A hierarchical Gauss-Pareto model for extreme precipitation

Application to storms in southern Sweden

Robert A. Yuen, University of Michigan and Peter Guttorp, University of Washington

Modeling spatial extremes

Why the standard geostatistical approach is undesireable

- Gaussian processes $\{W(s)\}_{s\in\mathbb{R}^2}$ are *independence* models for spatial extremes, i.e. the probability of two simultaneously extreme observations tends to zero:

$$\lim_{t\to\infty}\frac{\mathbb{P}(\min\{W(s_1),W(s_2)\}>t)}{\mathbb{P}(W(s_1)>t)}=0$$

- Insufficient in capturing aggregate effects of extreme precipitation leading to soil saturation and subsequent landslides and flooding.
- Marginals are not Pareto or generalized extreme value distributions that arise from exceedances or block-maxima.

We seek a statistical model that...

- Exhibits non-trivial tail dependence.
- Can be fit with established methods.
- Provides predictions at unobserved locations.

A non-trivial storm model

$$V(s):=Z\exp\{\varepsilon(s)+B(s)-\gamma(s)\}. \tag{1}$$

- Z - generalized Pareto distributed (GPD) with distribution function

$$G(z)=1-(1+\xi z)_+^{-1/\xi},$$

characterizes the overall intensity of the storm.

- $\{B(s)\}_{s\in\mathbb{R}^2}$ - Gaussian process, independent of Z with semi-variogram

$$\gamma(s)=(\|s-\omega\|/\lambda)^{\alpha},\ \lambda>0,\alpha\in(0,2),$$

where $B(\omega)=0$ a.s.

  - ω and λ are the *random* center and range of a storm.
  - α controls smoothness of the storm profile.

- $\{\varepsilon(s)\}_{s\in\mathbb{R}^2}$ - large scale trend surface, captures local effects such as elevation.

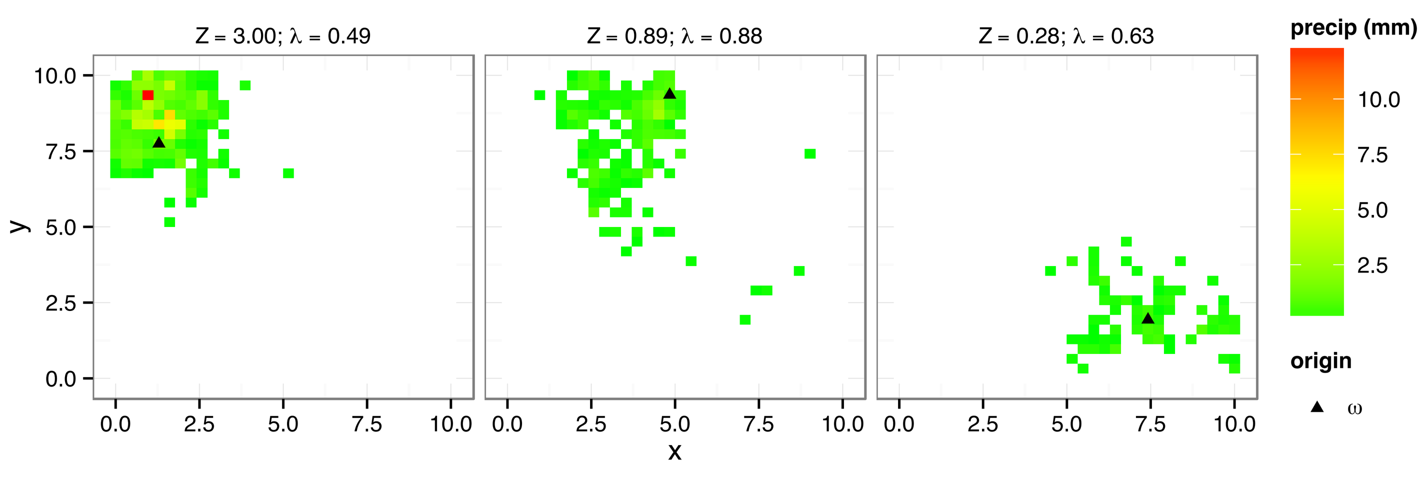

Figure 1: *Three simulated storms from the Gauss-Pareto model with $\alpha=1$ and $\varepsilon\equiv 0$. Precipitation below 0.1mm has been censored.*

Inference

An MCMC algorithm was developed to fit the model (1) using three hierarchies

$$[\text{Data}]\quad\times\quad[\text{Process}]\quad\times\quad[\text{Prior}]$$

$$p(\boldsymbol{y}|Z,\varepsilon,\lambda,\omega,\alpha)\times p(Z|\xi)p(\varepsilon|\theta)\times p(\lambda,\omega,\alpha,\xi,\theta|\vartheta)$$

where

$$Y(s):=\{\log V(s)|Z,\varepsilon,\omega,\lambda,\alpha,\xi\}_{s\in\mathbb{R}^2} \tag{2}$$

is a Gaussian process with covariance

$$\Sigma(s_1,s_2)=\lambda^{-\alpha}\{\gamma(s_1)+\gamma(s_1)-\|s_1-s_2\|^{\alpha}\},$$

and mean

$$\mu(s):=\log Z+\varepsilon(s)-\gamma(s).$$

Hence, d-dimensional projections $\mathbf{Y}=(Y(s_1),\ldots,Y(s_d))$ of (2) have density

$$p(\boldsymbol{y}|Z,\varepsilon,\lambda,\omega,\alpha)=N_d(\boldsymbol{\mu},\Sigma). \tag{3}$$

MCMC Details

- Gibbs sampler developed with a Gaussian process model for $p(\varepsilon|\theta)$.
- By construction $p(Z|\xi)$ is GPD.
- Independent proper priors specified for $p(\omega,\lambda,\alpha,\xi,\theta|\vartheta)$.
- Posterior predictive distributions at unobserved locations $\tilde{s}$ are generated by sampling from the Gaussian distribution $p(Y(\tilde{s})|\boldsymbol{Y})$ determined by (3).
- Censored observations can also be incorporated by sampling from truncated Gaussian at each MCMC iteration.

Storms in southern Sweden

- Extreme 24 hour precipitation recorded by the Swedish Meteorological and Hydrological Institute (www.smhi.se) from 1961-2011 at 42 synoptic stations with 21 locations held out for validation.
- Select $n=59$ dates comprising independent extreme 24 hour precipitation events during summer months June, July and August.

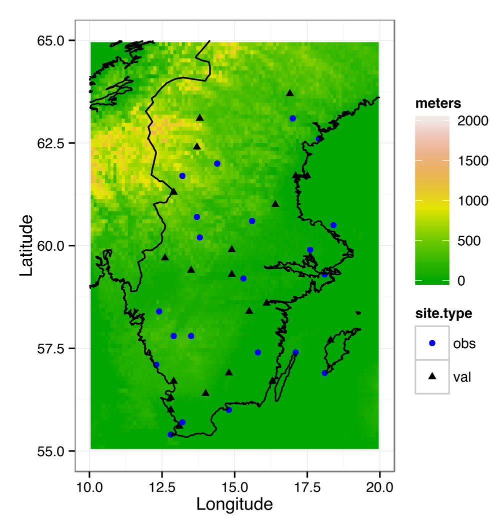

Figure 2: *Elevation map and location of 42 synoptic stations in southern Sweden. Blue indicates station was used in model fitting. Black stations were held out for validation.*

Results

- MCMC algorithm run for 60,000 iterations.
- After a burn-in period, parameter estimates are constructed using posterior means.
- We found α difficult to estimate. Hence, we determine α by comparing the negative log-likelihood (deviance) at the data level

$$D(\alpha)=-\log p(\boldsymbol{y}|Z,\varepsilon,\lambda,\omega,\alpha)$$

- Location of observed maxima $s^{\max}=\underset{s\in\{s_1,\ldots,s_{21}\}}{\arg\min}\ V(s)$ expected to be close to estimated storm center $\hat{\omega}$.
- Posterior predictive distributions $F_s(v)=\sum_{k=1}^{N}\mathbf{1}_{\{v\le v^{(k)}(s)\}}$ at validation sites are evaluated using probability integral transforms, i.e. if F_s is an ideal forecaster for $V(s)$ then $F_s(V(s))$ should be uniformly distributed.

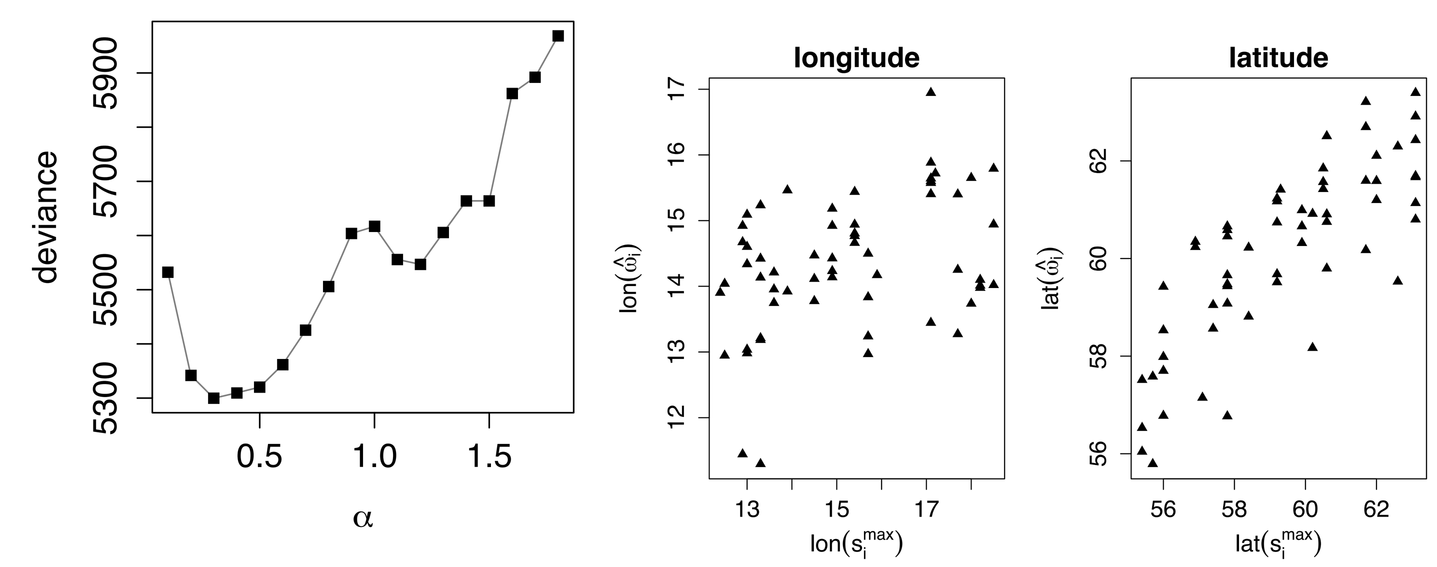

Figure 3: *(left panel) Deviance score for $\alpha\in\{0.1,0.2,\ldots,1.9\}$. (center and right panel) Location of observed maxima versus estimated storm center.*

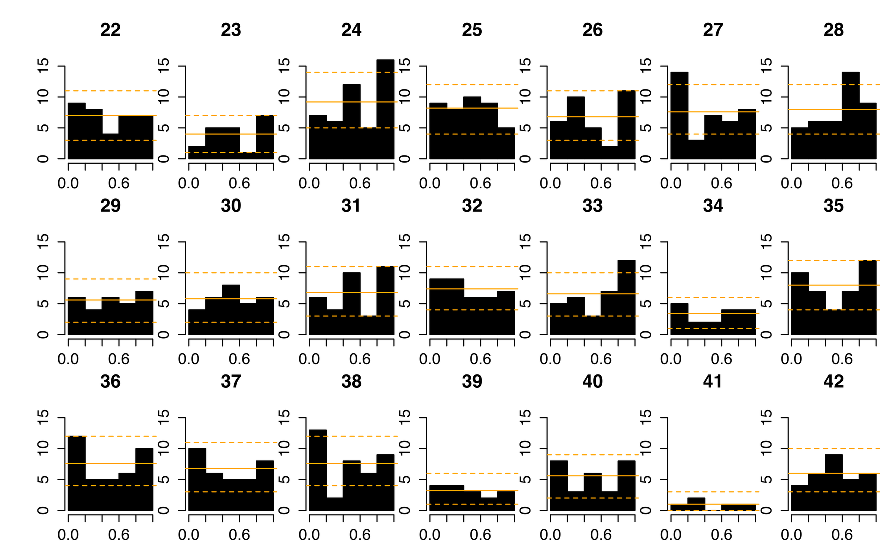

Figure 4: *Probability integral transform histograms for predictive distributions at 21 validation sites.*

For details on this research and other projects please visit:

http://www.stat.lsa.umich.edu/~bobyuen

This collaboration was made possible through the Research Network for Statistical Methods for Atmospheric and Oceanic Sciences (STATMOS) https://www.statmos.washington.edu.