A new spatial model for points above a threshold

August 15, 2014

1 Introduction

In most climatological applications, researchers are interested in learning about the average behavior of different climate variables (e.g. ozone, temperature, rainfall). However, averages do not help regulators prepare for the unusual events that only happen once every 100 years. For example, it is important to have an idea of how much rain will come in a 100-year floor in order to construct strong enough river levees to protect lands from flooding.

The extremal coefficient describes the pairwise dependence between spatial locations (Smith, 1990). Consider a spatial process $Y(\mathbf{s}) \in \mathcal{R}^n$ observed at locations $s \in \mathcal{D} \subset \mathcal{R}^2$. Then the bivariate extremal coefficient, $\theta(\mathbf{s}_i, \mathbf{s}_j) \in [1, 2]$, is defined as

$$\Pr(Y(\mathbf{s}_i) < c, Y(\mathbf{s}_j) < c) = \Pr(Y(\mathbf{s}_i) < c)^{\theta(\mathbf{s}_i, \mathbf{s}_j)}. \tag{1}$$

One way to characterize the dependence over the entire set of spatial locations is to calculate all of the pairwise extremal coefficients. Although this method provides information regarding the spatial structure of the observations, it does not fully characterize the joint spatial dependence.

Alternatively, one can use the χ coefficient as a measure of extremal dependence. The χ coefficient for the upper tail is given by

$$\chi = \lim_{c \to \infty} \Pr(Y(\mathbf{s}_1) > c | Y(\mathbf{s}_2) > c)$$

In a stationary spatial process, we can write the χ coefficient as

$$\chi(\mathbf{h}) = \lim_{c \to \infty} \Pr(Y(0) > c | Y(\mathbf{h}) > c)$$

where $\mathbf{h} = ||\mathbf{s}_1 - \mathbf{s}_2||$. If $\chi(\mathbf{h}) = 0$, then observations are asymptotically independent at distance $\mathbf{h}$. For Gaussian processes, $\chi(\mathbf{h}) = 0$ regardless of the the distance, so they are not suitable for modeling spatially dependent extremes.

Unlike multivariate normal distributions, it is challenging to model multivariate extreme value distributions (e.g. generalized extreme value and generalized Pareto distribution) because few closed-form expressions exist for the density in more than two-dimensions (Coles and Tawn, 1991). Given this limitation, pairwise composite likelihoods have been used when modeling dependent extremes (Padoan et al., 2010; Blanchet and Davison, 2011; Huser, 2013). One way around the multi-dimensional limitation of multivariate extreme value distributions is to use skew elliptical distributions to model dependent extreme values (Genton, 2004; Zhang and El-Shaarawi, 2010; Padoan, 2011). Due to their flexibility, the skew-normal and skew-t distribution offer a flexible way to handle non-symmetric data within a framework of multivariate normal and multivariate t-distributions. As with the spatial Gaussian process, the skew-normal distribution is also asymptotically independent; however, the skew-t does demonstrate asymptotic dependence (Padoan, 2011).

In this paper, we present a model that has marginal distributions with flexible tails, demonstrates asymptotic dependence for $\mathbf{h}$ small, and has computation on the order of Gaussian models for large space-time datasets. Specifically, we use a multivariate skew-t distribution that incorporates thresholding and random spatial partitions. The advantage of using a thresholded model as opposed to a non-thresholded model is that is allows for the tails of the distribution to inform the predictions in the tails (DuMouchel, 1983). The random spatial partitions increase the flexibility of the model by giving us a way to account for observations in spatial regions that may have higher variability due to certain regional climate variables (e.g. forest fires).

2 Statistical model

Let $Y_t(\mathbf{s}) \in \mathcal{R}$ be the observed value at location $\mathbf{s}$ on day t. To avoid bias in estimating tail parameters, we model thresholded data

$$\tilde{Y}_t(\mathbf{s}) = \begin{cases} Y_t(\mathbf{s}) & Y_t(\mathbf{s}) > T \\ T & Y_t(\mathbf{s}) \leq T \end{cases} \tag{2}$$

where T is a pre-specified threshold. Then, assuming the full data follow a skew-t distribution, we update values below the threshold using standard Bayesian missing data methods as described in Section 3.

2.1 Complete data

We assume the data can be modeled as skew-t (Zhang and El-Shaarawi, 2010)

$$Y_t(\mathbf{s}) = X_t(\mathbf{s})\beta + \alpha z_t + \sigma_t v_t(\mathbf{s}) \tag{3}$$

where $\alpha \in \mathcal{R}$ controls the skewness, $z_t \overset{ind}{\sim} N_{(0,\infty)}(0, \sigma_t^2)$ are a random effect from a half-normal distribution (see appendix A.3), $v_t(\mathbf{s})$ is a Gaussian process with mean zero, variance one, and Matérn correlation, and $\sigma_t^2 \overset{iid}{\sim} \text{IG}(a, b)$. When marginalizing over the z_t and σ_t^2 terms,

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

where μ is the location, $\Sigma^* = \frac{b}{a}\Sigma$, Σ is a Matérn covariance matrix, a, b are the parameters from the inverse gamma distribution for σ_t^2, and $\alpha \in \mathcal{R}$ controls the skewness. The skew-t process is desirable because it has a flexible tail that is controlled by both the skewness parameter α and the degrees of freedom. This process also has the desirable characteristic that $\chi(\mathbf{h}) > 0$ (Padoan, 2011).

2.2 Random daily partition

One problem with the skew-t distribution is that $\chi(\mathbf{h})$ does not suggest asymptotic independence as the distance increases. This occurs because all observations, both near and far, share the same z_t and σ_t^2 terms. We handle this problem with a daily random partition similar to Huser and Davison (2014) that allows z_t and σ_t^2 to vary by site. The model then becomes

$$Y_t(\mathbf{s}) = X_t(\mathbf{s})\beta + \alpha z_t(\mathbf{s}) + \sigma_t(\mathbf{s}) v_t(\mathbf{s}) \tag{4}$$

where α controls the skewness, $z_t(\mathbf{s}) \overset{ind}{\sim} \text{N}_{(0,\infty)}(0, \sigma_t^2(\mathbf{s}))$, $v_t(\mathbf{s})$ is defined as before, and $\sigma_t^2(\mathbf{s}) \overset{iid}{\sim} \text{IG}(a, b)$. Consider a set of daily knots $\mathbf{w}_{tk} \sim$ Uniform that define a random daily partition $P_{t1}, \ldots, P_{tK}$ such that

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

So, for $\mathbf{s} \in P_{tk}$, let

$$z_t(\mathbf{s}) = z_{tk} \tag{5}$$

$$\sigma_t^2(\mathbf{s}) = \sigma_{tk}^2. \tag{6}$$

Then within each partition, $Y_t(\mathbf{s})$ follows the distribution given in (3).

2.3 Hierarchical model

Conditioned on $z_t(\mathbf{s})$ and $\sigma_t^2(\mathbf{s})$, $Y_t(\mathbf{s})$ is a Gaussian spatial model. Thus, standard geostatistical methods can be used to fit the model, and predictions can be made by Kriging at unobserved locations. We model this with a Bayesian hierarchical model as follows. Let $w_{t1}, \ldots, w_{tK}$ be partition centers so that

$$P_{tk} = \{\mathbf{s} : k = \arg\min_\ell ||\mathbf{s}_t - w_{t\ell}||\}.$$

Then

$$
\begin{aligned}
Y_t(\mathbf{s}) \mid \Theta, z_t(\mathbf{s}) &= X_t(\mathbf{s})\beta + \alpha z_t(\mathbf{s}) + \sigma_t(\mathbf{s}) v_t(\mathbf{s}) && (7)\\
z_{tk} \mid \sigma^2_{tk} &\sim N_{(0,\infty)}(0, \sigma^2_{tk}) && (8)\\
\sigma^2_{tk} &\overset{iid}{\sim} IG(\alpha, \beta) && (9)\\
v_t(\mathbf{s}) \mid \Theta &\sim \text{Matérn } (0, \Sigma) && (10)\\
\alpha &\sim N(0, 10) && (11)\\
w_{tk} &\sim Unif(\mathcal{R}^2) && (12)
\end{aligned}
$$

where $\Theta = \{w_{t1}, \ldots, w_{tK}, \beta, \sigma_t, \alpha, \lambda, \rho, \nu\}$; $z_t(\mathbf{s})$ and $\sigma^2_t(\mathbf{s})$ are defined as in (5) and (6); $k = \arg\min_\ell ||\mathbf{s} - \mathbf{w}_\ell||$; and Σ is a Matérn covariance matrix with variance one, spatial range ρ, smoothness ν.

3 Computation

The MCMC for this model is fairly straightforward. First, we impute values below the threshold. Then, we update Θ using random walk MH or Gibbs sampling when appropriate. Finally, we make spatial precictions using conditional multivariate normal results and the fact that the distribution of $Y_t(\mathbf{s}) \mid \Theta, z_{tl}$ is the usual multivariate normal distribution with a Matérn spatial covariance structure.

3.1 Imputation

We can use Gibbs sampling to update $Y_t(\mathbf{s})$ for observations that are below T, the thresholded value. Given Θ, $Y_t(\mathbf{s})$ has truncated normal full conditional with these parameter values. So we sample $Y_t(\mathbf{s}) \sim N_{(-\infty,T)}(\mu(\mathbf{s}), \Sigma)$

3.2 Parameter updates

To update Θ given the current value of the complete data $\mathbf{Y}_1, \ldots, \mathbf{Y}_T$, we use a standard Gibbs updates for all parameters except for the knot locations which are done using a Metropolis update. See Appendix A.1 for details regarding Gibbs sampling.

3.3 Spatial prediction

Given $\mathbf{Y}_t$ the usual Kriging equations give the predictive distribution for $Y_t(\mathbf{s}^*)$ at prediction location $(\mathbf{s}^*)$

4 Data analysis

5 Conclusions

Acknowledgments

Appendix A.1: Posterior distributions

Conditional posterior of $z_{tl} \mid \ldots$

For simplicity, drop the subscript t and define

$$R(\mathbf{s}) = \begin{cases} Y(\mathbf{s}) - X(\mathbf{s})\beta & s \in P_l \\ Y(\mathbf{s}) - X(\mathbf{s})\beta - \alpha z(\mathbf{s}) & s \notin P_l \end{cases}$$

Let

$$\begin{aligned} R_1 &= \text{the vector of } R(\mathbf{s}) \text{ for } s \in P_l \\ R_2 &= \text{the vector of } R(\mathbf{s}) \text{ for } s \notin P_l \\ \Omega &= \Sigma^{-1}. \end{aligned}$$

Then

$$\begin{aligned} \pi(z_l | \ldots) &\propto \exp\left\{ -\frac{1}{2} \left[\begin{pmatrix} R_1 - \alpha z_l \mathbf{1} \\ R_2 \end{pmatrix}^T \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \begin{pmatrix} R_1 - \alpha z_l \mathbf{1} \\ R_2 \end{pmatrix} + \frac{z_l^2}{\sigma_l^2} \right] \right\} I(z_l > 0) \\ &\propto \exp\left\{ -\frac{1}{2} \left[\Lambda_l z_l^2 - 2\mu_l z_l \right] \right\} I(z_l > 0) \end{aligned}$$

where

$$\begin{aligned} \mu_l &= \alpha (R_1^T \Omega_{11} + R_2^T \Omega_{21}) \mathbf{1} \\ \Lambda_l &= \alpha^2 \mathbf{1}^T \Omega_{11} \mathbf{1} + \frac{1}{\sigma_l^2}. \end{aligned}$$

Then $Z_l | \ldots \sim N_{(0,\infty)}(\Lambda_l^{-1}\mu_l, \Lambda_l^{-1})$

Conditional posterior of $\beta \mid \ldots$

Let $\beta \sim \mathrm{N}_p(0, \Lambda_0)$ where Λ_0 is a precision matrix. Then

$$\begin{aligned} \pi(\beta \mid \ldots) &\propto \exp\left\{ -\frac{1}{2}\beta^T \Lambda_0 \beta - \frac{1}{2} \sum_{t=1}^{T} [\mathbf{Y}_t - X_t\beta - \alpha z_t]^T \Omega [\mathbf{Y}_t - X_t\beta - \alpha z_t] \right\} \\ &\propto \exp\left\{ -\frac{1}{2} \left[\beta^T \Lambda_\beta \beta - 2 \sum_{t=1}^{T} [\beta^T X_t^T \Omega (\mathbf{Y}_t - \alpha z_t)] \right] \right\} \\ &\propto \mathrm{N}(\Lambda_\beta^{-1}\mu_\beta, \Lambda_\beta^{-1}) \end{aligned}$$

where

$$\mu_\beta = \sum_{t=1}^{T} \left[X_t^T \Omega (\mathbf{Y}_t - \alpha z_t) \right]$$
$$\Lambda_\beta = \Lambda_0 + \sum_{t=1}^{T} X_t^T \Omega X_t.$$

Conditional posterior of $\sigma^2 \mid \ldots$

In the case where $L = 1$, then σ^2 has a conjugate posterior distribution. Let $\sigma_t^2 \overset{iid}{\sim} \text{IG}(\alpha_0, \beta_0)$. For simplicity, drop the subscript t. Then

$$\begin{aligned}
\pi(\sigma^2 \mid \ldots) &\propto (\sigma^2)^{-\alpha_0 - 1/2 - n/2 - 1} \exp\left\{ -\frac{\beta_0}{\sigma^2} - \frac{z^2}{2\sigma^2} - \frac{(\mathbf{Y} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{Y} - \boldsymbol{\mu})}{2\sigma^2} \right\} \\
&\propto (\sigma^2)^{-\alpha_0 - 1/2 - n/2 - 1} \exp\left\{ -\frac{1}{\sigma^2} \left[\beta_0 + \frac{z^2}{2} + \frac{1}{2} (\mathbf{Y} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{Y} - \mu) \right] \right\} \\
&\propto \text{IG}(\alpha^*, \beta^*)
\end{aligned}$$

where

$$\alpha^* = \alpha_0 + \frac{1}{2} + \frac{n}{2}$$
$$\beta^* = \beta_0 + \frac{z^2}{2} + \frac{1}{2} (\mathbf{Y} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{Y} - \boldsymbol{\mu}).$$

In the case that $L > 1$, a random walk Metropolis Hastings step will be used to update σ_{lt}^2.

Conditional posterior of $\alpha \mid \ldots$

Let $\alpha \sim N(0, \tau_\alpha)$ where τ_α is a precision term. Then

$$\begin{aligned}
\pi(\alpha \mid \ldots) &\propto \exp\left\{ -\frac{1}{2} \tau_\alpha \alpha^2 + \sum_{t=1}^{T} \frac{1}{2} [\mathbf{Y}_t - X_t \beta - \alpha z_t]^T \Omega [\mathbf{Y}_t - X_t \beta - \alpha z_t] \right\} \\
&\propto \exp\left\{ -\frac{1}{2} [\alpha^2 (\tau_\alpha + \sum_{t=1}^{T} z_t^T \Omega z_t^T) - 2\alpha \sum_{t=1}^{T} [z_t^T \Omega (\mathbf{Y}_t - X_t \beta) \right\} \\
&\propto \exp\left\{ -\frac{1}{2} [\tau_\alpha^* \alpha^2 - 2\mu_\alpha] \right\}
\end{aligned}$$

where

$$\mu_\alpha = \sum_{t=1}^{T} z_t^T \Omega (\mathbf{Y}_t - X_t \beta)$$
$$\tau_\alpha^* = t_\alpha + \sum_{t=1}^{T} z_t^T \Omega z_t.$$

Then $\alpha \mid \ldots \sim N(\tau_\alpha^{*-1} \mu_\alpha, \tau_\alpha^{*-1})$

Appendix A.2: MCMC details

Appendix A.3: Half-normal distribution

Let $u = |z|$ where $Z \sim N(\mu, \sigma^2)$. Specifically, we consider the case where $\mu = 0$. Then U follows a half-normal distribution which we denote as $U \sim HN(0, 1)$, and the density is given by

$$f_U(u) = \frac{\sqrt{2}}{\sqrt{\pi\sigma^2}} \exp\left(-\frac{u^2}{2\sigma^2}\right) I(u > 0) \tag{13}$$

When $\mu = 0$, the half-normal distribution is also equivalent to a $N_{(0,\infty)}(0, \sigma^2)$ where $N_{(a,b)}(\mu, \sigma^2)$ represents a normal distribution with mean μ and standard deviation σ that has been truncated below at a and above at b.

References

Blanchet, J. and Davison, A. C. (2011) Spatial modeling of extreme snow depth. *The Annals of Applied Statistics*, **5**, 1699–1725.

Coles, S. G. and Tawn, J. A. (1991) Modelling Extreme Multivariate Events. *Journal of the Royal Statistical Society: Series B (Methodological)*, **53**, 377–392.

DuMouchel, W. H. (1983) Estimating the stable index α in order to measure tail thickness: a critique. *The Annals of Statistics*, **11**, 1019–1031.

Genton, M. G. (2004) *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Statistics (Chapman & Hall/CRC). Taylor & Francis.

Huser, R. (2013) *Statistical Modeling and Inference for Spatio-Temporal Extremes*. Ph.D. thesis, École Polytechnique Fédérale de Lausanne.

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- and skew-normal distributions. *Journal of Multivariate Analysis*, **102**, 977–991.

Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-Based Inference for Max-Stable Processes. *Journal of the American Statistical Association*, **105**, 263–277.

Smith, R. L. (1990) Max-stable processes and spatial extremes.

Zhang, H. and El-Shaarawi, A. (2010) On spatial skewGaussian processes and applications. *Environmetrics*, 33–47.