit using the generalized extreme value distribution
- For a spatial analysis, max-stable processes give an appropriate limiting distribution

Standard analysis - Peaks over threshold

- Incorporates more data than block maxima
- Select a threshold, T, and use the Generalized Pareto distribution (GPD) to model the exceedances
- Temporal dependence may be an issue between observations (e.g. flood levels don't dissipate overnight)

Multivariate representations

- Multivariate distributions:
  - Assume common standardized max-stable marginal, like unit-Fréchet

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

  - 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 analysis

- Multivariate max-stable and GPD models have nice features, but they are
  - computationally challenging to work with
  - joint distribution only available in low dimension
- Bayesian hierarchical model (Reich and Shaby, 2012)
- Pairwise likelihood approach (Huser and Davison, 2014)

Model objectives

- Our objective is to build a model that
  - has marginal distribution with a flexible tail
  - has asymptotic spatial dependence
  - has computation on the order of Gaussian models for large space-time datasets

Censoring data

- 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.

χ coefficient

- The χ coefficient is a measure of extremal dependence
- Specifically, we focus on $\chi(\mathbf{h})$ 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 $\mathbf{h}$.
- We expect $\lim_{\mathbf{h} \to \infty} \chi(\mathbf{h}) = 0$.

Gaussian spatial model

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

Spatial skew-t distribution

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

$$Y_t(\mathbf{s}) = X_t(\mathbf{s})\beta + \alpha z_t + v_t(\mathbf{s})$$

  where

  - $\alpha \in \mathcal{R}$ controls the skewness
  - $z_t \overset{iid}{\sim} N_{(0,\infty)}(0, \sigma_t^2)$ is a random effect
  - $v_t(\mathbf{s})$ is a Gaussian process with variance σ_t^2 and Matérn correlation
  - $\sigma_t^2 \overset{iid}{\sim} \text{IG}(a, b)$

Spatial skew-t distribution

- Conditioned on z_t and σ_t^2, $Y_t(\mathbf{s})$ is a Gaussian spatial model
- Can use standard geostatistical methods to fit this model
- Predictions can be made through Kriging
- Marginalizing over z_t and σ_t^2 (via MCMC),

$$Y_t(\mathbf{s}) \sim \text{skew-t}(\mu, \Sigma^*, \alpha, \text{df} = 2a)$$

where

  - μ is the location
  - a, b are the IG parameters for σ_t^2
  - $\Sigma^* = \frac{b}{a}\Sigma$ is a scale matrix, and Σ is a Matérn covariance matrix
  - $\alpha \in \mathcal{R}$ controls the skewness

Spatial skew-t distribution

- Model properties
  - Has flexible tail controlled by skewness α and degrees of freedom $2a$
  - For a skew-t distribution $\lim_{c \to \infty} \chi(\mathbf{h}) > 0$ (Padoan, 2011)
  - Computation that is on the order of Gaussian computation
- For this distribution, $\chi(\mathbf{h})$ shows asymptotic dependence that does not approach 0 as $\mathbf{h} \to \infty$
- This occurs because all observations (near and far) share the same z_t and σ^2_t
- We deal with this through a daily random partition (similar to Huser and Davison)

Daily random partition

- Daily random partition allows z_t and σ_t^2 to vary by site

$$Y_t(\mathbf{s}) = X_t(\mathbf{s})\beta + \alpha z_t(\mathbf{s}) + \sigma(\mathbf{s}) v_t(\mathbf{s})$$

- Consider a set of daily knots $\mathbf{w}_{tk} \sim$ Uniform that define a random daily partition $P_{t1}, \ldots, P_{tK}$ such that

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

- For $\mathbf{s} \in P_{tk}$

$$z_t(\mathbf{s}) = z_{tk}$$
$$\sigma_t^2(\mathbf{s}) = \sigma_{tk}^2$$

- Within each partition $Y_t(\mathbf{s})$ has the same MV skew-t distribution as before

Example daily partition

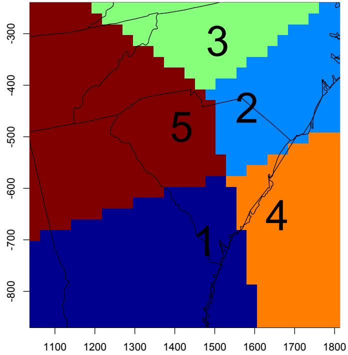
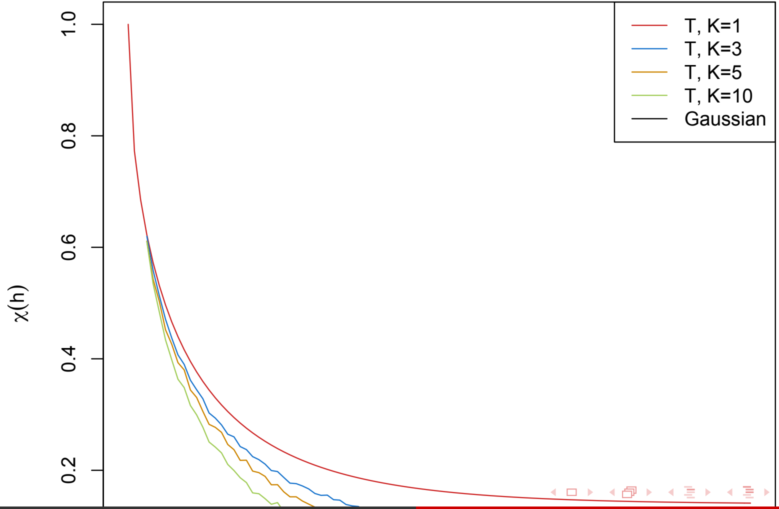


Figure: Two sample partitions (number is at partition center)

Simulated $\widehat{\chi}(h)$ plots



Sample simulated datasets

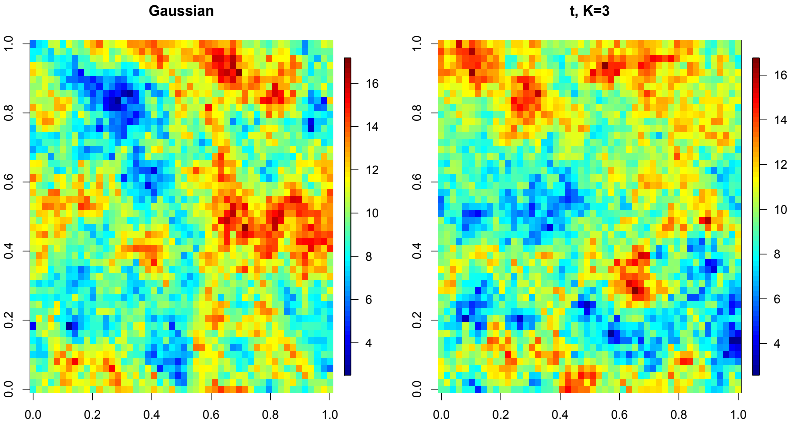


Figure: Gaussian and t with 3 partitions

Sample simulated datasets

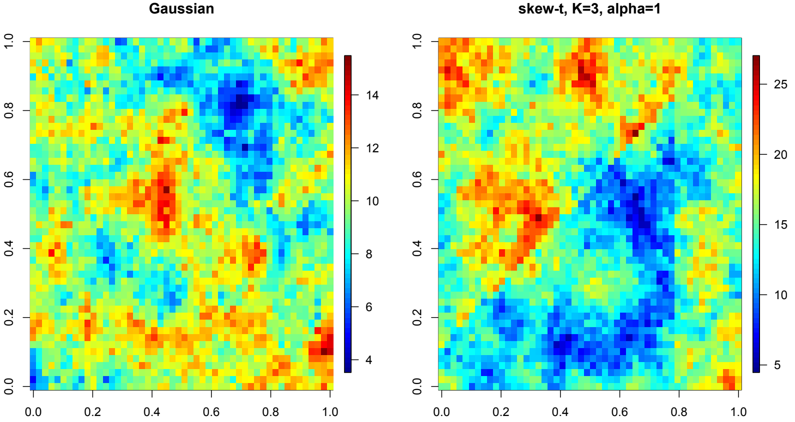


Figure: Gaussian and skew-t with 3 partitions

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 vs t vs skew-t marginal distribution
  - $K = 1$ partition vs $K = 5$ partitions

Brier score results

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

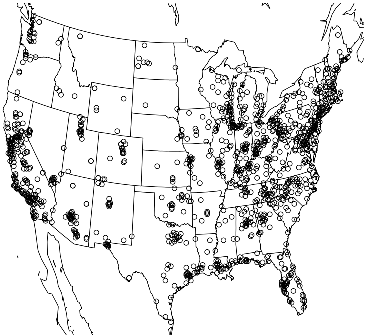
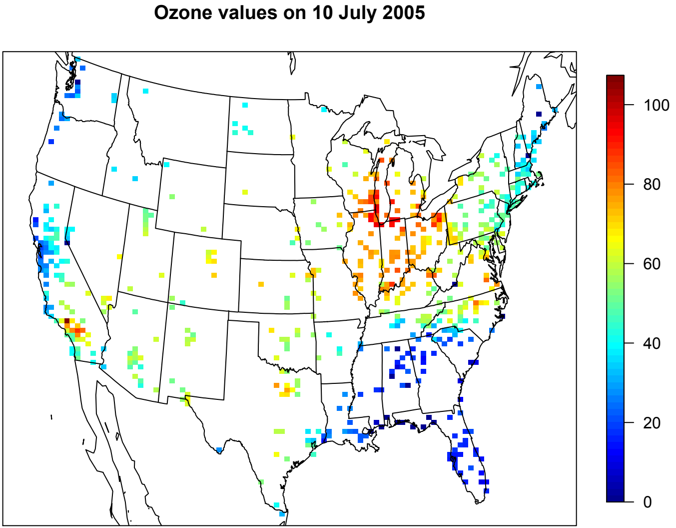


Figure: Ozone monitoring station locations

Data analysis



Exploratory data analysis

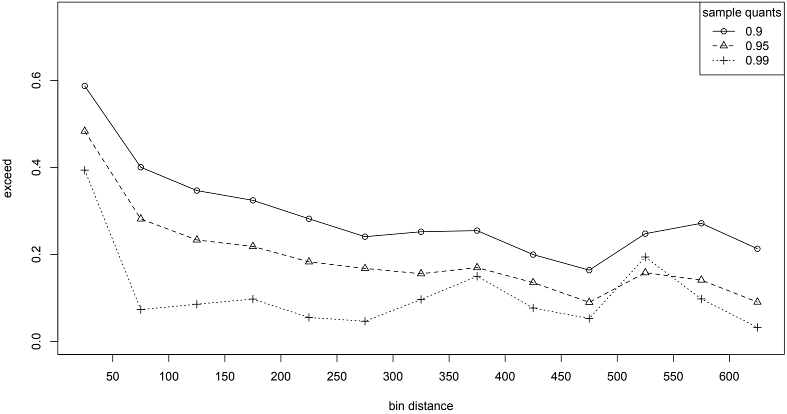


Figure: $\widehat{\chi}$-plot for sample quantiles of ozone observations

Model comparisons

- 9 different analysis methods incorporating
  - Gaussian vs t vs skew-t marginal distribution
  - $K = 1$ partition vs $K = 3$ partitions
  - No thresholding vs thresholding at $T = 0.90$ sample quantile
- All methods use a Matérn or exponential covariance ($\nu = 0.5$)
- Compare quantile and Brier scores using 5-fold cross validation (Gneiting and Raftery, 2007)
- Mean function modeled as

$$\beta_0 + \beta_1 \cdot \text{lat} + \beta_2 \cdot \text{long} + \beta_3 \cdot \text{lat}^2 + \beta_4 \cdot \text{long}^2 + \beta_5 \cdot \text{lat} \cdot \text{long}$$

Brier score

- 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

Five-fold cross-validation results

			τ				
Marginal	K	T	0.950	0.980	0.990	0.995	0.999
Gaussian	1	0	39.820	17.539	9.167	4.720	1.057
t	1	0	**31.008**	**13.898**	7.229	**3.405**	0.879
t	5	0	31.213	13.920	**7.218**	3.498	0.918
t	1	0.9	32.221	14.519	7.549	3.604	0.896
t	5	0.9	38.842	16.781	8.434	4.180	1.020
skew-t	1	0	31.845	14.542	7.533	3.645	**0.844**
skew-t	1	0.9	32.132	14.296	7.484	3.497	0.890
skew-t	3	0	33.653	15.453	8.119	4.338	1.188
skew-t	3	0.9	32.157	14.727	7.794	3.825	0.917

Table: Brier score for predicting exceedance of $c = \hat{q}(\tau)$ from five-fold cross-validation ($\times 1000$)

- Quantile score results are similar

Future Work

- Different ways to incorporate the temporal dependence
  - Three dimensional covariance model for $v_t(\mathbf{s})$ (e.g. Huser and Davison, 2014)
  - Use a temporal structure for $z_t(\mathbf{s})$:
    - AR(1)
    - Moving average
    - Association between $\mathbf{w}_{t,k}$ and $\mathbf{w}_{t+1,k}$
- Comparison with extreme value analysis methods

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.