# A new spatial model for points above a threshold

April 8, 2014

# 1 Introduction

# 2 Statistical model

Our approach is to take the construction of Schlather (2002) for a max-stable model for block maxima as a literal model for extreme daily data. We will show that his has nice theoretical and computational properties on the daily scale, and gives the popular Schlather (2002) as the limit of the block maximum data.

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 the thresholded data

$$\tilde{Y}_t(\mathbf{s}) = \begin{cases} Y_t(\mathbf{s}) & Y_t(\mathbf{s}) > T \\ T & Y_t(\mathbf{s}) \le T \end{cases} \tag{1}$$

where $T$ is a pre-specified threshold.

We first specify a model for the complete data, $Y_t(\mathbf{s})$, and then study the induced model for thresholded data, $\tilde{Y}_t(\mathbf{s})$. The full data model is given in Section **??** assuming a multivariate normal distribution with a different variance each day. Computationally, the values below the threshold are updated using standard Bayesian missing data methods as described in Section 3.

## 2.1 Complete data

Consider the model

$$Y_t(\mathbf{s}) = X_t(\mathbf{s})\beta + e_t(\mathbf{s}) \tag{2}$$

$$e_t(\mathbf{s}) = \sigma\delta|u_t(\mathbf{s})| + v_t(\mathbf{s}) \tag{3}$$

where $u_t(\mathbf{s})$ has marginal $N(0,1)$ distribution, $\delta \in (-1,1)$ controls skew, and $v_t(\mathbf{s})$ is a spatial process with mean zero and variance $\sigma^2(1-\delta^2)$.

### 2.1.1 GPD random effect

Let $r_t \overset{iid}{\sim} \text{GPD}(0, \sigma_r, \xi_r)$

### 2.1.2 Power law random effect

Let $r_t \overset{iid}{\sim} \text{GPD}(\sigma_r/\xi_r, \sigma_r, \xi_r)$. Then by marginalizing out the $r_t$ terms, we find that

$$f_Y(y) = \frac{1}{\sigma_r}\left(\frac{\xi_r}{\sigma_r}\right)^{-1/\xi_r - 1}(2\pi)^{-\frac{n}{2}}\beta^{-\left(\frac{1}{\xi_r}+\frac{n}{2}\right)}\gamma\left(\frac{1}{\xi_r}+\frac{n}{2}, \frac{\xi_r}{\sigma_r}\beta\right) \tag{4}$$

where $\beta = \frac{1}{2}(\mathbf{Y}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{Y}-\boldsymbol{\mu})$ and $\gamma(s,x)$ is the lower incomplete gamma function. We also find that marginally,

$$f(y) = \frac{1}{\sigma_r}\left(\frac{\xi_r}{\sigma_r}\right)^{-1/\xi_r - 1}(2\pi)^{-\frac{1}{2}}\beta^{-\left(\frac{1}{\xi_r}+\frac{1}{2}\right)}\gamma\left(\frac{1}{\xi_r}+\frac{1}{2}, \frac{\xi_r}{\sigma_r}\beta\right) \tag{5}$$

where $\beta = \frac{(y-\mu)^2}{2}$.

### 2.1.3 Gamma random effect

Let $r_t \stackrel{iid}{\sim}$ IG$(\xi_r, \sigma_r)$. Then after marginalizing out the random effect, we find that $Y$ follows a multivariate T distribution with $2\xi_r$ degrees of freedom, $\mu = \mu(\mathbf{s})$, and $\widehat{\Sigma} = \frac{\sigma_r}{\xi_r}\Sigma$ where $\Sigma$ is the correlation matrix from (**??**). Then, for a single location

$$P[Y_t(\mathbf{s}) > y] \quad = \quad 1 - F(y) \tag{6}$$

where $F(y)$ is the distribution function for a t-distribution with $2\xi_r$ degrees of freedom and $\mu = \mu$. The joint is

$$P[Y_t(\mathbf{s}_1) > y, Y_t(\mathbf{s}_2) > y] \quad = \quad 1 - F(y) - F(y) + F_{12}(y) \tag{7}$$

where $F_{12}(y)$ is the multivariate t-distribution described above. In this model,

$$\begin{aligned} P[Y_t(\mathbf{s}_1) > y | Y_t(\mathbf{s}_2) > y] \quad &= \quad \frac{1 - F(y) - F(y) + F_{12}(y)}{1 - F(y)} \\ &= \quad 1 - \frac{F(y) - F_{12}(y)}{1 - F(y)} = \chi(h) \end{aligned} \tag{8}$$

where $h = ||\mathbf{s}_1 - \mathbf{s}_2||$. It's not entirely obvious here, but the $h$ comes in from the $\widehat{\Sigma}$ in the t-distribution.

### 2.1.4 Addressing long-range dependence

In (8), $\chi(h)$ does not decrease to zero as distance $h = ||\mathbf{s}_1 - \mathbf{s}_2||$ goes to infinity. This is because all spatial locations share a common random effect $r_t$, and thus locations separated by very large distances are still dependent.

To remove long-range dependence, we allow the random effect $r_t$ to vary by subregion. We partition the spatial domain into $J$ subregions and allow the random effects to vary by subregion, $r_t(\mathbf{s}) = r_{tj}$ if $\mathbf{s}$ is in subregion $j$. The partitioning is treated as a random process. Let $\mathbf{v}_{t1}, ..., \mathbf{v}_{tJ}$ be spatial knots and assign

$$r_t(\mathbf{s}) = r_{tj} \text{ if } \quad j = \arg\min_l\{||\mathbf{s} - \mathbf{v}_{tl}||\}. \tag{9}$$

The knots $\mathbf{v}_{t1}, ..., \mathbf{v}_{tJ}$ are given a homogeneous Poisson process and $r_{tj} \stackrel{iid}{\sim}$ IG$(\xi_r, \sigma_r)$. Denote $\pi(h)$ as the probability that two points separated by distance $h$ are in the same subregion.

Given that two locations are in the same subdomain, their joint survival function is (6). If not,

$$P[Y_t(\mathbf{s}_1) > y, Y_t(\mathbf{s}_2) > y] \quad = \quad [1 - F(y)][1 - F(y)]. \tag{10}$$

Combining both cases gives

$$P[Y_t(\mathbf{s}_1) > y | Y_t(\mathbf{s}_2) > y] = [1 - \pi(h)][1 - F(y)] + \pi(h)\left[1 - \frac{F(y) - F_{12}(y)}{1 - F(y)}\right]. \tag{11}$$

Then

$$\lim_{y\to\infty} P[Y_t(\mathbf{s}_1) > y | Y_t(\mathbf{s}_2) > y] = \pi(h)\chi(h), \tag{12}$$

and

$$\lim_{y\to\infty, h\to\infty} P[Y_t(\mathbf{s}_1) > y | Y_t(\mathbf{s}_2) > y] = 0 \tag{13}$$

if $\lim_{h\to\infty}\pi(h) = 0$. A proof of this is given in Appendix A.1 Under this model, (12) shows there is asymptotic dependence, and (13) shows that long-range spatial dependence is eliminated.

### 2.2 Points above a threshold model

So, considering (2), the marginal distribution for the thresholded observations $\tilde{Y}_t(\mathbf{s})$ is

$$\begin{aligned} P(\tilde{Y}_t(\mathbf{s}) = T) &= \Phi\left(\frac{T - \mu(\mathbf{s})}{r_t \sigma_y}\right) \qquad (14) \\ P(\tilde{Y}_t(\mathbf{s}) | \tilde{Y}_t(\mathbf{s}) > T) &= \frac{1 - \Phi\left(\frac{Y_t(\mathbf{s}) - \mu(\mathbf{s})}{r_t \sigma_y}\right)}{1 - \Phi\left(\frac{T - \mu(\mathbf{s})}{r_t \sigma_y}\right)} \end{aligned}$$

where $\Phi(\mathbf{Z}(\mathbf{s}))$ is a multivariate normal distribution that follows $\text{Mat}(0, \sigma_W, \rho, \nu, \alpha)$. The tail dependence is the same as in (8). By allowing the location parameter $\mu(\mathbf{s})$ to vary spatially, the marginal probability of threshold exceedance vary by location. We let $\mu(\mathbf{s})$ be a Gaussian process, or maybe just a function of spatial covariates like CMAQ? We could also let $\sigma$ vary spatially to get an even more flexible model.

## 3 Computation

There are three main parts to the MCMC: (1) imputing the thresholded $Y_t(\mathbf{s})$; (2) updating the model parameters $\Theta = \{r_1, \ldots, r_m, \mu(\mathbf{s}), \sigma_Y, \rho, \nu, \alpha, \mu_r, \sigma_r, \xi_r\}$; and (3) making spatial predictions. All three require the joint distribution for complete data given $\Theta$. Denote $\mathbf{Y}_t = [Y_t(\mathbf{s}_1), ..., Y_t(\mathbf{s}_n)]^T$, and $Y_t(\mathbf{s}) = G_\Theta[W_t(\mathbf{s})]$ defined by solving (2). The distribution of $Y_t|\Theta$ is the usual multivariate normal distribution

$$f(Y_t|\Theta) = \phi_n[Y_t|\mu(\mathbf{s}), \sigma_Y, \rho, \nu_Y, r_1, \ldots, r_m] \qquad (15)$$

where $\phi_n$ is the multivariate normal pdf with Matérn correlation matrix.

### 3.1 Imputation

We can use Gibbs sampling to update $Y_t(\mathbf{s})$ for observations with $\tilde{Y}_t(\mathbf{s}) = T$. Given $\Theta$, $Y_t(\mathbf{s})$ has truncated normal full conditional with these parameter values. So we sample $Y_t(\mathbf{s}) \sim \text{TN}_{(-\infty, T)}(a, b^2)$. Imputation of the $Y_t(\mathbf{s})$ terms is done based upon the conditional multivariate normal. If we consider the full distribution of $Y_t(\mathbf{s})$, then

$$Y_t(\mathbf{s}) \sim \text{Mat}(\mu(\mathbf{s}), \sigma_Y, \rho, \nu, \alpha). \qquad (16)$$

where $\sigma_Y = r_t * \sigma_W$. So, to impute at each location below the threshold, the standard multivariate normal theory holds. So, conditional upon all other locations, at location $i$

$$Y_t(\mathbf{s}_{(i)}) | Y_t(\mathbf{s}_{(-i)}) \sim \text{TN}(a, b^2) \qquad (17)$$

where

$$\begin{aligned} a &= \mu_Y - \Sigma_{12}\Sigma_{22}^{-1}(y_{(-i)}(\mathbf{s}) - \mu_Y) \qquad (18) \\ b^2 &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \qquad (19) \end{aligned}$$

### 3.2 Parameter updates

To update $\Theta$ given the current value of the compute data $\mathbf{Y}_1, ..., \mathbf{Y}_{n_t}$, we use a standard Metropolis update using the likelihood (15).

### 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 Simulation study

Quantile scores. $T = 0.8$ is data that has been thresholded at the 80th sample quantile.

Table 1: Quantile Scores for $\alpha = .5$ settings

| | Setting 1 | | | Setting 2 | | | Setting 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\xi = -0.25$ | | $T = 0.8$ | $\xi = 0.25$ | | $T = 0.8$ | $\xi = -0.25$ | | $T = 0.9$ |
| | GPD | Gamma | Fixed | GPD | Gamma | Fixed | GPD | Gamma | Fixed |
| 0.900 | 0.182 | 0.181 | 0.182 | 0.271 | 0.270 | 0.280 | 0.121 | 0.121 | 0.127 |
| 0.950 | 0.131 | 0.131 | 0.133 | 0.202 | 0.202 | 0.202 | 0.095 | 0.095 | 0.095 |
| 0.970 | 0.098 | 0.098 | 0.105 | 0.154 | 0.154 | 0.157 | 0.075 | 0.074 | 0.077 |
| 0.990 | 0.050 | 0.049 | 0.066 | 0.077 | 0.077 | 0.092 | 0.039 | 0.038 | 0.050 |
| 0.995 | 0.032 | 0.031 | 0.051 | 0.048 | 0.048 | 0.067 | 0.025 | 0.024 | 0.040 |
| 0.999 | 0.012 | 0.011 | 0.032 | 0.015 | 0.015 | 0.037 | 0.008 | 0.008 | 0.026 |

Table 2: Quantile Scores for $\alpha = .9$ settings

| | Setting 4 | | | Setting 5 | | | Setting 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\xi = -0.25$ | | $T = 0.8$ | $\xi = 0.25$ | | $T = 0.8$ | $\xi = -0.25$ | | $T = 0.9$ |
| | GPD | Gamma | Fixed | GPD | Gamma | Fixed | GPD | Gamma | Fixed |
| 0.900 | 0.179 | 0.179 | 0.180 | 0.229 | 0.229 | 0.231 | 0.117 | 0.117 | 0.122 |
| 0.950 | 0.128 | 0.128 | 0.129 | 0.163 | 0.163 | 0.164 | 0.092 | 0.092 | 0.091 |
| 0.970 | 0.095 | 0.095 | 0.098 | 0.119 | 0.119 | 0.123 | 0.071 | 0.071 | 0.072 |
| 0.990 | 0.046 | 0.045 | 0.054 | 0.055 | 0.054 | 0.064 | 0.036 | 0.035 | 0.043 |
| 0.995 | 0.027 | 0.026 | 0.038 | 0.032 | 0.032 | 0.043 | 0.022 | 0.022 | 0.032 |
| 0.999 | 0.008 | 0.008 | 0.019 | 0.008 | 0.008 | 0.021 | 0.007 | 0.006 | 0.019 |

## 5 Data analysis

Posteriors of $\beta$ terms. $\beta_0$ is intercept, $\beta_1$ is x coordinate, $\beta_2$ is y coordinate, $\beta_3$ is CMAC coordinate.

Table 6 has quantile scores. $T = 0.8$ is data that has been thresholded at the 80th sample quantile.

Table 3: Threshold: 0.80 - GPD - Beta Posteriors

| | GPD | | | Gamma | | | Fixed | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0.025 | 0.500 | 0.975 | 0.025 | 0.500 | 0.975 | 0.025 | 0.500 | 0.975 |
| $\beta_0$ | 17.885 | 27.317 | 34.651 | 16.446 | 27.187 | 35.351 | 9.801 | 17.565 | 24.322 |
| $\beta_1$ | -4.000 | -1.352 | 1.209 | -3.993 | -1.380 | 1.123 | -3.894 | -1.240 | 1.152 |
| $\beta_2$ | -1.092 | 0.694 | 2.516 | -1.017 | 0.720 | 2.516 | -1.269 | 0.975 | 3.064 |
| $\beta_3$ | 0.456 | 0.569 | 0.707 | 0.444 | 0.570 | 0.727 | 0.631 | 0.726 | 0.834 |

Table 4: Threshold: 0.90 - GPD - Beta Posteriors

| | GPD | | | Gamma | | | Fixed | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0.025 | 0.500 | 0.975 | 0.025 | 0.500 | 0.975 | 0.025 | 0.500 | 0.975 |
| $\beta_0$ | 11.377 | 24.969 | 35.848 | 12.893 | 25.372 | 37.379 | -3.209 | 11.777 | 22.724 |
| $\beta_1$ | -4.299 | 0.060 | 5.691 | -4.513 | -0.171 | 4.965 | -4.983 | -0.801 | 4.534 |
| $\beta_2$ | -6.198 | -2.082 | 0.963 | -5.644 | -1.789 | 1.288 | -6.164 | -2.269 | 1.175 |
| $\beta_3$ | 0.514 | 0.663 | 0.829 | 0.492 | 0.649 | 0.825 | 0.696 | 0.845 | 1.028 |

Table 5: Threshold: 0.95 - GPD - Beta Posteriors

| | GPD | | | Gamma | | | Fixed | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0.025 | 0.500 | 0.975 | 0.025 | 0.500 | 0.975 | 0.025 | 0.500 | 0.975 |
| $\beta_0$ | -21.936 | 16.245 | 29.874 | -14.792 | 11.460 | 32.541 | -47.356 | -8.725 | 12.051 |
| $\beta_1$ | 2.948 | 12.280 | 26.863 | 3.497 | 13.160 | 27.162 | 2.788 | 13.244 | 29.326 |
| $\beta_2$ | -14.984 | -8.785 | -0.523 | -17.402 | -8.016 | 0.920 | -19.640 | -9.499 | -0.563 |
| $\beta_3$ | 0.599 | 0.773 | 1.164 | 0.567 | 0.807 | 1.073 | 0.807 | 1.057 | 1.454 |

Table 6: Quantile Scores for 5-fold cross-validation

| | $T = 0.80$ | | | $T = 0.95$ | | | $T = 0.99$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | GPD | Gamma | Fixed | GPD | Gamma | Fixed | GPD | Gamma | Fixed |
| 0.900 | 4.870 | 4.870 | 4.842 | 4.925 | 4.914 | 4.900 | 5.501 | 5.229 | 5.629 |
| 0.950 | 2.861 | 2.861 | 2.837 | 2.871 | 2.865 | 2.861 | 3.244 | 3.077 | 3.290 |
| 0.970 | 1.891 | 1.890 | 1.881 | 1.898 | 1.896 | 1.891 | 2.155 | 2.047 | 2.175 |
| 0.990 | 0.742 | 0.742 | 0.743 | 0.749 | 0.748 | 0.741 | 0.853 | 0.816 | 0.855 |
| 0.995 | 0.401 | 0.400 | 0.403 | 0.405 | 0.404 | 0.399 | 0.461 | 0.442 | 0.464 |
| 0.999 | 0.092 | 0.093 | 0.093 | 0.096 | 0.096 | 0.094 | 0.102 | 0.096 | 0.106 |

# 6 Conclusions

# Acknowledgments

# Appendix A.1: Proof that $\lim_{h\to\infty} \pi(h) = 0$

Consider two points $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{D} \subset \mathcal{R}^2$ with $h = ||\mathbf{s}_1 - \mathbf{s}_2||$, a set of random knots $\mathbf{v}_1, \dots, \mathbf{v}_J$. Define $g(\mathbf{s}_i) = \arg\min_j ||\mathbf{s}_i - \mathbf{v}_j||$ to be the partition in which $\mathbf{s}_i$ resides. Let $\pi(h) = \Pr[g(\mathbf{s}_1) = g(\mathbf{s}_2)]$. Then $\pi(h)$ is the probability that $\mathbf{v}_j$ is the only knot in $(B_a(\mathbf{s}_1) \cup B_b(\mathbf{s}_2) \cap \mathcal{D})$ where $B_r(\mathbf{s})$ is a ball of radius $r > 0$ centered at point $\mathbf{s}$. Assume that $g(\mathbf{s}_1) = g(\mathbf{s}_2) = j$, and let $a = ||\mathbf{s}_1 - \mathbf{v}_j||$ and $b = ||\mathbf{s}_2 - \mathbf{v}_j||$. Then

$$\pi(h) = \Pr\left[\mathbf{v}_2, \dots, \mathbf{v}_J \in (B_a(\mathbf{s}_1) \cup B_b(\mathbf{s}_2))^C\right] = \left(1 - \frac{|B_a(\mathbf{s}_1) \cup B_b(\mathbf{s}_2)|}{|\mathcal{D}|}\right)^{J-1} \tag{20}$$

where $|\mathcal{S}|$ is the area of $\mathcal{S} \cap \mathcal{D}$. We know that $\max(|B_a(\mathbf{s}_1) \cup B_b(\mathbf{s}_2)|)$ occurs when $B_a(\mathbf{s}_1) \cap B_b(\mathbf{s}_2) = 0$ and $\min(|B_a(\mathbf{s}_1) \cup B_b(\mathbf{s}_2)|)$ occurs when either $B_a(\mathbf{s}_1) \subset B_b(\mathbf{s}_2)$ or $B_b(\mathbf{s}_2) \subset B_a(\mathbf{s}_1)$. Therefore

$$\left(1 - \frac{2\pi(a^2 + b^2)}{|\mathcal{D}|}\right)^{J-1} \leq \pi(h) \leq \left(1 - \frac{2\pi \max(a^2, b^2)}{|\mathcal{D}|}\right)^{J-1}. \tag{21}$$

Then by the triangle inequality, $h \leq a + b$, so if $h \to \infty$, then $\max(a, b) \to \infty$. Then because $h \to \infty$ then $\max(a^2, b^2) \to \infty$ and

$$\lim_{h\to\infty} \pi(h) \leq \lim_{\max(a,b)\to\infty} \left(1 - \frac{2\pi \max(a^2, b^2)}{|\mathcal{D}|}\right)^{J-1} = \left(1 - \frac{|\mathcal{D}|}{|\mathcal{D}|}\right)^{J-1} = 0. \tag{22}$$

Therefore $\lim_{h\to\infty} \pi(h) = 0$.

# Appendix A.2: MCMC Details

## Priors

For a given day

$$\begin{aligned}
r_j &\overset{iid}{\sim} \text{IG}(\xi_r, \sigma_r) \\
\sigma_r &\sim \text{Gamma}(0.1, 0.1) \\
\xi_r &\sim \text{Discrete Uniform}(0.5, 30) \\
\mathbf{v}_j &\overset{iid}{\sim} \text{Uniform}(\mathcal{D}) \\
\mu(\mathbf{s}) &\sim \text{MVN}(0, \text{diag}(10)) \\
\log(\rho) &\sim \text{N}(0, 10) \\
\log(\nu) &\sim \text{N}(-1, 1) \\
\alpha &\sim \text{Unif}(0, 1)
\end{aligned}$$

where $v_j$ are the locations of the spatial knots over $\mathcal{D}$, $\alpha$ is a parameter controlling the proportion of $r_j^2$ that is attributed to the nugget and partial sill. If $\alpha = 0$, then $r_j^2$ can be entirely attributed to the nugget effect, and if $\alpha = 1$, then $r_{tj}^2$ can be entirely attributed to the partial sill. We use Gibbs sampling for $r_j, \sigma_r$, and $\mu(\mathbf{s})$. All other parameters are sampled using a random-walk Metropolis Hastings algorithm.

# References


Schlather, M. (2002) Models for Stationary Max-Stable Random Fields. *Extremes*, **5**, 33–44. URL`http://link.springer.com/article/10.1023/A:1020977924878`.