A Space-time skew-t model for threshold exceedances

Samuel A. Morris

North Carolina State University

August 3, 2016

Joint work with Brian Reich (NCSU), Emeric Thibaud (CSU), and Dan Cooley (CSU)

Spatial extremes

- Extreme Values Analysis (EVA) can benefit greatly from spatial methods
- Spatial methods can map risk and borrow strength over space to estimate rare-event probabilities
- Accounting for spatial dependence is necessary for valid inference
- Methods and software in this area are developing rapidly to meet a growing demand

Max-stable processes

- Let $Y_1(s)$, ..., $Y_m(s)$ be iid spatial processes
- The pointwise maximum process is

$$\tilde{Y}(s) = \bigvee_{l=1}^{m} Y_l(s)$$

- If there exist constants a_m and b_m so that

$$Z(s) = a_m + b_m \tilde{Y}(s)$$

converges to a valid process as $m \to \infty$, then Z is max-stable

Max-stable processes

- The marginal distribution of Z at each s follows generalized extreme value (GEV) distribution
- Distribution function is $F(z) = \exp\{-t(z)\}$ where

$$t(z) = \begin{cases} \left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}, & \xi \neq 0 \\ \exp\left\{-\frac{z-\mu}{\sigma}\right\}, & \xi = 0. \end{cases}$$

- Location: $\mu \in \mathcal{R}$
- Scale: $\sigma > 0$
- Shape: $\xi \in \mathcal{R}$

Challenges in EVA

- Covariance focuses on deviations around the mean and not the extremes
- Want dependence measure to capture likelihood of seeing values that are jointly extreme
- Two common measures of dependence:
  - Extremal coefficient $\vartheta \in (1, 2)$:

$$\text{Prob}[Z(\mathsf{s}_1) < c, Z(\mathsf{s}_2) < c] = \text{Prob}[Z(\mathsf{s}_1) < c]^{\vartheta(\mathbf{s}_1, \mathbf{s}_2)}$$

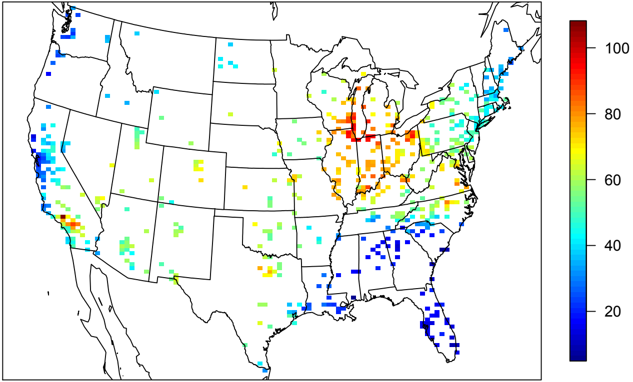
  - $\chi \in (0, 1)$:

$$\chi(\mathsf{s}_1, \mathsf{s}_2) = \lim_{c \to \infty} \text{Prob}[Z(\mathsf{s}_1) > c | Z(\mathsf{s}_2) > c]$$

Current approaches and limitations

- Theory suggests that a max-stable process is a good option for spatial extremes
- The max-stable process gives a complicated likelihood function with no closed-form except in trivial cases
- Current Bayesian approaches can handle only a moderate number of spatial locations
- This is limiting because most modern applications have hundreds or thousands of stations
- Because of these challenges advanced methods for e.g., multivariate or nonstationary data are limited

Application – Air pollution (ozone)



Ozone measurements on July 10, 2005, at 1,089 stations across the US

Back to a Gaussian process model

- The max-stable process is an elegant approach, but does that mean it's the right model?
- In reality, it is only an approximation
- There are less complicated approximations
- For example, we could model daily data as a Gaussian process (GP)
- If the goal is spatial interpolation, perhaps this is competitive

GP - asymptotic independence

- A GP leads to simple interpretation and computing, but asymptotic independence
- If $Y(\mathsf{s}_1)$ and $Y(\mathsf{s}_2)$ are bivariate normal then $\chi(\mathsf{s}_1, \mathsf{s}_2) = 0$, i.e., asymptotic independence
- This suggests Kriging will not capture extremes
- But so much is known for the Gaussian case: nonstationarity, multivariate, numerical approximations, ...
- Rather than toss it out, can we patch it up?

Spatial skew-t process

A spatial skew-t process (Azzalini and Capitanio, 2014) resembles a GP but exhibits asymptotic dependence

$$
\begin{aligned}
Y_t(\mathrm{s}) &= \mathrm{X}(\mathrm{s})^\top \boldsymbol{\beta} + \lambda \sigma_t |z_t| + \sigma_t v_t(\mathrm{s}) \\
z_t &\sim \text{Normal}(0, 1) \\
\sigma_t^2 &\sim \text{InvGamma}(a/2, b/2) \\
v_t &\sim \text{Spatial GP}
\end{aligned}
$$

- Location: $\mathrm{X}(\mathrm{s})^\top \boldsymbol{\beta}$
- Scale: $b > 0$
- Skewness: $\lambda \in \mathcal{R}$
- Degrees of freedom: $a > 0$

Good properties

- Flexible t marginal distribution with four parameters including the degrees of freedom which allows for heavy tails ($a = 1$ gives a Cauchy)

- Computation on the order of a GP; the only extra steps are z_t and σ_t which have conjugate full conditionals

- Asymptotic dependence: $\chi(\mathsf{s}_1, \mathsf{s}_2) > 0$ for all s_1 and s_2

Bad properties and remedies

- Modeling all the data (bulk and extreme) can lead to poor tail probability estimates if the model is misspecified
- Long-range dependence: $\chi(\mathsf{s}_1, \mathsf{s}_2) > 0$ for all s_1 and s_2 even if s_1 and s_2 are far apart
- This occurs because all sites share z_t and σ_t
- Remedies:
  - We use a censored likelihood to focus on the tails
  - We propose a local skew-t process

Censored likelihood

- Censored likelihood: We censor the data

$$\tilde{Y}_t(\mathrm{s}) = \begin{cases} T & \text{for } Y_t(\mathrm{s}) \leq T \\ Y_t(\mathrm{s}) & \text{for } Y_t(\mathrm{s}) > T \end{cases}$$

- Censoring is handled using standard Bayesian imputation methods

- The threshold T is chosen by cross-validation

- If T is moderately extreme in the distribution (e.g. $q(0.75)$), set $\lambda = 0$

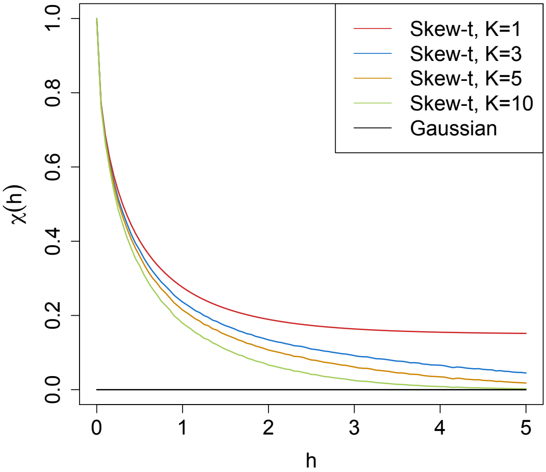
Local skew-t process

- Let the knots $\mathrm{v}_{t1}, ..., \mathrm{v}_{tK}$ follow a homogeneous Poisson process over the domain of interest (in practice we fix K)
- Associated with each is
  - $z_{tk} \sim \text{Normal}(0, 1)$
  - $\sigma^2_{tk} \sim \text{InvGamma}(a/2, b/2)$
- The knots partition the domain if we assign location s to subregion $k = \text{argmin}_l \, ||\mathrm{s} - \mathrm{v}_{tl}||$
- If s is in subregion k then

$$Y_t(\mathrm{s}) = \mathrm{X}(\mathrm{s})^T \boldsymbol{\beta} + \lambda \sigma_{tk} |z_{tk}| + \sigma_{tk} v_t(\mathrm{s})$$

- The marginal distribution remains a t, but partitioning breaks long-range spatial dependence

χ-statistic by $h = ||\mathbf{s}_1 - \mathbf{s}_2||$



Temporal dependence

- It may not be reasonable to assume that observations are temporally independent (e.g. flooding, high temperatures)
- Temporal dependence is handled through the z_{tk}, σ_{tk} and v_{tl}
- Method:
  - Use a copula to transform parameters to *nice* space (i.e. $\mathcal{R}$)
  - AR(1) structure imposed on parameters in transformed space
  - Transform back to original parameter space to preserve skew-t

Results of a simulation study

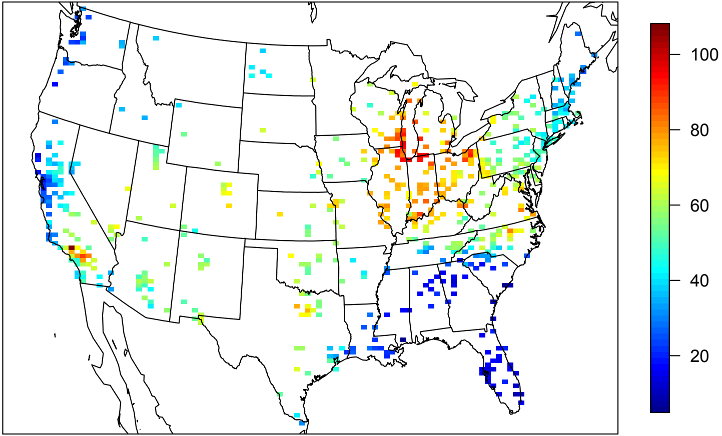
In terms of Brier scores for spatial prediction:

- Data generated as a GP:
  - skew-t is close to GP
  - max-stable is 15% – 30% worse than GP
- Data generated as a skew-t with multiple partitions:
  - skew-t is 15% better than GP
  - max-stable is 30% worse than GP
- Data generated as asymmetric logistic (max-stable):
  - skew-t is close to GP
  - max-stable performs 10% better than GP
- Data generated as Brown-Resnick (max-stable):
  - skew-t performs 40% – 60% better than GP
  - max-stable performs 40% – 60% better than GP

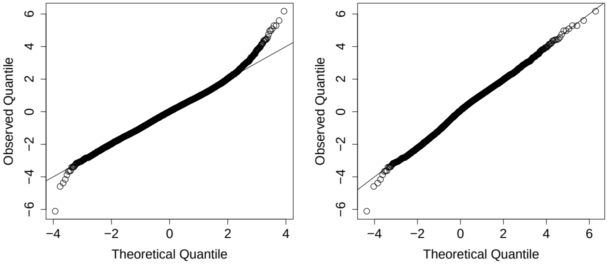
Application to ozone

- The USEPA has an extensive network of ozone monitors throughout the US
- We will analyze ozone for 31 days in July, 2005 at $n = 1,089$ stations
- Currently the EPA regulates the annual 99^{th} percentile
- Our objective is to map the probability of an extreme ozone event

Ozone on July 10



Q-Q plots

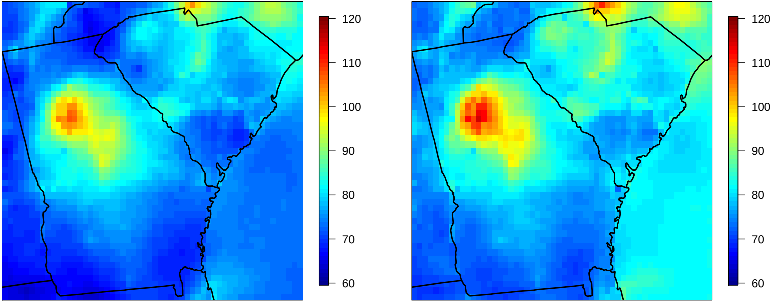


Gaussian Q-Q plot (left) and skew-t with $a = 10$ and $\lambda = 1$ Q-Q plot (right)

Cross-validation

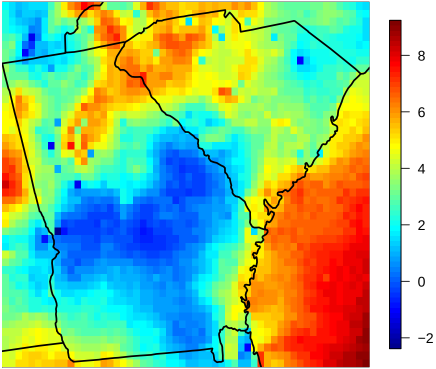
- We split the sites into training and testing
- We found that $K = 15$ knots and censoring at T equal to the median with no time series gave the best results
- Results were not sensitive to these tuning parameters
- This model was 5% more accurate (Brier score) than GP
- The max-stable model fit was 15% less accurate than GP

Fitted 99^{th} percentile - Gaussian



Gaussian (left) Symmetric-t, 10 knots, T = 75, Time series (right)

Difference (Thresholded t - Gaussian)



Difference between Symmetric-t, 10 knots, T = 75 and Gaussian

Summary

- Our proposed method can handle large datasets
- The proposed model gives a balance between theoretical properties and computational feasibility
- Work supported by NSF, NIH, DOI, and EPA
- Thanks!