A hierarchical Gauss-Pareto model for spatial prediction of extreme precipitation

Robert Yuen and Peter Guttorp

August 14, 2014
Technical Report #535

Abstract

We introduce a hierarchical Gauss-Pareto model for spatial prediction of 24 hour cumulative precipitation over south central Sweden, given that at least one observation is extreme. The model belongs to the max-domain of attraction of popular Brown-Resnick max-stable processes (Brown and Resnick, 1977; Kabluchko et al., 2009) and retains the essential dependence structure of their corresponding generalized Pareto processes (Ferreira and DeHaan, 2012). The hierarchical specification has flexibility to capture different range and intensities of various storms. An MCMC algorithm is developed for inference. The algorithm handles left censored data from precipitation that accumulates below reporting precision, which often happens despite nearby observations that are extreme.

1 Introduction

Some of the consequences of our changing climate is a change in the distribution of extreme rainfall (Field et al., 2012). In order to develop adaptive strategies for dealing with consequences such as insufficient drainage, soil saturation with consequent land slides, and various aspects of flooding, it is necessary to be able to estimate extremes at unobserved sites. The standard geostatistical approach is well suited to normally distributed spatial fields, but are not appropriate for fields whose marginal distributions are generalized extreme valued (GEV) or generalized Pareto (GPD), which result from, respectively, block maxima or exceedances (Coles, 2001). In order to attack this problem, climate literature suggests assimilating data into climate model output (Kharin and Zwiers, 2005). The difficulty of climate models, even on a regional scale, to reproduce extreme precipitation is well documented (e.g. Orskaug et al., 2011, showing that the high values are particularly poorly reproduced even in a regional model run in "weather forecasting" mode, using a reanalysis for boundary conditions).

Statistical tools for spatial extremes include copula methods (Sang and Gelfand, 2009; Fuentes et al., 2013) and max-stable processes (Padoan et al., 2010; Davison et al., 2012; Thibaud et al., 2013). Numerous difficulties in dealing with max-stable models hamper their use in practice, especially when spatial prediction is the goal (see e.g. Davison et al. 2012; Dombry et al., 2012; Wang and Stoev, 2011 and the references therein). Alternatively, generalized Pareto processes (Ferreira and DeHaan, 2012) have emerged as a flexible class of spatial models for extremes. Such processes arise as limiting conditional distributions given a threshold exceedance (See Section 2 below for a precise definition), and thus are natural models for spatial prediction given that nearby observations are extreme.

In this work, we propose a *Gauss-Pareto* process model for extreme precipitation which is closely related to the Pareto type models employed in recent manuscripts of Ferreira and DeHaan (2012) and Thibaud and Opitz (2013). However, there exist key differences between our methodology and previous approaches based on threshold exceedances. First, we do not assume exact asymptotic distributions, rather our Gauss-Pareto model belongs to the max-domain of attraction of limiting max-stable processes with certain spectral density governed by an underlying Gaussian distribution (i.e. *spectrally Gaussian*). The advantage is that the model can be fit using standard MCMC methods for hierarchical models with latent Gaussian structure. This greatly simplifies inference while retaining the essential dependence characteristics of the most commonly used models for spatial extremes. A second key difference is the nature in which we handle partial censoring. While most precipitation measurements are essentially left-censored due to cumulative precipitation falling below reporting precision, when working with threshold exceedance models it is common to partially censor marginal observations that fall below a much higher threshold than those arising in data collection. Justification lies in the fact that the model is derived asymptotically and thus including marginal observations that may not be approaching the asymptotic limit can lead to poor fit (Smith, 1994; Coles, 2001 Section 8.3.1). On the other hand, we found it very common that 24 hour cumulative precipitation at various locations fall

below desired thresholds even when nearby observations are extreme. This motivated us to consider a model that can account for such instances while maintaining essential tail dependence characteristics. We believe our methodology is new in this approach and allows us to consider larger spatial domains where there is greater chance of observing low cumulative precipitation given at least one extreme observation within the domain.

The rest of this paper is organized as follows: in the following Section 2 we provide some basic motivating theory, highlighting connections with existing max-stable and Pareto processes. In Section 3 we define our model, give a detailed construction of the model hierarchy and specify the MCMC fitting procedure. The main application: spatial prediction of extreme summer precipitation in south central Sweden is presented in Section 4. Finally, we conclude with summary and directions for future work.

2 Max-stable and Pareto processes

Let $\eta = \{\eta(s)\}_{s\in S}$ be a non-negative stochastic process corresponding to a physical or environmental process over a compact spatial region of interest $S \subset \mathbb{R}^2$. A classical extreme value theory approach to spatial extremes considers the limiting distribution of suitably scaled point-wise maxima

Condition 2.1. There exists a sequence of normalizing functions $a_n(s) > 0$ and $b_n(s)$ and a non-degenerate limit process $\zeta := \{\zeta(s)\}_{s\in S}$ for which the following holds

$$\left\{ \lim_{n\to\infty} \frac{\max_{i\le n} \eta_i(s) - b_n(s)}{a_n(s)} \right\}_{s\in S} \overset{f.d.d.}{=} \{\zeta(s)\}_{s\in S},$$

where η_i are independent copies of η and *f.d.d.* denotes equality in all finite dimensional distributions.

If Condition 2.1 holds, then it is well known (See e.g. Resnick, 1987 Chp. 5) that the limit process ζ must be *max-stable.* For simplicity and without loss of generality (Resnick, 1987 Prop 5.10(a)) we will assume that the max-stable process ζ is *simple max-stable,* i.e. ζ has identical margins that are standard Fréchet. All simple max-stable processes can be fully characterized by the following spectral representation

Proposition 2.2 (cf. Ferreira and DeHaan, 2012, Prop. 2.3)**.** *Let $\zeta := \{\zeta(s)\}_{s\in S}$ be a simple max-stable process then*

$$\zeta(s) \overset{d}{=} \max_{i\in\mathbb{N}} \Gamma_i^{-1} V_i(s),$$

where $\{\Gamma_i\}_{i=1}^{\infty}$ are the points of a unit rate Poisson point process on $(0,\infty)$ and $V_i(s)$ are independent copies of a stochastic process V with $EV(s) = 1$ and $E\sup_{s\in S} V(s) < \infty$ a.s.

The probability measure characterizing the distribution of V is known as the *spectral measure* of the max-stable process ζ. While max-stable models have been used recently in a variety of applications, inference suffers from a lack of tractable likelihoods (See e.g. Einmahl et al., 2012; Yuen and Stoev, 2014 and references therein). Furthermore, the complicated dependence structure imposed by taking point-wise maxima is often criticized as unrealistic because the point-wise maxima over a given time block likely occurs at different times for different locations. This can obfuscate the true space-time dependence structure of the underlying phenomena η. Furthermore, conditional sampling (prediction) with spectrally Gaussian max-stable models is not straightforward and can be quite computationally challenging (Dombry et al., 2012; Wang and Stoev, 2011).

While max-stable processes and point-wise maxima can still be a useful framework, the underlying scientific motivation for our work lies in the case where one wants to characterize the dependence structure of η given that $\eta(s)$ is *large* for some s belonging to a finite collection $\{s_1,\dots,s_d\}$ of observed locations. In statistical terms, this rather precise objective would be to characterize ρ_u a probability measure on $C^+(S)$ (the space of non-negative continuous functions on S) such that for a large threshold $u \gg 0$

$$P\left(\eta \in A | \max_{s\in\{s_1,\dots,s_d\}} \eta(s) > u\right) = \rho_u(A), \qquad (2.1)$$

for $A \in \mathcal{B}\left(C^+(S)\right)$ where $\mathcal{B}$ denotes the Borel sigma field. In practice, ρ_u can rarely be inferred directly, but it's characteristics can be approximated by limit distributions for the LHS of (2.1) in the following sense

Condition 2.3. There exists functions $a_u(s) > 0$ and $b_u(s)$, both continuous in s for every u, and for which the following holds

(i) For every $s \in S$, b_u is increasing in u.

(ii) $\lim_{u\to\infty} P\left(\eta(s) > b_u(s), \text{ for some } s \in S\right) = 0.$

(iii) There exists ρ, a non-degenerate probability measure on $C(S)$ such that for all $A \in \mathcal{B}(C(S))$ with $\rho(\partial A)=0$

$$P\left(T_u(\eta) \in A | \eta(s) - b_u(s) > 0 \text{ for some } s \in S\right) \to \rho(A), \qquad \text{as } u \to \infty \qquad (2.2)$$

where

$$T_u(\eta) := \left\{ \frac{\eta(s) - b_u(s)}{a_u(s)} \right\}_{s\in S}.$$

This suggests that if $T_u(\eta) \in A | \eta(s) - b_u(s) > 0$ converges, then under judicious normalization we may approximate of ρ_u by the limit measure ρ when u is large.

The task then becomes characterizing the class of possible limits ρ, this is the main subject of Ferreira and DeHaan (2012) and we summarize their result with the following

Theorem 2.4. *If Condition 2.3 holds, then there exists a sequence of normalizing transformations*

$$\widetilde{T}_u(\eta) := \left\{ \frac{\eta(s) - \tilde{b}_u(s)}{\tilde{a}_u(s)} \right\}_{s \in S}, \ \tilde{a}_u > 0,$$

such that for all $A \in \mathcal{B}(C(S))$

$$\lim_{u \to \infty} P\left(\tilde{T}_u(\eta) \in A | \sup_{s \in S} \left\{ \frac{\eta(s) - \tilde{b}_u(s)}{\tilde{a}_u(s)} \right\} > 1 \right) = P\left(ZX \in A\right), \quad (2.3)$$

where Z *is a Pareto random variable and* $X := \{X(s)\}_{s \in S}$ *is a non-negative stochastic process, independent of* Z*, with* $EX(s) > 0$ *for all* $s \in S$ *and* $\sup_{s \in S} X(s) = c > 0$ *a.s.*

Proof. cf. Theorem 3.2, Condition 3.2, Corollary 3.1 and Theorem 2.1(3) of Ferreira and DeHaan, 2012. □

Remark 2.5 (cf. Example 3.2 of Ferreira and DeHaan, 2012). Theorem 2.4 is equivalent to η belonging to the max-domain of attraction of a max-stable process $\zeta := \{\zeta(s)\}_{s \in S}$ with spectral representation $\zeta \overset{d}{=} \max_{i \in \mathbb{N}} \Gamma_i^{-1} V_i$, where the $\{V_i\}_{i=1}^n$ are independent copies of ZX appearing in (2.3).

Theorem 2.4 suggests that processes of the form $V = ZX$ comprise a theoretically justified class of models for threshold exceedances which share dependence characteristics of popular max-stable models. Of the max-stable processes that have been proposed to model precipitation, most have spectral measure determined by an underlying Gaussian law (See e.g. Davison et al., 2012; Thibaud et al., 2013; Yuen and Stoev, 2014). These *spectrally Gaussian* models are max-stable attractors of the Pareto processes specified by $X(s) = f(W(s))$ where f is a continuous non-negative function and $W = \{W(s)\}_{s \in S}$, is a Gaussian process. This broad specification, however, is not always useful with respect to inference or prediction.

The strategy we present here is to construct processes which share essential characteristics of popular max-stable and Pareto process models, yet remain amenable to straightforward MCMC techniques for Bayesian hierarchical models. For instance, we shall relax the requirement that $\sup_{s \in S} X(s) = c > 0$ a.s. To see why, consider the process $\bar{X} := X / \sup_{s \in S} X(s)$. Then $\sup_{s \in S} \bar{X}(s) = 1 > 0$ a.s., yet $\bar{X}$ retains the same spatial dependence structure as X. For convenience, we will call the processes $V(s) := Zf(W(s))$ *Gauss-Pareto* models. The reason for the Gaussian assumption is obvious in that it allows one to take advantage of the vast array of machinery developed for Gaussian processes including the use of parametric covariance functions, simulation of random fields and conditional sampling. Here we introduce two specifications of Gauss-Pareto models that belong to the max-domain of attraction of the two spectrally Gaussian max-stable processes that have prevailed within the literature. These include the *Schlather* or *extremal Gaussian* type (Schlather, 2002) defined by

$$V^{(SC)}(s) := Z \max\{W(s), 0\},$$

and *Brown-Resnick* (Brown and Resnick, 1977; Kabluchko et al., 2009) type models

$$V^{(BR)}(s) := Z \exp\left(W(s) - \gamma(s)\right),$$

where γ is the semi-variogram of a centered intrinsically stationary Gaussian process W. The terms *Schlather model* and *Brown-Resnick model* refer to max-stable processes as defined in Remark 2.5. Nonetheless we adopt identical names here for their associated Gauss-Pareto processes. Two previous works using max-stable versions of the Schlather and Brown-Resnick models for precipitation extremes include Davison et al. (2012) where block maxima of summer precipitation near Zurich Switzerland is considered and Thibaud et al. (2013) who fit max-stable models to threshold exceedances of precipitation in the Val Ferret catchment, which also resides in Switzerland. Davison et al. (2012) found the Brown-Resnick type models dominated the Schlater models in terms of goodness of fit criterion. Conversely Thibaud et al. (2013) found that the Schlather model produced a better fit over the small Val Farret region. This is possibly due to the fact that there exists a lower bound on the range of extremal dependence at large lags under Schlather models as characterized by the *extremal coefficient* (Davison et al., 2012). Indeed, our preliminary data analyses of extreme precipitation over south central Sweden (See Section 4 below) found that the Schlather type Gauss-Pareto model produced model fits that yielded overly strong spatial dependence between pairs of sites that were far apart. This led to spatial predictions that were biased towards heavier precipitation.

3 A log-Gauss-Pareto model for extreme precipitation

To develop our model, we began with an exploratory data analysis which revealed that the location and spatial range of extreme 24 hour cumulative precipitation over our region of interest varied greatly from storm to storm. Thus, in order to make accurate spatial predictions, it is necessary to capture both the spatial range

and profile of an extreme precipitation event as well as the approximate center. Consequently, we develop the following Brown-Resnick type model, which can be specified in a hierarchical manner to achieve such flexibility.

Definition 3.1. Let $Z \sim \text{GPD}(0,\sigma,\xi)$ and $W := \{W(s)\}_{s\in S}$ be a Gaussian process independent of Z. If

$$V(s) := Z \exp\{W(s)\}, \tag{3.1}$$

then $V := \{V(s)\}_{s\in S}$ a log-Gauss-Pareto process driven by W.

The following are immediate

1. $V(s)|W \sim \text{GPD}(0, \sigma e^{W(s)}, \xi)$, for all $s \in S$.
2. $Y|Z := \{\log(V(s))|Z\}_{s\in S}$ is a Gaussian process on S.

Without loss of generality, set $\sigma = 1$, otherwise replace $W(s)$ in Definition 3.1 with $W'(s) = W(s) + \log\sigma$. To capture varying location and range of 24 hour precipitation events, we consider processes driven by a *fractional Brownian surface* with drift. Specifically, we take the process W to be

$$W(s) = \varepsilon(s) + B(s) - (\|s-\omega\|/\lambda)^\alpha,\ \alpha \in (0,2),$$

where $B := \{B(s)\}_{s\in S}$ is a mean-zero Gaussian process with $B(\omega) = 0$ almost surely for some origin point $\omega \in S$, and covariance function given by

$$K(s_1,s_2|\boldsymbol{\theta}) = \lambda^{-\alpha} \times \{\|s_1-\omega\|^\alpha + \|s_2-\omega\|^\alpha - \|s_1-s_2\|^\alpha\} \tag{3.2}$$

with $\boldsymbol{\theta} := (\lambda,\alpha,\omega)$. Here $\|\cdot\|$ is Euclidean distance, λ determines range and α governs the smoothness of the process. Lastly, $\varepsilon(s)$ is a trend surface that captures the spatially varying scale. Hence our final model has the form

$$V(s) = Z \exp\{\varepsilon(s) + B(s) - (\|s-\omega\|/\lambda)^\alpha\}. \tag{3.3}$$

Note that the drift term $\gamma(s) := (\|s-\omega\|/\lambda)^\alpha$ is indeed the semi-variogram of the process B.

To interpret the model (3.3), consider $E(s) := \exp\{\varepsilon(s) + B(s) - (\|s-\omega\|/\lambda)^\alpha\}$ as the profile of extreme 24-hour cumulative precipitation which is roughly centered at ω. After controlling for scale within ε, the spatial maximum of the event profile is located, *near* the event center ω with high probability. Heuristically, this explained by the drift term $(\|s-\omega\|/\lambda)^\alpha$, which tends to infinity as one moves away from the origin ω. The intensity of the precipitation event is determined by Z, which exhibits the power law behavior that is characteristic of extreme precipitation. Simulated realizations from the log-Gauss-Pareto process driven by fractional Brownian surfaces are shown in Figure 1.

3.1 Model hierarchy

We now consider a series of independent extreme 24 hour precipitation events $V_i = \{V_i(s)\}_{s\in S}, i = 1,\dots,n$. The V_i are identically distributed according to (3.3), each with strength Z_i, range λ_i and center ω_i. We will assume that the overall smoothness α, and trend surface ε, remain constant across 'time' $i = 1,\dots,n$. At our disposal are measurements $V_i(s_j)$ at a sparse set of locations $s_1,\dots,s_d \in S \subset \mathbb{R}^2$. It is convenient to write the log transformation of model (3.3)

$$Y_i(s) := \log V_i(s) = \log(Z_i) + \log\left(E_i(s)\right), \tag{3.4}$$

where $\log(E_i(s)) = \varepsilon(s) + B_i(s) - (\|s-\omega_i\|/\lambda_i)^\alpha$. Observe that the origin ω_i and scale λ_i, of the fractional Brownian surface B_i vary with each independent event. Based on the the log transform (3.4), we formulate the model in three hierarchies, the *data generating level, process level,* and *prior*. Throughout, we use the notation $\boldsymbol{Y}_i = (Y_i(s_1),\dots,Y_i(s_d))$ to denote the vector of log transformed observations and $\Sigma_{\boldsymbol{\theta}_i}$ to denote the $d\times d$ matrix with entries $\Sigma_{\boldsymbol{\theta}_i}(j,k) = K(s_j,s_k|\boldsymbol{\theta}_i)$, where K is the covariance function defined in (3.2). Similarly, we denote $\boldsymbol{\varepsilon} = (\varepsilon(s_1),\dots,\varepsilon(s_d))$ and $\boldsymbol{\gamma}_{\boldsymbol{\theta}_i} = (\gamma_i(s_1),\dots,\gamma_i(s_d))$.

3.1.1 Data generation level

Let $\boldsymbol{\vartheta}_i = (Z_i,\boldsymbol{\varepsilon})$ for each event $i \in \{1\dots,n\}$. Following the model (3.4), we have that

$$\boldsymbol{Y}_i|\boldsymbol{\vartheta}_i,\boldsymbol{\theta}_i \sim N_d\left(\boldsymbol{\mu}_i, \Sigma_{\boldsymbol{\theta}_i}\right) \tag{3.5}$$

where $\boldsymbol{\mu}_i = (\mu_{i1},\dots,\mu_{id})$ is a mean vector with elements

$$\mu_{ij} = \log(Z_i) + \varepsilon(s_j) - (\|s_j-\omega_i\|/\lambda_i)^\alpha.$$

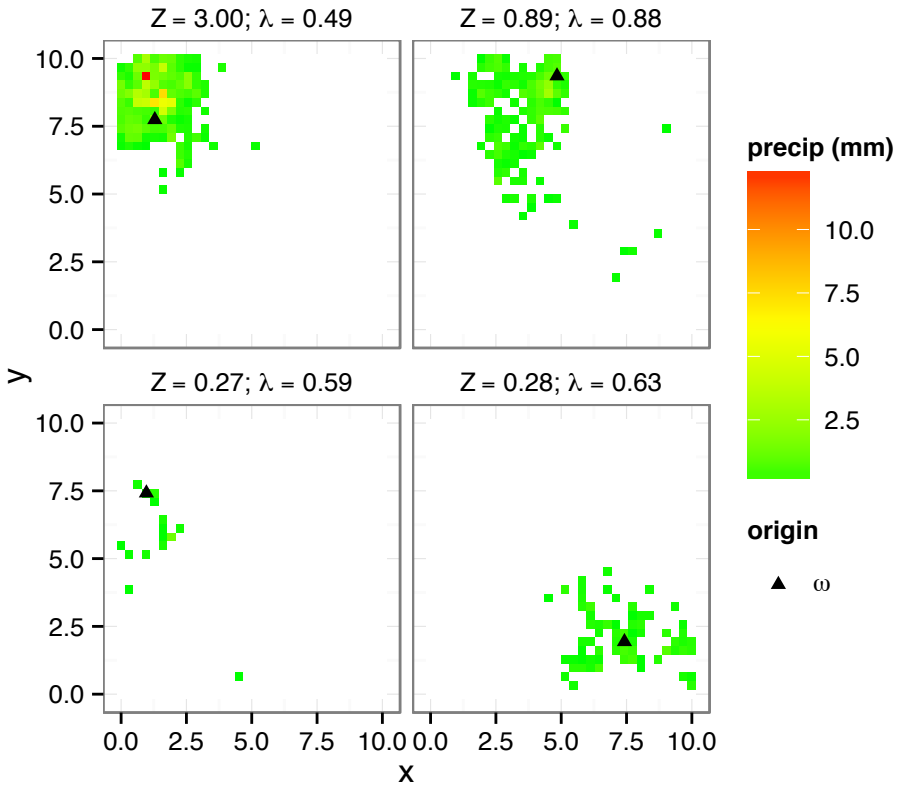
In our application, censoring limits arise from the data collection process where observations below a threshold are not available due to reporting precision. For a given precipitation event $i \in \{1,\dots,n\}$ we only observe the elements of $\boldsymbol{Y}_i$ that fall above a reporting threshold $l \in (-\infty,\infty)$. To be more concrete, we observe $\boldsymbol{Y}_{O_i}$ where $O_i \subset \{1,\dots,d\}$ indicates the subset of observations from event i that occur above the threshold l whereas $\boldsymbol{Y}_{C_i}$ with $C_i = \{1,\dots,d\}\backslash O_i$ are the censored data falling in the interval $(-\infty,l)$. Due to partial censoring, the observed information $\mathcal{D}_i$ for event i is

$$\mathcal{D}_i = \{Y_i(s_j); j \in O_i\} \cup \{Y_i(s_j) \le l; j \notin C_i\}$$

Hence, the likelihood for the process $\boldsymbol{\vartheta}_i$ given $\boldsymbol{\theta}_i, \mathcal{D}_i$ is

$$L(\boldsymbol{\vartheta}_i;\boldsymbol{\theta}_i,\mathcal{D}_i) = p(\boldsymbol{Y}_{O_i}|\boldsymbol{\vartheta}_i,\boldsymbol{\theta}_i)\int_{\boldsymbol{y}\le l\mathbf{1}} p(\boldsymbol{y}|\boldsymbol{Y}_{O_i},\boldsymbol{\vartheta}_i,\boldsymbol{\theta}_i)d\boldsymbol{y} \tag{3.6}$$

Figure 1: Four realizations from the log-Gauss-Pareto process (3.3) with $Z \sim \text{GPD}(\mu = 0, \sigma = 1, \xi = 0.5)$, $\lambda \sim \exp(1)$, $\omega \sim \text{uniform}(S)$, and $\alpha = 0.5$. The process has been censored below 0.1mm. The ▲ correspond to the process origin ω for each of the four realizations.



where $p(\boldsymbol{Y}_{O_i}|\boldsymbol{\vartheta}_i, \boldsymbol{\theta}_i)$ and $p(\boldsymbol{y}|\boldsymbol{Y}_{O_i}, \boldsymbol{\vartheta}_i, \boldsymbol{\theta}_i)$ are multivariate Gaussian densities derived from (3.5). To handle integration in (3.6) we follow the augmentation method of DeOliveira (2005) by embedding Monte-Carlo integration within our Bayesian MCMC. The exact algorithm is discussed in 3.2.

3.1.2 Process level

Here we specify the model for the process $\boldsymbol{\vartheta}_i = (Z_i, \boldsymbol{\varepsilon})^\top$. By construction, we have that $\{Z_i\}_{i=1}^n \overset{iid}{\sim} \text{GPD}\,(0, 1, \xi)$. The large scale trend ε is modeled as a Gaussian process whose d-dimensional projection is multivariate normal: $\boldsymbol{\varepsilon} \sim N_d\left(\mathbf{0}, \sigma_\varepsilon^2 \Lambda_{\phi_\varepsilon}\right)$, where $\Lambda_{\phi_\varepsilon}$ is a $d \times d$ correlation matrix with entries $\Lambda_{\phi_\varepsilon}(j,k) = \exp\left(-\|s_j - s_k\|/\phi_\varepsilon\right)$. Letting $\boldsymbol{\nu} = (\xi, \sigma_\varepsilon^2, \phi_\varepsilon)$, the density of the process model $\boldsymbol{\vartheta}_i$ given $\boldsymbol{\nu}$ is

$$p(\boldsymbol{\vartheta}_i|\boldsymbol{\nu}) \propto \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\boldsymbol{\varepsilon}^\top \Lambda_{\phi_\varepsilon}^{-1}\boldsymbol{\varepsilon}\right\}(1 + \xi Z_i)_+^{-1/\xi - 1}.$$

3.1.3 Prior

Prior distributions for the parameters $\boldsymbol{\nu} = (\xi, \sigma_\varepsilon^2, \phi_\varepsilon)$ and $\boldsymbol{\theta}_i = (\lambda_i, \alpha, \omega_i)$, are assumed to be independent. For σ_ϵ^2 we use the conjugate prior $\sigma_\varepsilon^2 \sim \text{IG}\,(a_\varepsilon/2, b_\varepsilon/2)$ where $\text{IG}\,(a, b)$ stands for the inverse Gamma distribution with mean $b/(a-1)$. We also specify $\phi_\varepsilon \sim \text{IG}(c_\varepsilon, d_\varepsilon)$. For the shape parameter ξ we use a normal prior $\xi \sim N(m_\xi, v_\xi^2)$ and we let $\{\lambda_i\}_{i=1}^n \overset{iid}{\sim} \text{Gamma}\,(a_\lambda, \beta_\lambda)$ where a_λ and b_λ may be determined by initial hypotheses about the spatial extent of storms. Lacking a priori information about the center of extreme 24 hour precipitation events, we let $\{\omega_i\}_{i=1}^n \overset{iid}{\sim} \text{Uniform}(S)$. Lastly, α is assumed to be a fixed constant. In practice it is difficult to estimate α unless the set of observations are very dense. For our analysis of Section 4, we fit the model by fixing $\alpha \in \{0.1, 0.2, \ldots, 1.9\}$ and select the corresponding α with the best fit as judged by goodness of fit criterion. Thus the prior distributions for $\boldsymbol{\nu}$ and $\boldsymbol{\theta}_i$ are given by

$$p(\boldsymbol{\nu}) \propto \exp\left\{-\frac{(\xi - m_\xi)^2}{2v_\xi^2}\right\}\left(\sigma_\varepsilon^2\right)^{-a_\varepsilon/2-1} \times \exp\left\{-\frac{b_\varepsilon}{2\sigma_\varepsilon^2}\right\}(\phi_\varepsilon)^{-c_\varepsilon-1}\exp\left\{-\frac{d_\varepsilon}{\phi_\varepsilon}\right\}$$

$$p(\boldsymbol{\theta}_i) \propto \lambda_i^{a_\lambda - 1}\exp\{-b_\lambda \lambda_i\}$$

3.2 MCMC sampling

For each $i \in \{1, \ldots, n\}$, the hierarchical model detailed above yields the following posterior density given the data $\mathcal{D}_i$

$$\begin{aligned} &p\left(\boldsymbol{Y}_{C_i}, \boldsymbol{\vartheta}_i, \boldsymbol{\theta}_i, \boldsymbol{\nu}|\mathcal{D}_i\right) \\ &\quad\propto p(\boldsymbol{\vartheta}_i|\boldsymbol{\nu})p(\boldsymbol{\theta}_i)p(\boldsymbol{\nu})p(\mathcal{D}_i|\boldsymbol{Y}_{C_i}, \boldsymbol{\vartheta}_i, \boldsymbol{\theta}_i)p(\boldsymbol{Y}_{C_i}|\boldsymbol{\vartheta}_i, \boldsymbol{\theta}_i) \\ &\quad\propto p(\boldsymbol{\vartheta}_i|\boldsymbol{\nu})p(\boldsymbol{\theta}_i)p(\boldsymbol{\nu}) \\ &\qquad\times |\Sigma_{\boldsymbol{\theta}_i}|^{-1/2}\exp\left\{-\frac{1}{2}\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right)^\top \Sigma_{\boldsymbol{\theta}_i}^{-1}\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right)\right\} \\ &\qquad\times \prod_{j \in C_i}\mathbf{1}_{\{Y_i(s_j) \le l\}}, \end{aligned}$$

where $\boldsymbol{Y}_i = (\boldsymbol{Y}_{O_i}, \boldsymbol{Y}_{C_i})$ denotes the 'full' vector of log transformed measurements. The censored components $\{y_{ij}, j \in C_i\}$ of the vector $\boldsymbol{Y}_i$ are initialized at the censoring limit l and then sampled individually at each iteration of the MCMC from their respective full conditionals which are univariate truncated normal:

$$\begin{aligned} &p\left(y_{ij}|\boldsymbol{\vartheta}_i, \boldsymbol{\theta}_i, \boldsymbol{\nu}, \mathcal{D}_i, j \in C_i\right) \\ &\qquad= p\left(y_{ij}|\boldsymbol{Y}_{i(j)}, \boldsymbol{\vartheta}_i, \boldsymbol{\theta}_i, \boldsymbol{\nu}, , j \in C_i\right) \\ &\propto N\left(\mu_{ij} + \boldsymbol{v}_{i(j)}^\top \Sigma_{\boldsymbol{\theta}_i(j)}^{-1}(\boldsymbol{Y}_{i(j)} - \boldsymbol{\mu}_{i(j)}), v_{ij} - \boldsymbol{v}_{i(j)}^\top \Sigma_{\boldsymbol{\theta}_i(j)}^{-1}\boldsymbol{v}_{i(j)}\right) \\ &\qquad\qquad\times \mathbf{1}_{\{y_{ij} \le l\}}. \end{aligned}$$

Here $\boldsymbol{Y}_{i(j)}$ is $\boldsymbol{Y}_i$ with the jth element removed, $\boldsymbol{v}_{i(j)}$ is the jth column of $\Sigma_{\boldsymbol{\theta}_i}$ with the jth element removed, $\Sigma_{\boldsymbol{\theta}_i(j)}$ is $\Sigma_{\boldsymbol{\theta}_i}$with the jth row and column removed and v_{ij} is the jth diagonal element of $\Sigma_{\boldsymbol{\theta}_i}$. This *data augmentation* for censored observations ensures that likelihood contributions given the data $\mathcal{D}_i$ follow (3.6). See DeOliveira (2005) for further details. Under the assumption that $\boldsymbol{Y}_i$ and $(\boldsymbol{Y}_{C_i}, \mathcal{D}_i)$ contain the same information, the full conditionals for $\boldsymbol{\varepsilon}$ and σ_ε^2 can be sampled directly

$$\begin{aligned} &p\left(\sigma_\varepsilon^2|\boldsymbol{\vartheta}_i, \boldsymbol{\theta}_i, \xi, \phi_\varepsilon, \boldsymbol{Y}_{C_i}, \mathcal{D}_i, i = 1, \ldots, n\right) \\ &\qquad= p\left(\sigma_\varepsilon^2|\boldsymbol{\vartheta}_i, \boldsymbol{\theta}_i, \xi, \phi_\varepsilon, \boldsymbol{Y}_i, i = 1, \ldots, n\right) \\ &\qquad= \text{IG}\left(\frac{1}{2}(d + a_\varepsilon), \frac{1}{2}(b_\varepsilon + \boldsymbol{\varepsilon}^\top \Lambda^{-1}\boldsymbol{\varepsilon})\right), \end{aligned}$$

$$\begin{aligned} &p\left(\boldsymbol{\varepsilon}|Z_i, \boldsymbol{\theta}_i, \boldsymbol{\nu}, \boldsymbol{Y}_{C_i}, \mathcal{D}_i, i = 1, \ldots, n\right) \\ &\qquad= p\left(\boldsymbol{\varepsilon}|Z_i, \boldsymbol{\theta}_i, \boldsymbol{\nu}, \boldsymbol{Y}_i, i = 1, \ldots, n\right) \\ &\qquad= N_d\left(\tilde{\boldsymbol{\mu}}_\varepsilon, \tilde{\Lambda}_\varepsilon\right), \end{aligned}$$

where $\tilde{\boldsymbol{\mu}}_\varepsilon = \sum_{i=1}^n \tilde{\Lambda}_\varepsilon \Sigma_{\boldsymbol{\theta}_i}^{-1}(\boldsymbol{Y}_i + \boldsymbol{\gamma}_{\boldsymbol{\theta}_i} - \log Z_i \mathbf{1})$, $\tilde{\Lambda}_\varepsilon = \Lambda_{\phi_\varepsilon}^{-1} + \sum_{i=1}^n \Sigma_{\boldsymbol{\theta}_i}^{-1}$ and $\boldsymbol{\gamma}_{\boldsymbol{\theta}_i} = \text{diag}\left(\Sigma_{\boldsymbol{\theta}_i}\right)/2$. The remaining full conditionals are non-standard and require Metropolis Hastings sampling. Letting

$$p(\boldsymbol{Y}_i|\boldsymbol{\vartheta}_i, \boldsymbol{\theta}_i) = N_d\left(\boldsymbol{\mu}_i, \Sigma_{\boldsymbol{\theta}_i}\right),$$

we have

$$p\left(Z_i|\boldsymbol{\varepsilon},\boldsymbol{\theta}_i,\boldsymbol{\nu},\boldsymbol{Y}_{C_i},\mathcal{D}_i\right) \propto p\left(Z_i|\boldsymbol{\varepsilon},\boldsymbol{\theta}_i,\boldsymbol{Y}_i\right) \propto \left(1+\xi Z_i\right)_+^{-1/\xi-1} p(\boldsymbol{Y}_i|\boldsymbol{\vartheta}_i,\boldsymbol{\theta}_i).$$

$$\begin{aligned} p\left(\xi|Z_i,\boldsymbol{\theta}_i,\boldsymbol{\nu},\boldsymbol{Y}_{C_i},\mathcal{D}_i,i=1,\ldots,n\right) &\propto p\left(\xi|Z_i,i=1,\ldots,n\right) \\ &\propto \left\{\prod_{i=1}^{n}\left(1+\xi Z_i\right)_+^{-1/\xi-1}\right\}\exp\left\{-\frac{(\xi-m_\xi)^2}{2v_\xi^2}\right\}. \end{aligned}$$

$$\begin{aligned} p(\phi_\varepsilon|\boldsymbol{\vartheta}_i,\boldsymbol{\theta}_i,\xi,\sigma_\varepsilon^2,\boldsymbol{Y}_{C_i},\mathcal{D}_i,i=1,\ldots,n) &\propto p(\phi_\varepsilon|\boldsymbol{\varepsilon},\sigma_\varepsilon^2) \\ &\propto |\Lambda_{\phi_\varepsilon}|^{-1/2}\exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\boldsymbol{\varepsilon}^\top\Lambda_{\phi_\varepsilon}^{-1}\boldsymbol{\varepsilon}\right\} \\ &\quad\times\left(\phi_\varepsilon\right)^{-c_\varepsilon-1}\exp\left\{-\frac{d_\varepsilon}{\phi_\varepsilon}\right\}. \end{aligned}$$

$$\begin{aligned} p(\lambda_i|\boldsymbol{\vartheta}_i,\alpha,\omega_i,\boldsymbol{Y}_{C_i},\mathcal{D}_i) &= p(\lambda_i|\boldsymbol{\vartheta}_i,\alpha,\omega_i,\boldsymbol{Y}_i) \\ &\propto \lambda_i^{a_\lambda-1}\exp\left\{-b_\lambda\lambda_i\right\}p(\boldsymbol{Y}_i|\boldsymbol{\vartheta}_i,\boldsymbol{\theta}_i). \end{aligned}$$

$$\begin{aligned} p(\omega_i|\boldsymbol{\vartheta},\lambda_i,\alpha,\boldsymbol{Y}_{C_i},\mathcal{D}_i) &= p(\omega_i|\boldsymbol{\vartheta},\lambda_i,\alpha,\boldsymbol{Y}_i) \propto p(\boldsymbol{Y}_i|\boldsymbol{\vartheta}_i,\boldsymbol{\theta}_i). \end{aligned}$$

For most parameters, we use normal proposals. For the non-negative parameters λ_i and ϕ_ε proposals are made with $\log\lambda_i' \sim N(\log\lambda_i,\psi_\lambda)$ and $\log\phi_\varepsilon' \sim N(\log\phi_\varepsilon,\psi_\varepsilon)$, where the candidate values $\lambda_i',\phi_\varepsilon'$ are accepted with probability

$$\max\left\{0,\frac{p(\lambda_i'|\boldsymbol{\vartheta}_i,\alpha,\omega_i,\boldsymbol{Y}_{C_i},\mathcal{D}_i)\log\lambda_i'}{p(\lambda_i|\boldsymbol{\vartheta}_i,\alpha,\omega_i,\boldsymbol{Y}_{C_i},\mathcal{D}_i)\log\lambda_i}\right\},$$

and

$$\max\left\{0,\frac{p(\phi_\varepsilon'|\boldsymbol{\vartheta}_i,\boldsymbol{\theta}_i,\xi,\sigma_\varepsilon^2,\boldsymbol{Y}_{C_i},\mathcal{D}_i,i=1,\ldots,n)\log\phi_\varepsilon'}{p(\phi_\varepsilon|\boldsymbol{\vartheta}_i,\boldsymbol{\theta}_i,\xi,\sigma_\varepsilon^2,\boldsymbol{Y}_{C_i},\mathcal{D}_i,i=1,\ldots,n)\log\phi_\varepsilon}\right\}.$$

Normal proposal distributions are also used for ξ and Z_i, taking care that the initialized values for ξ and $\{Z_i\}_{i=1}^N$ satisfy $\min_{i=1\ldots,n}\{1+\xi Z_i\} \geq 0$. Any proposals $\xi' \sim N(\xi,\psi_\xi), Z_i' \sim N(Z_i,\psi_Z)$ outside of the support for their respective full conditionals are accepted with nil probability. In the application below, proposals for the origin parameters ω_i were uniform on S which worked well, but a normal proposal could also be used.

3.3 Spatial prediction

Recall that our primary goal is to make predictions for $\{V_i(s)\}_{s\in S}$ given observations $\{\mathcal{D}_i\}_{i=1}^n$. In practice, posterior predictive distributions can be produced by sampling from the predictive distribution along a finite lattice $\widetilde{S}=\{\tilde{s}_1,\ldots,\tilde{s}_m\}$, conditional on $\boldsymbol{Y}_i,\boldsymbol{\vartheta}_i,\boldsymbol{\theta}_i$. In order to do this, we must first sample the scale trend $\tilde{\boldsymbol{\varepsilon}}=(\varepsilon(\tilde{s}_1),\ldots,\varepsilon(\tilde{s}_m)) \in \mathbb{R}^m$ at prediction sites conditioned on the vales $\boldsymbol{\varepsilon}$ at the observation sites. The distribution of the Gaussian process model for the scale surface ε at both the prediction and observation sites is

$$\begin{pmatrix}\tilde{\boldsymbol{\varepsilon}}\\ \boldsymbol{\varepsilon}\end{pmatrix}|\phi_\varepsilon,\sigma_\varepsilon^2 \sim N_{m+d}\left(\boldsymbol{0},\sigma_\varepsilon^2\begin{pmatrix}\Lambda_m & \Lambda_{dm}^\top\\ \Lambda_{dm} & \Lambda_\varepsilon\end{pmatrix}\right),$$

where Λ_m,Λ_{dm} are the corresponding matrices generated from the correlation function $\exp(-\|s-s'\|/\phi_\varepsilon)$. Hence, we sample from the conditional distribution

$$\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon},\phi_\varepsilon,\sigma_\varepsilon^2 \sim N_m\left(\tilde{\boldsymbol{\mu}}_\varepsilon,\sigma_\varepsilon^2\widetilde{\Lambda}_\varepsilon\right).$$

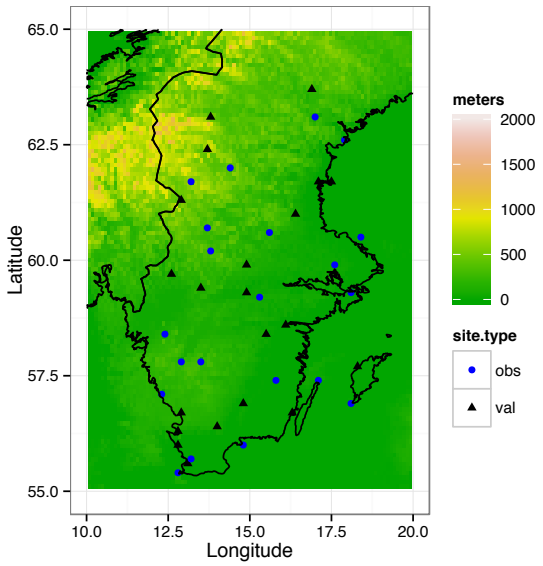
where $\widetilde{\Lambda}_\varepsilon=\Lambda_m-\Lambda_{dm}^\top\Lambda_\varepsilon^{-1}\Lambda_{dm}$ and $\tilde{\boldsymbol{\mu}}_\varepsilon=\Lambda_{dm}^\top\Lambda_\varepsilon^{-1}\boldsymbol{\varepsilon}$. Next, consider the distribution of the data generating model at the prediction and observation sites

$$\begin{aligned}\begin{pmatrix}\widetilde{\boldsymbol{Y}}_i\\ \boldsymbol{Y}_i\end{pmatrix}|\boldsymbol{\vartheta}_i,\boldsymbol{\theta}_i,\tilde{\boldsymbol{\varepsilon}} \\ \sim N_{m+d}\left(\begin{pmatrix}\boldsymbol{\mu}_{i,m}\\ \boldsymbol{\mu}_i\end{pmatrix},\begin{pmatrix}\Sigma_{\boldsymbol{\theta}_i,m} & \Sigma_{\boldsymbol{\theta}_i,dm}^\top\\ \Sigma_{\boldsymbol{\theta}_i,dm} & \Sigma_{\boldsymbol{\theta}_i}\end{pmatrix}\right),\end{aligned}$$

where $\boldsymbol{\mu}_{i,m}$ is a vector with entries $\mu_{ij,m}=\log Z_i+\varepsilon(s_j)-(\|\tilde{s}_j-\omega_i\|/\lambda_i)^\alpha$ and $\Sigma_{\boldsymbol{\theta}_i,m},\Sigma_{\boldsymbol{\theta}_i,dm}$ are the matrices generated from the covariance function $K\left(s,s'|\boldsymbol{\theta}_i\right)$. Consequently, we sample from the conditional distribution

$$\tilde{\boldsymbol{y}}_i|\boldsymbol{Y}_i,\boldsymbol{\vartheta}_i,\boldsymbol{\theta}_i,\tilde{\boldsymbol{\varepsilon}} \sim N_m\left(\tilde{\boldsymbol{\mu}}_i,\widetilde{\Sigma}_{\boldsymbol{\theta}_i}\right),$$

Figure 2: Map of synoptic stations over south central Sweden.



where $\widetilde{\Sigma}_{\boldsymbol{\theta}_i} = \Sigma_{\boldsymbol{\theta}_i,m} + \Sigma^{\top}_{\boldsymbol{\theta}_i,dm}\Sigma^{-1}_{\boldsymbol{\theta}_i}\Sigma_{\boldsymbol{\theta}_i,dm}$ and

$$\tilde{\boldsymbol{\mu}}_i = +\Sigma^{\top}_{\boldsymbol{\theta}_i,dm}\Sigma^{-1}_{\boldsymbol{\theta}_i}\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right).$$

The sampling is repeated for each iteration of the MCMC, yielding a posterior predictive distribution $\{\tilde{\boldsymbol{v}}_i^{(k)}\}_{k=1}^N := \{\exp(\tilde{\boldsymbol{y}}_i^{(k)})\}_{k=1}^N$. The resulting posterior predictive distributions can then be used to calculate point estimates from quantiles and uncertainty is based on the distribution spread.

4 Extreme Summer precipitation over south central Sweden

We apply our model to extreme Summer 24 hour precipitation totals gathered from synoptic stations of the Swedish Meteorological and Hydrological Institute. The data is open access and available at `http://www.smhi.se`. We focus our study to south central Sweden (below N 65° latitude) during the summer months June, July and August where precipitation falls almost exclusively as rain. Daily observations span years 1961-2011. We select only observations with highest quality control flags. With these data, non-zero precipitation below 0.1mm is reported as zero, hence we set the left-censoring limit to 0.1mm for all observations. Based on record length and completeness, 21 stations were selected as observation sites while an additional 21 sites with sparse observations were designated for validation. Figure 2 is a map of the region of interest and locations of the synoptic stations.

Table 1: Specification of hyper-parameters for MCMC.

σ^2_ε	ϕ_ε	ξ	λ
$a_\varepsilon = 5$ $b_\varepsilon = 2$	$c_\varepsilon = 3$ $d_\varepsilon = .25$	$m_\xi = .50$ $v^2_\xi = .03$	$a_\lambda = .10$ $b_\lambda = .10$

To designate observations that are extreme, with each day t in the record $\{1961 : 2011\}$ we first calculate $V_t^{\max} = \max_{s\in\{s_1,\ldots,s_d\}} V_t(s)$ where $\{s_1,\ldots,s_d\}$ is the set of $d = 21$ observation locations. Then we select dates t for which $V_t^{\max}$ exceeded the 95th percentile of daily maximum observations $\{V_t^{\max}\}_{t\in\{1961:2011\}}$. This subset of data displayed strong temporal dependence containing multiple clusters of consecutive dates. To rule out dependent observations we select the dates corresponding to the largest $V_t^{\max}$ within each cluster of consecutive days. This resulted in a final sample of $n = 59$ dates with no evidence of temporal dependence. There were no missing data in this sample.

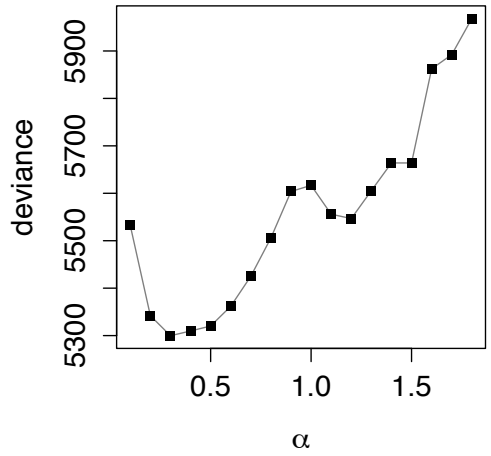
We fit our model using the MCMC sampling scheme described in Section 3.2 with 60,000 iterations. Proposal variances $\psi_\lambda, \psi_\varepsilon, \psi_\xi, \psi_Z$ were tuned using preliminary runs such that acceptance rates were between .25 and .40 (Gelman et al., 1996). Convergence of preliminary runs were monitored using trace plots. Posterior distributions for $\sigma^2_\varepsilon, \phi_\varepsilon$ and λ were not sensitive to prior specification and for these parameters vague priors were used. The specification of hyper-parameters is given in Table 1. Preliminary MCMC runs indicated that the shape parameter ξ is difficult to estimate which is typical for spatial extremes (see e.g. Berrocal et al., 2014; Thibaud et al., 2013). Hence, we chose an informative prior based on previous literature for extreme precipitation where analyses typically suggested $\xi \in (0, 1)$. In the final MCMC, the first 10,000 iterations were discarded as burn-in and then further thinned every 50 iterations to reduce serial correlations of the ω_i. This resulted in a posterior sample of size $N = 1000$. Due to the difficulty in estimating α, multiple chains were run for each $\alpha \in \{0.1, 0.2, \ldots, 1.9\}$. Because the number of parameters does not change as we vary α, we can compare model fit directly using the negative log-likelihood (deviance) at the data generating level

$$D = -2\sum_{i=1}^{n} \log\left\{\sum_{k=1}^{N} p(\boldsymbol{Y}_i^{(k)}|\hat{\boldsymbol{\vartheta}}_i, \hat{\boldsymbol{\theta}}_i)\right\}, \qquad (4.1)$$

where $\boldsymbol{Y}_i^{(k)} = (\boldsymbol{Y}_{O_i}, \boldsymbol{Y}_{C_i}^{(k)})$, represent MCMC draws for censored observations concatenated with uncensored observations. Note that (4.1) is derived from (3.6) and lower deviance scores correspond to better fit.

Figure 3 implies that the best model in terms of the

Figure 3: Deviance scores (4.1) of model fit versus smoothness parameter α.



deviance score corresponds to $\alpha = 0.3$. Hence, the remainder of the results are shown for $\alpha = 0.3$. Overall, it was difficult to detect significant differences in results for $\alpha \in (0.1, 1]$. Varying $\alpha \in (0.1, 1]$ had little effect on predictive performance and while the value of the range parameter λ_i adjusted accordingly, posterior distributions of the remaining parameters did not noticeably change. We did notice that performance deteriorated for $\alpha > 1$.

The MCMC procedure yields posterior point estimates of the event centers $\hat{\omega}_i := N^{-1} \sum_{k=1}^{N} \omega_i^{(k)}$. As an additional diagnostic, we examine the concordance between the estimated event center $\hat{\omega}_i$ and location of observed spatial maxima $s_i^{\max} := \arg\max_{s \in \{s_1, \ldots, s_d\}} V_i(s)$. While the true spatial maxima and even center ω_i are unknown, one would expect the observed $s_i^{\max}$ to be "close" to $\hat{\omega}_i$. Indeed, Figure 4 supports a positive association between location of the observed maxima and the estimated event centers, particularly in the meridional (N-S) direction.

The remaining fitted parameter estimates were constructed using posterior means. Specifically, $\hat{\xi} = 0.29$, $\hat{\sigma}_\varepsilon^2 = 3.70$ and $\hat{\phi}_\varepsilon = 26.67$. The distributions of the point estimates $\{\hat{\lambda}_i\}_{i=1}^n$ and $\{\hat{Z}_i\}_{i=1}^n$ are displayed in Figure 6.

4.1 Evaluating predictions

The MCMC algorithm generates posterior predictive distributions

$$\hat{F}_{\tilde{s}}^{(i)}(v) := N^{-1} \sum_{k=1}^{N} \mathbf{1}_{\{v \le v_i^{(k)}(\tilde{s})\}},$$

where $v_i^{(k)}(\tilde{s})$ represents the MCMC draw from the predictive distribution at location $\tilde{s}$ for iteration k. We eval-

Figure 4: Posterior mean of origin centers $\hat{\omega}_i$ versus location of maximum observation $s_i^{\max}$

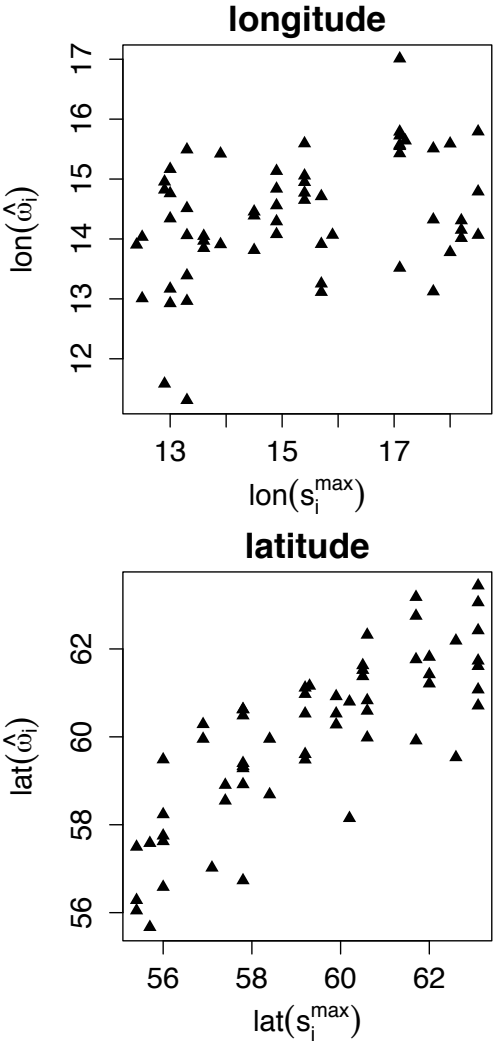
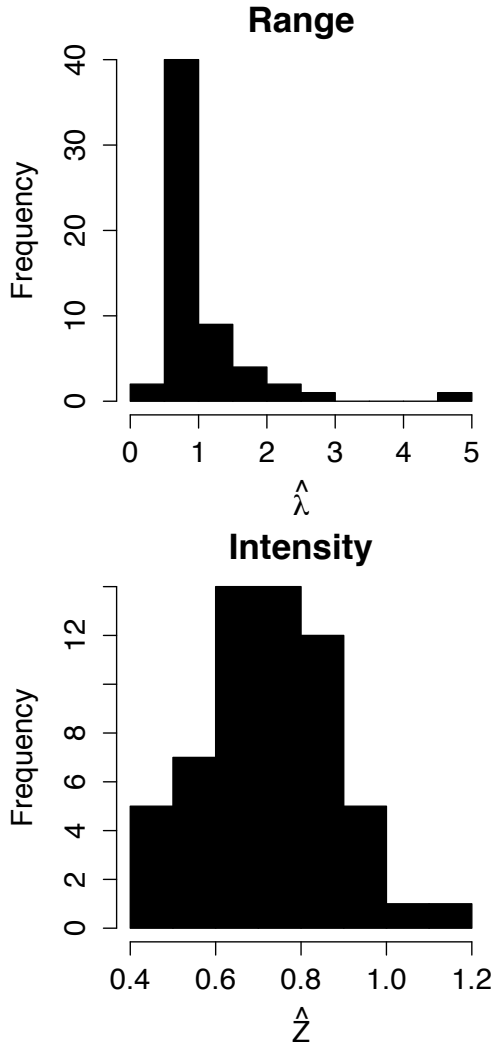


Figure 5: Distribution of posterior means for the range λ_i and intensities Z_i for each fitted date $i = 1, \ldots, 59$.



uate the predictive distribution using the probability integral transform (PIT). If the distribution $\hat{F}_{\tilde{s}}^{(i)}$ is ideal, then probability integral transforms $\{\hat{F}_{\tilde{s}}^{(i)}(V_i(\tilde{s})\}_{i=1}^n$ should be uniformly distributed on $(0, 1)$ (See e.g. Gneiting et al., 2007). Figure 6 displays the PIT histograms evaluated for the 21 validation sites. One should expect a reasonable amount of variance from the solid horizontal line indicating perfect uniformity. For reference, confidence bands indicating approximate 90 percent confidence intervals of the bar heights are shown. While some locations display a slight 'U' shape indicating underdispersion, overall the predictive distributions appear skillful when compared with random samples from a standard uniform distribution.

5 Discussion

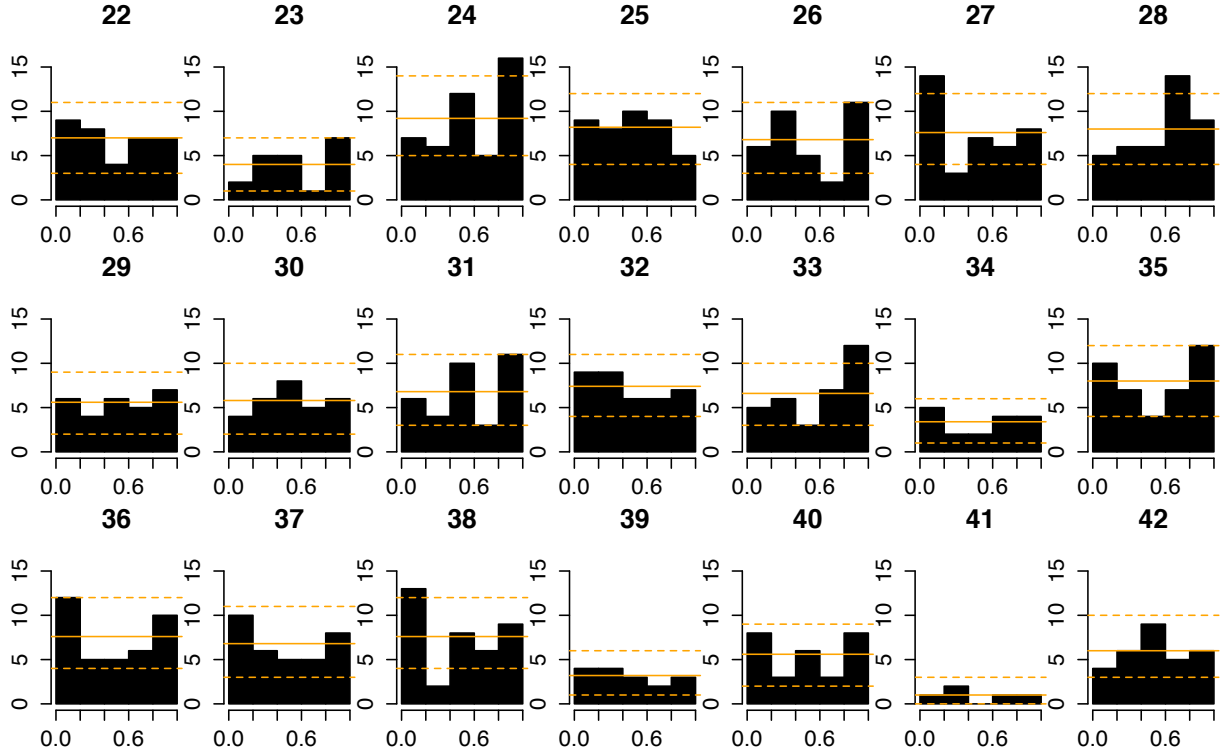
We introduced Gauss-Pareto processes as a flexible class of models for extreme precipitation that can be fit using standard MCMC techniques for Bayesian hierarchical modeling while retaining essential non-trivial dependence characteristics of popular max-stable and Pareto process models. Unlike the models used in Davison et al., 2012; Thibaud et al., 2013; Sang and Gelfand, 2009, spatial prediction is straightforward via conditional sampling from latent Gaussian processes and the hierarchical structure allows one to consider larger areas of interest by allowing the center and range of precipitation events to vary from storm to storm. Predictive distributions validated at holdout locations appear skillful and simulations from the process also appear realistic. There have been a very limited number of works for spatial prediction of extreme precipitation in the strict sense that we introduce in Section 2. and comparison with the alternative methods described in Ferreira and DeHaan (2012) and Thibaud and Opitz (2013) are still pending, nonetheless our results appear promising.

Our methodology can be used for several applications in climate modeling such as fusion of climate model output with observational data for downscaling and testing distributional concordance between observed extremes and climate models, these are considerations for future work. Another interesting extension is motivated by the fact that extreme 24 hour precipitation displays strong temporal dependence. Consider the spatio-temporal model

$$V(s,t) = Z(t)\exp\left(W(s,t)\right)$$

where W is now a spatio-temporal Gaussian process and Z is a Heavy-tailed time dependent process. Fitting a spatio-temporal model would allow the borrowing of information from time-dependent observations leading to the incorporation of more data, although characterizing the dependence structure of Z is more challenging. A

Figure 6: Probability integral transform histograms of the predictive distributions at 21 validation sites . Solid horizontal lines correspond to perfect uniformity. Confidence bands (dashed lines) are provided for reference, bin heights from random draws of a standard uniform distribution should fall outside of the dashed lines in approximately ten percent of cases.



spatio-temporal model allows probabilistic forecasts of future precipitation given past observations that are extreme which is perhaps of greater interest than the purely spatial prediction we consider in this work.

Acknowledgements

This collaboration was made possible the Research Network for Statistical Methods for Atmospheric and Oceanic Sciences (STATMOS) https://www.statmos.washington.edu. R.Yuen was partially supported by the National Science Foundation DMS-1106695 and the University of Michigan Rackham Merit Fellowship. The authors would like to thank Stilian Stoev and Veronica Berrocal for helpful suggestions and discussions.

References

V.J. Berrocal, A.E. Gelfand, and D.M. Holland. Assessing exceedance of ozone standards: a space-time downscaler for fourth highest ozone concentrations. *Environmetrics*, 25(4):279–291, 2014. ISSN 1099-095X. doi: 10.1002/env.2273. URL http://dx.doi.org/10.1002/env.2273.

B. Brown and S. Resnick. Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14: 732–739, 1977.

S. Coles. *An Introduction to the Statistical Modeling of Extreme Values*. Springer, 2001.

A.C. Davison, S. Padoan, and M. Ribatet. The statistical modeling of spatial extremes. *Statistical Science*, 2012.

V. DeOliveira. Bayesian inference and prediction of gaussian random fields based on censored data. *Journal of Computational and Graphical Statistics*, 14(1):95–115, 2005.

C. Dombry, F. Eyi-Minko, and M. Ribatet. Conditional simulation of max-stable processes. *Biometrika*, 2012. doi: 10.1093/biomet/ass067.

John H. Einmahl, Andrea Krajina, and Johan Segers. An M-estimator for tail dependence in arbitrary dimension. *Annals of Statistics*, 40(3):1764–1793, 2012.

A. Ferreira and L. DeHaan. The generalized pareto process: with a view towards application and simulation. *arXiv:1203.2551v2*, 2012.

C.B. Field, V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. Cambridge University Press, 2012.

M. Fuentes, J. Henry, and B. Reich. Nonparametric spatial models for extremes: Application to extreme temperature data. *Extremes*, 10:75–101, 2013.

A. Gelman, G. O. Roberts, and W. R. Gilks. Efficient Metropolis jumping rules. In *Bayesian statistics, 5 (Alicante, 1994)*, Oxford Sci. Publ., pages 599–607. Oxford Univ. Press, New York, 1996.

Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268, 2007. ISSN 1467-9868. doi: 10.1111/j.1467-9868.2007.00587.x. URL http://dx.doi.org/10.1111/j.1467-9868.2007.00587.x.

Z. Kabluchko, M. Schlather, and L. de Haan. Stationary max–stable fields associated to negative definite functions. *Annals of Probability*, 37(5):2042–2065, 2009.

V.V. Kharin and F.W. Zwiers. Estimating extremes in transient climate change simulations. *Journal of Climate*, 18(8):1156–1173, 2005. doi: 10.1175/JCLI3320.1.

E. Orskaug, I. Scheel, A. Frigessi, P. Guttorp, J. E. Haugen, O. E. Tveito, and Haug. O. Evaluation of a dynamic downscaling of norwegian precipitation. *Tellus A*, 63:746–756, 2011.

S. Padoan, M. Ribatet, and S. Sisson. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277, 2010.

S. I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York, 1987.

H. Sang and A.E. Gelfand. Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics*, 15: 49–65, 2009.

M. Schlather. Models for stationary max-stable random fields. *Extremes*, 5:33–44, 2002.

R.L. Smith. Multivariate threshold methods. In J. Galambos, J. Lechner, and E. Simiu, editors, *Extreme vault theory and applications*, pages 225–248. Kluwer, 1994.

E. Thibaud and T. Opitz. Efficient inference and simulation for elliptical pareto processes. *arXiv:1401.0168v1*, 2013.

E. Thibaud, R. Mutzner, and A.C. Davison. Threshold modeling of extreme spatial rainfall. *Water resources research*, 49:4633–4644, 2013.

Yizao Wang and Stilian A. Stoev. Conditional sampling for spectrally discrete max-stable random fields. *Adv. in Appl. Probab.*, 43(2):461–483, 2011.

Robert Yuen and Stilian Stoev. Crps m-estimation for max-stable models. *Extremes*, pages 1–24, 2014. ISSN 1386-1999. doi: 10.1007/s10687-014-0185-x. URL http://dx.doi.org/10.1007/s10687-014-0185-x.