PCA for extremes

Sam Morris[1], Brian J Reich[1], Emeric Thibauld[2], and Dan Cooley[2]

May 28, 2015

Abstract

words...

Key words: Max-stable process.

[1]North Carolina State University
[2]Colorado State University

1 Introduction

2 Model

Let Y_{it} be the observation at location $\mathbf{s}_i$ for $i \in \{1, ..., n_s\}$ and time $t \in \{1, ..., n_t\}$. To focus attention on the extreme values, we consider data above a threshold T. The marginal distribution of Y_{it} is then determined by the probability of exceeding the threshold and the distribution of the excursions. Denote the exceedance probability as Prob$[Y_{it} > T] = p_{it}$. Extreme value theory says that for sufficiently large T the excursion distribution can be approximated using a generalized Pareto distribution (GPD). Therefore we model $Y_{it}|Y_{it} > T \sim$ GDP(σ_{it}, ξ), where the GDP scale and shape parameters are denoted $\sigma_{it} > 0$ and ξ, respectively.

spectral, max-linear...finally we settle on... Spatial extremal dependence is captured using a max-stable copula (define). Let Z_{it} be a max-stable process with Fréchet marginal distributions (define GEV etc...). Our objective is to identify a low-rank model for spatial dependence in Z_{it}. Decompose Z_{it} as $Z_{it} = \theta_{it}\varepsilon_{it}$ where θ_{it} is a spatial process and $\varepsilon_{it} \overset{iid}{\sim}$ GEV$(1, \alpha, \alpha)$ is a nugget. The spatial component is written as a combination of L basis functions B_{il}

$$\theta_{it} = \left(\sum_{l=1}^{L} B_{il}^{1/\alpha} A_{lt} \right)^{\alpha}. \tag{1}$$

If $B_{il} > 0$, $\sum_{l=1}^{L} B_{il} = 1$, and the A_{lt} have positive stable (PS) distribution $A_{lt} \sim$ PS(α) (define), then Z_{it} is max-stable and has Fréchet marginal distributions.

The Z_{it} are conditionally independent given the spatial random effects, with conditional distribution $Z_{it}|\theta_{it} \sim$. As a result, the likelihood is $Y_{it}|\theta_{it} \overset{indep}{\sim} g(y; \theta_{it}, p_{it}, \sigma_{it}, \xi)$ where

$$g(y; \theta, p, \sigma, \xi) = \tag{2}$$

Therefore, the likelihood factors across observations which is computationally convenient. Marginalizing over the random effect θ_{it} induces extremal spatial dependence in the Z_{it}, and thus the Y_{it}.

Spatial dependence can be summarized by the extremal coefficient (EC) $\vartheta_{ij} \in [1, 2]$, where

$$\text{Prob}(Z_{it} < c, Z_{jt} < c) = \text{Prob}(Z_{it} < c)^{\vartheta_{ij}}. \tag{3}$$

For the PS random effects model the EC has the form

$$\vartheta_{ij} = \sum_{l=1}^{L} \left(B_{il}^{1/\alpha} + B_{jl}^{1/\alpha} \right)^{\alpha}. \tag{4}$$

In particular, $\vartheta_{ii} = 2^\alpha$ for all i. Since $\sum_{l=1}^{L} B_{il} = 1$ for all i, we have $\sum_{l=1}^{L}(\sum_{i=1}^{n_s} B_{il}/n_s) = 1$. Therefore, the relative contribution of term l can be measured by

$$v_l = \sum_{i=1}^{n_s} B_{il}/n_s, \tag{5}$$

with $\sum_{l=1}^{L} v_l = 1$. The order of the terms is arbitrary, and so we assume without loss of generality that $v_1 \geq ... \geq v_L$.

3 Estimating the extremal coefficient function

In this section we develop an algorithm to estimate the spatial dependence parameter α and the $n_s \times L$ matrix $\mathbf{B} = \{B_{il}\}$. Given these parameters, we plug them into our model and proceed with Bayesian analysis as described in Section 4. Our algorithm has the following 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 by minimizing the difference between model-based coefficients, ϑ_{ij}, and smoothed coefficients, $\tilde{\vartheta}_{ij}$.

To estimate the spatial dependence we first remove variation in the marginal distribution. Let $U_{it} = \sum_{k=1}^{n_t} I[Y_{ik} < Y_{it}]/n_t$, so that the U_{it} are approximately uniform at each location. Then for some extreme probability $q \in (0,1)$, solving (3) suggest the estimate

$$\hat{\vartheta}_{ij}(q) = \frac{\log[Q_{ij}(q)]}{\log(q)}, \tag{6}$$

where $Q_{ij}(q) = \sum_{t=1}^{n_t} I[U_{it} < q, U_{jt} < q]/n_t$ is the sample proportion of the time points at which both sites are less then q. Since all large q give valid estimates, we average over a grid of q with $q_1 < ... < q_{n_q}$

$$\hat{\vartheta}_{ij} = \frac{1}{n_q}\sum_{j=1}^{n_q} \hat{\vartheta}_{ij}(q_j). \tag{7}$$

Assuming the true B_{il} are smooth over space, the initial estimates $\hat{\vartheta}_{ij}$ can be improved by smoothing. Let

$$\tilde{\vartheta}_{ij} = \frac{\sum_{u=1}^{n_s}\sum_{v=1}^{n_s} w_{iu}w_{jv}\hat{\vartheta}_{uv}}{\sum_{u=1}^{n_s}\sum_{v=1}^{n_s} w_{iu}w_{jv}}, \tag{8}$$

where $w_{iu} = \exp(-\phi||\mathbf{s}_i - \mathbf{s}'_u||^2)$ is the Gaussian kernel function with bandwidth ϕ. The elements $\hat{\vartheta}_{ii}$ do not contributed any information as $\hat{\vartheta}_{ii} = 1$ for all i by construction. To eliminate the influence of these estimates we set $w_{ii} = 0$. However, this approach does give imputed values $\tilde{\vartheta}_{ii}$, which provides information about small-scale spatial variability.

The dependence parameters are estimated by comparing estimates $\tilde{\vartheta}_{ij}$ with the model-based values ϑ_{ij}. For all i, $\vartheta_{ii} = 2^{\alpha}$, and therefore we set α to $\hat{\alpha} = \log_2(\sum_{i=1}^{n_s}\tilde{\vartheta}_{ii}/n_s)$. Given $\alpha = \hat{\alpha}$, it remains to estimate $\mathbf{B}$. These estimate $\hat{\mathbf{B}}$ is taken as the minimizer of

$$m(\mathbf{B}) = \sum_{i<j}\left(\tilde{\vartheta}_{ij} - \vartheta_{ij}\right)^2 = \sum_{i<j}\left\{\tilde{\vartheta}_{ji} - \sum_{l=1}^{L}[B_{il}^{1/\hat{\alpha}} + B_{jl}^{1/\hat{\alpha}}]^{\hat{\alpha}}\right\}^2 \tag{9}$$

under the restrictions that $B_{il} \geq 0$ for all i and l and $\sum_{l=1}^{L} B_{il} = 1$ for all i.

The order of the B_{il} is not defined. Therefore, we sort the terms so that $v_1 > ... > v_L$.

4 Implementation details

The model has three tuning parameters: the quantile threshold q, the kernel bandwidth ϕ, and the number of terms L. How to pick? Say $q = 0.95$ or whatever seems to give GPD marginals. ϕ is something reasonable. For L, we start small and increase until the smallest proportion v_L is less than, say 0.05.

Given the estimates of α and $\mathbf{B}$, the hierarchical model is

$$
\begin{aligned}
Y_{it}|\theta_{ij} &\overset{indep}{\sim} g(y;\theta_{it},p_{it},\sigma_{it},\xi) \qquad (10)\\
\theta_{it} &= \left(\sum_{l=1}^{L}\hat{B}_{il}^{1/\hat{\alpha}}A_{lt}\right)^{\hat{\alpha}} \quad \text{where} \quad A_{lt}\overset{iid}{\sim} PS(\hat{\alpha})\\
\text{logit}(p_{it}) &= \mathbf{X}_{it}^T\boldsymbol{\beta}_1 \quad \text{and} \quad \log(\sigma_{it}) = \mathbf{X}_{it}^T\boldsymbol{\beta}_2
\end{aligned}
$$

where g is given in (2) and $\mathbf{X}_{it} = (X_{it1},...,X_{itp})^T$ is a vector of spatiotemporal covariates. To complete the Bayesian model, we select independent normal priors with mean zero and variance 100 for the components of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ and standard normal prior for ξ.

We estimate parameters $\Theta = \{A_{lt},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\xi\}$ using Markov chain Monte Carlo. Details...

5 Data analysis

5.1 Results

5.2 Model checking and sensitivity analysis

6 Conclusions

Acknowledgements

References